Australian Government
Productivity Commission

Data Availability and Use

Productivity Commission Draft Report

October 2016

This is a draft report prepared for further public consultation and input. The Commission will finalise its report after these processes have taken place.

An appropriate reference for this publication is:

Productivity Commission 2016, *Data Availability and Use,* Draft Report, Canberra.

Publications enquiries

Media and Publications, phone: (03) 9653 2244 or email: maps@pc.gov.au

The Productivity Commission

The Productivity Commission is the Australian Government's independent research and advisory body on a range of economic, social and environmental issues affecting the welfare of Australians. Its role, expressed most simply, is to help governments make better policies, in the long term interest of the Australian community.

The Commission's independence is underpinned by an Act of Parliament. Its processes and outputs are open to public scrutiny and are driven by concern for the wellbeing of the community as a whole.

Further information on the Productivity Commission can be obtained from the Commission's website (www.pc.gov.au).

Opportunity for further comment

You are invited to examine this draft inquiry report and comment on it by written submission to the Productivity Commission, preferably in electronic format, by **12 December 2016** and/or by attending a public hearing.

The final report will be prepared after further submissions have been received and public hearings have been held and will be forward to the Australian Government by the 21 March 2017.

Public hearing dates and venues

Location	Date	Venue
Melbourne	21 November 2016	Rattigan Rooms Level 12, 530 Collins Street
Sydney	28 November 2016	SMC Conference & Function Centre 66 Goulburn Street

Commissioners

For the purposes of this inquiry and draft report, in accordance with section 40 of the *Productivity Commission Act 1998* the powers of the Productivity Commission have been exercised by:

Peter Harris	Presiding Commissioner
Melinda Cilento	Commissioner

Disclosure of interests

The *Productivity Commission Act 1998* specifies that where Commissioners have or acquire interests, pecuniary or otherwise, that could conflict with the proper performance of their functions during an inquiry they must disclose the interests.

Ms Cilento has advised the Commission that she is a director of Australian Unity (which made a submission to this Inquiry) and of Woodside Petroleum (which is referred to in an included example in the Report).

Terms of reference

I, Scott Morrison, Treasurer, pursuant to Parts 2 and 3 of the *Productivity Commission Act 1998*, hereby request that the Productivity Commission undertake an inquiry into the benefits and costs of options for increasing availability of and improving the use of public and private sector data by individuals and organisations.

Background

The 2014 Financial System Inquiry (the Murray Inquiry) recommended that the Government task the Commission to review the benefits and costs of increasing the availability and improving the use of data. The 2015 Harper Review of Competition Policy recommended that the Government consider ways to improve individuals' ability to access their own data to inform consumer choices. The Government has agreed to pursue these two recommendations.

The Australian Government seeks to consider policies to increase availability and use of data to boost innovation and competition in Australia and the relative benefits and costs of each option.

Effective use of data is increasingly integral to the efficient functioning of the economy. Improved availability of reliable data, combined with the tools to use it, is creating new economic opportunities. Increasing availability of data can facilitate development of new products and services, enhance consumer and business outcomes, better inform decision making and policy development, and facilitate greater efficiency and innovation in the economy.

As in Australia, international governments are encouraging greater use of data through open data policies. This will increase the transparency and accountability of government processes.

Increased sharing of data across the public and private sectors could facilitate greater leveraging of technology to improve individuals' and entities' interactions with government, improve the integrity of systems and increase administrative efficiency.

In taking advantage of greater use of data, it is important to give appropriate attention to other interests such as privacy, security and intellectual property.

Scope of the inquiry

The Commission is to conduct a broad ranging investigation into the benefits and costs of options for improving availability and use of data. In developing recommendations, the Commission is to:

1. Examine the benefits and costs of options for increasing availability of public sector data to other public sector agencies (including between the different levels of government), the private sector, research sector, academics and the community. Where there are clear benefits, recommend ways to increase and improve data linking and availability. The Commission should:
   (a) identify the characteristics and provide examples of public sector datasets that would provide high-value to the public sector, research sector, academics and the community to assist public sector agencies to identify their most valuable data
   (b) examine legislation or other impediments that may unnecessarily restrict the availability and linking of data, including where the costs are substantial, and consider options to reduce or remove those impediments.
2. Examine the benefits and costs of options for increasing availability of private sector data for other private sector firms, the public sector, the research sector, academics and the community. Where there are clear benefits, consider ways to increase and improve availability. The Commission should:
   (a) identify the characteristics and provide examples of private sector datasets that would provide high value to the private sector, public sector, the research sector, academics and the community in developing or providing products and services and undertaking research and developing policy
   (b) identify the concerns of private sector data owners and provide recommendations on principles or protocols to manage these concerns
   (c) examine legislation or other impediments that unnecessarily restrict the availability of data, including where the costs are substantial, and consider options to reduce or remove those impediments
   (d) provide an update on existing data sharing initiatives in Australia, including the uptake of the credit reporting framework. Consider recommendations for improving participation in such initiatives.
3. Identify options to improve individuals' access to public and private sector data about themselves and examine the benefits and costs of those options. The Commission should:
   (a) examine how individuals can currently access their data, including data about them held by multiple government agencies, and develop recommendations to streamline access
   (b) identify datasets, including datasets of aggregated data on consumer outcomes at the product or provider level, that would provide high value to consumers in

making informed decisions and any impediments to their use. Develop guidance to assist in identification of other high value datasets

   (c) examine the possible role of third party intermediaries to assist consumers in making use of their data.

4. Examine the options for, and benefits and costs of, standardising the collection, sharing and release of public and private sector data.
5. Examine ways to enhance and maintain individuals' and businesses' confidence and trust in the way data are used. Having regard to current legislation and practice, advise on the need for further protocols to facilitate disclosure and use of data about individuals and businesses while protecting privacy and commercial interests and, if recommended, advise on what these should be. The Commission should:
   (a) balance the benefits of greater disclosure and use of data with protecting the privacy of the individual and providing sufficient control to individuals as to who has their information and how it can be used
   (b) benchmark Australia's data protection laws, privacy principles and protocols against leading jurisdictions
   (c) examine whether there is adequate understanding across government about what data can be made openly available given existing legislation
   (d) consider the effectiveness and impacts of existing approaches to confidentialisation and data security in facilitating data sharing and linking while protecting privacy
   (e) consider the merits of codifying the treatment and classification of business data.

In developing its recommendations, the Commission should take into account the Government's policy to improve the availability and use of public sector data (the *Public Data Policy Statement*) as part of its *National Innovation and Science Agenda* and to improve government performance through the *Efficiency through Contestability Programme*, as well as the findings of the *Public Sector Data Management Project*.

The Commission should consider domestic and international best practice and the measures adopted internationally to encourage sharing and linking of both public and private data.

Process

The Commission is to undertake an appropriate public consultation process, inviting public submissions and releasing a draft report to the public. A final report should be provided to the Government within 12 months from the date of receipt of the reference.

Scott Morrison
Treasurer

[Received 21 March 2016]

Contents

Abbreviations

ABS	Australian Bureau of Statistics
ACCC	Australian Competition and Consumer Commission
ACMA	Australian Communications and Media Authority
AGIMO	Australian Government Information Management Office
AIHW	Australian Institute of Health and Welfare
ALRC	Australian Law Reform Commission
ANDS	Australian National Data Service
API	Application Programming Interface
APP	Australian Privacy Principle
APRA	Australian Prudential Regulation Authority
ARA	Accredited Release Authority
ARC	Australian Research Council
ASAC	Australian Statistics Advisory Council
ATO	Australian Tax Office
AURIN	Australian Urban Research Infrastructure Network
BLADE	Business Longitudinal Analytical Data Environment
CCR	Comprehensive Credit Reporting
COAG	Council of Australian Governments
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DFAT	Department of Foreign Affairs and Trade
DHS	Department of Human Services
DPMC	Department of Prime Minister and Cabinet
DTO	Digital Transformation Office
FOI	Freedom of Information
GIS	Geographic Information System
G-NAF	Geocoded National Address File
GPS	Global Positioning System
HILDA	Household, Income and Labour Dynamics Australia
HREC	Human Research Ethics Committee
ICT	Information and Communications Technology

IDI	Integrated Data Infrastructure
IT	Information Technology
IoT	Internet of Things
JSON	JavaScript Object Notation
MADIP	Multi-Agency Data Integration Project
MBS	Medicare Benefits Schedule
MOG	Machinery of Government
MOU	Memorandum of Understanding
NAA	National Archives Australia
NCRIS	National Collaborative Research Infrastructure Strategy
NDC	National Data Custodian
NHMRC	National Health and Medical Research Council
NID	National Interest Dataset
NSS	National Statistical Service
NSW DAC	New South Wales Data Analytics Centre
OAIC	Office of the Australian Information Commissioner
OECD	Organisation for Economic Co-operation and Development
PBS	Pharmaceutical Benefits Scheme
PC	Productivity Commission
PHRN	Population Health Research Network
RBA	Reserve Bank of Australia
SURE	Secure Unified Research Environment
WWWF	World Wide Web Foundation

OVERVIEW

Key points

- Extraordinary growth in data generation and usability, fuelled by developments in computing power, Internet connectivity and algorithms, have enabled a kaleidoscope of new business models, products and insights to emerge. Individuals, businesses, governments and the broader community have all benefited from these changes.
- Frameworks and protections developed for data collection and access prior to sweeping digitisation now need reform. This is a global phenomenon and Australia, to its detriment, is not yet participating.
- The substantive argument in favour of making data more available is that opportunities to use it are largely unknown until the data sources themselves are better known, and until data users have been able to undertake discovery of data.
- Lack of trust and numerous barriers to sharing and releasing data are stymieing the use and value of Australia's data.
- Marginal changes to existing structures and legislation will not suffice. The Commission is proposing reforms to data availability and use, aimed at moving from a system based on risk aversion and avoidance, to one based on transparency and confidence in data processes.
- At the centre of proposed reforms is the introduction of a new *Data Sharing and Release Act*, a new National Data Custodian, and a suite of sectoral Accredited Release Authorities that will enable streamlined access to curated datasets.
- A key element of the recommended reforms is to provide greater control for individuals over data that is collected on them by defining a new Comprehensive Right for consumers. This right would mean consumers:
  - retain the power to view information held on them, request edits or corrections, and be advised of disclosure to third parties;
  - have improved rights to opt out of collection in some circumstances; and
  - have a new right to a machine-readable copy of data, provided either to them or to a nominated third party, such as a new service provider.
- Broad access to key National Interest Datasets should be enabled.
  - For datasets designated as national interest, all restrictions to access and use contained in a variety of national and state legislation, and other program-specific policies, would be replaced by new arrangements under the Data Sharing and Release Act.
  - Datasets would be maintained as national assets, access would be substantially streamlined, and linkage with other National Interest Datasets would be feasible.
  - Initial datasets that may be designated national interest and publicly released could include key registries of businesses, services or assets, and data on activity and usage in areas of substantial public expenditure.
- Secure sharing of identifiable data held in the public sector and by publicly funded research bodies should be formalised and streamlined. By pre-approving data uses, trusted users would have more timely access to identifiable data through Accredited Release Authorities and ethics committees.
- Reforming access to public sector data is a priority. Significant change is needed for Australia's open government agenda to catch up with achievements in competing economies.
- The incremental costs associated with more open data access and use — including possible impacts on individuals' privacy and willingness to share data — are expected to be minimal, but they will exist. But greater use of Australia's data can coexist with the management of these risks, including genuine safeguards and meaningful transparency to maintain community trust and confidence.

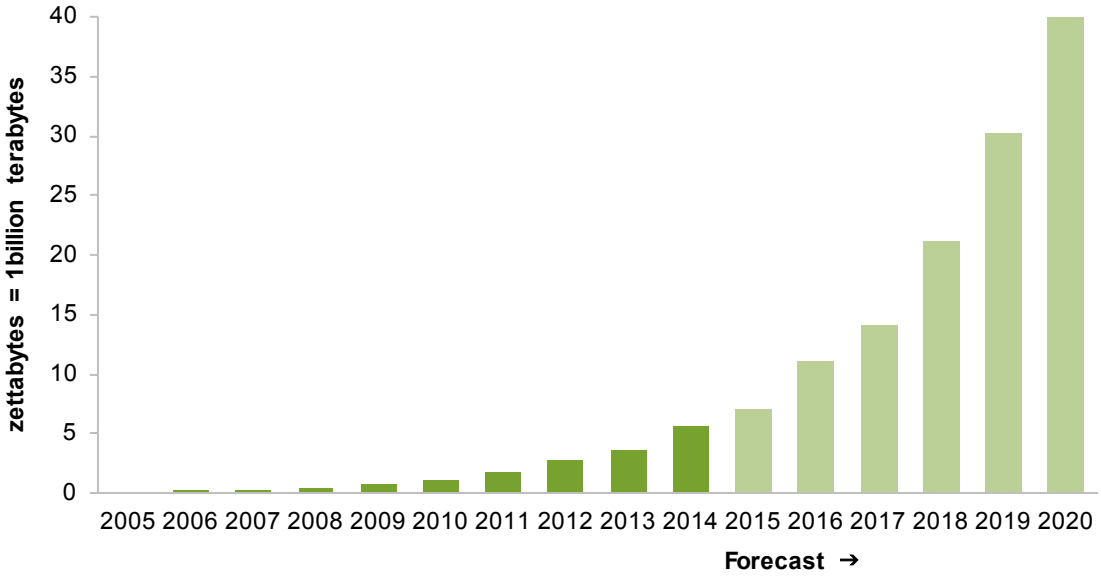
Overview

Thirty years ago, data for most people was primarily about details on paper. Data was largely collected and stored on paper (encyclopaedias, forms, bills, bank deposit slips and phone books); mail actually meant a letter in the letter box. Access to data was clear and locational (you needed keys to the filing cabinet); as was its destruction (via a shredding machine). With the mass digitisation of data, the capacity to collect data through everyday Internet activity and transactions, and through technologies such as sensors, cameras and mobile devices, means that what is 'data', and who can or should have a say in how it is collected, stored and used is no longer so simple.

Until this Inquiry, there has been no structured attempt to comprehensively review this matter in Australia, despite the enormity of the transformation under way.

Data now includes material (raw or processed) on: the characteristics, status, appearance or performance of an individual, product or service, or object (including infrastructure and environmental assets); and expressed or inferred opinions and preferences. The generation of data is seemingly heading upward on an unbounded trajectory (figure 1).

Figure 1 **Data generated (global)**



Source: United Nations Economic Commission for Europe (2015).

By some estimates, the amount of digital data generated globally in 2002 (five terabytes) is now generated every two days, with 90% of the world's information generated in just the past two years (IBM 2016). As we are now only in the very nascent stage of the Internet of Things (whereby our business equipment, vehicles, appliances and wearable devices can communicate with each other and generate data), the upward trend in data generated is more likely than not to accelerate into the future.

Falling costs (per record) of digital data storage, and the spread of low-cost and powerful analytics tools and techniques to extract patterns, correlations and interactions from within data, are also making data analytics more usable and valuable. Yet much of the data being generated remains underutilised. Some estimate that up to 80% of data generated globally may prove to have no value (numerous duplicative digital photos, for example). But still, less than 5% of the potentially useful data is actually analysed to generate information, build knowledge, and thus inform decision making and action (EMC Corporation 2014). And some data that was previously of limited value is becoming valuable as new uses for it emerge, analytical capabilities improve, new linkages are established, or investments made to improve its quality. There is enormous untapped potential in Australia's data.

Access denied — Australia's lost opportunities

With technological developments and advances in analytical techniques, not only is the volume of data being generated and collected growing, but so too is the scope to make use of data in innovative ways in every sphere of life.

Increased access to data can facilitate the development of ground-breaking new products and services that fundamentally transform everyday life. Many are widely known — apps that tell you in real time where to find vacant car parking places, the fastest route to travel to the city at the time you want to go, or which electricity provider offers you the best deal given your pattern of energy use, are all examples that rely on data analysis.

But better access to and use of data can also benefit business and government through improved operational processes and productivity. Examples abound of new found opportunities — in supply chain logistics, saving time and money; through more cost effective infrastructure and machinery maintenance and planning; improved safety and efficiency in aircraft engines; and in the capacity to better respond to and manage emergencies. And data is critical to building the evidence base to underpin incremental improvements, allowing governments and businesses to offer products and services that are more customised, coordinated or timely. The potential value of data is tremendous, but so too is the scope for Australia to forgo much of this value under the misconception that denial of access would minimise risks.

While this Report highlights some examples of where data is already being used to benefit the community, these are the tip of the iceberg of what could be achieved. What is already being done with data overseas is indicative of what is possible in Australia, if only more data could be released for use and the risks managed.

Health data exemplifies the problem

Australia's health sector exemplifies many of these opportunities, to date largely foregone, due to impediments and distrust around data use (box 1). Data from the sector that could be more widely used includes:

- broad level performance data on expenditure and activity at particular medical facilities (the number of available public and private hospital beds by state and territory) for particular medical conditions (the number of people diagnosed with asthma in each of the past 20 years and public expenditure on particular types of asthma treatment);
- finer level performance data on particular parts of the sector (the number of serious complications following orthopaedic surgery at each hospital, or how drugs prescribed for particular medical conditions vary across medical practitioners);
- data that relates to the health records of individual patients (documented reasons for visits to health professionals, the results from diagnostic testing undertaken, prescriptions received, private and public health insurance claimed); and
- data collected through personally controlled devices, such as smartphones and health monitors, that have an increasing potential to assist medical practitioners and patients themselves.

From the Commission's experience with its annual *Report on Government Services*, data that allows performance monitoring and comparison of government activities is a fundamental starting point for improving the delivery of those activities to the community. While data in that publication motivates a closer examination of practices within particular sectors and jurisdictions, the highly aggregated level limits its use by governments, businesses and the community in making better informed decisions about health products and services. Yet behind many of these thousands of aggregated data points are datasets, the equivalent of which capable, trusted researchers in nations — the United States, New Zealand and the United Kingdom — can and do actively analyse to enable discovery and solution to seemingly intractable problems. And in that context, we fall short.

Inquiry participants highlighted a range of health sector data that could underpin substantial long lasting benefits for the Australian community.

Using data to anticipate and prepare for community and individual health needs

Health data can help policy makers and researchers to:

- identify emerging health issues within communities and factors that contribute to particular medical conditions;
- assess the safety of pharmaceuticals and other treatment options on an ongoing basis; and
- evaluate the effectiveness and efficiency of health policy.

Box 1 Australia's health data — an underutilised resource that could be saving lives

Due to a multitude of legal, institutional and technical reasons, Australia stands out among other developed countries as one where health information is poorly used (OECD 2015c):

> The health sector is very good at generating and storing data. It is less effective at translating this data into useful information. It is poor at linking and sharing information between health professionals, where it could be used to improve health outcomes and system efficiency. Worst of all is the health sector's ability and willingness to share data with consumers (Medibank Private, sub. 98. p. 2).

The implications of this situation are significant. At the individual level, patients are required in many cases to act as information conduits between the various health care providers they see. Inadequate information can lead to errors in treating patients (Joint Council of Social Service Network, sub. 170). At the system level, inefficient collection and sharing leads to data gaps and unnecessary expenditure:

- [H]ealthcare providers largely operate in disconnected silos, hindering continuity of care. Doctors often do not know what medications and tests have been given to patients by other doctors, even when they are members of the same care team. It is even more difficult to bring relevant medical knowledge to the point of care, to create integrated care plans, to monitor a patient's progress against the care plan, or to alert care providers when a patient's condition requires intervention. (Georgeff 2007, pp. 6–7)
- A Parliamentary Committee in Western Australia reporting on data portability problems at one hospital stated "the Health Services Union indicated that the ICU CIS was not compatible with the systems in use on the general wards. According to the HSU, this meant that patient's records must be printed and scanned when they transfer from the ICU to a general ward." (Education and Health Standing Committee (Western Australia) 2015, p. 23)

Furthermore, the lengthy approval process for researchers requesting access to personal data limits their ability to make potentially life-saving discoveries:

- Nearly five years after requesting the data, researchers at the University of Melbourne received de-identified information about CT scans and cancer notifications. Their work showed there was an increased cancer risk for young people undergoing CT scans, and led to changes in medical guidelines for the use of scans. "Had [the] study been approved sooner, and been able to proceed at an earlier date…, we would have had results sooner, with potential benefits in terms of improved guidelines for CT usage, lesser exposures and fewer cancers" (John D Mathews, sub. 36, p. 13).
- Since 2008, the Australian Research Council and other government bodies have been providing funding to the Vaccine Assessment Using Linked Data Safety Study. Among other objectives, this study examines whether there is a relationship between vaccination and admission to hospital or death. The study requires data from both the Australian and State Governments. Obtaining data from the Australian Government has taken six and a half years; state data has not yet been linked. According to Research Australia (sub. 117), linkage is expected to occur in late 2016, eight years after the project commenced.

Electronic health records, for example, could incorporate and use data from monitoring devices to help to identify patients most likely to benefit from particular interventions, and predict those patients whose condition is likely to worsen (which would allow for targeted interventions by healthcare providers).

In the UK, administrative hospital records linked (via unique patient health service number) with a number of cancer screening registries have been used to improve how and when cancer is diagnosed (to increase early detection and survival). Undertaking similar analysis in Australia would require linking of data held by a range of groups, including data from Medicare Australia, the Australian Government Department of Health and its counterparts in the states and territories, various cancer registries and other organisations.

There is already strong support for using Australia's health data in research. A recent survey revealed that over 90% of Australians were willing to share their de-identified health data to advance medical research and improve patient care (Research Australia 2016). Yet more effective use of data is not being sufficiently enabled. Inquiry participants noted a wide range of further medical advances and health sector transformations that could be made possible through the linkage of administrative data with large scale health data collections (such as the 'Busselton Health Study', the 'Australian Atlas of Healthcare Variation', and '45 and Up'), and private sector health insurance data.

Data that allows improved service provision

Inquiry participants flagged the potential for data relating to health service provider costs and performance, as well as de-identified linked data about health service recipients, to be used for more effective and targeted service interventions and improved health outcomes.

The New Zealand Treasury has used longitudinal data from anonymised linked administrative datasets (in this case, mental health program usage and pharmaceuticals) to identify young people who are at risk of poor outcomes in adulthood. By identifying a number of key characteristics that appear predictive of poor future outcomes, the analysis provided valuable insights into the effectiveness of various policies and interventions. The separation of data holdings across three levels of government and across different agencies within each of these jurisdictions, and the distrust that inhibits sharing of this data for linkage purposes, means that such analysis is not yet feasible in Australia.

Yet opportunities are emerging. Integrating data across clinical systems is becoming feasible with greater adoption of electronic health records in Australia. This would enable more effective and holistic healthcare for patients who receive treatment from a range of healthcare providers. While some duplication of diagnostic processes may be necessary for certainty or for alternative treatment plans, roughly 10% of pathology and other tests have been found to be unnecessary duplicates (CBO 2008). As using data to alert practitioners to duplicate radiology tests has been estimated to reduce the number of tests by up to 25% and test waiting time by up to 50% (Chaudhry et al. 2006), there are substantial gains in service efficiency to be had from reducing duplicative effort and integrating health data.

To allow new services to emerge in response to community demand and compete with existing product offerings, potential providers need geographic information on current use of health services. The Australian Dental Association highlighted that access to private health insurance data could allow for new dental practices to be established in areas of high demand.

Data that empowers individuals in managing their use of health services

Patient access to their own medical history (wherever they are, instantly) would not only improve professionals' knowledge of their patients' medical condition and reduce the number of diagnostic tests, but enable the ready and secure sharing of health information to other healthcare providers.

Some private sector services are already developing in Australia to allow consumers to manage their health data. Health&, for example, allows consumers to manually input and store their health data, including medical records and data from fitness devices, in a centralised location to allow better preventative health care and simpler sharing of health information between health service providers. How much more efficient and less error-prone would such transfer be if this could be done at a key-stroke? And it can, but not in Australia. That such services exist, even though they rely on manual rather than electronic input of information, is indicative of the appetite of some consumers for more control over the management of their own health data.

The risks are real but manageable

Allowing and enabling data to be available and used more widely would provide enormous benefits, but there are risks involved. These risks vary with the nature of the data holding, and the environment and purpose for which it is used. Release of aggregated data on government regulatory activities, for example, may pose a very low risk of adverse consequences. Release of data that identifies individuals who have attended a particular medical facility could, in contrast, be highly detrimental to both the individuals concerned and the reputation of the facility. Thus, the risk of harm needs to be assessed based on both the likelihood and scale of harm associated with data being more widely available. Where the adverse consequence of increased data access are considered high, the availability of the data needs to be carefully managed.

The types of risks that Inquiry participants pointed out as being most significant — related to the potential to identify persons or businesses within datasets — were:

- discrimination
- loss of control over the boundaries around the 'you' that the world sees
- reputational damage or embarrassment
- identity fraud
- other criminal misuse of the data
- commercial harm.

That these risks exist is undeniable, but it is important not to fall victim to fear. Some, indeed most, apply to every form of data management, including pen and ink. Identity theft affected 126 000 Australians in 2014-15 (ABS 2016e). Most personal information used in identity theft is obtained online, either through theft, hacking or from information sent by email or

placed on a website, rather than through data release or sharing. Some victims have suffered financial losses; others have reported being refused credit or accused of a crime.

Risks of identification can increase with the linkage of separate pieces of data about an individual. Matching data across individuals can also reveal more information about the activities and associations of those individuals.

These risks — and the desire for privacy and confidentiality — should not be downplayed or trivialised. They are real and important. But, many of them are able to be managed with the right policies and processes. The likelihood of unintended or inappropriate release needs to be carefully considered alongside the likelihood of any genuine harm or costs to the individuals or organisations concerned. Systems and processes can and should be developed to identify, assess, manage and mitigate risks related not just to data release and sharing, but also data collection and storage. Where it is not possible to reduce risks to an acceptable level, the approach being advocated by the Commission would not support release of the data.

Even with data that has never been about individual persons or businesses — such as data on the use of publicly funded facilities — increased availability can come with risks of misuse, where the quality, meaning or context of the data is not understood by users.

Giving data away

Australians give away a lot of personal information online (figure 2). For many, the information gate is (often consciously) wide open. In innumerable ways, individuals deliberately or inadvertently provide information about themselves for one purpose, which then is, or has the potential to be, used for other purposes.

- Some 68% of Australian Internet users have a social media profile, with one quarter accessing their account more than five times per day. The most popular of these sites, Facebook, soaks up information from users' computers and uses it to earn 96% of its revenue through targeted advertising. Only 12% of Internet users avoid social media for security or privacy reasons.
- Similarly, around 84% of Australians are enrolled in at least one customer loyalty program — with an average of 3.8 program memberships. While 47% recognised that a primary reason for loyalty programs is data collection by the company, less than 2% were concerned about their privacy or felt the business knew too much about them.
- Australians have a relatively big appetite for technologies that generate or collect data (we are typically early adopters). For example, at 13% of the population, Australia has the second highest take-up rate of fitness band devices in the world. Wearable technologies, such as Fitbits, transfer data on the physical wellbeing and location of individuals back to the device provider and may be reused by it.

Some 47% of Australians report altering personal information provided online in an attempt to make themselves less identifiable (ACMA 2013a), but ignore the fact that fragments of correct information on them from a wide variety of sources are being compared and matched by intelligent algorithms to form a complete and accurate picture of them.

Figure 2 **The risks that Australians take with data**

Source: Directivity et al. (2015); ACMA (2012, 2013a); OAIC (2013a); Sensis (2015)

That privacy is often said to be a concern but individuals still willingly and readily hand over personal information may seem a paradox. Because much of the data that is being generated is a byproduct of other activities, it was once easy for individuals to dismiss it as being of secondary importance. Today, that should not be the case. If you are using a product or service and not paying for it (or sometimes even when you are), then you are the product. This is perhaps most obvious by the 'all or nothing' nature of personal data requested in exchange for typically free online products and services. What you are consuming, how and when you are consuming it, is all being collected as data that is likely of more value to the supplier than whatever it is they are offering you.

Individuals typically have less choice about providing personal information to governments and may see a less immediate or personal benefit from doing so. Despite claims of a few privacy advocate groups, this Inquiry has not been presented with evidence to suggest widespread concern about the provision of personal information to governments. Indeed, the Office of the Australian Information Commissioner has found that 70% of Australians trust governments in the handling of personal information (only health service providers and financial institutions were rated higher). If individuals do have concerns about provision of personal information to governments, we would welcome hearing these

views for further consideration in the Inquiry final report. The Productivity Commission website is established to receive comments (http://www.pc.gov.au/inquiries/current/data-access/make-submission#lodge).

Increasing data use does not necessarily increase risk

In reality, most risks of data misuse arise not through the public release of robustly de-identified data, but rather from poor or outdated data collection, storage and management practices, often coupled with malicious intent to gain access and use data that would otherwise not have been available. The other avenue made possible by increased online activity is misuse of personal information that individuals have readily made public, to access other information that is not public (essentially a form of identity theft). As the value of data rises, the incentives for such exploitation rise, underscoring the need for all data collectors to remain vigilant and up-to-date in technology around data collection, handling and de-identification.

Given these sources of risk, the main factors stopping breaches of privacy are safeguards around data handling — prompted by the desire of most large private sector data holders, ethics committees, and public sector data custodians to maintain trust and reputation — and inaction on dataset release to avoid potential legal recrimination, given profound uncertainty about privacy and secrecy requirements.

Yet this inaction not only denies discovery (and perhaps innovation), it also takes no account of incentives — for example, there is a profound lack of interest amongst most researchers in government and academia in identifying particular individuals from large datasets; for them, de-identified datasets about large *groups* of people hold the answer to many pivotal questions.

That most data breaches are inconsequential and go largely unnoticed is hardly the point. It only takes one major breach to destroy public confidence. But tightening privacy legislation will not prevent human error and is, at best, a small disincentive to criminal intent.

Greater use of data does not mean Australians should be put at greater risk of harm. In fact, it is vital for Australia's data future that the risks of data handling are managed. The re-identification problems with the recent release of de-identified linked MBS-PBS data underscore the clear need for a robust framework, including expert technical support, for the management of Australia's high value public interest datasets. Against the background of an ocean of personal data that is already public, there remains a need for continued community acceptance and trust in the handling of personal data by governments and business. Built through genuine safeguards, meaningful transparency, and effective management of risk, such acceptance and trust will be vital for the implementation of any reforms. This should be the overarching objective of any reform agenda.

Fundamental change is needed

The legal and policy frameworks under which public and private sector data is collected, stored and used (or traded) in Australia are ad hoc and not contemporary. The impetus for changes in governance structures around data — changes that deal head-on with the fact that data is increasingly digital, revealing of the activities and preferences of individual people or businesses, and held in the private sector — will not diminish. It is a global movement and, to its detriment, Australia is not participating.

Tweaking existing structures and legislation will not suffice. Rather, fundamental and systematic changes are needed to the way Australian governments, business and individuals handle data. This conclusion is based on a number of findings:

- The nature of data sources and data analytical techniques are evolving rapidly and moving away from any effective control by individuals, and will continue to do so — doing nothing is no longer an option.
- As data standards and metadata improve, digital data will be able to be transferred across the economy, between sectors and across national boundaries with increasing ease. To ensure coverage is comprehensive and understandable, data management frameworks need to be consistent across the economy.
- Incremental changes in the data management framework to date have failed to deliver a culture of making data available for widespread use. The range and volume of datasets now held in the public or private sector, that *could* potentially be made more widely available and the associated opportunities are monumental. While there have been noticeable increases in the sharing and release of certain data in recent years, these releases remain a 'pimple on the pumpkin' of data release possibilities.
- There are key unanswered questions that go to the fundamental rights of individuals to data held about them, and how individuals can use data more effectively for their own benefit, that lie at the heart of data availability and use. These questions necessitate an across-the-board rethink of the way data is managed.

The Commission's recommended approach incorporates recent progress in policy and practices around data management but is deliberately aimed at creating a new, comprehensive framework that should, by design, be capable of enduring beyond current policies, personnel and institutional structures. It takes account of the significant differences in data types and associated risks and uses of each, and recognises that while the incremental risks of making data more available might appear very small (given how much data is already in the public domain), incentives and trust nevertheless have to be maintained. Crucially, the proposed reforms take Australia beyond the stage of viewing data availability solely through a privacy lens, in recognition that there is much more than privacy at stake when it comes to data availability and use.

The new Framework

A key issue in balancing access and trust is consideration of the level of data required for different uses. Near real time data that identifies individual persons or businesses carries the highest risks to privacy and security. Access to this level of data by those other than the parties to a transaction — while useful for the enforcement of some regulations (for example, traffic speed limits) and for inducing timely changes in consumer behaviour (for example, price responsive household electricity consumption) — is not necessary in order to obtain much of the benefits of data use. For analysis of market opportunities, scenario development, policy evaluation or improved delivery of many products and services, de-identified data can be sufficient, and indeed, desirable. And, of course, there is considerable data that is non-personal and non-confidential, which also needs to be made more accessible for use and reuse.

The Commission's Framework (figure 3) recognises the spectrum of risk associated with different types and uses of data, and the corresponding need for different risk controls and approaches to apply.

Where the risks associated with release cannot be effectively mitigated, the Commission's approach would not involve release in the case of *genuinely* commercial in confidence data, or data that is integral to the security of the country. For the remaining bulk of data, the recommended approach to improving sharing or release is detailed below, and reflects a sliding scale of release strategies and controls commensurate with the potential risks and benefits of potential release.

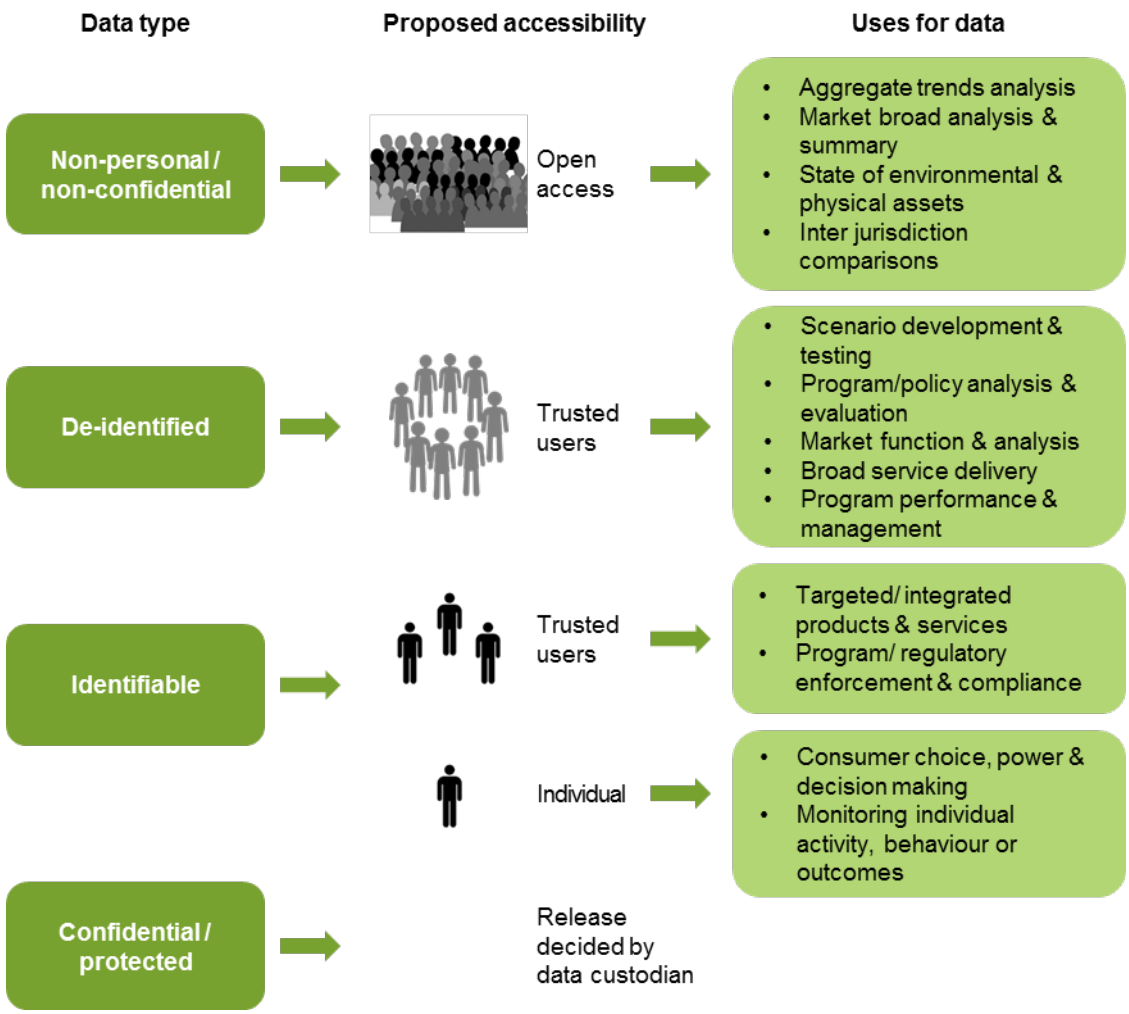
Broad criteria shaping the Framework are that it must: deliver net benefits to the community; increase the availability and usefulness of data; engender community trust and confidence in how data is managed and used; and preserve commercial incentives to collect, maintain and add value to data.

There are four key elements to the recommended approach, consistent with the Inquiry terms of reference, and underpinned by changes to legislative and governance structures:

1. Giving individuals more control over data held on them
2. Enabling broad access to datasets (public and private sector) that are of national interest
3. Increasing the usefulness of publicly funded identifiable data amongst trusted users
4. Creating a culture in which non-personal and non-confidential data gets released by default for widespread use.

Each of these elements is discussed in detail below.

Figure 3 **Framework of the recommended approach**



New legislation and governance structures for data access

Although this Inquiry would have preferred to find solutions that are non-regulatory, it is a clear conclusion that legislative changes are needed to implement the Commission's recommended reforms. These primarily involve changes to existing Commonwealth privacy legislation as well as the creation of new legislation — a new Data Sharing and Release Act — to facilitate data sharing and release. This Act would be a Commonwealth piece of legislation *applying across Australia to all digital data*. It would therefore be 'umbrella' legislation, and would make redundant some clauses in other dataset-specific and program-specific legislation around privacy, secrecy and other matters. Apart from giving entities 'permission' to publicly release data while managing risks, it forms the basis for a new lens through which to view data availability and use: the lens of a valuable asset being created, not merely a risk or an overhead.

Further consideration will be particularly needed in regard to the interface between the new Act and the Privacy Act. A primary intention would be to retain the key protections within the latter legislation, particularly as they apply to the use of personal information, whilst also ensuring that the new Act facilitates a more open and effective approach to data management.

Implementing this legislated Framework would be a central government agency with data responsibility, a new national statutory office holder — the National Data Custodian (NDC) — and a suite of sectoral Accredited Release Authorities (box 2). These Accredited Release Authorities (ARAs) would be funded and tasked with assisting data custodians to improve the curation and quality of datasets to be released (including de-identifying data where necessary), and facilitate timely updates and ongoing dataset maintenance.

The ARAs would ultimately, on advice from original dataset custodians, be responsible for deciding whether a dataset is available for public release or limited to sharing with 'trusted users'. Trusted users would *potentially* include any individual working in an entity that is covered by privacy legislation, with necessary governance structures and processes to address the risks of inappropriate data use or release associated with particular datasets, including access to secure computing infrastructure. Trusted user access to higher risk identifiable data would necessarily be more tightly controlled, with more stringent requirements on the means and purpose of data use to be satisfied before the ARA could grant access.

In addition to these new structures, additional roles may be afforded in governance to the Office of the Australian Information Commissioner, the Australian Competition and Consumer Commission, the Administrative Appeals Tribunal, and relevant industry ombudsmen. Bodies such as the Australian Bureau of Statistics, Australian Institute of Health and Welfare, Sax Institute and CSIRO Data61 may also take on additional technical advice and support roles with regard to data linkage and integration, de-identification, and safe sharing and release.

Giving individuals more control over their data

Australian consumers have little capacity to choose how personal data about them is used; and too often, organisations and governments make decisions about the use of individuals' data on behalf of the individuals concerned. In the face of the ubiquity of data collected, the scope to provide consumers with a greater say — within limits — on the handling of data that is sourced from them, is considerable.

Box 2 Key institutions and roles in the new Framework

The Commission's recommended reforms require government to take up important new functions to enable the opportunities from data to be realised. This will require the establishment of a new national position, and the authorisation, through the new Data Sharing and Release Act, of some additional functions by existing institutions.

The National Data Custodian (NDC)

The new position of the NDC creates a role that parallels for data access and use, the role of Australia's Information Commissioner for data protection. The NDC will have responsibility for broad oversight of the operation of the national data system, be involved in designation of datasets of national interest, potentially with designation by disallowable instrument. The NDC would also accredit release authorities and trusted users within the reformed data system.

Accredited Release Authorities (ARA)

ARAs will largely be existing public sector agencies (Australian Government or state/territory government) that already release data but would now be funded to take on additional responsibilities as an ARA (the Australian Institute of Health and Welfare may be a workable model for an ARA). ARAs would play an important role in deciding whether a dataset is available for public release or limited sharing with trusted users, curating datasets and assisting dataset custodians with curation and the development of metadata, ensuring the timely update and maintenance of datasets, and supporting the linkage of national interest and other datasets for release. Given the emphasis on sectoral expertise, these entities would have a long track record of trusted data management in their particular areas of focus. It is envisaged that ARAs would also perform an important advisory role on technical matters, both to government, and to the broader community of data custodians.

Other existing institutions with additional roles

Existing institutions with important new roles in the reformed data management system include:

- Australian Competition and Consumer Commission — oversee the Framework around consumer access to data
- Office of the Australian Information Commissioner — continue its fundamental role regarding privacy and handling of privacy related data complaints
- Administrative Appeals Tribunal — possible role to assess disputes and appeals regarding data sharing and release
- Australian Bureau of Statistics, Australian Institute of Health and Welfare and CSIRO Data61 — key advisory and support roles, in addition to their existing functions within the data infrastructure.

Control and access to personal information can help encourage information sharing because it builds individuals' confidence that their personal information will be used in a way that reflects their preferences. Increased access and greater control over their ability to use data collected about them also affords individuals more choice about the products and services they consume (and the providers of those), and is an avenue to improve market competitiveness. Currently, it is a business or service provider that determines how to extract value from an individual's data. We propose to explicitly create the opportunity for

individuals also to choose to trade or reuse their data. While consumers arguably already have some access power, there are severe practical constraints in Australia at present, on how to exercise this. But this need not be the case.

A new right to data access

Under the Commission's proposed plan, consumers would retain the existing ability to view what information a business or government agency holds on them and request edits or corrections (related to accuracy). The capacity to have data edited would be a right to *request* specific edits, not a right to compel data holders to change their datasets unless incorrect; however, it would generally be in the data holder's interests to ensure data holdings are as accurate and up-to-date as possible. This design feature is necessary to ensure that inaccurate corrections are not made: the Wikipedia experience is best not applied to health data.

Consumers would also have a right to be informed of disclosure of data by a data holder to third parties; and a right to appeal automated decisions, such as those based on statistical profiling. *The Commission's recommendation increases powers of individuals and formally defines them as a 'right', while maintaining safeguards*. This access right would be enforceable in the same way that existing powers are — via complaint to the Office of the Australian Information Commissioner (OAIC) or the relevant industry ombudsman.

Under the Commission's recommended approach, consumers would also have an explicit new right to require that a data holder stop collecting information on them (that is, they can 'opt out' of a collection process). This capacity to opt out at any time would be subject to a number of exceptions, including that individuals would not be able to have collection cease if the collected information is necessary for public benefit purposes (such as the maintenance of public health and safety, or administrative purposes such as tax collection) or forms part of a National Interest Dataset (described later).

In the private sector, opting out of data collection may well mean that a particular product or service is no longer available or no longer free to that consumer. But consumers must, in the ever-expanding world of data opportunities, be able to make that call for themselves.

Nor would the right to stop data collection include having historical data deleted or use of it cease. This provision recognises that once data is integrated into a dataset or analysis, it can be costly or infeasible to extract it, and on occasion, damaging to the interests of others using the dataset. It is also intended to ensure that individuals' opt out decisions do not decimate the investment that data holders have made in datasets and, in some cases, ensure that information on past activity is available to inform future activity (information on past medical procedures of an individual would be necessary for future medical treatment, for example).

More scope for individuals to use data about themselves

The new Comprehensive Right of individuals over their data would extend to include the ability to direct that a copy of their data be transferred safely from one data holder to another. This is a key additional power afforded to individuals under the new Framework.

The capacity for individuals, as consumers, to copy their data between service providers is an integral part of facilitating competition in markets and reducing barriers to market entry. In some circumstances, the consumer may see benefits in having a copy of their data provided to an entity that is not a competitor (for example, provision of medical records to a life insurance company or provision of utility payment information to a credit provider). In other cases, it will be to form a new customer relationship, or obtain a quote that may lead to one, at the consumer's discretion.

Underlying this right, and to maintain incentives for existing data holders and collectors, is the idea that the right to data is a *joint right*, shared between the individual and the businesses or agencies that hold the individual's data. The individual's decision to switch service providers would not alter the right of the initial service provider to retain the data that they had already collected while providing a service to the individual.

All businesses, government agencies and government business enterprises should be subject to this new Comprehensive Right. In some sectors, transfer of data may be achieved by the use of application programming interfaces; in others, the transfer of files. Either way, standards around data formats and definitions will be necessary. We consider that participants in each sector, rather than governments, are best placed to develop these standards. But we propose a process to achieve this that draws on existing standards development practice in Australia.

Any charges levied by data holders for access, editing, copying and/or transferring of data should be monitored by the Australian Competition and Consumer Commission (ACCC), with the methodology transparent and reviewable on request by the ACCC.

While the changes proposed aim to enable consumers to exercise more control over the collection and use of data on them, the onus remains on individuals to make responsible choices about whom they provide personal information to in the first instance and for what purposes.

Comprehensive credit reporting

In some circumstances, collating consumer data may offer net public benefit in making markets more efficient. A specific case is covered in the terms of reference for this Inquiry: comprehensive credit reporting. The Productivity Commission has previously found comprehensive credit reporting to be desirable and, consistent with the approach of New Zealand, the United Kingdom and the United States, a voluntary approach to data input

should continue to be pursued, unless it becomes clear that a critical mass of accounts is not achievable on that basis.

Broad access to datasets of national interest

We have given considerable thought to establishing an element in the Framework to enable wider access to high value, National Interest Datasets. The intention is to promote the development of a valuable suite of datasets — some of which are released publicly; others that will be shared with a smaller group of trusted data users. Designating datasets as national interest collections will also signify their value as resources collected in the national interest, not merely (as today) for compliance, record-keeping or audit.

The term national interest in our Framework necessarily covers data with a significant public interest element that is collected by Australian, state/territory or local government agencies and publicly funded research bodies. In specific cases, private sector data deemed to have a public interest element will also form part of a National Interest Dataset — private health insurance data is a potential example.

Governments across Australia hold enormous amounts of data, but lag behind other comparable economies by typically not exploiting it beyond the purposes for which it was initially collected, nor allowing others access to do so. Australia's private sector data holders are more innovative in their use of data. But even so, the extent to which their data holdings are available more broadly — when data are collected through public funding or to meet a public interest objective — remains constrained by limitations on data linkage, ad hoc frameworks to facilitate release and commercial incentives. Wider availability of such data (public and private sector) would likely trigger significant investment and improvements in national welfare.

The extent to which use of a particular dataset could provide benefits to the broader community (beyond those derived by just the data holders and data contributors), is important in considering how widely available it should be made — that is, whether the objective is ultimately to release the data, or to share it more effectively.

While there are some important and obvious initial examples of likely National Interest Datasets (such as those that provide registers of businesses, services or assets; or those that record activity in key areas of public expenditure), others may be less immediately obvious but will become clear candidates over time. This element of the Framework is designed to allow such evolution, with public scrutiny in each case.

Datasets with national interest characteristics may be identified: in prior research or program evaluation within the relevant sector; through use of datasets with comparable features or circumstances in other sectors or overseas; or may be inferred from the interest in or demand for access to particular datasets over an extended period.

To be a valuable resource, the suite of National Interest Datasets must extend beyond the low hanging fruit of spatial data and aggregated activity data to include access to de-identified datasets that are integral to service delivery and decision making, as well as key privately held datasets. The Framework established is intended to promote the inclusion of such data to enable its broader use, while not dis-incentivising data collection and value adding activities.

Extensive community and stakeholder consultation is expected to be an important aspect of the dataset designation process. To enhance community consultation in the process and ensure ongoing input, a deliberative forum — a parliamentary committee, in our current thinking — could be established to review nominations made and the level of access granted, and make proposals for future designations. A mechanism of this kind would ensure that there is detailed consideration of the existing pool of datasets from which nominated sets can be drawn. It would also open to public scrutiny arguments against designation.

Why designate a dataset?

The Commission's recommendations provide a Framework for public and private datasets to be nominated and designated as National Interest Datasets (NIDs). For those datasets that are so designated, all restrictions to access and use that are contained in a variety of national and state/territory legislation, and in other program-specific policies, would be replaced by new access and use arrangements under the proposed Data Sharing and Release Act (enabled in the states and territories by the Australian Government's powers under section 51(v) of the Constitution). This would ensure ongoing dataset curation as a national asset, substantially streamline access to the dataset and, where relevant, enable linkage to other datasets.

The process to designate datasets as NIDs should be open to the states and territories and to private sector entities, to allow them to similarly benefit from having their data curated (to the extent that is in the public interest) and accessible under the new Act, and more readily allow linkage of their data with other datasets. Where states and territories opt in to have datasets designated as NIDs, separate state/territory legislation may be required to enable release of data held by state government bodies and some unincorporated entities.

Having a system for identifying and funding the ongoing maintenance of national interest data assets would help create consensus and cooperation between sectors and between the Commonwealth and the states and territories. This would build on existing work at COAG to identify a spine of essential public sector assets.

A listing of all NIDs that have been publicly released or are potentially available to share, the relevant dataset custodian and ARA for that dataset, and a contact point, would be included on a central website, such as on data.gov.au. This would enable potential users of these to know of the dataset's existence and how to gain access to it.

What would access look like?

Under the Commission's Framework, a valuable suite of datasets would be developed. At the discretion of the ARA, these datasets would either be released publicly or, alternatively, shared with all Australian and state and territory government agencies and other trusted users, under rules to be developed by the NDC.

The approach represents a marked expansion in data access in Australia that would provide significant opportunities for research and innovative market development and improve delivery of public services.

In contrast to existing arrangements for access to significant datasets, the approach recommended aims to expand:

- the availability and quality of that group of high value datasets — in the private and public sector, across all levels of government — that are of national interest
- the range of data users that would be considered 'trusted' to access de-identified data
- the types of uses to which the data can be put — by allowing unlimited use of data that is not about individual persons or businesses, and approved access to de-identified data to be ongoing (not project specific) and limiting use only where the risks of re-identification cannot be effectively managed.

The special case of higher risk data shared with trusted users

Some publicly funded data that identifies individual persons or businesses is already shared in a very limited way with trusted users within government and/or the research community. This data is typically used for targeted program and product/service delivery, regulatory compliance, and for research (such as rare medical conditions) where there are very small populations involved.

The current process is, however, costly to data custodians, those who endeavour to gain access to the data, and also for the public, who ultimately fund the activities for which the data is used. Depending on the particular dataset, access requests (even from within the same government) can require separate and duplicative agreement of multiple dataset owners, custodians and stewards, integration units, ethics committees, other advisory bodies, and the individuals about whom the information was collected. Each policy and approval step is intended to ensure privacy and confidentiality are maintained, but in combination they create major obstacles to data access.

The Commission recommends streamlining access to identifiable data within and between Australian governments, and for the limited range of other trusted users with which such data is shared. It is intended that identifiable data that could be shared would include both that collected by, or on behalf of, government agencies and that collected through publicly funded research bodies and projects.

In addition, there is a need for the research community to put its house in order when it comes to data sharing. Just as government data custodians should consider that they hold data not solely for their own purposes but in the public interest on behalf of citizens, so too should the data of publicly funded research be available beyond the initial researchers. And where it is not, much better justification and record-keeping is needed, to at least enable other researchers to learn what data has already been collected.

Key features of the recommended approach include that:

- access be granted on a *project-specific* basis to approved personnel in either Australian or state/territory government agencies and to approved researchers
- projects for which the data could be used would be subject to a pre-approved list of public interest purposes and require approval of the ARA and, if relevant, an ethics committee (but not the data custodian)
  - the NDC, in consultation with dataset custodians, would develop and provide the list of approved purposes for the dataset to an ARA
  - if a project does not satisfy the list of approved purposes, the applicant would be able to apply to the ARA for special access
- existing exceptions on the need to obtain consent of individuals for use of personal information for health and medical research purposes, would be extended to cover public interest research more generally
- access would occur in a specified secure computing environment with output from the dataset reviewed by an automated process prior to project completion to ensure confidentiality requirements have been satisfied
- responsibility for appropriate use of datasets would rest with trusted users, with clear and significant consequences for any breach of this trust, to provide additional incentives for maintaining the security of data and appropriate data use.

To the extent that pre-approved data purposes cover the range of uses to which data will be reasonably applied, transferring final approval from initial data custodians over to the ARA and ethics committees will substantially streamline access to data.

Making other data readily available to all

Governments have proven to be poorly equipped in understanding consumer and business demand for data and in making non-personal datasets, and those that are not genuinely commercial-in-confidence, widely available.

While the reasons for governments' inability to derive value from their data holdings may at times be understandable — governments are not entrepreneurial nor would we necessarily want them to be — they are at other times disappointing. Risk aversion is not desirable where it results in the public interest being poorly served.

There needs to be a shift in emphasis from only releasing data on request for particular projects, toward actively pushing data out in a coordinated way. In principle, all datasets in fields where there are burgeoning opportunities and capability would be opened up and released, as resources and sectoral demand allow.

This would mean that all data that is non-confidential and not related to individual people or businesses would be routinely available for use by governments, consumers, businesses and the research community. This includes information that, while it may identify individuals, is already in the public domain in some form (property ownership, for example). A realistic assessment of the risks associated with public release of identifiable information that is already public in a less accessible form, should be undertaken.

Such an approach has the potential to make a marked difference to the range and volume of data available for decision making, innovative activity and improved service delivery in the community.

The challenges in achieving this should not be underestimated. There is a very real culture of risk aversion and risk avoidance in the public sector when it comes to data release. Changing this will require strong and consistent leadership, backed up by policies that clearly spell out objectives and expectations. In addition, releasing data (and the costs associated with that) presents questions regarding whether and how access should be charged and the extent to which government agencies should or should not seek to add value to datasets.

Simply put, the Commission recommends that governments should adopt a zero or low (genuinely marginal) cost approach for data release, consistent with increasing access and the achievement of public interest benefits related to that. The exception is when a clear and compelling commercial demand and proposition for a particular dataset exists, in which case a more commercial or market-driven approach to pricing may be desirable and consistent with public interest objectives. Similarly, any value adding should only be undertaken where there is a demonstrable willingness and capacity to pay on the part of the user, and where a number of criteria related to agency capability are able to be met.

In other words, it is expected that most released datasets would be curated with metadata and necessary provenance, but otherwise be free, plain vanilla, fit for release and timely.

While release of public sector data would be the focus of governments, it is anticipated that once governments start to more actively push data out, this will encourage private entities to do likewise and to profit from doing so. That is, across the economy the value will shift from being embodied in the data itself, to being derived from the clever analysis and use of data.

Implementation

The staging of reforms will be important — with major reforms such as these, establishment of the Framework in full should be viewed from the outset as a project in need of a firm implementation plan.

Negotiation and consultation will be required with state and territory governments (as significant data holders); some parts of the private sector (for similar reasons); and with sectoral groups where NIDs are sectoral in nature (health or education).

Reforming public sector data access is a strong first step, and Australia's governments need to make significant changes if open government agendas are to catch up with those in competing economies.

The Comprehensive Right for consumers proposed will need discussion and information campaigns to help people use it to the fullest extent. Firms affected will need to be involved. Although some may have doubts, even in the course of this Inquiry's first six months some key firms (a number of banks) have recognised that allowing consumers to discover and realise benefits from their data is a key driver in building community trust.

Implementation of the proposed arrangements for providing access to NIDs and to identifiable data may take slightly longer to progress, will require consideration of transition arrangements for existing data users, and require the finer details with regard to key roles, mechanisms and technical approaches to be bedded down and based on a process of extensive consultation.

Technological developments and solutions may ensure that, over time, some of the changes required to facilitate improved and secure sharing and release of data will become easier to attain.

Yet this is not advice that implies de-coupling of parts of the Framework. In a project that aims to create new opportunity for both public and private benefit, each element supports the others. Removal of one (or more) will imbalance the opportunities and reduce the prospect for broad community acceptance.

Findings and recommendations

Addressing specific impediments to *public* sector data access

DRAFT FINDING 3.1

Australia's provision of open access to data is below comparable countries with similar governance structures — including the United States and the United Kingdom. There remains considerable scope to improve the range of datasets published (and, correspondingly, the diversity of agencies and research bodies publicly releasing data) and the usability of open data portals.

DRAFT FINDING 3.2

Data integration in some jurisdictions (particularly Western Australia and New South Wales) has made good progress in some fields, but highlights a lack of action in equivalent fields at both national and state/territory level, and reveals the large unmet potential in data integration.

DRAFT FINDING 3.3

Despite recent statements in favour of greater openness, many areas of Australia's public sector continue to exhibit a reluctance to share or release data. The entrenched culture of risk aversion, reinforced by a range of policy requirements and approval processes, greatly inhibits data discovery, analysis and use.

DRAFT RECOMMENDATION 3.1

All Australian Government agencies should create comprehensive, easy to access data registers (listing both data that is available and that which is not) by 1 October 2017 and publish these registers on data.gov.au.

States and territories should create an equivalent model where one does not exist and in all cases should make registers comprehensive. These should in turn be linked to data.gov.au.

The central agencies responsible for data should:

- set measurable objectives, consistent with best practice, for ensuring that available data and metadata are catalogued and searchable, in a machine-readable format
- improve accessibility of data for potential data users.

Limited exceptions for high sensitivity datasets should apply. Where they do, a notice indicating certain unspecified datasets that have been assessed as Not Available should be published by the responsible department of state, on the relevant registry.

DRAFT FINDING 3.4

There is a clear public interest in having research-oriented data widely available to trusted researchers in a timely manner. A corresponding presumption that it be released needs to be balanced against a number of potentially competing interests, including:

- the need for the researcher to benefit from their own research
- interests in commercialisation of research — for example, if the research was partly privately funded
- specific legislative or ethics approval restrictions
- privacy or confidentiality considerations
- capacity to provide access through secure sharing environments, where privacy or confidentiality considerations cannot be managed to enable the release of data.

DRAFT RECOMMENDATION 3.2

Publicly funded entities, including the Australian Research Council, should publish up-to-date registers of data holdings, including metadata, that they fund or hold.

Publication of summary descriptions of datasets held by funded researchers but not released, and an explanation of why these datasets are not available, are also essential and would provide far greater transparency about what is being funded by taxpayers but withheld.

Addressing specific impediments to *private* sector data access

DRAFT RECOMMENDATION 4.1

The Australian Government should adopt a minimum target for voluntary participation in Comprehensive Credit Reporting of 40% of accounts. If this target is not achieved by 30 June 2017, the Government should circulate draft legislation to impose mandatory reporting by 31 December 2017.

DRAFT RECOMMENDATION 4.2

All Australian governments entering into contracts with the private sector, which involve the creation of datasets in the course of delivering public services, should assess the strategic significance and public interest value of the data prior to contracting. Where data is assessed to be valuable, governments should retain the right to access or purchase that data in machine readable form and apply any analysis that is within the public interest.

The conundrum of personal data

DRAFT FINDING 5.1

The boundaries of personal information are constantly shifting, in response to technological advances and community expectations. The legal definition of personal information, contained in the *Privacy Act 1988* (Cth), gives rise to uncertainty. This uncertainty will only increase in future, as new technology continues to emerge.

DRAFT RECOMMENDATION 5.1

In conjunction with the Australian Bureau of Statistics and other agencies with data de-identification expertise, the Office of the Australian Information Commissioner should develop and publish practical guidance on best practice de-identification processes.

To increase confidence in data de-identification, the Office of the Australian Information Commissioner should be afforded the power to certify, at its discretion, when entities are using best practice de-identification processes.

DRAFT RECOMMENDATION 5.2

The *Privacy Act 1988* (Cth) exceptions that allow access to identifiable information for the purposes of health and medical research without seeking individuals' agreement, should be expanded to apply to all research that is determined to be in the public interest.

The Office of the Australian Information Commissioner should develop and publish guidance on the inputs required to establish a public interest case.

DRAFT FINDING 5.2

A wide range of more than 500 secrecy and privacy provisions in Commonwealth legislation plus other policies and guidelines impose considerable limitations on the availability and use of identifiable data. While some may remain valid, they are rarely reviewed or modified. Many will no longer be fit for purpose.

Incremental change to data management frameworks is unlikely to be either effective or timely, given the proliferation of these restrictions.

DRAFT FINDING 5.3

Although parts of the government view community expectations as a factor that limits the use of data, reliable surveys have shown that most individuals believe sharing personal information between government departments can be beneficial, and indeed is occurring without damage.

However, individuals expect to remain in control of who data on them is shared with.

DRAFT FINDING 5.4

Large volumes of identifiable information are already published online by individuals or collected by various organisations, with or without explicit consent.

In this context, the incremental risk of allowing increased access to formerly identifiable data by public and private sector organisations, using security protocols and trusted user models, is likely very small.

Breaches of personal data, often enabled by individuals' unwary approach to offering data, are largely dominated by malicious or criminal activity. By comparison, breaches due to sharing or release are far fewer in number and reach.

DRAFT RECOMMENDATION 5.3

The Australian Government should abolish its requirement to destroy linked datasets and statistical linkage keys at the completion of researchers' data integration projects.

Data custodians should use a risk-based approach to determine how to enable ongoing use of linked datasets. The value added to original datasets by researchers should be retained and available to other dataset users.

INFORMATION REQUEST

The Commission seeks further views on the most practical ways to ensure improvements to linked datasets are available for subsequent dataset uses.

DRAFT RECOMMENDATION 5.4

To streamline approval processes for data access, the Australian Government should:

- issue clear guidance to data custodians on their rights and responsibilities, ensuring that requests for data access are dealt with in a timely and efficient manner;
- require that data custodians report annually on their handling of requests for data access;
- prioritise funding to academic institutions that implement mutual recognition of approvals issued by accredited human research ethics committees.

State and territory governments should mirror these approaches to enable use of data for jurisdictional comparisons and cross-jurisdiction research.

DRAFT RECOMMENDATION 5.5

In light of the Australian Government's commitment to open data, additional qualified entities should be accredited to undertake data linkage.

State-based data linkage units should be able to apply for accreditation by the National Data Custodian (Draft Recommendation 9.5) to allow them to link Australian Government data, and the intention of 'open by default' should apply to these exchanges.

Making data more useful

DRAFT FINDING 6.1

The lack of public release and data sharing between government entities has contributed to fragmentation and duplication of data collection activities. This not only wastes public and private sector resources but also places a larger than necessary reporting burden on individuals and organisations.

DRAFT RECOMMENDATION 6.1

Government agencies should adopt and implement data management standards to support increased data availability and use as part of their implementation of the Australian Government's Public Data Policy Statement.

These standards should:

- be published on agency websites
- be adopted in consultation with data users and draw on existing standards where feasible
- recognise sector-specific differences in data collection and use
- support the sharing of data across Australian governments and agencies
- enable all digitally collected data and metadata to be available in commonly used machine readable formats (that are relevant to the function or field in which the data was collected or will likely be most commonly used), including where relevant and authorised, for machine to machine interaction.

Policy documents outlining the standards and how they will be implemented should be available in draft form for consultation by the end of 2017, with standards implemented by the end of 2020.

Agencies that do not adopt agreed sector-specific standards would be noted as not fully implementing the Australian Government's Public Data Policy and would be required to work under a nominated Accredited Release Authority (Draft Recommendation 9.6) to improve the quality of their data holdings.

DRAFT RECOMMENDATION 6.2

The private sector is likely to be best placed to determine sector-specific standards for its data sharing between firms, where required by reforms proposed under the new data Framework.

In the event that voluntary approaches to determining standards and data quality do not emerge or adequately enable data access and transfer (including where sought by consumers), governments should facilitate this, when deemed to be in the public interest to do so.

INFORMATION REQUEST

The Commission seeks more information on the benefits and costs of a legislative presumption in favour of providing data in an application programming interface (API) format, specifically:

- *In which sectors would consumers benefit from being able to access data in an API format?*
- *What are the main costs and barriers to implementing APIs?*

DRAFT FINDING 6.2

Data standards should aim to ensure that the content produced is usable by those who seek access to their own data. To achieve this, available data needs to be published in machine readable and commonly used formats that are relevant to the function or field in which the data was originally collected or will likely be most commonly used.

Valuing and pricing data

DRAFT RECOMMENDATION 2.1

In determining datasets for public release, a central government agency with policy responsibility for data should maintain a system whereby all Australian governments' agencies, researchers and the private sector can, on an ongoing basis, nominate datasets or combinations of datasets for public release, with the initial priority being the release of high value, in-demand datasets.

A list of requested datasets should be published. Decisions regarding dataset release or otherwise, and access arrangements, should be transparent. Agencies should provide explanations where priority datasets are not subsequently released on legitimate grounds. Where there are not legitimate reasons for withholding requested data, remedial action should be undertaken by the Australian Government's central data agency to assist agencies to satisfy data requests.

Existing government data initiatives, such as data.gov.au, should be leveraged as part of this system.

DRAFT RECOMMENDATION 7.1

Beyond achieving a 'fit for release' standard (Draft Recommendation 6.1), government agencies should only value add to data if there is an identified public interest purpose for the agency to undertake additional value adding, or:

- the agency can perform the value adding more efficiently than either any private sector entities or end users of the data; and
- users have a demonstrable willingness to pay for the value added product; and
- the agency has the capability and capacity in-house or under existing contract; and
- the information technology upgrade risk is assessed and found to be small.

DRAFT FINDING 7.1

There is no single pricing approach that could act as a model for guiding public sector data release decisions. The identification by agencies of the grounds for undertaking each release will have a direct bearing on the choice of price approach. Cost recovery, long considered to be the default option in the public sector, is only one of a range of approaches and not necessarily to be preferred.

DRAFT RECOMMENDATION 7.2

The pricing of public sector datasets to the research community for public interest purposes should be the subject of an independent review.

DRAFT RECOMMENDATION 7.3

Minimally processed public sector datasets should be made freely available or priced at marginal cost of release.

Where there is a demand and public interest rationale for value-added datasets, agencies should adopt a cost recovery pricing approach. Further, they should experiment with lower prices to gauge the price sensitivity of demand, with a view to sustaining lower prices if demand proves to be reasonably price sensitive.

DRAFT RECOMMENDATION 7.4

For datasets determined through the central data agency's public request process (Draft Recommendation 2.1) to be of high value and have a strong public interest case for their release, agencies should be funded for this purpose. Funding should be limited and supplemental in nature, payable only in the event that agencies make the datasets available through release or sharing.

Aside from this additional funding, normal budgetary processes should apply for all agencies' activities related to their data holdings.

Fundamental reform is needed

DRAFT FINDING 8.1

It is important governments and businesses maintain a social licence for their collection and use of data. This can be built through enhancement of consumer rights, genuine safeguards, transparency, and effective management of risk. Community trust and acceptance will be vital for the implementation of any reforms to Australia's data infrastructure.

DRAFT FINDING 8.2

There is no shared vision amongst public sector data holders in Australia on how to consistently deliver widespread data sharing and release. The community — current and future — is entitled to expect such a vision. Comprehensive reform of Australia's data infrastructure is needed to signal that permission is granted for active data sharing and release and that data infrastructure and assets are a priority. Reforms should be underpinned by:

- clear and consistent leadership
- transparency and accountability for release and risk management
- reformed policies and legislation
- institutional change.

DRAFT FINDING 8.3

By applying a risk-based approach to data access, government agencies can establish a sound basis for where further risk mitigation effort is necessary and for moving early to the sharing or release of low risk data, while building and retaining the trust and confidence of users and the wider community.

DRAFT RECOMMENDATION 9.1

The Australian Government should introduce a definition of consumer data that includes:

- personal information, as defined in the *Privacy Act 1988* (Cth)
- all files posted online by the consumer
- all data derived from consumers' online transactions or Internet-connected activity
- other data associated with transactions or activity that is relevant to the transfer of data to a nominated third party.

Data that is transformed to a significant extent, such that it is demonstrably not able to be re-identified as being related to an individual, should not, for the purposes of defining and implementing any Comprehensive Right, be defined as consumer data.

The definition of 'consumer data' should be provided as part of a new Act regarding data sharing and release (Draft Recommendation 9.11). Given the need for this definition to have broad applicability, it should also be included within the *Acts Interpretation Act 1901* (Cth). Consequential amendments to other Commonwealth legislation would ensure harmonisation across federal laws.

INFORMATION REQUEST

Further views are sought on the effects of providing access to consumer data, as defined. In particular, views are sought on the potential creation of incentives for deliberate de-identification of data holdings to avoid providing access, and whether effective and low cost remedies to such behaviour could be introduced.

DRAFT RECOMMENDATION 9.2

Individuals should have a Comprehensive Right to access digitally held data about themselves. This access right would give the individual a right to:

- continuing shared access with the data holder
- access the data provided directly by the individual, collected in the course of other actions (and including administrative datasets), or created by others, for example through re-identification
- request edits or corrections for reasons of accuracy
- be informed about the intention to disclose or sell data about them to third parties
- appeal automated decisions
- direct data holders to copy data in machine-readable form, either to the individual or to a nominated third party.

Individuals should also have the right, at any time, to opt out of a data collection process, subject to a number of exceptions. Exceptions would include data collected or used as:

- a condition of continued delivery of a product or service to the individual
- necessary to satisfy legal obligations or legal claims
- necessary for a specific public interest purpose (including archival)
- part of a National Interest Dataset (as defined in Draft Recommendation 9.4).

The right to cease collection would not give individuals the capacity to prevent use of data collected on the individual up to the point of such cessation.

INFORMATION REQUEST

The Commission seeks views on what methods of disclosure would be most likely to result in consumers making a meaningful choice about how their personal information is being used, and how these disclosure requirements might best be implemented.

DRAFT RECOMMENDATION 9.3

The Australian Government should provide for broad oversight and complaints handling functions within a reformed framework for individual data access. Key roles should be accorded to the Australian Competition and Consumer Commission (ACCC) the Office of the Australian Information Commissioner (OAIC), and to existing industry ombudsmen.

Any charging regimes, policies or practices introduced to address costs associated with data access, editing or transferability should be transparent and reasonable. The ACCC should be responsible for monitoring and assessing the reasonableness of charges applied. The ACCC, supported by state and territory Fair Trading Offices, should also educate and advise consumers on their new rights in regard to data access and collection.

For specified datasets (such as in banking) the relevant ombudsman scheme would need to be expanded to deal with disputes.

INFORMATION REQUEST

The Commission seeks further views on datasets that are of national interest and that could feasibly be designated as such under the process proposed.

DRAFT RECOMMENDATION 9.4

The Australian Government, in consultation with state and territory governments, should establish a process whereby public and private datasets are able to be nominated and designated as National Interest Datasets (NIDs).

Datasets (across the public and private sector) designated as NIDs would satisfy an underlying public interest test and their release would be likely to generate significant community-wide net benefits. Designation would occur via a disallowable instrument on the recommendation of the National Data Custodian.

NIDs that contain non-sensitive data should be immediately released. Those NIDs that include data on individuals would be available initially only to trusted users and in a manner that retains the privacy of individuals and/or the confidentiality of individual businesses. The in-principle aim should be for these de-identified datasets to be publicly released in time.

The process to designate datasets as being of national interest should be open to the states and territories in order to cover linked datasets, with negotiations undertaken to achieve this.

For community confidence, consideration should be given to use of a deliberative forum, such as a parliamentary committee, to take community input on and review nominations made, and to make proposals for future designations.

DRAFT RECOMMENDATION 9.5

The Australian Government should establish an Office of the National Data Custodian, as a new function within the Government to have overall responsibility for the implementation of data management policy.

Specifically, the National Data Custodian (NDC) would have responsibility for broad oversight and monitoring of Australia's data system, recommending the designation of National Interest Datasets, and accrediting Release Authorities and trusted users within the reformed data system.

DRAFT RECOMMENDATION 9.6

Selected Australian and state/territory government agencies should be accredited as Release Authorities by the National Data Custodian. In considering applications for accreditation, the National Data Custodian should consult a wide range of parties and ensure Accredited Release Authorities (ARAs) have sectoral expertise. The current model used by the National Statistical Service for appointing data linkage authorities should be considered in developing a model upon which to base this process.

ARAs will be responsible for:

- deciding (in consultation with initial data custodians) whether a dataset is available for public release or limited sharing with trusted users
- collating, curating and ensuring the timely updating of National Interest Datasets.

ARAs will also perform an important advisory role in regard to technical matters, both to government, and to the broader community of data custodians and data users.

DRAFT RECOMMENDATION 9.7

Trusted users should be accredited by the National Data Custodian for access to those National Interest Datasets (NIDs) that are not publicly released. Trusted users should be drawn from a wide range of potential entities, including: all Australian Government and state and territory government agencies; all Australian universities; and other entities (be they corporations, not-for-profit organisations or research bodies) that are covered by privacy legislation.

The default position should be that someone from one of these organisations would be approved for access unless the National Data Custodian transparently specifies a reason, on consideration, of why this should not occur.

For trusted users of NIDs, trusted user status should provide an ongoing access arrangement, with few restrictions on what could be done with the data. Trusted user status for NIDs should cease when the user leaves the approved organisation or be suspended if a breach occurs by any other trusted user in that same organisation and/or working on the same project.

INFORMATION REQUEST

The Commission seeks further views on the establishment of a Parliamentary Committee to take community input on possible National Interest Datasets, to review nominations made, and make proposals for future designations. Views are also sought on practical alternatives.

DRAFT RECOMMENDATION 9.8

Arrangements for access by trusted users to identifiable data held in the public sector and by publicly funded research bodies should be streamlined and expanded by the Australian Government. The National Data Custodian should be given responsibility to:

- develop, in consultation with data custodians, a list of pre-approved uses for a dataset, and make decisions on access to data for projects not consistent with the pre-approved uses list
- grant, on an approved project-specific basis, trusted user access to personnel from a range of potential entities, including: all Australian Government and state and territory government agencies; all Australian universities; and other entities (be they corporations, not-for-profit organisations or research bodies) that:
  - are covered by privacy legislation
  - have the necessary governance structures and processes in place to address the risks of inappropriate data use associated with particular datasets, including access to secure computing infrastructure.

Access would be granted for the life of the specific approved project. Trusted user status for use of identifiable data would cease when the user leaves the approved organisation; a project is completed; or if a breach occurs in that same organisation and/or project.

DRAFT RECOMMENDATION 9.9

Public research funding should be prioritised on the basis of progress made by research institutions in making their researchers' data widely available to other trusted researchers on conclusion of research projects.

DRAFT RECOMMENDATION 9.10

All non-sensitive public sector data should be released, consistent with release priorities and as resources allow, with curation, provision of metadata and adherence to agreed standards resourced as specified in Draft Recommendation 7.4. A realistic assessment of the risks associated with public release of identifiable information that is already public in a less accessible form, should be undertaken by all governments.

Data that could be used for program or agency performance management purposes should not be withheld from release.

DRAFT RECOMMENDATION 9.11

The Australian Government should introduce a *Data Sharing and Release Act* which includes the following:

- Provisions requiring government agencies to share and release data with other government agencies and requiring sharing between government agencies and other sectors.
  - These provisions would operate regardless of all restrictions on data sharing or release contained in other legislation, policies or guidelines.
  - The provisions may be waived in limited exceptional circumstances, and the Act should specify what these circumstances are.
- Strengthened provisions on access to data by individuals, including rights to access and edit data about them, a right to have data copied and transferred, and a right to request that collection cease.
- Provisions establishing the Framework for the governance of Comprehensive Rights of consumers, access to National Interest Datasets, approval of trusted users, and accreditation processes for Release Authorities.

1 Data, data everywhere

Key points

- New sources of data — as varied as social media sites, smart mobile devices and sensors fitted to physical objects (the Internet of Things) — continue to emerge and expand. Digital data is being collected ubiquitously — the amount of digital data collected in one year in the early 2000s is, according to some estimates, now being collected every two days — and is a source of considerable potential value.
- Data refers simply to a collection of material, which might include characters, text, words, numbers, pictures, sound or video. Digitisation enables data to be copied, stored and transferred very readily.
- The extraordinary capabilities of data analytics and the increasing ability to link previously separate datasets is compounding the usefulness of new data sources — offering important opportunities for better-informed decision making by individuals, businesses and governments, and for research breakthroughs.
- Frameworks and protections developed for data collection and access prior to sweeping digitisation require re-examination — privacy law is not the only lens through which to view the use of data. This shift in how to view data — as an opportunity, not necessarily a threat — is a global phenomenon.
- There can be many different competing interests in a particular dataset: the subject of the data (such as an individual); the parties who collect, aggregate and analyse the data; and those who commission these actions. Clarity about these interests is essential to allow Australia to harness the full value of its data.
  - The line between what is 'personal' data and what is not is blurred, as shown by the increased preparedness of individuals to share information about themselves on social media.
- A common misperception is that privacy laws give individuals an ownership right over their data.
  - In Australia, no one 'owns' data, although copyright law may apply in limited circumstances.
  - Instead, privacy legislation regulates how personal information is collected, used and disclosed.
  - Unaddressed questions of ownership and perceived and actual rights are important to future data access.
- This Report offers guidance on ways to generate community acceptance of the costs and risks associated with data use, and to do so where benefits may be most evident. These are lofty goals, but it would be poor play on our part, and not consistent with the rationale for the Productivity Commission's existence, if we simply conceded that such a task was too difficult.

1.1 About the Inquiry

The Inquiry has its origins in the 2014 Financial System Inquiry (the Murray Inquiry) (Murray et al. 2014) and the 2015 Competition Policy Review (the Harper Review) , both of which highlighted the potential to improve data access and use in Australia.[1] The Productivity Commission too has repeatedly, in a variety of inquiries and reports, drawn attention to issues around data access and use — such as the cost of not allowing researchers to access Australia's rich administrative data holdings (PC 2013).

The Australian Government has requested that the Commission conduct a broad ranging investigation into the benefits and costs of increasing the availability and use of public and private sector data by Australian individuals and organisations (see the Terms of Reference at the front of this Report). The Commission's processes for the Inquiry are set out in appendix A.

Inquiry scope

In this Report and in the literature, a distinction is made between 'data', 'information' and 'knowledge' — although, in the context of data collection, storage and use, the difference between data and information may often appear to be not particularly important:

- *Data* is representations of facts that are stored or transmitted as qualified or quantified symbols. It comprises material such as characters, text, words, numbers, pictures, sound or video. However, without being organised and put into context, data has little, if any, inherent meaning.

  A *dataset* is a collection of related data points or records with a common context (such as the collation of credit card records across a bank's customer base) that can be manipulated as a unit by a computer.
- *Information* is the meaning resulting from the interpretation of facts conveyed through data (and other sources). Information can be derived from a set of data after it has been presented in context and interpreted. For example, a student's score in a test is a piece of data. The average test score of a class is information that can be derived from the given data.
- *Knowledge* is information and experience that has been internalised or assimilated through learning (OECD 2015b).

[1] The final report of the Financial System Inquiry and the Government's response were released in December 2014 and October 2015 respectively. The final report of the Competition Policy Review and the Government's response were released in March 2015 and November 2015 respectively. The Murray Inquiry recommended that the Government task the Commission to review the benefits and costs of increasing the availability and improving the use of data. The Harper Review recommended that the Government consider ways to improve individuals' ability to access their own data to inform consumer choices.

The terms of reference for this Inquiry include examination of data collected, stored and used in both the public and private sectors.

Public sector data is defined in this Report to be data held by government agencies (at all levels of government) and entities that receive public funding (chapter 3). This includes data held by Government Business Enterprises and bodies such as universities and research institutes. As recognised by the *Public Sector Data Management* (2015, p. 6) report, public sector data may be:

- personal data — identifiable information used by agencies for service delivery
- research data — de-identified data securely shared with trusted users for research
- open data — aggregated data made publicly available with minimal restrictions on use
- security data — data where national security or other compelling public interest considerations tell against its release.

We consider that the first three types of public sector data are in scope for this Inquiry, but we do not consider security data — the Australian Government is considering changes to the regime governing this data, and the Protective Security Policy framework, as part of its implementation of the Belcher Review recommendations. As such, security data is only referenced for clarity — that is, where we believe it may appear to be captured by our draft recommendations, we clarify that it is not.

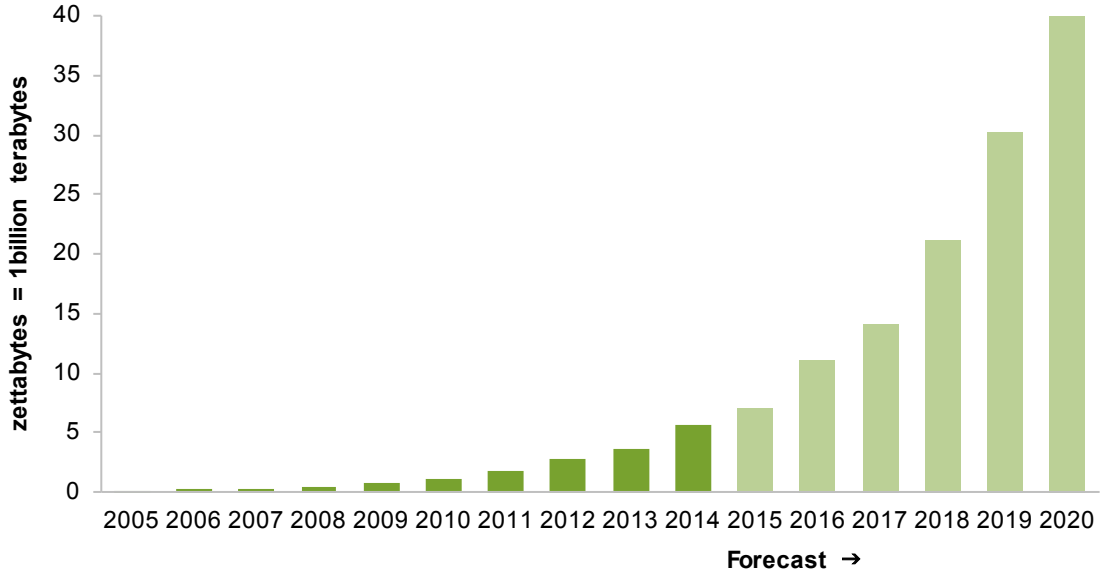
Private sector data is defined in this Report as data held by businesses and not-for-profit organisations (but not necessarily collected by them). We recognise that data has significant commercial value and sensitivity, and that there are a range of commercial incentives to consider when contemplating increases in its access. Further consideration of these interests is given in chapter 4.

1.2 Why data matters

The growth in the volume and variety of data

The amount of data being generated across the world is growing exponentially (figure 1.1). By some estimates, the amount of digital data generated globally in 2002 (five exabytes) (Lyman and Varian 2003) is now being generated every two days, while other estimates suggest that 90% of the world's information was generated in just the past two years (IBM 2016). Structured data — that is, organised data, such as that found in databases and spreadsheets — accounts for about 20% of all data (Nemschoff 2014).

Figure 1.1 **Data generated (global)[a,b]**



[a] Forecast from 2015. [b] Four zettabytes (in 2013) is equivalent to two quintillion jpg images, 456 billion hours of digitally recorded music, one trillion high definition digital films, or 166 billion 32 gigabyte iPads (Larson 2014).

Source: UN ECE (2015).

The Internet has become pervasive in economic activity and social interactions and its use is still growing rapidly. In the year to June 2016, the volume of data downloaded by fixed line and wireless broadband in Australia increased by 51% (ABS 2016). Data is being generated from a multitude of transactions, production activities and communications through the Internet and a range of other information and communication technologies. In particular, over the last couple of decades, massive amounts of data have been generated from three main emerging sources:

- social media posts, video and audio files, and emails (appendix F)
- mobile devices, such as mobile phones and fitness trackers (appendix F)
- physical objects (buildings, vehicles, machinery) that have been embedded with sensors (the Internet of Things) — that is, the computerisation of 'things' as varied as cars and tractors, airplanes and dishwashers, turbines and dog collars.

At the same time there has been enormous growth in connectivity within and between these data sources, and substantial increases in the availability of data (driven by data generation and data sharing).

Data digitisation and the growing power of data analytics

Digitisation — the conversion of data into electronic format and the direct creation of digital data — has increased the capacity to collate individual records into datasets and to copy and transmit these datasets without diminution of their quality. At the same time, data analytics — the techniques and tools that extract useful information from data — are enhancing the ability to reveal the patterns, correlations and interactions among data (OECD 2015a). The falling cost of data storage and processing (with the advent of cloud computing) and the spread of low-cost analytics tools, are increasing the affordability of data analytics and making it accessible to small and medium-sized enterprises. The ability to link datasets is also expanding and this growing connectivity is enhancing the usefulness and value of individual datasets — the whole is greater than the sum of the individual parts.

Data and the provision of digital services are integrally related. Data enables the production of digital products and services but it can also be an output from the use of these products and services.

New digital services, such as the numerous online consumer services that have emerged over the last two decades (for example, Amazon, Netflix and Uber), have been highly disruptive to the industries they have entered. The financial sector, conservative (some would say historically resistant) in the face of disruption by technology, is also currently experiencing a wave of innovation driven by digital technology and the use of existing and new sources of data. New innovative 'fintech' companies are capitalising on these developments along with the incumbent companies in the sector (appendix E).

Big data underpins greater knowledge and enhanced decision making

Big data is essentially data that has 'high volume, high velocity and high variety' characteristics, while its significance stems from 'a belief in the power of finely observed patterns, structures, and models drawn inductively from massive datasets' (Lane et al. 2014, p. 46). 'Big data' has become synonymous with the potential of data in the digital age.

The rapid growth in the generation, availability and connectivity of big data and the ever-expanding power of data analytics are delivering two primary benefits:

- greater insights and knowledge with which to better understand, influence or control the subjects of these insights and knowledge (such as natural phenomena, health, social systems and behaviour of individuals)
- more effective, better informed decision making — leading to efficiency gains and productivity growth (OECD 2015a) (chapter 2).

Data remains underutilised

Despite the rapid technical developments and the large potential benefits, much of the data being generated remains underutilised. Admittedly, a significant portion may prove to have no value (numerous duplicative digital photos may be an example). However, according to one estimate, in 2013 around 22% of the digital data generated globally was potentially useful as an input into subsequent analysis (to generate information and build knowledge and thus inform decision making and action) but less than 5% of that data was actually analysed (EMC Corporation 2014). Further, some data that was previously of limited value may become valuable as analytical capabilities improve or with investments made to improve its quality.

The conundrum is knowing which category data falls into, particularly when considering broad policies such as 'open by default', and the costs that such policies imply. Judgements by governments on what data is valuable for further use will surely prove hit and miss.

There is clearly scope to increase the availability of data and make more productive use of it. The appetite of governments for releasing their data holdings has been constrained by the associated resource costs and a general culture of risk aversion (chapter 3). For the private sector, commercial considerations can generally be left to drive business decisions on managing their data holdings — whether to sell, freely share or retain exclusive control — but there nevertheless appears to be untapped potential, on which we comment (chapter 4). While individuals have displayed a growing willingness to share data about themselves, it has largely been motivated by social considerations, notwithstanding the fact that the data they share has value.

One of our aims in this Report is to improve individuals' access to data. Greater access would empower individuals to make more-informed decisions about the products and services they choose, help drive the development of new products and services and improve the efficiency of markets. It would also shine a light on the activities of government and thus help to improve its efficiency and accountability.

This aim is consistent with a number of Commission reports produced in recent years. These have highlighted a lack of public access to government administrative data — along with a range of other barriers to data sharing, linking and use — and detailed the benefits of increased access to it.[2]

With such broad scope to improve data availability and such significant potential for associated benefits, data use should be a key plank in any contemporary government reform agenda. This Report offers guidance on ways to generate community acceptance of the costs and risks associated with data, and to do so where benefits may be most evident.

---

2 See, for example, *Gambling* (PC 2010), *Disability Care and Support* (PC 2011b), *Caring for Older Australians* (PC 2011a), *Annual Report 2012–13* (PC 2013), *Childcare and Early Childhood Learning* (PC 2014) and *Housing Assistance and Employment in Australia* (PC 2015).

These are lofty goals. But it would be poor play on our part, and not consistent with the rationale for the Productivity Commission's existence, if we simply conceded that the task was too difficult. We acknowledge the challenge of fashioning a Report of this nature, and expect some criticism. But the policy debate is not assisted by risk aversion (itself a central topic of this Report).

The economic properties of data

Data as a form of capital

Today, much more than ever in the past, data is a form of capital essential to the production of goods in some cases and, almost without exception, essential to the production of services.

However, it has several features that distinguish it from other forms of capital.

First, one person's use of a piece of data does not detract from the capacity of others to use it at that time — rather, a single piece of data can be used by multiple people at one time and in a variety of ways.

Second (and somewhat related), in contrast to many capital assets, while the information value of a piece of data may increase or decrease over time, data itself does not wear out with use.

Third, data in a digital format is virtually costless to reproduce.

Finally, some data is non-fungible — that is, it cannot be perfectly substituted for other data. This means that while the overall volume of data is expanding exponentially, many datasets will have significant scarcity value — notwithstanding that the creation of new datasets will sometimes lower the value of others.

The unique properties of data mean that despite its ever-growing abundance and its durability and reproducibility, it is now seen by many governments and businesses not as a burden, but as a keenly sought-after resource.

These properties, combined with ever lower costs of using data and of combining various datasets, have increased the benefits of re-using data. They have also increased the importance of clarity around data access and capacity to repurpose it. Individuals too are likely to appreciate the growing importance of their data, but they may be the last to realise its value.

Data can be used to address market failure, particularly information asymmetries

The economic value of data is largely reaped when it is used to better inform decision making of individuals, businesses and governments (chapter 2). In particular, the

information derived from data analysis can alleviate information asymmetry and reduce inefficiencies in market operation (box 1.1). The extent to which data can be used to improve market operations will be constrained, however, by any restrictions on data access and use.

Box 1.1 Examples of where data can alleviate information asymmetries in markets

- Comprehensive credit reporting seeks to address the information asymmetry between lenders and borrowers — borrowers typically having more information on their credit worthiness than lenders (appendix E), leading to instances where relatively creditworthy applicants are denied credit or are priced out of the market while less creditworthy applicants are able to access credit, potentially at an inefficiently low interest rate.
- Applicants for jobs typically have more information about their competencies, level of commitment and the accuracy of their resume than potential employers. Because of this information asymmetry, an employer may hire an unsuitable employee.
- Similarly, there is often an information asymmetry between providers of insurance and those purchasing insurance. For instance, the latter typically has more information about the riskiness of their behaviour — for example, their health and safety consciousness and their likelihood of attempting to defraud the insurance provider.
- Conversely, in many instances a consumer will know less than the organisation they are dealing with — for example, a customer may not have a good understanding of a phone plan or insurance policy they are considering. Data services that provide comparisons of alternative product offerings can help address this asymmetry.

Source: Lane et al (2014).

Different data, different uses and different risks

The same set of data can be presented in various different forms — for example, a dataset may identify particular individuals or the same dataset can be de-identified. Both forms of presentation may contain valuable information. Similarly, some data may be recorded in near real time — and the same data can be reviewed at some future time, perhaps for a different purpose. Both of these uses may be valuable.

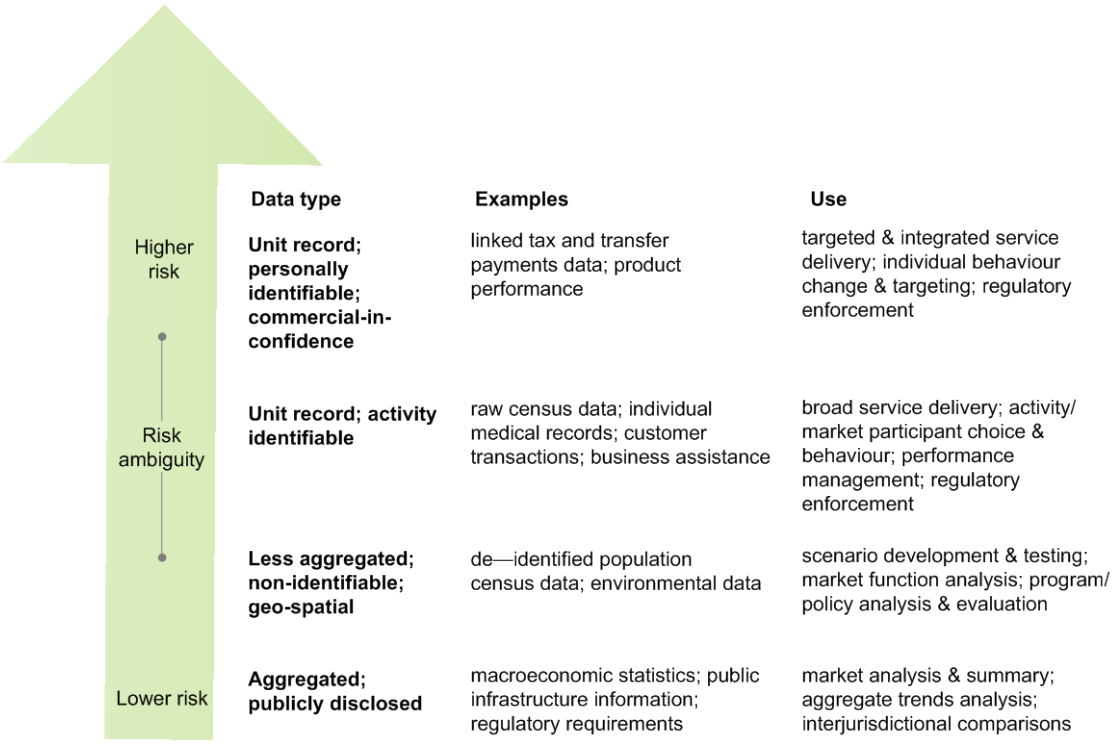
Near real time data that identifies individual persons or businesses carries the highest risks to privacy and security. Yet these risks can be exaggerated: open access to this level of data — useful for the enforcement of some regulations (such as traffic speed limits) and for inducing timely changes in consumer behaviour (for example, price responsive household electricity consumption) — is often not necessary in order to obtain many of the benefits of data use (figure 1.2). For analysis of market opportunities, scenario development, policy evaluation and improved delivery of many products/services, de-identified, non-real time data can be perfectly adequate.

Within each broad data category, there are also significant differences between data with regard to attendant risks. So, for example, personal health data that is identifiable (or if only simply de-identified, could be easily re-identified), would fall at the high end of a risk spectrum. At the other end of the spectrum, routinely collected data of an administrative or program nature would, in many instances, be classifiable as low risk.

Risks are not just related to the characteristics of a particular dataset and how it is used, but may also vary considerably with who is using the data. For instance, an individual may be quite relaxed about their doctor reviewing their health records but quite uncomfortable about such information being publicly available, where it could be used to embarrass or otherwise adversely affect the individual. In between these extremes are opportunities for both individuals and the wider public interest — for example, the Australian company Health& offers to store your medical data and use it 'to intelligently remind you when to see your doctor … based upon your risks' (Health& 2016, p. 1).

There are undoubtedly areas of ambiguity where it is debatable who should have access to what types of data (figure 1.2). In these areas, and throughout the spectrum of data types and uses, a rigorous assessment of genuine risk is needed.

Figure 1.2 **Different data, different uses and different risks**



1.3 Stakeholders in data management and access

Any proposals for ways to increase data availability and use must take into account the various existing rights, obligations and interests associated with data.

Multiple interests in a dataset

Assessing the rights and interests over data is complex, as there are potentially many competing interests in, and claims to, a particular dataset (box 1.2). An individual's personal information may be held by an organisation, but that organisation may have added value to the data, such as by collating separate records or applying a proprietary algorithm.

To further illustrate this point, transactions data — detailing the supplier of the good or service, the purchaser of the good or service, the date, what was exchanged, and the price — may be viewed as being owned by both the supplier and purchaser. In many cases, however, only one of the parties involved in the creation of data will collect, store and use the data, even though sometimes both parties may have the opportunity to do so. Often only a business will retain a comprehensive record of transactions even though their customers are generally issued with a receipt or periodic statement.

Box 1.2 **Data — a range of potential stakeholders**

There will often be multiple interests in a dataset, such as two or more of the following parties:

- data collector — a party that instigates data collection, such as a business or government agency
- data subject — the party that is the subject of data, such as an individual
- data user (data consumer) — a party that uses data that they have collected themselves or attained from other parties
- data compiler — a party that compiles existing data from different sources and adds value to it by tailoring it for a particular market or their own use – for example, a data broker or a data aggregator
- data funder — a party that commissions the creation of data
- data transformer — a party that transforms data with the intent of adding value, such as by de-identifying personal data
- data purchaser — a party that buys data; this party may also be the ultimate data consumer.

Source: OECD (2015a); Loshin (2001).

In many cases, no single data stakeholder will have an exclusive claim. Different stakeholders will typically have different powers and interests in the data depending on their role. For example, in the context of smart meters for electricity consumption, the Victorian Government Department of Economic Development, Jobs, Transport and Resources noted:

A number of consumer groups we interviewed suggested that there was uncertainty over who *owns* data generated by smart meters, and that this uncertainty should be resolved … No clear legal principles exist as yet for property rights over data (and some mooted mechanisms such as intellectual property protection are somewhat controversial and unattractive in online consumer affairs).

Moreover, in the Australian privacy regime, businesses are obligated to safeguard Personal Information regardless of who 'owns' it. The concept of information ownership does not figure in the Privacy Act. (DEDJTR 2015)

How the law treats interests in data

There is no proprietary right over data or information itself in Australia — that is, no one 'owns' data. However, in limited circumstances, copyright law can protect the form in which information is expressed, and privacy law regulates how personal information is collected, used and disclosed. Furthermore, in some circumstances it may be possible to claim ownership over a processed dataset.

How copyright law applies to data

Copyright law does not protect the underlying data, but it can give particular rights over its form or expression if it is sufficiently 'original', and expressed in 'material form'.[3] Copyright operates automatically — unlike patents, no registration of the interest is required, and it operates for a set period of time.

Copyright has relevance to a range of information covered by this Inquiry. Medical consultation notes, for example, are copyright of the doctor as the doctor has created them by virtue of her skill in, for instance, interpreting test results (*Breen v Williams* (1996) 186 CLR 71). For databases, the operation of copyright law is complicated and recent cases have shifted the Australian position. Now, to be 'original' enough to be covered by copyright, a data compilation must:

- not be copied
- originate from an identifiable human author
- result from independent intellectual effort directed to expressing the work in its final form.

For instance, *Telstra v Phone Directories 2010* held that phone directories (such as the White and Yellow Pages) that were compiled automatically by a computer did not have sufficient human effort of a literary nature to be covered by copyright. Changing technologies can also affect the operation of copyright — while films are protected by

[3] Data compilations fall within the 'literary works' category of works protected under the Copyright Act. Literary works are defined as including a table or a compilation, expressed in words, figures or symbols (whether or not in a visible form). A factual compilation will be a literary work if it provides intelligible information, as opposed to a random collection of data (*Hollinrake v Truswell 1894)*.

copyright because they are fixed in material form, the way digital broadcasts are transmitted (via impermanent data fragments) does not appear to afford them the same protection (*Commissioner of Taxation v Seven Network Limited 2016*).

While the recent decisions excluding automated data compilations from copyright protection bring Australian copyright law more in line with that of the United States, some Inquiry participants have argued that these developments in copyright law leave commercially valuable data unprotected (Winn and Wrathall (2000), ANZ (sub. 64, pp. 3–4)).

Where there is copyright, the holder of the copyright can, if they wish, assign rights to other parties to use that database or software under a license. In recent years, some public sector and some large search entities (for example, Google) have supported a move away from releasing data under restrictive licenses to releasing data under more permissive creative commons licenses that allow the data to be reused (appendix C). The terms of the use are set out in the conditions of the license. Creative Commons can, however, potentially impede some reuse activities (Bureau of Meteorology, sub. 198, p. 17).

How the *Privacy Act 1988* applies to data

The *Privacy Act 1988* (Cth) regulates the collection, use and disclosure of individuals' personal information — that is, information that could identify, or reasonably identify, an individual. It applies to private sector entities with an annual turnover of at least $3 million and to all Commonwealth Government agencies, subject to certain exemptions. Similar legislation exists in most state and territory jurisdictions (applicable to state/territory government data and entities not regulated by Commonwealth legislation), the exceptions being Western Australia and South Australia (appendix C). The rationale for any overlap of privacy legislation will be examined in the course of this Inquiry.

As noted earlier, privacy legislation does not give ownership rights — although this perception appears common and it is understandable that the right to have one's data protected would translate in many people's minds to a continuing ownership right to that data.

In effect, however, the *Privacy Act 1988* and common law have substituted for any fit-for-purpose structure to define rights, including anything to do with 'ownership' — these being the only legal measures dealing with the generalised collection of data before the advent of the Internet and the explosion of digital data collection and analysis.

Instead of ownership rights, privacy legislation imposes obligations on collectors of data to ensure that the personal information they hold is managed in accordance with the principles set out in the legislation. In general, these include obligations to:

- ensure personal information is collected fairly and lawfully
- inform the individual that their information is being collected and how it will be handled (under a privacy policy)

- only use personal information for the purpose it was collected for, a directly related purpose, or another purpose that the individual has consented to
- keep personal information safe from unauthorised access and misuse
- take reasonable steps to delete or de-identify personal information when they no longer need it.[4]

Privacy legislation also gives individuals the power to request access to, or correction of, their personal information.

Recent years have seen criticisms of privacy legislation's reliance on consent (see below). Evidence shows that many individuals are unable to understand how their information is being used and/or do not read use disclosures.

Protecting the identity of individuals and organisations

Given the way digital data is collected, even data that does not directly relate to individual people or organisations — for example, traffic patterns — can often be used to identify individuals. The sharing and release of such data gives rise to concerns about the potential for unwanted reductions in the privacy of individual people. Businesses too may feel some exposure or suffer detriment as a result of commercially sensitive information being released, although often they are better placed to manage this.

Personal data and privacy

Data about people — that is, personal data or information — is a special case amongst the volumes of data being generated because of its implications for privacy. The claim to privacy differs with context — it can include the right to a private life, and the right to freedom from unwarranted interference (Warren and Brandis 1890). The social norms around the treatment of personal information (that is, what privacy means in practice) vary with who the subjects and recipients of data are. For example, in a health care context, 'informational norms' — some of which are embedded in legislation and some of which are primarily socially or culturally accepted practices — govern the flow of personal data between patients, doctors, nurses, insurance companies and pharmacists. In turn, there are informational norms covering which of these parties has access to the various types of data — ranging from patient symptoms, diagnoses, prescriptions and biographical information (Lane et al. 2014). Thus, privacy may be achieved by having control over information about oneself — that is, control over who one shares personal information with and the terms under which one agrees to share it.

[4] These obligations are set out in the Australian Privacy Principles and ss. 16A and 16B of the Privacy Act. They are open to interpretation by organisations that collect and manage data. These interpretations often tend to focus on minimising real and perceived legal risks rather than identifying ways to enable appropriate access to data (chapter 5).

The costs and benefits of privacy are inextricably linked and complicate the task of determining where public interest needs might lie in a data-driven environment. For instance, an individual may benefit from disclosing certain personal information to a company in return for services (for example, access to social media) and resent any regulatory restrictions that would limit the service received. On the other hand, the same individual may wish to keep other personal information (such as their location) private, to reduce the risk of adverse consequences if conveyed to unintended third parties. There may be an expectation that such risks would be managed by a public policy response.

Moreover, even defining when data is private can be complicated by the opportunities presented by modern communication. Think of Twitter accounts: the line between personal and impersonal information is blurred, with individuals sharing information about themselves on social media under simple user names, but with limited ability to prevent re-identification and, possibly, subsequent harassment.

Individuals are often called upon to make judgements about privacy in their adult lives (making such judgements as children is quite a different matter) and it can be argued that such responsibility must be relied upon in defining any response to the burgeoning development of big data and data analytics. Nevertheless, circumstances have arisen — and will continue to arise — when such a position seems inadequate.

Consent

Privacy legislation is of limited value once data is shared legitimately — that is, with consent. Consent is one of the main mechanisms by which individuals can influence their privacy but once consent has been given, a large array of data transactions beyond the awareness of most individuals can — and often will — take place. A considerable amount of data is collected about individuals without their explicit consent and, sometimes, without their knowledge.

Personal data on individuals is generated and collected as:

- *volunteered data* — when an individual actively and deliberately shares data about themselves, such as by creating a social network profile or entering credit card information for online purchases (OECD 2015a)
- *observed data* — when an individual's action or activity is recorded. Examples of such passive data generation are location data from cellular phones, Internet browsing preferences and some data generated when an individual interacts with a government service agency, such as Centrelink or the Australian Taxation Office
- *inferred data* — from the analysis (including linking) of data about an individual. An example is an individual's credit score based on their observed payment history.

Individuals can grant *express consent* for organisations to collect, use and share data about them — for example, by accepting the terms and conditions for using some licenced software (such as Facebook) (appendix F).

Individuals implicitly agree to share some personal details every time they complete a transaction using a credit card (*implicit consent*).

The Australian Privacy Principles Guidelines (APPs) state that *informed consent* — which can be either *express consent* or *implied consent* — involves several key elements, including that the individual is adequately informed before giving their consent, the individual gives consent voluntarily, and the consent is current and specific (OAIC APP guidelines).

Meaningful consent — in the context of consenting to an organisation's privacy policy and/or terms of use of its services — involves the organisation seeking consent achieving a balance between providing the necessary information and avoiding 'notice fatigue' on the part of the consent giver. In this regard, the OAIC recommends short notices that explain what data will be collected and any third party data sharing practices, as well as highlighting to the reader any collection, use or disclosure that they would not otherwise reasonably expect. However, the overwhelming evidence is that the majority of those granting such consent neither read nor understand the terms and conditions (chapter 4).

Confidentiality

Confidentiality is what is bestowed on an individual (or an organisation) — through a set of rules or a promise — by restricting disclosure of information about that individual (or organisation). Such restrictions on disclosure to unauthorised individuals, organisations or processes helps to preserve the privacy of data subjects. One method of providing confidentiality is to de-identify the data — that is, removing any links to an individual or organisation. This action can be strengthened by 'confidentialising' the data — that is, taking the additional step of assessing and managing the risk of indirect re-identification occurring (appendix B).

'Data sharing' versus 'data release'

Data sharing

Usually referred to in the context of public sector data, data sharing is the provision of data to restricted organisations or individuals. Data sharing can also involve constraints on the use of the data, the timeframe over which it is used, and the technology on which it is analysed.

The concept of 'trusted parties' — approved individuals or organisations who have differential access arrangements — is often cited in the context of data sharing. Access tends to be restricted to trusted parties because the data being shared is sensitive — for example, it may be personal information and its general release could cause embarrassment to some of the individual subjects of the data. Such data might be shared with the data subjects (for example, an individual gaining access to their health records), within

government, or outside government (for example, with researchers or businesses). In some countries, governments have initiated programs to encourage private data sharing, such as the UK's midata program (which encourages businesses in selected sectors to allow customers to download data about their use of specific products (chapter 4)).

Data release (open data)

By definition, open data is available to all potential users without restrictions on what they may do with it. The corollary is there cannot be any legal restrictions on making that data available. For example, personal data can generally only be open data if restrictions imposed by the *Privacy Act 1988* do not apply — such as it is written in law that it must be published, or if the individual subjects of the data have given their permission for the data about them to be released.

Legal restrictions may also include licence arrangements (appendix C and chapter 6)

1.4 The challenges for governments and society

Dealing with the unknown

The arrangements and policies for collecting, analysing, linking and passing data from one party to another in the public and private sector, and the obligations, rights and opportunities between parties have developed reactively to digital technology improvement — in Australia and elsewhere.

This reactive approach has been inevitable. Designing for the advent of the World Wide Web and mobile computing could hardly have occurred in the 1980s, or for Facebook and Google's near-ubiquitous nature, even as recently as 2000. But a more strategic and holistic approach is now feasible and appears necessary.

Although we cannot see the shape of future data development in detail, we can discern its direction sufficiently to benchmark current corporate, government and individual rights, expectations and standards (to the extent they exist) with the increasingly evident direction of change. Some countries have already taken steps to bring their policy settings up to date, with a view to unlocking the value of data through better access and use arrangements. Moves are afoot in many other countries.

Data has the potential to deliver analytic insights that will add to the stock of knowledge, improve decision making and enhance the awareness and understanding of individuals. However, where the data collected records details about individual behaviour, it can be perceived by some as a threat to fundamental values, including 'everything from autonomy, to fairness, justice, due process, property, solidarity, and, perhaps most of all, privacy' (Lane et al. 2014, p. 44).

Such fears are unsurprising in an era of rapid social change. Yet societal values and priorities can be context-specific and do alter over time. This is already observable in today's data-rich areas, as quality and capability increases in an extraordinary fashion and people's preparedness to trade (knowingly or unknowingly) data for services increases.

In practical terms, these evolving and shifting societal values may reasonably be regarded as unknowable.

For policy-makers, who are the core focus of this Report, precautionary preventative behaviour is often the order of the day when dealing with the unknowable. But where benefits are large and change is rapid — both are observable today — being reactive on data access is certain to be costly.

Whether data is held in the public or private sector can matter

Governments have considerable scope, within political and constitutional limits, to change the environment in which public sector data is collected, stored, shared, released and used. Their scope to influence these aspects of private sector data collection is — wisely — usually more constrained. Investment in the aggregation, development and analysis of data is providing durable private incentives to innovate with data. And the results of this — on-line search, social media, on-line retailing, remote sensing, medical or public health analyses, to name a few — are clearly highly valued.

The willingness of individuals to share personal data with governments as opposed to private sector entities will be different. The benefit for an individual from an on-line transaction or search is usually relatively clear, but is less obvious in the case of filling out an ABS survey or submitting information to the Australian Taxation Office.

Similarly, the incentives and legislative constraints for public sector data holders to then share and release data will differ (chapter 3). Sharing adds a third party to what can already be a difficult sell for a government data collector; whereas for a private sector data collector, the story may be quite different — third parties are often a substantial part of the reason many of the most popular sites are successful — Facebook being a stand-out example (chapter 4).

While much of the analysis and many of the findings and recommendations in this Report apply to both public sector and private sector data — consistent with the way the Terms of Reference are structured — there are instances where these distinctions translate into different policy approaches.

The identifiable nature of some data raises challenges

Historically, much data held by governments and the private sector was about 'things' — products, services, environmental features and capital assets.[5] This is changing — a significant proportion of data being created has an aspect to it that identifies individual people or businesses, or activities undertaken by them.

For example, seemingly innocuous digital photos of landscapes or objects loaded onto platforms and websites will often contain metadata that reveals the device used to take the photos and the time and location each photo was taken. Further, facial recognition technology may be used to reveal the identity of any people in the photos, and the identity of the person uploading the photos may also be apparent or deducible.

The identifiable nature of some data poses challenges for individuals, governments and businesses (which depend on corporate reputation) in finding ways to increase access to data while not inadvertently revealing the identity of the individual people or businesses to which the data relates. This Report will examine the issues around how access to certain types of data can impact on the privacy of individuals and the competitive advantage of businesses (chapter 5).

Trade-offs are involved

The challenge for policy makers is to provide a structure that enhances societal wellbeing — that is, one that fosters the beneficial economic, environmental and social uses of data while simultaneously trying to manage the potential costs. This requires achieving, at a broad societal level, the 'right' appetite for risk, not least across government but also for businesses and individuals.

In most aspects of life, risk can never be completely removed, and the sphere of data is no exception. However, with a realistic assessment of the risks, the mitigation of these risks where practicable, and a recognition of the enormous potential for benefits, the Commission's view is that we, as a society, should err on the side of data openness.

A key to achieving this outcome will be building and retaining community trust in how data is managed and used. Trust is particularly important in the area of personal data — if individuals can see benefits arising from the use of data about themselves and feel they have control over how such data is being used, a virtuous cycle of greater trust and larger benefits can be established. While government can be part of the solution — particularly where it is a data collector in its own right — the policy mix that embeds trust necessarily involves commitment and awareness across individuals, governments and businesses.

---

5 Notwithstanding that many governments have, going back centuries, compiled data collections on births, deaths and marriages. However, these collections were sometimes less about the efficient and effective delivery of public services, and more about recording the existence and growth of society.

The Commission proposes a policy framework (chapter 9) that attempts to achieve this outcome — one that reinforces the prospects for societal benefits and gives individuals more control and influence over data. The framework seeks to embed some basic rules, principles and understandings about how data is collected, managed and used that, when embedded, will foster openness and enable trust to be retained, even in the face of the inevitable risks.

The approach is deliberately one for the medium term.

The Commission is not judging short-term initiatives by government other than where they offer lessons (good or bad) for the future public interest of a society swimming in data, and in need of all the useful input — lessons, lifeguards, tide forecasts — that can go to ensure a better outcome than drowning or being taken by the occasional shark. At present we have one input only — on privacy. It seems insufficient.

We invite feedback from interested parties on the recommended plan for Australia's data future.

2 Opportunities enabled by data

Key points

- In less than two decades, extraordinary developments in computing power, data generation, and algorithms that can detect patterns and preferences previously indiscernible, have enabled new business models and opportunities. Individuals, businesses, governments and the broader community have all benefited from these changes.
- Examples of innovative uses of data are offered by this Report and submissions to it. Many such innovations would be unimaginable without the ability to collect and interpret large volumes of data.
- Private sector data owners are leading the way in finding innovative uses for data. Governments across Australia also hold lots of data, but are typically not using it beyond the purposes for which it was initially collected.
- Wider release of data held by governments would likely trigger significant investment (private as well as public) and improvements in national welfare. However, determining which datasets would most likely lead to such improvements remains a serious practical issue for governments.
- Opportunities to use and link datasets, and the benefits that could be achieved by doing so, are largely unknown until the data sources themselves are made known and a wide range of users have had opportunities to experiment with the data. This underscores the substantive argument in favour of greater data access.
- Traditional mechanisms for demonstrating the value of an asset — such as price signals and revenue streams — either do not apply to data in government hands or are derived from cost recovery goals completely unrelated to the *prospective* value of data in the hands of innovators.
- Only after potential data users are appraised of what data governments are holding, and are given the opportunity to advise data holders of its usefulness — such as by submitting proposals or requesting access to datasets — can specific datasets be determined as being of high or lesser value.
- Existing government data initiatives, such as data.gov.au, should be leveraged as part of these efforts. A framework should be developed to formalise how such efforts will be implemented and managed.
- A priori, datasets of high value have a number of distinct characteristics, including that they are unique (or cannot be replicated), are of high quality, have a high degree of coverage in the relevant population, and are up to date or updated regularly.

Over recent decades, increases in computing power, data analytics, and the quantity and availability of data have coincided with, and contributed to, the emergence of so-called big data. This has led to new and innovative models (within businesses and governments) for using data. While the benefits stemming from such initiatives are typically uncertain, they have the real potential to be materially large and widespread. Along with this chapter, appendixes to this Report illustrate the scope for innovative uses of data in the areas of health and finance.

This chapter examines the opportunities that greater availability and use of data would enable and the ways in which data could be used to generate social and economic benefits. Examples presented are 'best case scenario' benefits from improved data access — they highlight what Australia forfeits through not making data available for wider use.

For some data sets, greater access will also involve greater risk — occasionally, risk that is not able to be managed fully. For other data sets, risks will not alter even after release. Privacy may not be the biggest such risk. Costs of a financial or reputational nature may exceed today's focus on privacy. As such, data management techniques will need to evolve over time.

In deciding which datasets to make more available, the risks and costs of wider release need to be carefully considered, along with approaches that might be adopted to mitigate these (examined in detail in chapters 5, 6 and 8). But as with other decisions that governments and individuals make, the potential benefits of wider data use should not be dismissed in the belief that locking up data will minimise risks from its existence.

The focus of this chapter is on:

- the ways that society can benefit from increased data access and use, with several illustrative case studies (section 2.1)
- issues in determining high value datasets and prioritising dataset release (section 2.2).

The term 'high value datasets' should not be conflated with datasets that are of 'national interest' which are addressed in detail later in this Report. High value datasets include a broader range of datasets that are valuable to particular users or sectors, but which might not be of national interest. In this sense, national interest datasets can be thought of as a subset of high value datasets.

2.1 What can be done with data?

Increased access to data provides opportunities for innovative uses that benefit individuals, companies, researchers and/or the public more broadly. As noted by many of the 211 submissions received, data access is crucial for realising a range of private and social benefits — many presently unimaginable.

Estimates of the value of data

There have been numerous studies in recent years that have attempted to place quantitative estimates on the benefits that could arise from greater availability and use of public sector data in particular (table 2.1). Estimates for the value of Australian public sector data vary 100-fold from $625 million per year to up to $64 billion per year. But, the wideness of this range reflects that:

- vastly different measures of benefit or value are being estimated
- estimates are based on different types of data (some use spatial data only, some use a broader range of public sector data)
- there are a variety of assumptions made about how data is used and the associated benefits
- there are considerable structural differences between the countries for which the estimates were initially made (prior to being converted into an Australian estimate).

Some of the smaller estimates are based on a subset of data (the ACIL Tasman estimates use only spatial data) or focus just on direct impacts; some of the larger estimates make what may be extravagant assumptions about economy-wide uses and impacts, or ignore the costs of achieving wider data access (the McKinsey estimates of 'gross output'). Nevertheless, the estimates — based on the value of just public sector data — could easily be double were the value of private sector data included as well.

While these factors mean the estimates are not comparable, they do not negate two obvious conclusions: first, the potential quantitative value of data, by some estimates, is immense; second, that you cannot be definitive about this value, particularly when it requires speculation about possible current and future uses.

There may be cases, however, where inferences can be drawn about the value of a particular dataset. For example, the Australian Securities and Investment Commission maintains several registers related to Australian businesses, which generated $61 million of search-related revenue (including from information brokers) in 2013-14 (Davis 2015). While it is not possible to quantify the exact value of this registry, the expected *annual* value to the community is inferred to be at least $61 million (in 2013-14 prices).[6] Alternatively, the annual costs to data users if the business registry dataset was not available or did not exist (that is, the losses avoided), would likely be much larger than $61 million.

Generally speaking, opportunities from data release will, by and large, beget further opportunities and, subject to consideration of the risks and costs involved, the balance should generally lie in favour of releasing data where the benefits are uncertain.

6 Although as noted later, standard cost recovery or charging principles generally do not reflect the true value of data held in government hands.

Table 2.1 **Estimates of the value of public sector data[a]**

	Measure of value	*Examples (value per year)*
Gross output	Total *potential* change in economic output if all public data was made open (setting aside the costs of using data)	McKinsey Global Institute (2013) ($64 billion)
Wider net economic benefits	Economy-wide impacts from the use of public sector data[b]	ACIL Tasman (2010) ($7.6-$14.8 billion) Vickery (2010) ($25 billion) Deloitte (2013) ($7 billion) Lateral Economics (2014) ($34 billion)
Direct net economic value	Consumer and producer surplus from the collection and sale of public sector data[c]	DotEcon (2006) ($625 million - $1.2 billion)
Direct use value	Value added plus market value	Deloitte (2013) ($1.9 billion)
Value added	Value added by entities that use public sector data to generate other goods and services[d]	Pira (2000) ($22 billion) ACIL Tasman (2010) ($682 million Vickery (2010) ($4.5 billion)
Market value	Net profits from sale of public sector data by government agencies and from the re-sale of public sector data by brokers[e]	MEPSIR (2006) ($3.9 billion)
Investment value	Costs of making public sector data available to the public Costs of collecting public sector data	Pira (2000) ($3 billion)

[a] Estimates are based on the studies listed and were converted to Australian dollars (2013 prices) by Lateral Economics on the basis of relative GDP (between Australia and the economies in which these studies were undertaken). ACIL Tasman and Lateral Economics were Australian studies with Australian dollars estimates. [b] Wider economic estimates generated by applying an average return on investment coefficient to the total costs of public sector data collection and management, or by plugging productivity shocks into a computable general equilibrium model. It includes value added by firms using data and market value of data at initial point of sale. [c] Consumer surplus estimated using elasticities of demand (to derive willingness to pay estimates). Producer surplus estimated based on revenue data from the sale of public sector data and demand for public sector data. [c] Estimating value added typically relies on surveys and/or case studies to assess the extent to which public sector data is an input — estimates are then extrapolated more widely to other users and sectors. [e] Estimation based on revenue from sale of data and surveys of data brokers. Because much public sector data is released free of charge, this could under estimate actual market value.

Source: Lateral Economics (2014).

Data can be used to improve consumer choices

An often cited benefit of increased access to data is that it can enable consumers to make better choices about the products and services that are suitable for them. In particular, this could occur if individuals were better able to access, and securely share, data about their

own activities. For example, streamlining access to consumers' transactional data has the potential to improve personal financial decisions (Centre for International Finance and Regulation (CIFR), sub. 9). These types of consumer data could also be leveraged by budgeting services and apps — such as Mint and Spending Tracker — which aim to help individuals make better spending and budgeting decisions (Manyika et al. 2013).

In the United Kingdom, the midata program encourages businesses in several sectors to allow customers to download data about their use of specific products (such as their banking transaction history) (chapter 4). A stated aim of this program is to give individuals the ability to provide this data to third parties, who — with the right incentives (such as the ability to charge a fee or to better compete with incumbents) — would be able to recommend a suitable product for the individual based on analysis of their data and the fees and charges of products in the marketplace. This could be particularly beneficial in relation to products with complex pricing structures, such as those in banking and telecommunications.

In sectors where comparisons between products are difficult for Australian consumers, third-party businesses are emerging and putting pressure on businesses to make data more available to customers. Energy Tailors in Victoria, for example, obtains customer consent to access and analyse smart meter and other data to suggest better deals on electricity plans. Health& allows consumers to manually input and store their health data — including medical records and data from fitness devices — in a centralised location, to allow better preventative health care and simpler sharing of health information between health service providers.

A related benefit is that open data could provide guidance to businesses on the needs of consumers, and therefore might lead to businesses offering consumers an improved suite of products and services. Moreover, to the extent that businesses can use data to 'segment' customers, there would also be scope for providing personalised products and services.

The music service provider Pandora uses self-reported customer data (such as age and gender), together with information provided by customers about songs they 'like', to tailor the selection of songs streamed to them. The more data that users provide, the better the service. In addition, in its free version, Pandora uses customer data to target advertising — customers are provided with the music they like and presumably advertising that is more relevant to their interests. The trade-off seems to work — there are nearly 80 million active subscribers to the free service (Morey, Forbath and Schoop 2015).

Finally, increased corporate transparency can empower consumers to make decisions about which companies to purchase from. In recognition of this, Nike, for example, publishes a range of information and data as part of its corporate responsibility reporting, including a full list of contracts with its suppliers, data on working conditions (including pay and hours worked) in supplier factories, and estimates of its carbon footprint (Nike 2016).

Data use can offer commercial benefits

The rise in big data has largely been driven by commercial entities who recognise the substantial value that can be generated by applying novel analytic techniques to rich datasets. Google and Facebook are prominent examples of businesses that have monetised the collection and use of data (chapter 4).

Broadly speaking, data can create commercial value by facilitating innovation, and by increasing efficiency and productivity within businesses. It enables firms to create new products and services, enhance existing ones, and introduce entirely new business models.

Data on the quality or performance of inputs used in production can be used to improve the efficiency of production processes and data on consumer demand can be used to more closely align product specifications with what consumers want. GE Aviation uses data from sensors in deployed aircraft engines to evaluate fuel efficiency, which forms the basis of advice to airlines on how to minimise fuel costs (Porter and Heppelmann 2014).

Data can also be used internally to drive cultural change within businesses. Woodside Energy is using the IBM Watson data analytics platform to digitise and make searchable its library of project documents. This improves the ease with which previous projects can be searched and understood, which in turn is helping to build the skills of its workforce and transfer corporate knowledge (Head 2016).

In a similar vein, data has allowed logistics businesses to increase efficiency and cut costs, including through:

- real-time route optimisation
- crowd-based pick-up and delivery
- strategic network planning
- operational capacity planning (Jeseke, Grüner and Weiß 2013).

Wal-Mart established a data sharing system in the early 1990s, in which sales data — by item, store and day — was provided to all of its suppliers. Access to such data ' … translated to lower merchandising costs for Walmart, and also saved suppliers time and expense in planning their production and distribution' (Waller and Boccasam 2013).

UPS, a major package delivery service, uses data from telematics sensors in its delivery vehicles, along with map data from other sources, to monitor daily driver performance and to optimise a driver's pickups and drop-offs in real time. It was estimated that in 2011, this approach saved more than 8.4 million gallons of fuel. Moreover, UPS estimates that it saves about US$30 million per annum when it reduces the distance driven per driver by a mile per day (Davenport and Dyché 2013).

Big data is also helping liquefied natural gas companies identify plant problems and avoid shut-downs that can cost upwards of US$25 million a day (Hunn 2016).

Large amounts of data are also a requisite input for businesses deploying algorithmic decision making tools which can also lead to improved decision making and risk management. The Financial Services Inquiry highlighted the potential for small businesses, in particular, to benefit from increased availability of data for benchmarking and decision making purposes (Murray et al. 2014).

Analysis of large quantities of data on what does and does not work can be a crucial part of the development and testing of innovative new products and services. In this regard, open publication of research data can benefit the commercial sector by facilitating the transfer of knowledge from researchers to businesses (Beagrie, Lavoie and Woollard 2010). The Surveying and Spatial Sciences Institute (SSSI) noted that the release of data held in the public sector can also spark innovation in businesses:

> … the release of the Geocoded National Address File as open data, as well as the availability of other public datasets as open data has provided the opportunity for businesses to expand into new and emerging technology markets. (sub. 101, p. 2)

Geospatial data has been used in a range of commercial applications in recent years (box 2.1).

Data use can enable more efficient markets

While possessing data and information that few others have can be a source of market power, widespread data release can facilitate market entry and, more generally, reduce information asymmetry and improve the efficiency of markets.

An example specific to this Inquiry's terms of reference is credit reporting. Credit reporting is a systematic approach to sharing data in order to address the information asymmetries between borrowers and lenders that can lead to inefficient allocation and pricing of credit (chapter 4; appendix E). In a similar vein, information asymmetry between insurers and the insured can inhibit the efficiency of insurance markets, increasing premiums and the likelihood that it is primarily the highest risk consumers who take out insurance. In recent years, insurance companies have started applying big data techniques to improve the efficiency of insurance underwriting and to identify fraudulent claims (Bharal and Halfon 2013).

Inquiry participants have also highlighted the potential for data access to increase competition, particularly within the financial sector. For example, FinTech Australia (sub. 182) and Tyro Payments Limited (sub. 7) suggested that providing individuals with the means to share verifiable data about themselves in a machine-readable format would erode the competitive advantage held by the major banks due to their extensive data holdings, inducing them to price more competitively and service customer segments that may currently be underserviced.

Box 2.1 The commercial uses of spatial data

Spatial data is information about a physical object, such as its location and metric relationship to other objects, that is represented by a numerical value. When referenced against physical geography, it is often termed geospatial data. There are various sources of spatial data, including satellite-based global positioning systems and images, and geographical information systems.

Because of its many applications, spatial data is often held up as a valuable resource for industry, governments and researchers. Specific uses of spatial data across a range of industries include:

- agriculture: controlled traffic farming; yield monitoring; natural resources management; pest and disease management
- forestry: inventory management; remote assessment of forest attributes; yield management; canopy health mapping; operations management
- fisheries: recording fishing tracks; fisheries management; habitat mapping
- mining and resources: explorations; planning and management; spatially enabled robotic mining; environment compliance
- property and services: surveying; advertising and market research; planning; retail and trade; property and infrastructure development
- construction: surveying; planning and design; maintenance
- transport and storage: logistics; route selection/itinerary planning; transport planning; vehicle tracking; traffic and congestion management; transport operations in air and rail; intelligent transport systems
- utilities: asset management; supply/demand management; planning and construction of infrastructure
- communications: network planning; asset management; address management; route planning (for postal services)
- government: natural resources and environmental management; biosecurity; defence and security; air and sea navigation safety; search and rescue; land development administration; policy formation; service delivery.

Source: ACIL Tasman (2008).

Greater access to data can also help to reduce search costs in a range of markets. Seek, an online employment advertiser, publishes a range of information about Australian employers, including employee reviews, with the aim of helping job seekers identify companies that would be a suitable match for them (Seek nd).

Mobile app Parkopedia utilises user-provided data to help individuals search for the closest and cheapest parking option (Parkopedia nd), which could reduce the time and fuel costs of searching for a parking spot. Similarly, the ACT Government has implemented a 'smart parking' trial in Manuka, which uses in-ground sensors to produce real-time information on available parking spaces. This information is relayed to drivers via a smartphone app (ACT Government Chief Minister and Treasury 2016). The use of traffic data to improve road management in Singapore is a similar example (box 2.3).

Ebay, as an intermediary between buyers and sellers, collates data from sellers (such as prices and ratings from buyer feedback) into a central repository, which can reduce search costs for buyers (Frontier Economics 2008). Ebay subsidiary, PayPal, has begun using data it collects from sellers as a basis for providing 'working capital' credit to those sellers (appendix E).

Data use can improve social outcomes

Improvements in lifestyle and living environments

Deloitte (2013) identified that real-time availability of government traffic data would allow third parties to deliver apps that could reduce congestion, therefore cutting fuel consumption and carbon emissions.[7] In 2011, New York City began releasing detailed energy and water use data for commercial buildings, with the datasets used by building owners to benchmark the energy efficiency of their buildings and to prioritise future energy-reducing investments (Chui, Farrell and Jackson 2014).

In Australia, the City of Melbourne is using a range of data — including geospatial data as well as data from strategically located weather stations — to increase canopy cover, improve the health and diversity of the 'urban forest', improve water quality, and enhance the urban ecology (City of Melbourne nd). More broadly, spatial data has been used by policy makers and other groups to contribute to improved outcomes in Australia, such as those stemming from improved natural resource and environmental management (box 2.1).

Improvements in health service delivery

The Telethon Kids Institute highlighted the importance of data for health outcomes — including a range of linked data on educational, economic, geographic and racial factors (sub. 5). Australian Unity (sub. 95) suggested that access to an individual's health and clinical data (including administrative Pharmaceutical Benefits Scheme data) would enable private health insurers to identify health risk triggers and develop more timely interventions.

The experience in Western Australia — which has had a data linking unit for many years — has been illustrative in this respect (appendix D). The ability to link various datasets has facilitated a range of uses, including:

- a review of the safety of specific surgical procedures
- an investigation of a cancer cluster at Royal Perth Hospital
- demand and supply modelling for hospital services

[7] Reduced congestion would also directly benefit individuals through reduced transport times and fuel costs.

- influenza impact modelling (Data Linkage Branch (Dept of Health WA), sub. 13, attachment 4).

As noted by PricewaterhouseCooper (2014), the data sharing that would be made possible by greater adoption of electronic health records could improve healthcare by:

- helping to identify the patients most likely to benefit from particular interventions
- facilitating the use of algorithms to predict potential readmissions, which would allow for targeted interventions by healthcare providers
- improving the management of patients and allocation of resources through the use of triage algorithms
- allowing for evaluation of data from monitoring devices to pinpoint those patients whose condition is likely to worsen
- providing a basis for integrating data across clinical systems to provide a better standard of healthcare to patients who receive treatment from different healthcare providers.

Moreover, health data can help policy makers and researchers to identify a range of factors that contribute to illnesses, assess the safety of pharmaceuticals on an ongoing basis, and evaluate the effectiveness and efficiency of health policy (appendix D; PHRN 2016). In the United Kingdom, administrative health records have been used to improve how cancer is diagnosed (box 2.2).

Getting more value out of research activity

Inquiry participants highlighted that making data more widely available has the potential to significantly improve the productivity of, and hence the benefits flowing from, research activity (for example, Telethon Kids Institute, sub. 5; Centre for Big Data Research in Health – University of NSW, sub. 21; Commonwealth Grants Commission, sub. 58; Department of Industry, Innovation and Science, sub. 69; Judy Allen and Carolyn Adams, sub. 106; AURIN – The University of Melbourne, sub. 116). Improvements in research can occur through increased availability of data created by researchers, or because researchers have improved access to data generated outside of the research sector by governments and the private sector. From reviewing a range of studies, Houghton (2011) identified that open research data can:

- create opportunities for repurposing and reusing data
- stimulate new research networks and collaborations, including by creating greater opportunities for downstream research
- facilitate knowledge transfer to industry
- allow for verification or correction of previous study findings.

Box 2.2 Improving cancer survival rates in the United Kingdom

Motivated by observations that cancer survival rates in England were lower than in Europe, a team of researchers linked several administrative and cancer diagnostic datasets to estimate routes to diagnosis, and to evaluate whether different routes to diagnosis were associated with different survival rates.

Prior to this analysis, researchers were able to observe the path taken from cancer screening to cancer treatment, but not the path to cancer screening. The analysis was based on the linking — via the unique National Health Service number assigned to each patient in England — of:

- *the Administrative Inpatient and Outpatient Hospital Episodes Statistics dataset*
- *the National Cancer Data Repository*
- *the National Cancer Waiting Times Monitoring dataset*
- *data from the National Bowel Screening programme*
- *data from the National Breast Screening programme.*

This analysis determined that, of the cancer diagnoses in the linked datasets, on average almost 25 per cent were diagnosed following presentation at an emergency department (rather than through other routes, such GP referral). The analysis also found that for all cancer types, 1-year survival rates (from the date of diagnosis) were significantly lower for emergency presentations than for other routes to diagnosis — the magnitude of the difference was typically in the range of 20-40 per cent.

Policy interventions implemented on the basis of this research were successful in reducing the proportion of diagnoses through emergency presentations to about 20 per cent.

Undertaking similar analysis in Australia would require linking of data held by a range of groups, including data from Medicare Australia, the Commonwealth Department of Health and its counterparts in the states and territories, various cancer registries and other organisations (appendix D). Given these challenges, it is perhaps not surprising that similar research has not yet been undertaken in Australia.

Source: Elliss-Brookes et al. (2012).

Data use can enable better government

Open data policies can enhance transparency of governments, leading to improved policy outcomes and providing the incentives and means for governments to be more efficient. The UK Cabinet Office (2013) noted the possibility to improve governance:

> Providing access to government data can empower individuals, the media, civil society, and business to fuel better outcomes in public services such as health, education, public safety, environmental protection, and governance.

Widespread release of government data can lead to more engaged and empowered citizens, resulting in greater participation and improved public debate (Bureau of Communications Research 2016; UK Cabinet Office 2013).

The Commission's Report on Government Services (SCRGSP 2016) is an example in this respect. It collates and presents a range of information (across thousands of data points)

related to the delivery of government services, and among other things provides a basis for cross-jurisdictional comparisons to be made.

The Commission's assessment is that this transparency has led governments to persistently consider ways to improve their performance and service delivery.

Improved transparency

Greater scrutiny of government spending can lead to significant social benefits. Data can facilitate comparative performance monitoring of agencies and agency employees, and — knowing that performance and quality of services will be publicly observable — encourage improvements in the quality of services provided. This is particularly the case where data allows citizens to benchmark and compare the performance of different service providers. The My School website in Australia is an example of where data is used to benchmark different service providers (in this case schools) (PC 2012). Not surprisingly though, the downside of this increased capacity to observe performance is that it can result in extreme reluctance to release agency-level data.

Transparency can also lead to increased scrutiny of individual lawmakers. In the United Kingdom, TheyWorkForYou uses open data and information from official UK Parliamentary sources to follow and track the activities of members of Parliament, including their comments in debates and their legislative voting record, which is made available in an easily understandable format. Similar organisations have sprung up in other countries, including Sunlight Foundation in the United States, and OpenAustralia in Australia (Chui, Farrell and Jackson 2014; OpenAustralia nd).

Data on public sector contract procurement can help to expose corruption. The Brazilian government releases a range of data, including that related to government expenditures, expenses of elected officials, and companies that are blacklisted from public contracts. This data has been used by journalists and activist groups to expose corruption (Chui, Farrell and Jackson 2014).

Better decision making

Data can be used by governments to improve policy and decision making. The Grattan Institute (sub. 12) suggested that the provision of standardised data related to major infrastructure projects — such as costs, benefits, timing, funding and financing arrangements, risk allocation, and procurement approaches — could substantially improve decision making and facilitate cost benchmarking. Similarly, public release of business cases for individual infrastructure projects could create incentives for better decision making. Another example is the use of data to improve management of transport networks (box 2.3).

Box 2.3 Toll road data in Singapore — setting an example for Australia

In 1998, Singapore replaced its existing coupon-based road pricing system with an electronic road pricing (ERP) system, which facilitated the introduction of variable pricing (Olszewski and Xie 2006).

In 2006, the Land Transport Authority (LTA) began working with IBM to improve the accuracy of its traffic predictions by trialling IBM's Traffic Prediction Tool in Singapore's CBD (Singapore LTA 2008). An aim of the trial was to explore the feasibility of highly variable pricing within the ERP system (ITS International 2010).

The system analysed data from a number of sources, including video surveillance cameras, GPS devices (including those installed in Singapore's taxi fleet), road charges, and street embedded sensors (ITS International 2010). The prediction tool was able to predict traffic volumes and travel speeds ten minutes ahead with an accuracy exceeding 85 per cent, and above 90 per cent in peak periods (ITS International 2010; Singapore LTA 2008). This was complemented with an algorithm designed to 'fill in' traffic data on sections of the road network not monitored, to allow for accurate and detailed route guidance across Singapore's entire road network (ITS International 2010).

This data also forms the basis for quarterly reviews of ERP rates, which are set for 30 minute blocks and set differently for different roads. This allows the LTA to not only reduce total traffic volumes, but also reduce the severity of peak period congestion by spreading traffic volumes (Murray 2012).

In Australia, the issue of variable, time of day road pricing has been raised in previous inquiries (for example, the New South Wales Inquiry into Road Access Pricing). To date, only two toll roads (the Sydney Harbour Bridge and Tunnel roads) have implemented variable time of day pricing. A comprehensive road access pricing system would have the potential to improve equity, decrease congestion, improve infrastructure investment decisions, and lead to lower levels of air pollution. The onus, however, would be on governments to set access charges with such objectives in mind.

In this sense, the Singapore ERP system provides a possible model of how toll road data, along with data from other sources, could be used to improve the management of road infrastructure in Australia.

Improved service delivery and data management

Governments can utilise data to generate insights that enable them to reduce the costs of, and improve efficiency and productivity in, the provision of services. In its submission to this Inquiry, the Telethon Kids Institute (a funder of medical research) stated that:

> Obtaining and analysing public sector data can enable a proper evaluation of whether services are of value, are cost effective, and are useless or even harmful. (sub. 5, p. 2)

One way in which this occurs is through benchmarking the costs of different programs and policies (Manyika et al. 2013). Following the 2014 Productivity Commission Inquiry into the costs of infrastructure, the Bureau of Infrastructure, Transport and Regional Economics published data from the states and territories on the capital costs of new infrastructure for

benchmarking purposes (BITRE 2015). This provides a basis for states and territories to improve procurement processes and lower procurement costs.

Big data techniques could also assist governments to better understand the needs of different groups of citizens, and thereby provide personalised services (Australian Government Department of Finance and Deregulation 2013). Data can be used, for example, to assess whether individuals are eligible for government entitlements (regardless of whether the individual has applied to receive them) (Australian Government Department of Finance and Deregulation 2013). New Zealand has used its integrated data holdings to improve outcomes in the delivery of human services (box 2.4).

Box 2.4 Using integrated data to deliver better support for at risk youth in New Zealand

The New Zealand Treasury is using longitudinal data from the Integrated Data Infrastructure to identify youth at risk of poor outcomes in adulthood, based on analysis of a specific cohort of young people. Specifically, the researchers used anonymised linked administrative datasets containing cohort-specific data about:

- welfare benefits
- interactions with the Department of Child, Youth and Family (related to care and protection)
- corrections sentencing
- schooling and tertiary participation and achievement
- births and deaths
- usage of mental health and addiction services and the use of mental health pharmaceuticals
- salaries and wages
- movements in and out of New Zealand.

Researchers linked youth cohorts with members of an adult cohort by matching characteristics such as receipt of welfare benefits, correction sentencing rates, gender and ethnicity to estimate the likely longer term outcomes.

By determining key characteristics that appear predictive of poor future outcomes, the analysis has provided valuable insights into the effectiveness of various policies and interventions — a necessary first step to improve the outcomes possible for at risk youth.

Source: McLeod et al. (2015).

In some sectors, understanding the needs of citizens may require governments to supplement administrative data holdings with information held by private entities, such as private health insurers.

Broader sharing or public release of public sector data holdings could lead to cost savings flowing from reductions in data collection efforts and improved data quality (chapter 6), by:

- reducing data management costs where agencies could reuse data (from other agencies) rather than individually collecting and maintaining data themselves (Australian Government Department of Finance and Deregulation 2013)
- reducing the burden on individuals providing information to governments (Department of Social Services, sub. 10)
- improving the accuracy of data by crowdsourcing correction of errors (Government 2.0 Taskforce 2009).

There are also examples of agencies using new sources of data to improve their processes. The ABS in recent years has collected comprehensive data on grocery prices for the Consumer Price Index (CPI) in an electronic format directly from retail outlets (such as supermarket chains), rather than via its traditional survey approach. This has allowed the ABS to improve the accuracy of CPI estimates, which are particularly important as a basis for monetary policy.

2.2 High value datasets

Amongst other aspects, the terms of reference for this Inquiry asked that we consider the determination of high value datasets. While high value datasets may be held by either the public or private sector, the emergence of data markets and brokers (such as Quantium and Data Republic in Australia) and profit incentives for private sector data holders to make their data available (where doing so is in their commercial interests) means that there are unlikely to be systemic issues in determining high value private sector datasets. Those seeking to share data face price signals (or value exchanges) that can allow them to assess which of their data holdings are potentially valuable to other parties. Entrepreneurs and investors can trigger reassessments of the value of private data holdings, such as the value of patents and research holdings of firms in receivership (Nortel and Kodak are examples in this regard).

While there may be a role for governments in encouraging private sector data holders to make their data available, the far bigger challenge for governments lies in determining what is a high value public sector dataset and what is a priority for public release.

Proponents of open data have expressed views that the default position of governments, at least, should be to release all non-sensitive data holdings (see, for example, OECD 2014). Setting aside the resource costs of doing so, an *a priori* view that governments should make publicly available all data holdings, subject to privacy and other concerns (such as national security), would appear to be a reasonable position — and the Australian Government's *Public Data Policy Statement* provides a mandate for agencies to do just that (Department of Prime Minister and Cabinet, sub. 20).

A counter viewpoint recognises that there is a range of factors that create challenges for governments seeking to increase the availability of government data (box 2.5) and so

governments should prioritise for release those datasets that are likely to contribute significant value to the economy and society (so-called high value datasets). This raises the question of how governments can determine which datasets are likely to be of most value.

Box 2.5 Why don't governments just release all non-sensitive data?

Non-sensitive data that governments could release potentially includes data of very high value, data of very low value, and everything in between. Releasing low-quality data can swamp release sites and make it more difficult to search and find useful datasets. Moreover, bulk release can diminish the apparent value of data, particularly if agencies simply release whatever they have whenever they have it.

Apart from this challenge, Inquiry participants (such as the Office of the Information Commissioner – QLD (sub. 42), the NSW Government (sub. 80) and the Cancer Council Australia (sub. 141)) suggested that culture and risk aversion within government agencies, as well as concerns about data being misinterpreted and misrepresented, means that agencies are reluctant to release data and thus it can be challenging for users to secure access to government data holdings (see chapters 3 and 5 for more details).

While some data cannot be released because it identifies individuals or businesses (chapter 5), the *Privacy Act 1988* (Cth) is often unwarrantedly cited as a barrier to releasing government data (Office of the Australian Information Commissioner, sub. 200). More often, it is specific provisions in individual Acts — or the cautious interpretation of them — that are responsible (chapter 5).

Finally, agencies can face costs in preparing and cleaning datasets for release (chapter 6), which can further inhibit the availability of government-held data. Failure to prioritise high value datasets means time and money will be spent on easier, low cost datasets, which in turn are likely to 'underwhelm' on release.

This inevitably raises the prospect that if release becomes the preference for a particular dataset, rather than retention for administration, then the method of collection may need to change rather than the dataset being constantly curated for release. The cost of such a move then falls not only on the agency but on the submitters of data. Overall, there is little doubt that significant costs may be implied in a move towards open by default.

The challenge of determining high value public sector datasets

In comparison, private sector business models are rarely used by (or even usable for) governments as part of their data management policies. Consequently, valuing public sector datasets is often very difficult.

Public sector data holders do not generally receive market-determined price signals to assist them in identifying high value datasets. Most agencies would not regard seeking such price signals, or valuing datasets, as 'core' business. Cost recovery for releasing data is practiced at times, but by definition is not related to value but to cost; very occasionally efforts are made to price according to the benefit a user might make of the data (Bureau of Meteorology, sub. 198) but these cases are the exception rather than the norm. The

implication is that the price of data sold by the public sector in Australia is not broadly indicative of the value of that data or even of its priority for release.

For both private and public data holders, however, many innovative uses of data cannot be valued now as they are not yet envisioned. Likewise, the way in which a particular dataset is able to be used can influence the value of subsequent uses of that dataset and of other datasets. The value of a particular dataset will also be influenced by the ways in which it can be linked and/or integrated with other datasets (Telethon Kids Institute, sub. 5; Health Research Institute – University of Canberra, sub. 115).

In other words, the value of data is not necessarily about the revenue streams it generates. The potential to trigger innovative investments or opportunities for better governance are all considerations that could easily outweigh any revenue generation, even if government datasets were priced according to market forces. Moreover, some opportunities are not likely to be amenable to monetary valuation alone (such as those related to health research). The benefits associated with those uses can only truly be understood once data has been released and users have had opportunities to discover its potential.

These issues raise a fundamental question about how governments can determine, and prioritise for release, high value datasets. While assessment of which datasets are high value clearly relies on input from end users, a necessary precondition for determining the most valuable datasets is an awareness of what data actually exists.

Any method of determining what is of high value must accordingly start with giving parties external to government much better information on what datasets are held by government. This requires data users to be able to access comprehensive and complete information (including metadata) on what datasets are held by government (chapter 3). In the absence of efforts in this area, researchers, not-for-profits and commercial entities cannot know enough to effectively participate in the determination of high value datasets.

What have stakeholders said they want access to?

Inquiry participants nominated a range of datasets as having the potential to contribute significant value to society. Broadly speaking, data on the administration of government policies and programs (so-called administrative data) was held up as being particularly valuable, since, as noted earlier, it is comprehensive and is not characterised by some of the shortcomings of survey data (such as narrow geographical coverage).

In the context of health policy:

- The Telethon Kids Institute (sub. 5) pointed to a range of administrative data types that would be useful for policy assessment, including health, education, training, employment, housing and environmental data. In particular, it proposed that linking data from the Pharmaceutical Benefits Scheme with other health data would provide a more effective method of detecting adverse effects from the use of specific pharmaceuticals. This was supported by others (for example, the Centre for Big Data

Research in Health – University of NSW, sub. 21, and the Australian Institute of Tropical Health and Medicine (AITHM), sub. 52).

- The Telethon Kids Institute (sub. 5) also pointed to the value of privately held data, such as that held by private health insurers and supermarkets (which could be useful for deriving insights into the consumption habits of individuals).
- The Australian Institute of Tropical Health and Medicine (AITHM) (sub. 52) highlighted the importance of linking administrative data held by:
  - Queensland Health, including patient, emergency department, perinatal and cancer registry data
  - health and hospital services, including patient flows and discharge medication data
  - the Australian Department of Health, including data from the Medicare Benefits Schedule and Pharmaceutical Benefits Scheme.
- Health insurers, such as Australian Unity (sub. 95) and Medibank Private (sub. 98), echoed the importance of being able to access data from the MBS and PBS.
- Australian Unity also flagged the potential for data relating to service provider cost and performance, as well as de-identified linked data about service recipients, to lead to more effective and targeted interventions (from health insurers), lower premiums and improved health outcomes. An example of such an intervention is Australian Unity's Mindstep mental health program, which has shown strong results in preventing hospitalisation due to anxiety and depression.
- The Australian Dental Association (sub. 8) highlighted that access to granular private health insurance data could allow for new dental practices to be established in areas of high demand, and thereby enhance competition.

Several financial businesses, including a number of fintech firms, nominated a range of private and public sector data sources that could increase competition in the financial sector (such as by facilitating customers switching between service providers), enhance consumer choice, help businesses meet responsible lending obligations, assist them to satisfy regulatory requirements to identify users of financial services, and improve the efficiency of financial market risk pricing:

- Comprehensive credit reporting data, customer transaction data, and data held by regulators (such as data on the loans underpinning Residential Mortgage Backed Securities held by the Reserve Bank of Australia) would help inform decision making in the financial sector and lead to more efficient pricing of risk (appendix E).
- Administrative data held by government agencies, including that in drivers' licence datasets and electoral rolls, could be used more effectively to help financial firms verify their customers' details (on an ongoing basis) with a higher degree of certainty (Australian Bankers' Association, sub. 93).

- The Australian Retail Credit Association (ARCA) (sub. 87) noted that data related to individuals' government debts would provide a richer picture to credit providers of an individual's credit obligations.
- Similarly, the Insurance Council of Australia (sub. 66) noted several types of government held data that would be valuable to insurers, including data related to natural hazards, building standards, mental health, policies and claims, and driving records (such as demerit points).

A broad range of data that would enhance the value of research in Australia was nominated:

- The Australian Urban Research Infrastructure Network (AURIN) (sub. 116) pointed to the value of fine-grained spatial data (related to specific properties, neighbourhoods or individuals) across a range of topics, including transport, economic planning, population health and wellbeing, energy and water use, and innovative urban design.
- The Cooperative Research Centre for Spatial Information (CRCSI) (sub. 43) echoed the value of spatial data, drawing attention to the value that can be added to a range of datasets through geocoding (a technique for adding or linking spatial data to datasets).
- Monash University (sub. 133) pointed to the value of health, census, welfare, justice, environmental and education data, as well as data from the electoral roll, with specific examples including the Australian Cancer Database and the National Death Index.

Participants reported a number of datasets that would be particularly valuable for improved delivery of government functions and programs:

- The Federation of Ethnic Communities' Councils of Australia (sub. 16) suggested that public policy and service delivery to migrants could be improved if government agencies collected, and linked with a range of indicators (such as income and gender), data on: country of birth; the primary language spoken at home; religious background; ethnicity; and English language proficiency.
- The Commonwealth Grants Commission (sub. 58) suggested that access to a linked set of hospital and Census data would allow them to better understand the hospital funding needs of different states.
- The NSW Government (sub. 80) pointed to a range of data that could be useful for various government agencies, including real estate data to guide the provision of public housing, spatial data to assist emergency preparedness and recovery, and financial data to detect and prevent fraudulent activities.
- The Department of Social Services (sub. 10) highlighted that data created by a proposed single touch payroll system (to be managed by the Australian Tax Office) would contribute significant value to other Australian government agencies — such data could be useful for detecting undeclared employment income (which influences social security payments).

We note that there was a greater focus among Inquiry participants on high value datasets held by the public sector (as opposed to the private sector datasets). The nominated datasets span a broad range of areas and include datasets related to health, natural hazards, education and welfare. AURIN (sub. 116) noted that most of the datasets nominated as being of a high value to the research sector are held by governments. A key point to note from the above, of course, is that simply asking the question produces an already long list of opportunities.

Characteristics of high value datasets

In forming a view on the characteristics of high value datasets, it is important to recognise that there is no universal definition of what constitutes a high value dataset (Department of Social Services, sub. 10; University of Melbourne, sub. 148). What is highly valuable to one party might not be as valuable to another.

While acknowledging the inherent subjectivity of the notion of 'high value', the concept can be characterised as having both a 'use' element (what data can be used for will impact its value) and a 'quality' element.

It is possible to discern a number of likely characteristics around 'use' that high value datasets *might* possess, including that they:

- are unique (in the sense that there are no suitable substitutes or that they could not be easily replicated) (University of Sydney, sub. 35)
- contain unit record level data (which can be particularly useful for evaluating the effectiveness of particular policies) (Telethon Kids Institute, sub. 5)
- have a high degree of coverage in the population of interest — which minimises issues around sampling bias and allows for analysis of small and vulnerable groups (Telethon Kids Institute, sub. 5; Curtin University, sub. 41)
- have been designed for linking with other datasets (Queensland Government, sub. 207), or use identifiers to allow linking with other datasets (Telethon Kids Institute, sub. 5) (though there are techniques that can be used to link datasets without relying on a unique identifier — see appendix B)
- are central to service delivery and/or core decision making (Victorian Department of Treasury and Finance 2015)
- contain time-specific data that allows for comparisons to be made over time (Queensland Government, sub. 207; Geoscience Australia, sub. 211)
- have a high potential for use and reuse, and a large potential user base (Archer et al. 2014).

Moreover, characteristics that are indicative of quality could include that datasets:

- are current (real-time) and/or updated regularly (University of Sydney, sub. 35; Queensland Government, sub. 207)

- are accurate and complete (Telethon Kids Institute, sub. 5)
- contain clear, consistent definitions (Telethon Kids Institute, sub. 5)
- provide details on data quality, lineage and provenance (Queensland Government, sub. 207).

As data analytics continues to evolve, the characteristics that make a particular dataset valuable must also evolve. Adherence to central agency guidance, for example that from the Department of Prime Minister and Cabinet on the use and updating of 'dataset identifiers' with machinery of government changes, could be a worthwhile part of any framework for determining high value datasets. That said, consistent use of sufficiently descriptive metadata could be all that is required for data users to understand the properties of particular datasets, and thus draw conclusions about their value.

While it is possible to broadly determine characteristics of valuable datasets, as noted above, the myriad ways in which datasets can be used, and therefore the benefits associated with those uses, can only truly be understood once data has been released and users have had opportunities to experiment.

Datasets that are of national interest

While there might be many datasets that are of a 'high value', not all of them will necessarily be used to generate benefits beyond particular groups in the community. There will be some high value datasets, however, that may be used to generate substantial benefits across such a broad swathe of the Australian population that they may be thought of as being of 'national interest'. Datasets that are of national interest will be a subset of high value datasets.

There are several criteria that could signal that a specific dataset — whether already publicly available or not — might be of national interest.

First, the dataset is of interest to a broad range of users and its use would be likely to generate broad economic and social benefits beyond those accruing to data users, holders and contributors. This could be because the dataset directly enables innovative and beneficial uses, or because the dataset enables other datasets to function more effectively. For example, the Queensland Government (sub. 207) noted several specific datasets that are enablers for linkage of other datasets, including the Australian Statistical Geography Standard and the Geocoded National Address File. Geoscience Australia (sub. 211) highlighted the importance of data from the national positioning infrastructure, noting that it underpins all spatially referenced data in Australia.

Second, the dataset (or datasets) can be used as a basis for comparison between states and territories. Selected datasets used by the Productivity Commission to prepare the Report on Government Services could be illustrative in this respect. Other examples could include datasets pertaining to health outcomes — such as data collected by the Australian

Government as part of administering the Pharmaceutical Benefits Scheme and the Medicare Benefits Schedule. Datasets collected and maintained by the states and territories, such as those related to education outcomes, could similarly be of national interest.

Finally, datasets that have a national focus, such as those related to immigration or key macroeconomic indicators, could be of national interest. The ABS Consumer Price Index is an obvious example.

While created on the basis of different criteria, the ABS's register of Essential Statistical Assets for Australia may include some examples of national interest datasets. The intent of the ABS in developing the register was '… to identify the core set of essential statistical assets that are critical for decision making for the nation' (ABS 2013, p. 1). The set of criteria applied by the ABS to develop a preliminary list of essential statistical assets included:

- application in public policy
- importance to key national progress measurement
- domestic electoral or legislative requirement
- international reporting obligation and/or being critical for international comparability.

The list consists of 74 essential statistics, with 178 distinct datasets, including those related to:

- business performance and structure — such as business demography and freight movement
- competitiveness — such as productivity and the exports and imports of goods and services
- household economic wellbeing — such as income and wealth
- housing — such as housing activity and affordability (ABS 2013).

The approach adopted by the ABS in developing the register is a practical illustration of how governments could determine some high value datasets of national interest. Governments could also draw on international developments to guide them in determining which datasets are of national interest.

Methods to understand demand for public datasets

Whether of national interest or of particular focus to a group of external users, understanding value as a way of prioritising action (and expenditure) is essential. And value is generally not equivalent to cost, as discussed further in chapter 7.

The level of demand for datasets can be used to infer their likely value. There are different methods governments can use to gauge demand. These include:

- Surveys of known user groups (Headd 2016) — for example, as part of the Australian Open Data initiative, the Australian Government surveyed Australian businesses and non-government organisations to provide a basis for assessing the value of government open data, and to identify how government data could be made more useful.
- Reviews of previous data requests, including those made under freedom of information legislation, to determine frequently requested datasets (Headd 2016; Victorian Department of Treasury and Finance 2015).
- The use of feedback mechanisms on data portals that allow users to suggest and vote on datasets for release (Victorian Department of Treasury and Finance 2015) — for example, the Australian Government's data portal (data.gov.au) provides this functionality.

While it may be useful to employ methods that do not directly involve feedback from users — such as active monitoring of government websites to understand what users are searching for — methods that facilitate direct feedback from users are much more likely to reveal valuable datasets. For example, many jurisdictions publish a list of government data holdings on dedicated open data websites (such as data.gov.au or data.vic.gov.au) and allow users to suggest and vote on datasets to be prioritised for release.

Emerging data sources could also provide insights into which datasets should be considered for prioritised release. For example, Google's Public Data Explorer — which allows users to link and visualise data from a range of sources — could provide insights into datasets that would add value through linkages with other data.

In addition to the methods outlined above, governments should also liaise internally to assess which datasets are required to resolve particular policy issues. The NSW Data Analytics Centre provides a model of how such an approach could work in practice (chapter 3). In some instances, Commonwealth, state and territory, and local government open data platforms are linked through search result sharing (Department of Prime Minister and Cabinet, sub. 20).

There is room for improvement

Despite these elements of progress, and as noted by various parties (such as the Department of Prime Minister and Cabinet, sub. 20), there is considerable room for improvement in how governments determine datasets that could be valuable to users and the community more broadly.

First, not all government data holdings are discoverable through open data platforms. Many Australian Government entities do not make their data holdings discoverable on data.gov.au, the Australian Government's primary data portal. As noted in chapter 3, many agencies do not even publish registers of their data holdings on their own websites.

Second, some government datasets are not published in fully structured formats (which is important for automated exchange and use of data). For example, the ABS continues to publish key data only in semi-structured formats (Centre for Policy Development, sub. 11; Queensland Government, sub. 207). This approach is brittle — small spreadsheet formatting changes can lead to significant errors in data accuracy when automated processes are used to collect and process data.

Third, data custodians often lack the skills or expertise to best determine which datasets would be useful to address outstanding policy issues (NSW Government, sub. 80).

Finally, while governments employ various strategies for identifying the needs of government and non-government users, it appears that they are largely applied in an ad hoc manner, and it is not clear that they account for the value created through the linking of disparate datasets, including those held by different agencies and in different jurisdictions.

A framework for assessing candidate datasets for release

The need for a systematic framework to determine which government datasets should be released was previously recognised by the Department of Prime Minister and Cabinet (2015). The Department is currently developing a framework to '… assist [Australian Government] entities and data custodians in identifying high-value datasets for priority release' (sub. 20, p. 10), and intends to draw upon the conclusions made in this Report on the characteristics of high value datasets as part of developing the framework.

A suitable framework should incorporate measures to improve discoverability of datasets and elicit feedback from data users and the community more broadly (Office of the Information Commissioner – QLD, sub. 42). It should also ensure high levels of transparency and accountability of agencies and data custodians.[8] This approach can loosely be thought of as leveraging 'crowdsourcing' to determine high value datasets.

The framework outlined below would be suitable for determining which data should be *considered* for prioritised release. However, risk management would also need to be considered prior to datasets being made publicly available.

Getting data user input to determine priority releases

Government agencies have been using, with limited success, a range of measures to understand the value of datasets that may be released, including facilities for potential users to request data on open data platforms and periodic surveys of users.

[8] There are recommendations in chapter 3 of this Report related to government agencies creating registers of their own data holdings (described with appropriate metadata) and broader measures to make sure those data registries are discoverable.

The Australian Government recently simplified governance around data sharing and release, which included the formation of two bodies — an External Reference Group made up of experts from academia, non-government organisations and business, and a Deputy Secretaries Data Group. These bodies provide clear channels for some data users within and outside of government to signal which government datasets should be considered for prioritised release.

State, territory and local government agencies do not appear to be directly engaged with these bodies. It would be desirable for that to be addressed, given the value of data holdings in those entities.

Moreover, it may not be the case that the External Reference Group would be an effective mechanism for identifying *all* datasets that would have a high commercial value, since it might not encompass all data users (including new innovative start-ups). An expansion of current consultation efforts, to address these matters, would strengthen the capability of the Australian Government to determine valuable datasets.

Accordingly, the Australian Government should develop a systematic framework for broadening the range of mechanisms used to determine valuable datasets, with a particular focus on mechanisms that seek feedback directly from data users.

On data.gov.au, users are already able to suggest datasets for release and vote for datasets suggested by others. This offers a mechanism for ranking datasets for release. However, popularity is not necessarily the best indicator of the value of a particular dataset — it might simply reflect that there are many potential users, not that the intended uses are of significant value. As such, we consider that while data.gov.au provides an important avenue for feedback from data users, there is potential to expand on this concept by allowing data users to provide greater context when requesting access to specific datasets.

One approach is to establish a formal framework within which prospective data users advocate for release by describing to governments the scope for gains from release. This could involve data users (groups or individuals) submitting detailed proposals on how — if curated and released[9] — the data in question would be used and the expected benefits (private and public), which would provide a basis for agencies to determine which datasets to release first.[10]

Each of these approaches can be thought of as inherently competitive — data users are vying for public agencies to invest their limited resources to curate and release specific datasets. In assessing these proposals, and ultimately determining which datasets should be

9 Chapter 7 discusses the factors that government agencies should consider when deciding to curate datasets before release.

10 As noted earlier, quantifying the benefits of data use is difficult. In assessing such proposals, government agencies should carefully assess whether the estimated benefits are reasonable, and place a greater weight on estimates that directly relate to the intended use, rather than estimates of the wider economic benefits (which tend to be more speculative).

released first, government agencies might also consider whether data users have sufficient resources to use the datasets, as outlined in the detailed proposal.

Such processes might be viewed as too slow by some entrepreneurial spirits (such as tech start-ups). However, while streamlining of data release may be warranted in some limited circumstances, on the whole, seeking views more broadly from companies, researchers and other users on what data to release (and why), and authorising agencies to consider these in a systematic way, is less likely to lead to clunky processes and poor data release outcomes.

In the interests of transparency, proposals should be published openly (such as on open data forums). However, in some cases, businesses and other parties (such as academics) might wish to submit confidential proposals (such as due to commercial confidentiality). Governments should accommodate data users in circumstances where there are legitimate confidentiality concerns.

As a matter of principle though (and to remain consistent with an open by default approach), where government agencies release a dataset to a particular data user, the dataset should also be released publicly on data.gov.au, unless there are legitimate reasons for withholding public release.

While feedback mechanisms are a vital component of this framework, the success of such feedback approaches in leading to data release is likely to be limited if not led by an accountable, effectively resourced entity that is supported by substantial political will.

Adoption of this framework by government agencies would signal a commitment to a significant cultural shift, in which agencies embrace exposure of their data holdings and engage with motivated data users to understand what data is of value. It does not oblige agencies to respond to external demands but it does advertise and require that they are open to those ideas.

An important aspect of such a framework is that it is applied on an ongoing basis — users should be able to submit and vote on data requests on data.gov.au and submit detailed data request proposals whenever they wish.

Supplementing this ongoing expanded feedback system could be periodic data 'competitions', in which interested parties *publicly* and formally submit data requests (box 2.6). A central authority could be responsible for undertaking data competitions on behalf of the Australian Government and its agencies, and have responsibility for publishing the results of this process.

Under *current* Machinery of Government arrangements, responsibility for obtaining feedback from data users (on either an ongoing basis or through periodic events) would reside with the Department of Prime Minister and Cabinet. State and territory governments could consider developing their own frameworks to be implemented by a central agency at the jurisdictional level.

Box 2.6 Data competitions

Formalised data competitions could be useful for drawing attention to the Australian Government's high value dataset framework, demonstrating the benefits that can accrue from innovative uses of government data, exposing innovative ideas for using data, seeing winners declared and ultimately resulting in the release of high value datasets. Additional funding could be provided, as required, to relevant agencies for the purposes of releasing the 'winning' datasets.

These competitions would be centred on proposals submitted by data users, detailing the datasets required (including proposed linkages), intended uses, and anticipated outcomes and benefits. Proposals would be released publicly, which would allow others to learn from the ideas put forward.

Given the benefits associated with public sector data use (such as in public policy design), Australian, state and territory government agencies should be able to submit their own proposals alongside other data users, such as those from the commercial and research sectors.

Current initiatives, such as GovHack, provide a potential template for running such competitions, with a few key differences. First, GovHack involves a pre-determined list of datasets, whereas the competitions proposed here would allow users to suggest datasets. Second, the timeframes involved would necessarily differ in order to allow participants sufficient time to develop and submit detailed proposals, and for organisers to assess submitted proposals.

Transparency and accountability

There is a range of legitimate reasons for governments withholding datasets, including privacy and commercial-in-confidence concerns. That said, Inquiry participants have also pointed to government inertia and risk aversion as unproductive and addressable barriers to greater access to government data. And the experience of the Productivity Commission itself in seeking data access, as it does regularly, is strongly consistent with this failure of culture.

If treated as a project led by a central government agency, a framework for prioritising datasets for release would see agencies resourced to deliver datasets or linkages within programmed timeframes. Failure to release high value datasets on time, or in line with the original request, would be considered as a failure of the agency, and would trigger remedial efforts to address this. Exceptions would apply only if there were legitimate reasons for not providing particular datasets.

As with other large cultural and technological shifts in response to digitisation, setbacks are inevitable in some early projects. The new structure should account for this and encourage agencies to learn from these setbacks in order to transition to a culture of openness.

Data custodians should also have the remit to consider alternative ways to satisfy the purpose of a data request, including through the provision of appropriately de-identified data (chapter 5). Where a data request is not met due to a legitimate reason (such as

privacy), agencies should consider whether there are other data assets that could be useful in meeting the purpose of the request.

Given the broad and often unforeseeable ways in which data can be used, a shift away from current legislative arrangements that incentivise data custodians to be wary — if not resistant — to data release (or data access for research) that is not squarely within the lore of each agency's operations may be necessary to fully prompt a broad cultural shift in public sector data management (chapter 9).

DRAFT RECOMMENDATION 2.1

In determining datasets for public release, a central government agency with policy responsibility for data should maintain a system whereby all Australian governments' agencies, researchers and the private sector can, on an ongoing basis, nominate datasets or combinations of datasets for public release, with the initial priority being the release of high value, in-demand datasets.

A list of requested datasets should be published. Decisions regarding dataset release or otherwise, and access arrangements, should be transparent. Agencies should provide explanations where priority datasets are not subsequently released on legitimate grounds. Where there are not legitimate reasons for withholding requested data, remedial action should be undertaken by the Australian Government's central data agency to assist agencies to satisfy data requests.

Existing government data initiatives, such as data.gov.au, should be leveraged as part of this system.

3 Public sector and research data collection and access

Key points

- The public sector holds a vast quantity of data collected from individuals, businesses, and other government departments, and generated from publicly funded services and scientific research.
  - Governments make use of this data for activities such as providing community services; administration of payments; response to emergencies; compliance monitoring and law enforcement; research; and on occasion, policy development and evaluation.
- Australian governments lag other comparable countries in sharing and release of public sector data. There is scope for governments to do more to increase access to their data.
  - In terms of *open* access (or public release), there are some encouraging developments in Australia but significant gaps remain for many types of data, particularly in the health sector. A much larger number of agencies should be publishing and curating their information online in machine-readable formats.
  - In terms of *restricted access* between government agencies and certain trusted parties, sharing has been limited at best. There are many serious examples of data siloing, despite the obvious benefits from sharing.
  - Researchers and research funding bodies also unnecessarily restrict access to their data.
- The extent to which the usefulness of datasets is enhanced by linking or integrating them with other datasets varies substantially across jurisdictions. New South Wales and Western Australia have made good progress on this, whilst other jurisdictions lag. Integration of Australian Government datasets has been limited but improving with the Multi-Agency Data Integration Project and the Business Longitudinal Analytical Data Environment.
- In addition to impediments due to confidentiality and data interoperability, data sharing and release are limited by piecemeal bureaucratic processes, intellectual property issues, disruption from machinery of government changes and poor incentives facing the public sector to manage the risks while receiving little by way of reward or recognition.
- Decisions about public sector data release are not typically based on an assessment of the potential value of that release or a realistic assessment of likely risks. There is a culture of risk aversion that requires strong leadership and clear communication of aspirations, expectations and intent, if the value of public sector data is to be unlocked through greater access and use.

Public sector agencies collect and store a wide array of data, ranging from individual health records to Australia-wide maps. Much of this data remains largely confined to the agencies that collect it, and as a result it has substantial unrealised value. There is strong potential for the benefits of data use to be expanded through broader sharing and release.

This chapter examines the collection, use and extent of access to public sector data. It also discusses some factors that are limiting data sharing and release.

3.1 What data is collected and what is done with it?

What information is collected by public agencies?

Data from individuals

Australians provide a wide variety of information to a multitude of government agencies and service providers ranging from government line departments — such as the Department of Human Services (DHS) — to publicly funded institutions, including research institutions and government business enterprises. Among other things, information is collected on:

- identity (such as names, addresses, dates of birth, and family relationships)
- ownership of physical and intellectual property and other assets (such as vehicles and animals)
- activities undertaken and services used (such as public libraries, immunisation, transport, border control services and licensing)
- personal and family wellbeing (such as medical and educational facilities used)
- employment and income (such as wages and salaries, investment income, tax paid and income support).

Much of this information — such as that related to identity, education, work or travel history — is often provided many times over, even where the information relates to documents that have been issued by the government (for example, proving your identity using a passport and birth certificate to multiple different government service providers). Indeed, individuals are required to separately inform the Department of Foreign Affairs and Trade (DFAT) and the Department of Social Services (DSS) of their overseas travel for security and social service payment purposes.

Inquiry participants have commented that information sharing arrangements within government and 'tell us once' initiatives are not employed enough (Queensland Government, sub. 207). However, some attempts have been made. In 2011 the Australian government introduced a 'tell us once' pilot to allow changes in contact details to flow across agencies (Kennedy 2011) and more recently, the Digital Transformation Office (DTO) began implementing a similar 'tell us once' feature for myGov (Carrasco 2015).

Some data collection on individuals is undertaken specifically for research, such as: the Household, Income, and Labour Dynamics in Australia survey (a household-based panel data survey administered by the Melbourne Institute and funded by the DSS) and the Longitudinal Study of Australian Children (conducted by the Australian Institute of Family Studies in partnership with the DSS and the Australian Bureau of Statistics (ABS)). These

surveys collect information on individuals relating to a broad range of demographic, social and financial fields. The framework for the collection of this data is development by the National Centre for Longitudinal Data (box 3.1).

Box 3.1 National Centre for Longitudinal Data

The Department of Social Services operates the National Centre for Longitudinal Data to support the management of critical national data assets. A major review is being undertaken to determine what Australia's longitudinal data needs are for the future and the best way to meet these needs. A report on the review findings is expected to be finalised in October 2016.

To date, the review has identified several models for systematic management of longitudinal data assets into the future. The leading option allows for a coordinated approach to longitudinal data management while leaving custodians responsible for their data collections. A small team would be formed to manage the "system." This would include coordinating access arrangements advising Government on funding for the key longitudinal data assets under the guidance of a board of expert council.

Source: Department of Social Services (sub. 10).

Data from businesses

Businesses similarly provide a wide range of information to governments over the course of their operation.

- At their inception, businesses register names, identification numbers and addresses, as well as details of owners and directors.
- Businesses provide information on an ongoing basis for tax purposes, including data on sales, purchases and wages paid. For imports and exports, detailed information on products and suppliers may be required for regulatory purposes. If seeking government assistance, subsidies or contracts, further information on business activities will be provided.
- At closure, businesses cancel their names and registrations, providing governments with information on length of operation and reasons for closure, such as insolvency or personal bankruptcy. This information in turn contributes to registers such as those of disqualified business owners and companies.
- Many businesses have to comply with industry-specific regulatory reporting requirements. Financial institutions have to report particular transactions to AUSTRAC, and authorised deposit taking institutions, general insurers and registered financial corporations are all required to report information on financial performance to the Australian Securities and Investment Commission.

Much of the information that businesses provide to governments is provided many times over. The Commission (2009) has previously noted wide-spread burdensome, duplicative and redundant reporting requirements.

The Australian Government has adopted policies of 'tell us once' and/or 'no wrong entry point' for businesses to enable the sharing of basic information (such as business name, address, or owners' identities) across government agencies under some schemes. Such an approach is a key feature of Standard Business Reporting, for example, which provides an information standard and allows tax-relevant information to transfer across multiple agencies.

Businesses operated under a government-funded contract may also have particular obligations under the terms of their contract to provide data or information to government (chapter 4).

Other data sources

Governments and publicly funded institutions collect a range of data beyond that relating to individuals and businesses. In addition to the important role of the ABS, many other agencies and government bodies have a data collection and dissemination role. A wide range of macroeconomic statistics (including, for instance, foreign exchange information) are provided by the Reserve Bank of Australia (RBA), and Tourism Research Australia conducts regular surveys on Australia's tourism industry.

Some government bodies collect and release datasets containing spatial and geospatial information on the natural environment. Geoscience Australia publishes a wide range of data, including, for instance, hydrogeological maps of Australia. With regard to the natural environment, this includes collecting information on temperature, rainfall, humidity and other aspects of weather. Much of this data is collected through the Bureau of Meteorology (BoM), while water resources, resource and energy statistics, fisheries, forestry and agriculture statistics, and data on Australia's flora and fauna is also collected by other institutions. For example:

- *The Living Atlas of Australia* lists 201 data collections on fauna, insects, microorganisms and plants from 78 institutions, including some based in other countries.
- The Australian Government Department of Agriculture and Water Resources (DAWR) holds information on a wide range of forest, fishery and land-related data that is relevant to its policy remit.
- The Australian Government Department of Environment and Energy (DEE) holds data on environmental matters and biodiversity as a result of assessments and approvals it conducts under the *Environment Protection and Biodiversity Conservation Act 1999* (Cth).

With regard to infrastructure, information such as that on aviation, road usage, movement of cargo and freight is also monitored and reported on. In all of these areas, improving sensor technology and the ability for small devices to connect to the Internet (the so-called Internet of Things) can be used to supplement existing datasets and/or provide new sources

of insight. There are many examples of such data collection. Whilst forecasting the growth of new technology is difficult, projections for the number of devices comprising the Internet of Things by 2020 range from just under 20.8 million to around 38.5 billion (Gartner 2015; Juniper Research 2016). Examples include:

- The NSW Roads and Maritime Services has live traffic cameras in various locations of Sydney, with traffic images updated every 60 seconds.
- Similarly, VicRoads provides near real time access to information related to travel times, delay and congestion for several popular travel routes within Melbourne. The information is collected by BlueTooth receivers located throughout the traffic network.
- A network of seven motion sensor buoys (Waverider), located several kilometres off the NSW coastline, helps to determine onshore wave conditions and identify associated coastal hazards. Data from the buoys is regularly used by the BoM to issue marine warnings and by the NSW State Emergency Services to inform its response to coastal storms. The data is also publicly available on the Manly Hydraulics Laboratory Website (Gardiner 2015).
- The Commonwealth Scientific and Industrial Research Organisation (CSIRO) has developed Zebedee, a lightweight handheld 3D laser mapping system, which can create 3D laser maps as the operator walks through a site, and has also improved the way flying foxes are tracked via sensors — improving modelling of, for example, the spread of disease.

Regulatory functions and reporting obligations also generate significant amounts of data — for example, Australia's greenhouse gas emissions are reported and published under the *National Greenhouse Reporting Act 2007* (Cth), and IP Australia has a range of open data on application and applicant information across four intellectual property rights — patents, trademarks, design rights and plant breeders' rights. International agreements and treaties can also require Australia to provide information on particular matters — for instance, the World Trade Organisation requires Australia to report information on its temporary trade barriers.

In addition, governments and publicly funded institutions generate data from their own operations and service delivery. This can include information about their own performance, the performance of markets and the economy, research data, and personal information related to receipt of support payments, use of government services or taxation receipts.

Uses of public sector data

Without attempting to be comprehensive, governments' uses of data relates broadly to:

- *Administering payments and other government services* — Australian governments administer billions of dollars in payments each year (welfare payments totalled over 230 billion dollars in the 2015-16 budget). Given Australia's heavy reliance on

means-testing, administering these payments accurately requires (and generates) significant volumes of data.

- *Informing service provision* — the services that governments provide vary across jurisdictions and levels of government, but typically rely on data to enable targeting and effective service delivery. Some examples of this include:
  - delivering welfare services using information on names, addresses and family relationships
  - monitoring patient diagnoses and treatment history in hospitals to inform funding decisions under the activity-based funding arrangements, and further investment in hospitals
  - tracking student outcomes to inform teaching methods (VCAA 2016)
  - sharing patient data between service providers — for example, through MyHealth Record (box 3.2)
  - designing early interventions for troubled families by linking Commonwealth and state data to enable place-based insights into troubled families (NSW Department of Families and Children has expressed interest in being involved with this project) (DPMC 2015)
  - using data analysis to target government service provision to address long term welfare dependency (Doran 2016).

Box 3.2 Using data in the provision of health services

The Australian Government has recently introduced the MyHealth Record. This service creates an account that outlines patient information such as allergies, prescription of medicines, Medicare claims history, diagnostic imaging reports and pathology reports. This information is partially controlled by patients, but also used by health care providers to inform their care. (The sharing, use and transfer of data within the health sector is outlined in detail in appendix D.) It has been reported that, as at July 2016, more than 4 million Australians had signed up.

Source: Australian Digital Health Agency (2016b).

- *Performance monitoring* — governments use data to assess the effectiveness of government provided or funded services, such as public transport, and the operation of courts and prisons. This can involve measuring outcomes against benchmarks and program costs.
  - An example is the Commission's public reporting of equity, effectiveness and efficiency for six separate broad categories of government services: childcare care, education and training; justice; emergency management; health; community services; housing and homelessness.
- *Responding to emergency situations* — data is critical to emergency response, where states and territories usually have primary responsibility. Fire, police, ambulance and state emergency services all rely on data when preparing for, and responding to, natural disasters and other emergency situations — sometimes this requires sharing

information across jurisdictional boundaries. Access to data can be essential to allow respondents to communicate the location of incidents, and in many cases, assess the nature of the emergency. Examples include:

  - Mapping the spread of floods and bushfires for public safety — for example, to notify individuals which roads are covered by floodwaters. In Victoria, Vic Emergency assembles information from a wide range for sources to map the location of emergencies, along with additional information.
  - Using river flow, terrain and weather data for flood impact assessment (Geoscience Australia 2014). As noted in the Queensland Floods Commission of Inquiry, the Brisbane City council has a flood model that informs the majority of response and recovery activities (Holmes 2012).
  - Using spatial data on vegetation and weather patterns to predict fire spread (Bushfire CRC 2014). In Victoria, Queensland, New South Wales, South Australia and Tasmania, state bushfire respondents use a computer program, Phoenix Rapidfire, to assess the likely spread of fire and deploy resources based on that prediction.
  - Data on people who are immobile or otherwise at risk (such as existence of aged care facilities) can be used to prioritise evacuations.

- *Enforcing the law and regulatory schemes* — governments use data to not only develop regulatory regimes (risk-based approaches to regulation in particular rely on outcomes data), but also to monitor and investigate compliance and implement enforcement actions. For example:
  - sharing information on criminal offenders, such as fingerprint data and DNA records (Mobbs 2001)
  - assessing the likelihood of visa fraud using information on immigrants entering Australia (Tay 2012)
  - monitoring suspicious financial transactions to detect instances of money laundering and counter-terrorism financing (AUSTRAC 2016)
  - ASIC monitors data on all orders and trades in Australian equity markets and provides data on market characteristics (including volatility), measures of market concentration and measures of market efficiency.
- *Research* — researchers within government and publicly funded institutions use data for investigation across a broad range of fields, ranging from basic research in the natural science to policy-relevant research in the social sciences. In some cases, this involves using data collected specifically for research, whilst in other cases it involves using repurposed administrative data. Some examples research data uses include:
  - examining drivers of school attendance rates among children (Daraganova, Mullan and Edwards 2014)
  - increasing understanding of dynamics in Indigenous families (Butler et al. 2010)
  - modelling the efficiency of various taxation approaches (Cao et al. 2015).

3.2 The state of open access in Australia

Open data policies

Greater sharing and release of public sector data has the potential to bring a wide range of benefits to the community (chapter 2). Data provided without restriction and at no cost is often referred to as 'open data'. Open data portals often contain non-personal, aggregated, or heavily de-identified data (further discussion about de-identification processes is in chapter 5). For the purposes of this chapter, 'open access' and 'release' are used synonymously to mean allowing all members of the public to access data, potentially at a cost.

Open access to public sector data is receiving increasing attention by governments in Australia and internationally. A key driver of the open data movement has been the Open Government Partnership (box 3.3). Australia and a number of countries within the Open Government Partnership have set the goal of making public sector data 'open by default' (Turnbull 2015). In Australia, this has been the culmination of a series of initiatives (box 3.4).

Box 3.3 **The Open Government Partnership**

The Open Government Partnership is a multilateral initiative involving 69 countries. It aims to promote transparency and accountability of governments, while also facilitating technological innovation using government data. To join the Open Government Partnership, governments must endorse an open government declaration, deliver a country action plan and commit to independent reporting on the progress of their open data initiatives.

The open government declaration is a letter in which governments commit to upholding the principles of open government: increasing the availability of information about governmental activities; supporting civic participation; and increasing access to new technologies for openness and accountability.

National action plans outline commitments to open government initiatives and span a two year period. Governments must develop National Action Plans through consultation with the wider community and are recommended to include between 5 and 15 ambitious commitments. (At the time of writing, Australia's National Action Plan was under development.)

Independent reporting of commitments takes places through the Open Government Partnership Independent Reporting Mechanism. The Independent Reporting Mechanism annually examines the development and implementation of national action plans.

Source: Open Government Partnership (2015).

Box 3.4 Key Australian Government open data initiatives

For the Australian Government, key developments since late 2009 include:

- In December 2009, the Australian Government released *Engage: Getting on with Government 2.0*, which recommended declaration of open government and making public sector information open by default (Government 2.0 Taskforce 2009)
- In July 2010, Australia declared open government and announced plans to join the Open Government Partnership, however progress on this stalled soon after.
- In July 2013, Australia created data.gov.au, a portal for the publication of open data (Waugh 2013).
- In December 2015, the Australian Government released the Australian Government Public Data Policy Statement. This declared open access as default for non-sensitive data. The policy also outlined standards for accessing data, including: providing availability through application programming interfaces with descriptive metadata; using open data standards; and publishing under a Creative Commons Attribution license (Turnbull 2015).
- In the same month, the Australian Government announced plans to finalise Australia's membership application with the Open Government Partnership, after progress had stalled following the announcement in July 2010 (Turnbull 2016b). The government's initial schedule listed a launch date in July 2016 for Australia's National Action Plan, however, to date, the plan remains under development (Australian Government 2015, 2016b).

Data releases are fragmented

The Australian Government and all state and territory governments (other than the Northern Territory) have adopted policies of making non-sensitive public sector data open by default. In New South Wales and Queensland, this has been supported by the adoption of a 'push' model in their freedom of information legislation, designed to encourage the proactive release of public sector information and make a formal freedom of information application a last resort. To support open access data, the Australian Government, and most states and territories (with the exception of Tasmania and the Northern Territory), have designated open data websites (such as data.gov.au) to provide a platform for open data discovery and release. In addition, some councils also host open data websites — such as the City of Melbourne and Glenorchy City Council in Tasmania. A lead agency (generally based within the Prime Minister's/Premier's department) is responsible in each jurisdiction for driving the implementation of this open data policy (appendix B).

Public sector data is also released on a number of other websites. Spatial data in particular tends to have separate arrangements. In Queensland, for example, data.qld.gov.au is the state government open data website, QSpatial is the Queensland spatial data catalogue, and MinesOnline is a host for maps that provide spatial information relevant to the mining and resources industry. In Victoria, the main open data portal, data.vic.gov.au, coexists with spatial data websites, including Data.Vic, Spatial Datamart, Geovic (search and map geospatial data), Forest Explorer (maps forests and recreation tracks), and the new GDA

2020 initiative (appendix B). Open data is also hosted by VicRoads and the City of Melbourne.

Additionally, individual government bodies such as Tourism Research Australia, Geoscience Australia and the Reserve Bank of Australia publish a large amount of data on their websites.

In addition to fragmentation resulting from the existence of multiple portals, fragmentation can occur within a single portal. This is the case where datasets could be stored together but are not. For example, Geoscience Australia lists 540 separate GIS datasets of the same format connected to various towns. These datasets are all stored separately with no indication that they are related. This can make data less discoverable and less accessible.

Whilst fragmentation is the norm, some jurisdictions have better coordinated their releases than others — a leading example is Western Australia, which has data.wa.gov.au for its open data, and Landgate SLIP Enabler for location-based information. Most data published through SLIP is also searchable through data.wa.gov.au. Moreover, DataSA collaborates with data.gov.au to share metadata about government datasets, which allows search of the two sites to be interoperable. However, most jurisdictions have not done this (appendix B) and, overall, Australia's the current open data framework remains fragmented. Reasons for this fragmentation could include:

- *The focus on releasing datasets as a metric:* this comes at the expense of releasing comprehensive datasets. As a result, the absolute number of datasets released may not provide a good measure of Australia's open data progress.
- *Australia's federal system*: because there are different levels of government with responsibility for different matters, portals have arisen that contain data that reflects these responsibilities. This contributes to fragmentation. At the same time, different levels of government can have responsibility within the same sector (for example, health and education), resulting in duplication.
- *Difficulties in standardising the data or metadata* (chapter 6): poor metadata reduces the discoverability of datasets which, in turn, contributes to their fragmentation and duplication.

Indeed, the CSIRO has observed substantial fragmentation and duplication:

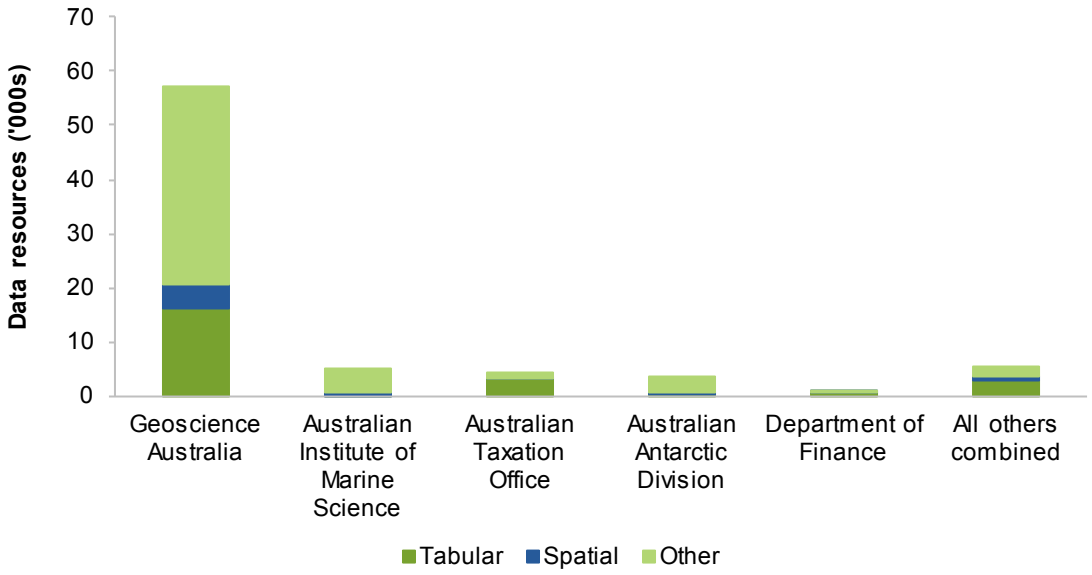
> There is also considerable duplication of effort in the creation of various access points such as portals into data of various types. Having multiple access points can be advantageous – it increases exposure and pulls relevant data together for a specific purpose. Reinventing the technology and re-doing the description is less desirable. There should be an ability to discover infrastructure in place, understand it and re-use or modify it for purpose where possible. Data should only ever be described once and those descriptions shared and transformed where appropriate. The model for resource sharing employed by the National Library may be worth examining. (CSIRO sub. 161, p. 37)

Most of Australia's open data is scientific information

Participation in data.gov.au is dominated by a limited number of major public sector contributors. While the count of almost 114 000 resources is substantial, it suggests a broader coverage of agencies than further analysis reveals. The top five publishers (Geoscience Australia, Australian Institute of Marine Science (AIMS), the Australian Tax Office (ATO) and the Australian Antarctic Division) account for around 95% of datasets on the site (see figure 3.1). This means that most of Australia's open data is scientific or spatial information.

More broadly, there is currently limited ability for private sector data users to contribute. In some cases, these users can add value to datasets by cleaning or restructuring them. For example, one member of the public has provided code to load and simplify the G-NAF. This code, however, is provided on a separate Internet platform and requires substantial expertise from users. Such processed datasets would benefit the public more if they were available directly from data.gov.au.

Figure 3.1 **Most open data is scientific information[a]**
2016



[a] This count has been limited to federal contributors to facilitate comparison.

Source: Australian Government (2016a).

What data is missing?

Compared to other countries, Australia registers particularly low scores on measures of availability of spending, legislation and health data (table 3.1). Spending data refers to

government spending recorded at a transactional level on specific items (generally above some threshold value, for example $1 million). Legislation data refers to information on laws at the federal level — in many jurisdictions this is provided in a machine-readable format. Health data refers to statistics generated from administrative data that indicates performance of specific service, or the health system as a whole. Whilst many of these datasets exist and are available in Australia, lower scores result from poor update frequency and poor formatting.

Table 3.1 **Open data availability for specific datasets across comparable countries[a]**
2015

Datasets	*Australia*	*United Kingdom*	*Canada*	*United States of America*	*New Zealand*
Spending	**5**	95	5	80	5
Legislation	**15**	100	80	85	80
Health	**60**	95	80	70	80
Map	**65**	100	100	95	80
Environment	**65**	95	95	85	65
Land	75	100	95	15	85
Census	80	90	95	100	80
Transport	80	95	90	65	15
Crime	90	95	95	70	65
Budget	95	90	95	95	80
Company	95	85	45	5	15
Trade	95	95	95	95	80
Education	95	95	55	80	80
Elections	95	95	80	70	80
Contracts	95	80	95	80	15

[a] Scores listed in the table are in raw form and calculated using a formula that awards points according to whether the data exists, how it is made available and whether it is up to date (WWWF 2015b). Unlike the scores presented in figure 3.1, these are not scaled with respect to the best performing country. Bolded numbers highlight datasets for which Australia has a comparatively low score.

Source: World Wide Web Foundation (2015a).

Other countries have performed much better than Australia in these areas. Indeed, the United Kingdom and the United States perform highly on most measures, and highlight that there are many potentially high value datasets that other countries release that Australia does not. For example:

- In the United Kingdom, the Health and Social Care Information Centre — a non-departmental public body — has released general practice prescribing data that lists all medicines, dressings and appliances that are prescribed and dispensed each

month. This has allowed analysis of prescribing practices at both the nation and practice level.

- The United Kingdom also releases data on all transactions over £25 000 for every agency. At least six software applications have been created to allow members of the public to analyse this spending.
- The United States releases datasets containing over 100 measures of performance for over 4000 hospitals. The measures range from operational measures — for example, the average number of minutes before outpatients with chest pain or possible heart attack received an electrocardiogram — to patient survey information — for example, the proportion of patients who responded that their nurses 'always' communicated well.

The lack of such datasets in Australia partly contributes to Australia's relatively low overall open data barometer score of 10th internationally (the most widely-cited measure of open data progress) (figure 3.2). Whilst Australia has a high implementation rank of 4th, which gauges the range of datasets available and the extent of their accessibility, weaknesses in other areas are apparent both in the index and in the course of this Inquiry — Australia is ranked 10th in 'readiness' and 19th in 'impact'.

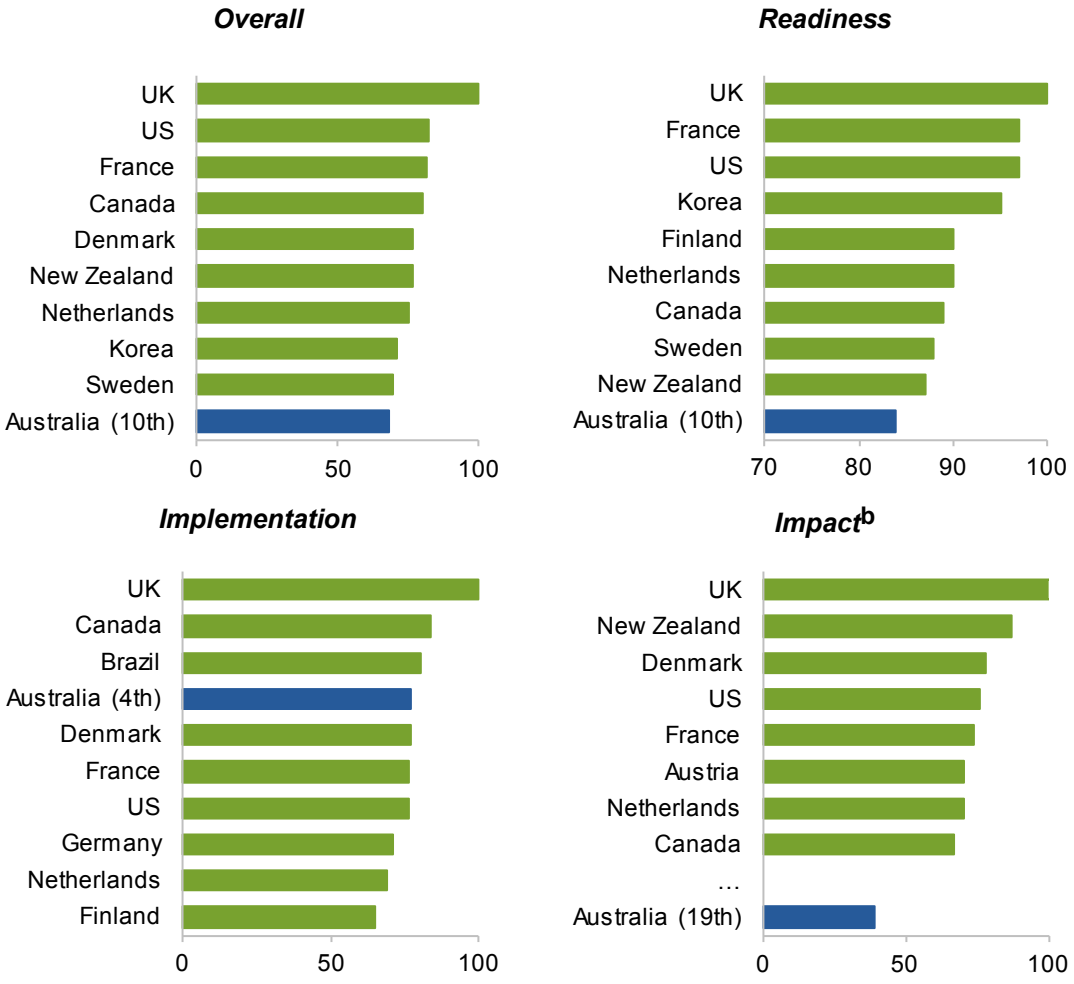
Australia's overall score is indexed to the United Kingdom, which is widely recognised to be a world-leader in open data due to the significant progress it has made data due to, among other things, the strong political commitment to open data and public pressure in favour of reform (box 3.5). Indeed, the UK's overall Open Data Barometer score is almost 20 points above the nearest country, with leading scores for all sub-indexes: readiness, implementation and impact. In sum, while it is difficult to assess Australia's absolute open data performance from these rankings alone, Australia consistently scoring below other countries in its open data performance suggests there is considerable room for improvement. As the Australian Data Archive (2016, p. 3) notes:

> Australia is well behind the UK, US and most of Europe on open data. This is impacting Australia's ability to be competitive [in research] and its standing in the HASS discipline.

DRAFT FINDING 3.1

Australia's provision of open access to data is below comparable countries with similar governance structures — including the United States and the United Kingdom. There remains considerable scope to improve the range of datasets published (and, correspondingly, the diversity of agencies and research bodies publicly releasing data) and the usability of open data portals.

Figure 3.2 **Measures of Australia's open data performance[a,b]**
2015



[a] Open data barometer scores are indexed to the leading country for each measure. As the United Kingdom leads each category, Australia's scores reflects its position with respect to the United Kingdom. [b] Impact score measures online, mainstream media and academic publications about open data impacts as a proxy for existence of impacts. Whilst this measure should be interpreted with caution, it suggests a relatively lower level of use open data in Australia than comparable countries, including those in which less open data is available.

Source: World Wide Web Foundation (2015a).

Box 3.5 Open data policies and progress in the UK

The United Kingdom is ranked 3rd in the OECD's OURdata Index (OECD 2015) and 1st in the widely-cited World Wide Web Foundation's Open Data Barometer (WWWF 2015a).

In terms of specific datasets, the UK receives particularly high scores for land, legislation, mapping and government spending categories.

As outlined in table 3.1, compared to Australia the United Kingdom has released substantially more data on government spending, legislation, health and the environment. A number of factors appear to have contributed to this progress, including:

- *Global influences* — the UK's initial steps toward open data were influenced by the EU Directive on the Re use of Public Sector Information 2003 which encouraged the release of as much information as possible. As a member state of the EU, the UK was required to transpose the EU directive into its domestic law.
- *Growing political will and public pressure* — public pressure on the UK government to improve transparency and the release of data increased significantly in the mid-2000s. The Guardian's 'free our data campaign', for example, which ran from 2006 to 2010, expressed dissatisfaction with the release of public data, arguing that members of the public should not pay to access data collected with public funds.
- *Increasing awareness of open data benefits* — in 2009, the Power of Information Taskforce examined open data benefits and recommended: marginal cost pricing (which in practice equates to free release); a 'Crown Commons' style approach to licensing; and the creation of an online repository to ensure all datasets are easy to find.
- *Political leadership* — within a month of forming government in 2010, David Cameron implemented several policies outlined his party's pre-election technology manifesto, circulating a letter to departments that called for the release of several government datasets, including: publication of all government spending over £25 000; publication of the names and salaries of all central government and Quango managers earning over £150 000 per year; monthly online publication of local crime data on a street by street basis, education and health performance data; and detailed information about all of DFID's projects and spending programmes (GOV.UK 2010).
- *Extensive consultation* — the UK Government's open data policy was implemented through extensive public consultation and clear communication. Indeed, extensive consultation took place through: data requests within data.gov.uk; the Making Open Data Real consultation; Open Data White Paper; and the Shakespeare Review of Public Sector Information.
- *A supporting institutional framework* — several institutions were created to drive implementation of the UK's open data initiatives. For example, the Public Sector Transparency Board was announced in David Cameron's 2010 letter to department heads. The Board was set up to drive forward the governments transparency agenda, chaired by the Minister for the Cabinet Office and comprising a mix of public sector data specialists and data experts.

Sources: Rogers (2013); GOV.UK (2010).

3.3 Restricted access to public sector data

Restricted access involves limiting data access to specific, 'trusted' individuals or organisations. Data subject to restricted access is sensitive, often containing information for which there is the risk of identification, re-identification, or some other confidentiality or government sensitivity concern (the role of de-identification is discussed in chapter 5). Restricted access is required for a large portion of datasets shared and released by publicly funded service providers. These datasets often contain personal information, relating to: health (for example, diagnoses, treatment, and services provided); education (for example, records of attendance and test scores); aged care (for example, attendance and health status) among many other areas. Whilst sharing these datasets carries potentially high risk, for the individuals, the reputation of the collecting agency and the data user; there is equally potential for substantial value to be derived from doing so.

Agencies may share data with other government bodies for a wide variety of reasons. Sharing may be:

- necessary for effective service delivery (for example, ensuring families struggling to make ends meet are identified and offered community support)
- required by law (for instance, law enforcement or national security purposes) or necessary to achieve the implementation of a law (for instance, preventing social security or tax fraud)
- or for research and policy development purposes.

There is no one metric that measures the extent of data sharing in the public sector — rather any assessment of sharing must rely on an accumulation of evidence by example. Overall, the evidence examined in this Inquiry suggests that — whilst some jurisdictions are clearly better than others — data remains systematically siloed in the public sector with little sharing between agencies or with researchers. As the Centre for International Finance and Regulation (sub. 9) states:

> One metaphor for the current state would be to liken the data assets of the nation as residing upon well-tended continental plates that occasionally bump up against one another … With this, perspective, the principal challenges of policy are to smooth out some of the points of friction for: privacy protections; data licensing; data linking and sharing; data skills development and the agility of public consultation and development processes. (sub. 9, p. 4)

Sharing data within the public sector

Information sharing within jurisdictions — some sharing, but far from enough

One indicator of the extent of data sharing within the public service is the difficulty that agencies encounter when attempting to transfer data. A number of participants have noted significant difficulty associated with sharing data.

For example, the Queensland Government (sub. 207) noted:

> In many public bodies data sharing is formalised via a Memorandum of Understanding (MoU) agreement. This requires multiple legal departments to be engaged on projects, along with external legal counsel. These are often drafted by policy and legal teams with little or any knowledge of where the data was captured, or what is the end-to-end journey of the data across a service change. Therefore, these MoUs are often complex, difficult to understand, unrealistically constrained to where the data can come from, or be used for, and bear little relation to what data is really required. (p. 8)

Similarly, after reviewing a number of sectoral case studies (box 3.6), the UNSW Social Policy Research Centre (2015) also noted that sharing is a complex process, with many data holders reluctant to share:

> Research suggests that sharing information is often perceived to be complex by front line workers and agency managers. Many practitioners are reluctant to share information even when they have the legal authority to do so, and many agencies have a risk-averse attitude to information sharing even when this may be in the interests of clients. There is often a disparity between the actual legal and policy context and the perceptions of those involved. (p. 1)

Moreover, the Department of Foreign Affairs and Trade (sub. 202) stated:

> Data sharing across government agencies can be challenging and the costs of finding and obtaining data onerous. Privacy and secrecy legislation means that agencies, particularly those handling large administrative datasets, often make data either unavailable, or at the least very difficult to obtain; or provide it in unsuitable formats. (p. 24)

More broadly, there are minimal ongoing arrangements for information sharing between agencies — information sharing often occurs on a once-off basis, which is ineffective. As noted by the Australian Institute of Health and Welfare (sub. 162):

> It is widely recognised that the current data sharing system which relies on once-off linkages has proven to be slow and cumbersome to the point where it has not been effective at enhancing data sharing activities. (p. 11)

Other stakeholders and participants to this Inquiry also provided a number of examples where data sharing does not occur at all. One example provided by the Department of Prime Minister and Cabinet (DPMC) (2015) is that, in some cases, when individuals move from one departmental program to another (for example, from the Department of Defence to the Department of Veterans' Affairs (DVA)), data and client history are not transferred. The same report noted that it can take several years and multiple memoranda of understanding to establish data sharing arrangements between government agencies.

Box 3.6 Opportunities for information sharing in NSW — a summary

Child welfare

Information is exchanged between statutory agencies such as Health, Family and Community Services, Education and Police, between statutory and non-government organisations, and between NGOs when services are provided to children and families. While there are no legislative barriers to sharing information, and many organisational barriers around information sharing have been addressed, there is still a risk-averse culture in many human service agencies, and many people did not understand their legislative obligations and did not know who to ask for advice. Most barriers occurred in the interpretation of the legal and policy constraints rather than in the actual legislative and policy provisions.

Schools

Information is exchanged between schools when students move from one school to another, and between schools and other agencies that are providing services to students. When a child moves between school sectors, the parent is responsible for completing the information required by the new school. Most parents provide comprehensive information to the new school, but some parents withhold important information about their child from the new school. While generally information is shared appropriately, there are gaps between legislative and policy obligations as well as different constraints and perceptions arising from inadequate understanding or poor practice. No significant legal, policy or political barriers to information sharing between schools were identified.

Housing support

The Housing and Accommodation Support Initiative (HASI) for people with mental illness is a partnership program between Housing NSW, NSW Health, NGO accommodation support providers, and community housing providers. The HASI program does not have a specific legislative framework governing information sharing — seeking and obtaining consent for sharing personal information is the primary way information is shared. When exchanging information without consent, HASI program staff must establish that an exception applies. As such, practice was variable around information sharing across agencies and locations. In some cases there was a lack of clarity around information sharing both where consent had been provided and where consent had been withdrawn. There appeared to be a lack of awareness of the policy and legislative frameworks. Practice around information exchange was poorer where interagency meetings were not being held regularly.

The UNSW Social Policy Research Centre concluded that effective and appropriate information sharing can only take place in a context where:

- there is a clear legal and policy framework, and policies and procedures specify the appropriate processes but are flexible enough to allow for these processes to be tailored to individual situations
- organisational cultures facilitate appropriate information sharing and collaborative practice while taking into account people's rights to privacy and confidentiality
- the human services workforce has knowledge of the legal and policy framework and is trained and supported in delivering good practice.
- workers and agencies trust each other to use the information appropriately.

Source: UNSW Social Policy Research Centre (2015).

Evidence suggests that even where there is a legal requirement to share information this does not always happen due to legislative opacity, risk aversion, concerns about privacy, overly bureaucratic requirements, and other sectoral and jurisdictional barriers (these barriers are discussed later). For example, there is, to varying degrees across the Australian states and territories, legislative provision for the sharing of information related to the safety, welfare, and wellbeing of children and young people between prescribed bodies (such as health departments, schools and police) and between these bodies and the child protection agency. In some jurisdictions this is voluntary, in other cases this is mandatory. However, even where there is an express requirement to disclose, this may not occur. For example, in New South Wales, Adams et al. (2016) noted:

> Recent research indicated that, although the passing of [the NSW legislative provisions] (and training and organisational support for these changes) has improved information sharing, many jurisdictions still experience considerable difficulties regarding information sharing, and practice is variable across agencies, sectors and geographical locations. The process for sharing is perceived as being cumbersome and bureaucratic and there is still considerable anxiety about sharing information and the reluctance of some agencies and professionals to do so. (p. 64)

Similar concerns were expressed to a recent Inquiry into child protection services in Victoria, where child protection workers were confused about what information they could and could not share with the police, despite an express legislative provision governing the disclosure (Victorian Ombudsman 2009). The legislation has since been amended.

Clearly, there are broad, systemic issues with data sharing between government agencies.

However, there are some instances of good practice. New South Wales in particular appears to be a leader. It has adopted legislation granting whole-of-government authorisation to sharing within the New South Wales public sector. This facilitates the movement of data such as that relating to New South Wales Fire and Rescue, local government construction and the electricity grid with the state's Data Analytics Centre (box 3.7), although the system is not perfect (discussed below). South Australia also has a bill before parliament that aims to facilitate public sector data sharing, and Western Australia has a longer history of achievement than any jurisdiction, but limited to a single area (health) where the passion and commitment of individuals is more responsible for its success than a policy or structural level of support.

As at state level, particular Commonwealth agencies can at times exhibit good practice and demonstrate the unmet potential of equivalent entities. The DSS is one agency active in data sharing, within the limits of fairly anachronistic legislation and varying support for evidence-based policy. It operates the National Centre for Longitudinal Data to support the management of critical national data assets (Department of Social Services, sub. 10), is developing its own trusted access arrangements (discussed below) and is working on releasing information through a synthetic dataset.

While noting that opportunities exist to encourage more sharing of data, the Australian Tax Office (ATO) (sub. 204, p. 5) similarly states that, within legislative strictures, it transmits to and receives a significant amount of data from other agencies — individuals are able to

get pre-filled tax returns as a result of information sharing processes between the DSS and the ATO; and the ATO is a participant in myGov's 'tell us once' approach to sharing contact information between relevant departments. These are positive but isolated examples of good practice that many more agencies should be looking to emulate.

Box 3.7 NSW Data Analytics Centre

The NSW Data Analytics Centre (DAC) was established in 2015 under the auspices of the NSW Department of Finance, Services and Innovation (DFSI) to facilitate data sharing between NSW Government agencies, with the aim of leveraging data to inform more efficient, strategic, whole-of-government decision making. As noted by the Australian Government Department of Industry, Innovation and Science (DIIS):

> The DAC will be used to gain insights and reduce challenges in areas such as crime prevention, childhood obesity, pollution, and sustainable urban planning. The DAC will source expertise from research, industry and from NSW government agencies (2015, p. 10).

A key part of this facilitation involves identifying which datasets, if shared and linked, would improve government policy and subsequently, social outcomes.

The DAC's approach to data analytics involves several steps, including:

- consultation with agencies to identify the policy questions that agencies would like answers to, which forms the basis of projects undertaken by the Data Analytics Centre
- assessment of the likely value of the project
- evaluation of NSW Government data holdings to identify which datasets would be useful in answering the specified question
- liaison with agencies to secure datasets
- design of an appropriate analytics framework to provide insights to the relevant agencies.

An advisory board was established to provide strategic advice and support to the DAC, made up of representatives from industry, government and academia.

All projects undertaken by the DAC require Cabinet endorsement.

Source: DIIS (2015).

Sharing data between sectors and jurisdictions can be fraught

Additional difficulties exist when data is required to be shared between jurisdictions and sectors. This often occurs in publicly funded service provision. The Australian and state and territory governments fund a wide range of services — sometimes these services are provided by public bodies (such as public hospitals), and in other cases, these publicly funded services are provided by private or not-for-profit bodies (such as private hospitals).

Sharing of information related to publicly funded services appears piecemeal and limited, particularly where services are delivered by the private or not-for-profit sectors and multiple sectors are involved. This failure to share information can be due to a range of factors, from genuine legislative impediments to technical barriers. For instance, in an

Inquiry into *Service Coordination in Communities with High Social Needs*, the NSW Standing Committee on Social Issues (2015) concluded:

> The complexity and persistence of disadvantage in many communities necessitates a collaborative approach to ensure sustainable and positive change. There remain a number of barriers to overcome, including organisational silos, a lack of access to relevant and timely data and a knowledge gap between what is permitted by privacy law and what is practiced by service deliverers. There are also significant constraints within the funding environment, including the short-term length of funding periods. (p. ix)

Similarly, the Department of Prime Minister and Cabinet (sub. 20), referring to data on Indigenous outcomes, stated that:

> Ideally, information on all government services and programmes by location should be brought together. To be useful, this would cover information about Commonwealth programmes and state and territory programmes. This information would not only assist citizens to know what services are available in their area it would also assist governments when they make decisions about what programmes and services to fund. At the moment decisions on funding are not always based on a full understanding of the programmes and services that are already available in each location. (p. 32)

Information exchange *between service providers* can remain severely limited even where there is a clear public interest in doing so. For instance, the Australian Law Reform Commission (2010) examined information sharing on domestic violence within the legal system (which involves multiple jurisdictions and sectors such as federal courts, state agencies and police). It noted:

> Throughout the course of this Inquiry, the Commissions have heard about the problems that arise because of the gaps in information flow between the family law system, the family violence system and the child protection system. In many circumstances, important information is not being shared among courts and agencies and this is having a negative impact on victims, impeding the 'seamlessness' of the legal and service responses to family violence. (p. 78)

Similar examples were noted by the Magistrates Court of Victoria (2015) in its submission to the Victorian Royal Commission into Family Violence:

> The capacity to share appropriate information quickly and securely across courts, police, and family violence legal and support services is a critical prerequisite for any systemic reform aimed at improving the efficiency and responsiveness of the court's approach to meeting the needs of families experiencing family violence.
>
> Currently, even if the requisite legislative frameworks were in place to facilitate comprehensive information and data sharing across jurisdictions and agencies, the case management systems of [the Magistrates' Court of Victoria and the Children's Court of Victoria] do not have adequate functionality to support this aim. (p. 52)

The Victorian Crime Statistics agency has noted that data on the use of police, homelessness and victims assistance programs is not shared across services. Even within the health sector, emergency, outpatient and inpatient services do not currently exchange information on patients. (This can affect data availability for service provision *and* for

researchers. The value of bringing such data together was a key consideration of the United Kingdom's National Health Service in developing *care.data*, however this program was ultimately abandoned due following an unfavourable privacy assessment (box 3.8)).

Box 3.8 **care.data**

Care.data was an initiative that begun within the United Kingdom's National Health Service. It was introduced to integrate data from general practices and hospital data to enable more comprehensive research. Information to be contained with *care.data* included a range of factors such as medical diagnoses, prescriptions, family history, vaccinations, test results and body mass index data among others. However, before the rollout of care.data, a privacy review by Dame Fiona Caldicott argued that the case for data sharing was yet to be made to the public and made a range of recommendations to increase security of health data. Whilst the review did not explicitly recommend abandoning care.data, the initiative was subsequently cancelled.

Sources: Boitten (2016), Caldicott (2016).

Another example of poor data sharing occurs in the field of child protection. Legislation in all the states and territories, with the exception of Queensland and Victoria, provides that the head of the child protection agency may disclose to an interstate officer any information that they consider necessary to allow the person to administer the law. A protocol for doing so has been agreed to by all states and territories, and New Zealand. Yet evidence suggests that there has been reluctance to make use of these provisions due to risk-aversion and uncertainty over their application and scope (Adams and Lee Jones 2016). Finally, implementing data sharing initiatives within the education sector has proven difficult (box 3.9)

In sum, there appear to be serious systemic issues with sharing information between sectors and jurisdictions, even where the legislative framework facilitates or even requires sharing. While some isolated examples of good practice exist (such as CrimTrac), these are hardly commonplace. And while the Commission's Report on Government Services, which combines data from all states and territories to report on the performance of six broad service categories, is a useful product, it remains unable to obtain information on particular services due to some jurisdictions' reluctance to share data.

Service providers often do not have access to the information they need

As a result of this failure to share data, service providers report difficulties accessing information that is necessary for them to provide their services effectively. In a recent New South Wales Parliamentary Inquiry into *Service Coordination in Communities with High Social Needs*, the Benevolent Society (2015) noted that:

> [E]ffective planning and service coordination requires access to information about the full range of services which are being funded and delivered in a given area …[u]p-to-date data and information which is accessible to communities and service providers is currently not available for many of the communities in which The Benevolent Society works. (p. 9)

Box 3.9 An example of cross-jurisdictional data sharing challenges

An example of the challenges faced in sharing education data between school systems and jurisdictions with inconsistent privacy legislation is the Trans-Border Attendance Strategy. The Strategy commenced as a pilot in 2009 and included 45 Northern Territory, South Australian and Western Australian remote schools, and was intended to address the issue of absenteeism and significant mobility amongst Indigenous students. The Strategy gradually progressed to include 399 public, Catholic and Independent schools in 2013.

A key feature of the Strategy was an information technology platform — the Central Schools System — which consolidated and merged attendance data from participating systems. Through the use of the platform, schools were able to share attendance, enrolment and learning information across education sectors (public, Catholic and Independent) and across jurisdictions. As a result of the availability of this information, school staff could determine (on or before student arrival) the student's past enrolment and attendance history, and access other information necessary to progress individual learning or develop behaviour plans for transient students.

The project initially drew considerable interest from other jurisdictions seeking to be involved. However the project was hampered by issues with uptake and usability of the data platform by schools. These issues were compounded by legislative privacy barriers in some jurisdictions in relation to data sharing. Expansion of the strategy has been halted and ongoing participation of existing jurisdictions remains uncertain. According to the Northern Territory Government (sub. 77):

> Along with the legislative issues that can arise in sharing data between jurisdictions and school sectors, this project highlights the importance of ensuring data projects of this kind are designed to be valuable and accessible for schools. (p. 5)

Source: PC (2016).

Similarly, in their submission to this Inquiry, the Brotherhood of St. Lawrence (sub. 186) stated:

> Unfortunately, very few of the datasets identified above are available in the public domain without complex administrative processes; and many require knowledge of how they are constructed. Our experience with the Education First Youth Foyers program has seen eight separate application processes to be completed, including the use of a data intermediary to ensure anonymity. This process is administratively burdensome and costly for not-for-profits. (p. 5)

Moreover, the Joint Council of Social Service Network (sub. 170) note that, whilst significant benefits would arise from improved data sharing, government agencies are often reluctant to make data available to service providers, and that negotiations to access data can be long and time consuming. Where data is provided, it is often in formats that are difficult to work with or require specialised software.

Providing information to government and the community is piecemeal

Some information generated by service providers is shared with governments for reporting requirements, and may be required under legislation. Reporting requirements across health,

education, aged care and community services are too numerous to list. A limited example in education is illustrative:

- There are a number of legislative requirements under the *Australian Education Act 2013* (Cth) for independent schools to provide information through national data collections and via National Assessment Program testing
- Non-government schools are required to complete an annual census, which collects information on schools, staff and students
- The Australian Government Department of Education and Training (DET) collects annual financial information from independent schools via a financial questionnaire.

Sometimes government service providers are required under contract to provide information, yet contracting practices vary widely. Many examples provided to us outline situations where services have been privatised and the service provider has not been required under the contract to provide information to that government, which would enable more robust performance assessment and could subsequently be useful for policy development or infrastructure planning purposes. This issue is discussed further in chapter 4.

It is still rare for government-funded service providers to report information to *the community* in a comprehensive way, other than through annual reporting requirements. Even where data is provided to government (for example, on private hospital performance — Australian Private Hospital Association, sub 183, p. 3) it is invariably not developed into data series for the public to use. Participants in this Inquiry — Australian Unity (sub. 95) and Medibank Private (sub. 98) — have stated that, despite the reporting requirement to government community access to information on the performance of individual health service providers remains severely limited, hindering the public's ability to make informed decisions about their healthcare compared to the information that is available overseas.

One stark exception to this is the National Assessment Program — Literacy and Numeracy (NAPLAN), where data on school performance is publicly available through the collection of performance data via participation in NAPLAN. This data is a significant asset. As the University of Western Australia Faculty of Education (2016, p. 1) notes, the investment in NAPLAN is 'likely to provide important new insights into the effective delivery of educational services in Australia'.

In sum, most evidence suggests that there appears to be serious systemic problems with information sharing between sectors and jurisdictions. Legislative complexity and a lack of trust combine to stymie information sharing, leading to less efficient delivery of services, and very often, poorer outcomes for the community.

Sharing public sector data with researchers

The extent of data sharing for linkage and integration

Data can be shared between different government agencies for linkage and integration purposes (appendix B discusses these processes in more detail).

The value of linking data has long been recognised. It was noted in 1946, for example, that 'sometimes it is necessary to examine all of an individual's important records simultaneously' and that 'at times such a collection is of sufficient value that it is made at considerable cost in time and money.' (Dunn 1946, p. 1). Data linkage and integration offer large potential benefits by substantially increasing the scope of questions that can be answered with existing datasets. As the Telethon Kids Institute (sub. 5) notes:

> If we really want to improve [health] outcomes, it is likely that the solutions may come from multiple agencies; siloed thinking and policies do not work in terms of prevention. Hence the first huge benefit to the community from joining up public sector data is to provide an understanding of the causal pathways and the effective ways to prevent these problems. Many of the health, educational, mental health, child maltreatment and crime outcomes arise from similar sets of risk factors and hence there is power in bringing these data together. (p. 2)

Similarly, the National Health and Medical Research Council (sub. 126) states:

> Linkage of data between datasets allows new discoveries to be made in a highly cost-effective way, and by combining datasets the power to answer questions about health is increased — effectively leveraging the Commonwealth's investment in health and medical research. Gertig *et al.* (2013) showed through data linkage that human papillomavirus vaccinations in Australia significantly reduced cervical abnormalities while Mathews et al. (2013) successfully investigated cancer incidence rates for people exposed to CT scans. (p. 2)

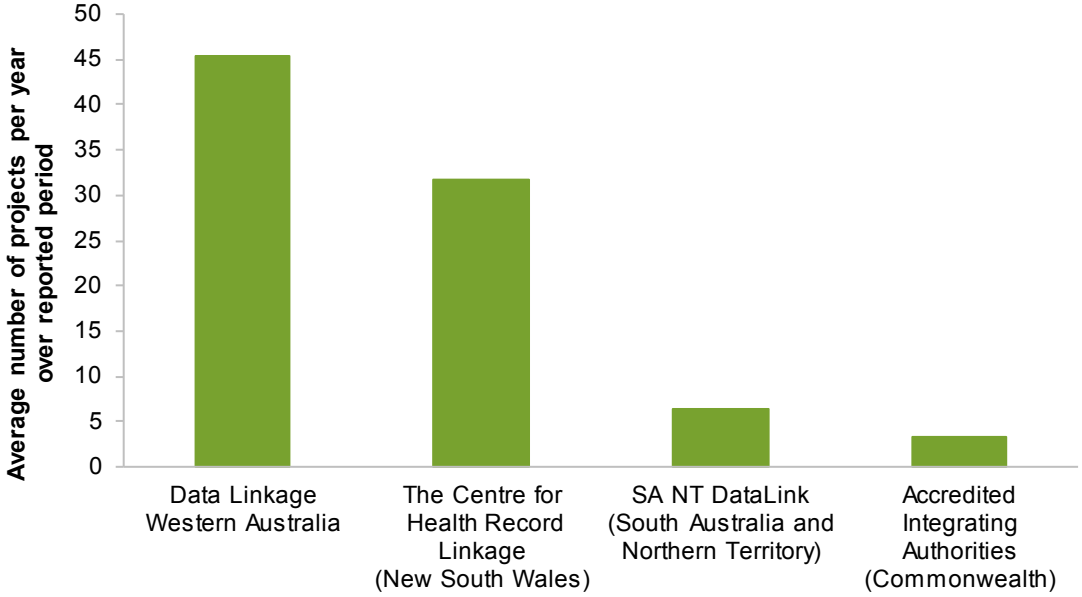
The potential for data linkage and integration, however, is far from realised in Australia. The Centre for Big Data Research in Health (sub. 21, p. 2), for example, stated that whilst there has been some recent positive developments, 'real improvements in access and use of the linked Commonwealth health data have yet to be realised', while AIHW (sub. 162, p. 7) notes 'the level of Commonwealth datasharing today is generally below the level required to appropriately harness the value of Commonwealth data holdings.' More critically, a recent study on data linkage capability found that 'complexity, duplication and lack of cohesion undermines any attempts to conduct research involving national record linkage in a timely manner'. (Mitchell et al. 2015, p. 1).

Indeed, reported data linkages among relevant institutions suggests only a few jurisdictions are undertaking a substantial number of linkage projects (figure 3.3):

- Data Linkage Western Australia has been integrating unit record administrative datasets since 1995, averaging 45 integration projects per year.
- Similarly, the Centre for Health Record Linkage (New South Wales) has been integrating datasets since 2007 and has averaged 32 projects per year.

- SA-NT DataLink (South Australia and the Northern Territory), however, has completed only 7 projects per year on average since 2011 (though this has been trending upwards with 18 projects completed from mid-2015 to mid-2016).
- Commonwealth accredited integrating authorities have registered only 3 projects on average per year from 2005 to 2015.[11]

Figure 3.3 **Data linkage is occurring most often in Western Australia and New South Wales**



Sources: Centre for Health Record Linkage (2016); Data Linkage Western Australia (2016); National Statistical Service (2016); SA-NT DataLink (2016a).

Of course, the number of projects is only a partial proxy for benefit. It is possible that the benefits delivered by a number of small projects were significant and larger than a collection of many small projects.

By international standards, Australian Government data integration is well behind world leading initiatives such as Statistics New Zealand's Integrated Data Infrastructure (box 3.10), which has commenced 73 data integration projects over the last five years. And although the UK has traditionally lagged behind other jurisdictions in data integration, some promising developments are progressing (box 3.11). More generally, compared to the policy and institutional reforms in other jurisdictions designed to facilitate data integration, Australia's institutional arrangements (appendix B) also appear to lag international best practice.

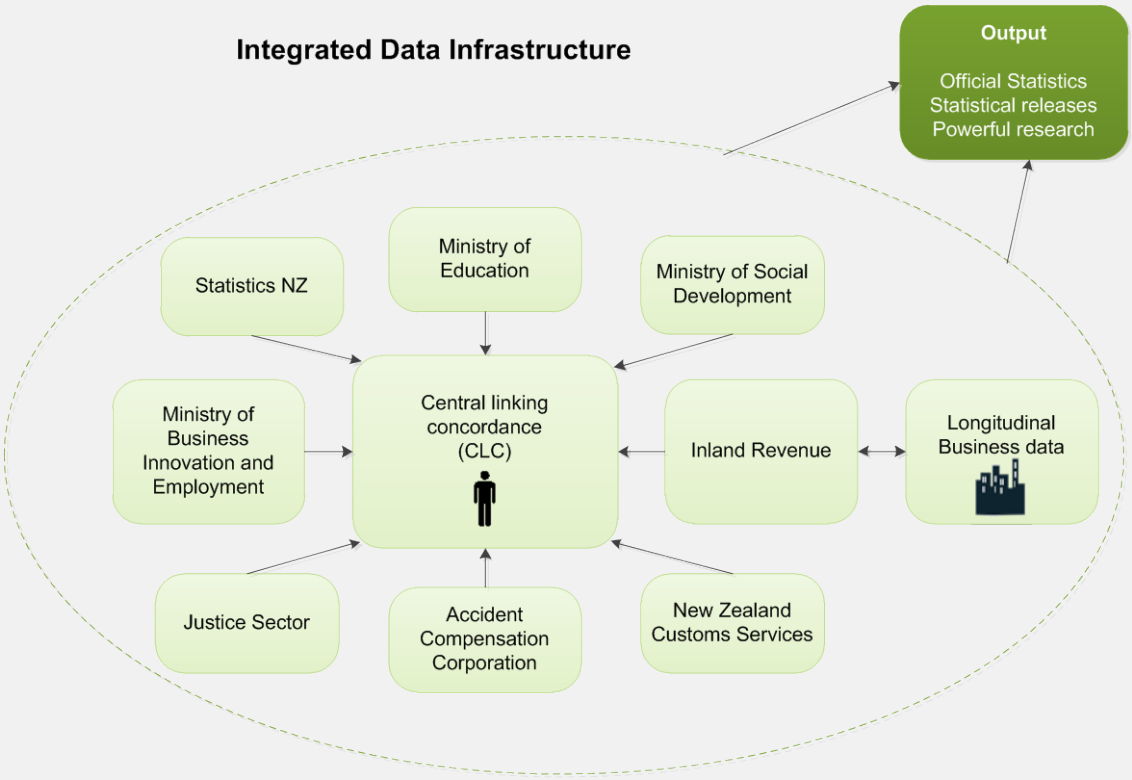
11 The remaining data linkage centres do not report projects.

Box 3.10 The New Zealand Integrated Data Infrastructure

The New Zealand Integrated Data Infrastructure (IDI) is an integrated, longitudinal dataset developed and maintained by Statistics New Zealand. Data from different sources are linked using both deterministic and probabilistic linking, depending what information is common across datasets. The IDI includes unit record data on: health and safety; justice; benefits and social services; tax and income; education and training; student loans and allowances; travel and migration; and family and household factors. The datasets that contribute to the IDI are primarily administrative datasets, although some survey data is also integrated.

Access to IDI data involved a trusted user model, with access limited to approved researchers. These researchers gain access using either facilities located within Statistics New Zealand, or through a secure remote connection. Whilst users are able to analyse the data with little constraint, Statistics New Zealand vets any results that researchers intend to share with others.

To date, researchers have used the IDI for a broad array of research projects. Some examples include measuring impacts of social policy, undertaking micro-simulation modelling of the New Zealand tax and welfare system, examining the relationship between labour mobility and productivity and investigating the increase in earnings associated with tertiary education.



Source: Statistics New Zealand (2016).

However, the policy field at Commonwealth level is not completely barren. The Multi-Agency Data Integration Project (MADIP) has linked data from five separate Commonwealth agencies: Department of Social Services; Department of Health; Department of Human Services; Australian Taxation Office; and the Australian Bureau of Statistics. It intends to extend its scope to states (Australian Statistical Advisory Council,

sub. 25). Moreover, the Business Longitudinal Analytical Environment is under development, integrating data on financial and business characteristics for over two million active businesses, with data from the Australian Tax Office, the Australian Bureau of Statistics and IP Australia.

Box 3.11 Administrative Data Research Centres in the United Kingdom

The UK government launched an Administrative Data Taskforce in 2012. The taskforce's report provided multiple examples of significant barriers that researchers face in accessing data. Key barriers included difficulty in identifying legal pathways, ensuring confidentiality and designing robust methods for linkage to other datasets. Overall, the taskforce found that:

> The United Kingdom has lagged behind some countries in Europe in the use of administrative data within government and while there have been a number of initiatives outside government these have been limited. (Administrative Data Taskforce 2012, p. 1)

The Administrative Data Taskforce report recommended a number of changes centring around the creation of an Administrative Data Research Network — a body responsible for linking and releasing administrative data. Its key recommendations were that:

- an Administrative Data Research Centre (ADRC) should be established in each of the four countries in the United Kingdom
- legislation should be enacted to facilitate research access to administrative data and to allow data linkage between departments to take place more efficiently
- a single UK-wide researcher accreditation process, built on best national and international practice, should be established
- a strategy for engaging with the public should be instituted
- sufficient funds should be put in place to support improved research access to and linkage of administrative data.

The UK government has enacted the bulk of task force's recommendations, establishing four Administrative Data Research Centres. Whilst some key recommendations that increase the ability of government agencies to share data are not enacted, the United Kingdom has a plan and action is under way.

Source: Administrative Data Taskforce (2012).

In sum, despite these promising developments, there appears to be significant differences in the appetite for and resourcing of data integration projects across jurisdictions, with the performance of some jurisdictions being well below any reasonable expectations. A number of Inquiry participants highlighted this, as does an examination of progress to date, compared to the potential for comprehensive measures.

Significant barriers to implementation are evident — including bottlenecks in integrating authorities at the Commonwealth level, explicit legislative restrictions on data linking, and requirements to delete linked data after use. Such barriers are discussed in chapter 5.

DRAFT FINDING 3.2

Data integration in some jurisdictions (particularly Western Australia and New South Wales) has made good progress in some fields, but highlights a lack of action in equivalent fields at both national and state/territory level, and reveals the large unmet potential in data integration.

What is the extent of public data access for researchers?

There is strong interest among researchers in accessing government and publicly funded research data.

- Levels of access in Australia have been encouraging for datasets *collected primarily for release*.
  - The Household, Income and Labour Dynamics (HILDA) dataset (of which the DSS is custodian) has been very popular among academics, with 547 academic researchers accessing the dataset in 2014. Similarly, the ABS Confidentialised Unit Record Files (CURFs) have had extensive use among academics, with universities representing around one third of organisations accessing the data.
- However, research institutions have reported substantial difficulty in accessing the vast bulk of public sector *administrative data typically not originally collected for sharing or release*. Examples of such difficulty were presented in the Senate Select Committee on Health's report, *Big health data: Australia's big potential*, where Dr Heather Gidding outlined the difficulty navigating access procedures:

> It took us a long time to find the data custodian for that ACIR [Australian Childhood Immunisation Register] data. When we did find someone who did understand the dataset, they are very hard to access. They are overworked and hard to get a hold of. It certainly does make the research quite difficult. (SSCH 2016, p. 42)

Frustration at limited discoverability and access more generally has been voiced by a number of private sector organisations. Archerfish Consulting (sub. 30) note:

> In our experience, public data, optimistically characterised as a taxpayer funded public good is being redefined by stealth as an excludable good, increasingly available to a select few. Whether deliberately or otherwise, State and Territory agencies are able to extinguish data requests made by all but the most determined non-state actors. Practically, only other public sector organisations and university-based researchers supported by publically funded research grants are able to withstand the obfuscation and demands for fees adopted by State and Commonwealth Governments. (p. 1)

Similarly, 'The Hive' (a group of individuals from the Mount Druitt Community, local service providers, government agencies and business) noted the difficulty accessing data on the number of pre-school aged children in the Mt Druitt area (LCSCSI 2015). As a

result, the organisation had to door knock the area to gain this information directly from households.

Data sharing via trusted user models

A promising approach to sharing sensitive data is the trusted user model, which involves allowing data users to access more detailed datasets in highly secure settings (appendix B). Key features of this approach are: remote access to a virtual computer run on controlled hardware; the inability for data used to be exported without passing through a gateway which assesses disclosure risk; and close governance and recording of all file movements. One trusted user model is that based on the 'five safes' principle which involves examining control over: the people that access the data; the projects for which data is accessed; the settings in which data is accessed; the outputs the users have access to and are able to publish; and the characteristics of the data itself.

Implementation of trusted user models in Australia remains limited — few institutions have taken them up, there is limited cooperation between jurisdictions and they have only been applied to a limited range of (mainly health) data. Some trusted user model implementations include:

- *The Secure Unified Remote Environment (SURE):* a computing environment that allows researchers to analyse linked health data in manner that prevents confidentiality breach whilst maintaining flexibility in analysis. The Sax Institute (sub. 56) reports hosting data for over 60 projects with more than 300 researchers accessing the data. Moreover, the institute notes that it is working with over 80 agencies at the federal and state levels, including the Centre for Health record Linkage (CHeReL), a data linkage unit managed by the New South Wales Ministry of Health.
- *The ABS Virtual DataLab:* a similar remote access environment is under development at the ABS. The ABS has commenced trials of a virtual DataLab, including providing several agencies with access to data created by the Multi-Agency Data Integration Project. As at July 2016, over 60 users from around 10 agencies were accessing data under this framework (Australian Bureau of Statistics sub. 94).
- *DSS Trusted User Model:* the Department has collaborated with the Australia Institute of Health and Welfare (AIHW) to enable researchers to query selected social services data remotely, producing aggregated data. Work is now underway to provide a platform through which trusted users can access administrative data (Department of Social Services sub. 10)
- *The Australian Longitudinal Income File (ALife):* the ATO has committed to developing a de-identified longitudinal taxpayer file for research. It plans to release this file in a secure trusted access environment, through which approved users (located within Australian government or universities) can undertake non-commercial research. To access the data, researchers are required to sign a deed and have ethics approval (Australian Tax Office, Canberra, pers. comm., 7 October 2016).

But significant gaps remain. Much remains to be done before all agencies with researchable databases are able to point to their standard for providing trusted user access; the vast majority of sensitive administrative remains data largely inaccessible for reuse. A shift towards a more risk-based approach via greater adoption of trusted access models could improve outcomes for data custodians and researchers. It would allow data custodians to have the confidence that risk is being managed in a holistic way, rather than placing all emphasis on de-identification, and researchers to have access to more useable data. Given the significant benefits of trusted user models, they appear to be a leading practice that could be applied more broadly across the public sector under central agency direction. However, we recognise that trusted access models can require significant capability to implement — options for addressing this are evaluated in chapter 8.

Private firm access to public sector research data is too limited

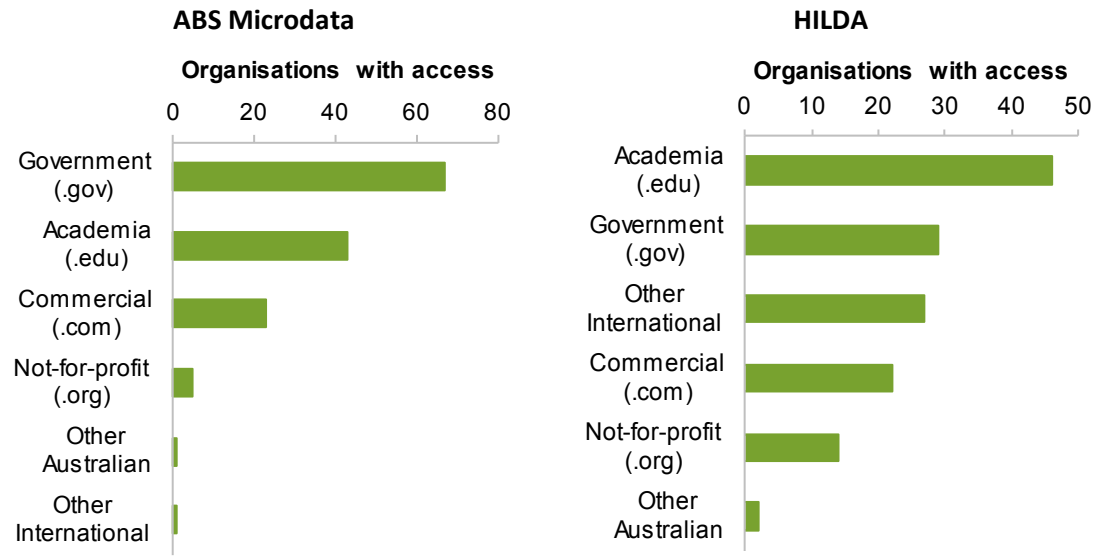
Access to public sector data appears to favour known, trusted parties such as other government agencies and academics. Figure 3.4 presents a measure of access for two groups of well-known restricted access public sector datasets — ABS Microdata and HILDA. These datasets are leading examples of Australian practice, offering sound standards for developing trust, and clear specifications for users.

Public sector data offers significant potential value to private and not-for-profit sector organisations, based on opportunities that have arisen in more entrepreneurially driven environments such as the United Kingdom and United States. For example, the Australian Data Archive (2016) argues that better open data in the United Kingdom and United States has contributed to their better social science research performance than Australia. And *restricted* data access flows between the public sector and private sector firms also appear to be weak in Australia. Only 25 private sector organisations have access to ABS micro data and 22 have access to HILDA, comprising 18% and 16% of all user organisations, respectively (figure 3.4). The majority of private sector users include consulting firms, peak bodies and actuarial firms.

While some training is necessary to use these datasets (particularly HILDA), there is no reason to suspect these skills would be less abundant in the private sector than in government and academia. This would suggest that the comparatively low private sector access rates reflect obstacles posed to private sector entities in getting access to the datasets.

Public access versions of comparable datasets *are* available overseas. The US National Longitudinal Surveys (including seven cohorts), the US Panel Study of Income Dynamics, the Russia Longitudinal Monitoring Surveys, and the Korean Labour Income Panel Study are all openly available, requiring only that users register online.

Figure 3.4 **Top-level domains of organisations accessing restricted data**
2016



Sources: ABS (2016); Melbourne Institute (nd).

Public access of such datasets can add substantial value. Excessive paper work can be a significant deterrent to potential data users, reducing the extent to which studies are replicated and inhibiting researchers from accessing the data for small or exploratory projects. John Daley from the Grattan Institute notes that: '[m]ore unit data needs to be released that would enable us to do more longitudinal research: specifically social security census and tax longitudinal data' in the DPMC's (2015, p. 16) report, *Public Sector Data Management*.

Access to data about yourself

Individuals can seek access to public sector data about themselves for a range of reasons, including to assess (and correct) its accuracy. Several instruments give individuals the right to access information held on them in the public sector, in both federal and state jurisdictions.

At the federal level, such access is provided by both the *Privacy Act 1988* (Cth) and the *Freedom of Information Act 1982* (Cth) (FOI Act), and similar provisions apply in all state and territory jurisdictions (appendix C). Within the Privacy Act, Australian Privacy Principle 12 (APP 12) requires that an entity holding personal information about an individual must give the individual access to that information on request (generally within 30 days). However, entities are not required to provide access to information where: it would pose a serious threat to health; the request is frivolous or vexatious; access would contravene another law, or it would prejudice law enforcement activities (among other

reasons). Additionally, agencies are not able to charge individuals for the request or for access to the data (but private sector organisations can — chapter 4).

APP 13 states that an entity must take reasonable steps to correct personal information to ensure that, having regard to the purpose for which it is held, it is accurate, up-to-date, complete, relevant and not misleading. There is no simple process for individuals to make use of these access rights under the privacy legislation, although the OAIC (and relevant state privacy regulators) are empowered to handle complaints about access requests.

The FOI Act operates alongside the Privacy Act, providing individuals the right to access copies of documents held by agencies, including those with information pertaining to the individual. The FOI Act sets a deadline of 45 days within which agencies must provide the information. Like the Privacy Act, there are a number of exceptions regarding the requirement to provide information. Similar legislation operates in each of the states and territories — for example the *Queensland Information Privacy Act 2009*, the *Victorian Freedom of Information Act 1992* and the *NSW Government Information (Public Access) Act 2009*. These are outlined in further detail in appendix C.

Concerns about the difficulty, complexity and cost of FOI processes are commonplace. For example, in 2006, the Victorian Ombudsman criticised unnecessary delays in processing FOI application and poor quality of assistance for applicants. Others have argued that exceptions to the FOI Act are too broad. Chris Doulton (sub. 145), for example, states that in most cases access to data can be prevented on grounds of an overriding public interest. In 2012, the Office of the Australian Information Commissioner reviewed charging under the FOI Act and recommended simplifying the current charging framework. In 2013, Hawke (2013) recommended changes to simplify FOI procedures and increasing the ability of agencies to manage their FOI workload. Some jurisdictions (New South Wales, Queensland and Tasmania) have introduced reforms designed to make reliance on formal FOI requests a last resort — hoping to encourage a 'push' rather than a 'pull' model of government data release.

Additionally, many countries have legislated access to personal information held by government agencies and private organisations via 'data protection' or privacy laws. In others, specific statutes give rights of access to information in such areas as health and the environment. There are also codes of practice that spell out the procedures for access. Some have Internet portals through which individuals can access (though not necessarily correct) government-held information about themselves. In Australia, this is currently available for the several government services that are linkable through the myGov platform. Usage of the myGov platform has increased significantly in recent years (not least because completing a digital individual income tax return now requires use of myGov).

Other initiatives such as the MyHealth Record may ultimately prove to help individuals to access and manage their health information. Adoption has been slow (appendix D), although it has improved in recent times (with about 4 million users reported in 2016).

3.4 What is holding the public sector back?

There is a wide variety of uses for public sector data beyond government, and much value in its release (chapter 2). But most public sector data in Australia is retained and used only by original data custodians. Despite high aspirations, public sector data remains best characterised as 'closed by default'.

Different datasets have different benefits and costs of release. For instance, there may be large quantities of data that are likely to have limited value for purposes other than those for which they were originally collected — perhaps data on the breeds of licensed cats in a particular local council area. On the other hand, release of some other public sector datasets — such as data on the likelihood of failed hospital procedures — may carry more risk, but also have the potential to provide much greater benefits to the community. It is important that risks of any data release be properly managed, but it is equally important that the severity of the risks be balanced against the potential value in releasing the dataset. Of course, it can be difficult for governments to estimate the value of the data before it is released, but governments should not abandon attempts to do so — instead, novel solutions may be needed (chapter 2).

Many factors contribute to Australia's public sector data being more or less closed by default. Those relating to confidentiality protections for identifiable information and data interoperability apply broadly and are discussed at length in chapters 5 and 6, respectively. This section examines some more narrow limitations on data sharing and release, including:

- incentives for data sharing within the public sector
- a lack of whole of government approach to data sharing and release, and disruptions from machinery of government changes
- intellectual property issues.

The role of incentives and culture in the public sector

Government agencies vary in their approach to releasing data. Some, like Geoscience Australia, have been responsible for releasing the vast bulk of Australian Government open data to date. The BoM is another leading example. But these agencies, and other agencies dealing with scientific, technical and spatial information (which makes up the bulk of the Australian Government open data — figure 3.1) are often dealing with non-sensitive data. For such data, there are often limited risks associated with release and a clear public benefit.

Where releasing data carries with it more risk — for instance, data that, if not properly de-identified, could potentially identify an individual or business, or otherwise carries with it some political or commercial sensitivity — governments have traditionally preferred to not release the data at all rather than take steps to *manage* the risk.

While in some cases genuine legislative barriers to sharing data exist, a public sector culture that is risk-averse and creates reluctance to share or release data has been described by both members of governments and the general public as perhaps posing an even greater barrier than the actual legislation.[12] For example, the DPMC (2015, p. 7) has cited the lack of 'a culture releasing data or widespread use of data for administration, policy development, service delivery or regulatory functions' as one of the five main barriers to the more effective use of public sector data. The Government 2.0 Taskforce (2009) described the existence of this culture:

> [There is] a public sector decision making culture which focuses on avoiding mistakes or embarrassment and achieving consensus rather than the seizing opportunities. … officials and politicians will also be considering how information might be 'spun' by the media, their opponents or those with direct commercial interests or an axe to grind. … a practical obstacle may be an agency's concern about the real or perceived potential for organisational, professional or personal embarrassment. (p. 50)

Other independent bodies have pointed to a culture of reputational risk-aversion in the public sector. The OAIC (2013) reported cultural resistance to open access, particularly in some larger agencies and in sections of agencies that operate independently. The Australian Law Reform Commission (2010) has noted that problems with data sharing in the context of family violence services are not simply legislative, but also cultural and administrative. Chris Doulton (sub. 145, p. 2) commented that:

> Currently there seems to be a Govt wide culture of denial and inhibiting access. There appears to be far too many varied opinions of the Privacy Act and how to interpret it within Govt Departments. Default position is to say no as this is SAFE. …
>
> Our experience in small business has been that the standard response from 99% of Government agencies is to say no ! — It's a NO without thinking, really reading, listening or considering. Everyone wants to quote The Privacy ACT — without really knowing if the current legislation is really applicable. Perhaps they are just scared of making a mistake whilst making a decision — so the default position is to say no ! and therefore do nothing ! — all at the Tax Payers expense.

Moreover, the Attorney-General's Department (sub. 209) states:

> Agencies regularly point to concerns such as privacy or security, but the real issue can be a reluctance to make data available. This reluctance can be driven by a concern for how an agency's information will be used by other entities (loss of control), concerns about the cost of changing systems and processes to enable sharing of data, and concerns about exposure to criticism and/or legal risk. Public servants are primarily focussed on the actions and performance of their own agency, whereas sharing is likely to be for the purposes of other

[12] Such a culture is not confined to the Australian public sector. Schrier (2014) describes a tendency of reputational risk aversion in the United States, and in a survey of government officials from several different countries by Martin (2014), around 75% of survey participants broadly agreed with the statement 'government organisations tend to have risk-averse cultures and so presume that access to data should be restricted'.

agencies. In addition, public servants have a positive obligation to control access (the need to know principle), which makes increasing accessibility counter-cultural. (p. 1)

This culture of risk aversion has a small kernel of justification — getting things right, even when they are high risk, draws little attention, but mistakes are magnified for public servants when things go wrong, as the recent Census showed.

But such risk aversion is also reinforced by an array of other incentives faced by government agencies and their employees, which militate against data sharing or release. Overwhelmingly, the public sector does not seek to accept risk, it prefers to *avoid* it altogether leading to suboptimal outcomes.

If this is to change, leadership regarding aspirations, expectations and intent must be clearly and consistently communicated by agency heads and at the political level.

Risk is often a manageable concern

Public sector agencies face a range of risks when sharing or releasing data. In addition to confidentiality concerns (discussed in chapter 6), departments risk revealing poor data management practices and losing control over evaluation of their policy and its implementation (Stewart 2015; Mathews, sub. 36).

Whilst these risks are primarily reputational, there is also potential for real harm where inaccurate conclusions are drawn from poorly documented data collections. As discussed by the DIIS, such risk-aversion may arise where data custodians are concerned about possible misinterpretation:

> When working with administrative data, it is important to understand how the data was collected and how the nature of that source shapes the information and conclusions that can be drawn from it. … It is common for this knowledge to be tacit rather than documented, meaning that often only people who routinely work with the datasets are familiar with the range of limitations that apply to this data. When the keepers of such data are approached by others wishing to use it, awareness of the limitations of the data coupled with the magnitude of work required to adequately document these creates a bias against release.
>
> Furthermore, even when documentation is available, some keepers of data do not trust others to understand the limitations of data, and fear any potential misinterpretation. The risk of potential misinterpretation is high for some of the department's programme data holdings, as stakeholders may interpret results from analytical work in ways that the data stewards may not. (sub. 69, pp. 4–5)

In many cases, improved curation of datasets would reduce or eliminate concerns about data misinterpretation.

Concerns about a loss of control also appear to underpin the lack of data release. Some government agencies expressed to the OAIC, in its 2013, survey a concern relating to the loss of control of the data they release:

> The more information you release, the more publicly available the material is, and the less control we have over the information … this is where our reputation or how we are reflected as the Australian Government might be in danger. (Large government agency quoted in OAIC 2013, p. 24)

In a similar vein, Stewart (2015, p. 114) describes 'continuing resistance within the public service, at least within some departments at senior levels, over promoting disclosure at the possible expense of the involvement of the department in policy-setting.' That is, release of portfolio data means the agency also shares with other data analysts the capacity to comment on, and evaluate, policy options for their portfolio. In other instances, attempts to coordinate the release of datasets appear to delay release. We have been made aware of several datasets — including performance indicators of the school systems — that were withheld from public release until publication in the Commission's Report on Government Services, resulting in unnecessary delay.

As discussed in greater detail in the next section, concerns about control across jurisdictional boundaries are also apparent. Commonwealth government agencies have demonstrated an ongoing general reluctance to share data with state and territory agencies, even in areas such as education and social security, where sharing would demonstrably improve service delivery. Similarly, state and territory agencies have often been reluctant to share jurisdictional data with the Australian Government; this may be because of state concerns that an uninformed analysis would reflect adversely on perceptions of their performance and/or funding, or because the state is seeking to use the data as leverage for additional funding from the Commonwealth (Mathews, sub. 36).

DRAFT FINDING 3.3

Despite recent statements in favour of greater openness, many areas of Australia's public sector continue to exhibit a reluctance to share or release data. The entrenched culture of risk aversion, reinforced by a range of policy requirements and approval processes, greatly inhibits data discovery, analysis and use.

Lack of a whole of government approach to data sharing and release

While culture and underlying incentives are important, a further cause of the present malaise regarding data sharing across the public sector is an absence of clear enabling frameworks to deliver a more cohesive, whole of government approach.

The present approach is not providing coherent outcomes

Currently, data sharing between departments, and/or open data policies, generally proceeds via a patchwork of mechanisms. In the Commonwealth and most of Australia's states and territories, sharing non-released data (even when not personal in nature) with other

agencies is too often a piecemeal process with a mix of different types of data sharing agreements including MOUs, contracts, deeds, letters of exchange, undertakings, licences, head of agency/ministerial agreements, and public interest certificates (NSS 2009, p. 4). Arrangements might be one-off or ongoing, may or may not involve a payment for costs, are typically long and complex and involve negotiation.

Moreover, even though they are not legally binding, MOUs often involve legal advisers in their drafting, adding complexity and cost when it is almost never needed — agencies are unlikely to ever be authorised to sue each other over data sharing. As appendix B explains in more detail, apart from MOUs and situations where sharing is required by legislation, data sharing between two government entities is still unhelpfully considered to be sharing data externally, even when the two entities are under the same government.[13]

In March 2016, the DPMC released guidance for Commonwealth public sector data sharing that recommended a move away from MOUs to less cumbersome letters of exchange between entities. This was due to the view that MOUs were unnecessarily complicated and time consuming, taking several years and multiple agreements to establish while not being legally binding (a previous DPMC (2015) report cited one agency that reported having up to 11 data access MOUs simultaneously with the same department). The guidance document also stated that government entities were to foster a culture of trust and collaboration with each other, and should provide data in high-quality machine-readable formats that comply with agreed open standards, with as few restrictions on use as possible.

However, many agencies continue to publish data in non-machine-readable formats, in some cases in an apparent effort to preserve confidentiality.

A dense web of legislative requirements

Additionally, departments must deal with a complex array of legislation. This often requires individual departments to consider their enabling legislation (where relevant) and a range of other legislative instruments, including the Privacy Act and regulations specific to their portfolio or program area.

Participants have noted the complexity involved with ensuring compliance with law across all jurisdictions, managing inconsistencies in data collection practices and coordinating permissions across multiple and diverse data custodians (Department of Social Services, sub. 10; Department of Environment and Energy, sub. 120; Department of Prime Minister and Cabinet, sub. 20). Indeed, WA Data Linkage (sub. 13, attachment 3) list 22 separate policies, relevant documents and pieces of legislation with which they must comply when undertaking data linkage. Similarly the Attorney-General's Department (sub. 209) states:

[13] For situations where government entities do reach an agreement to share data within or between jurisdictions, the Australian Government operates a physical infrastructure for secure, encrypted information sharing, called FedLink. See Appendix B for more detail.

> Government data is subject to a complex legislative and administrative regime that includes regulation of general application, such as the Privacy Act 1988 and the information security requirements of the Protective Security Policy Framework, and a plethora of subject specific regulation, including more than 500 secrecy provisions across the Commonwealth statute book. (p. 1)

There is no overarching legislation at the Commonwealth level or for most states and territories that addresses, in a whole-of-government way, how data is made available and used. Several Inquiry stakeholders argued that a dedicated omnibus piece of legislation, covering data release and use at the Commonwealth level and, potentially, between the Commonwealth and other entities, including state governments, researchers and the private sector, is needed. For example, WA Data Linkage (sub. 13, p. 5) have called for the enactment of legislative enablers of data linkage, both in Western Australia and nationally. The Queensland Government (sub. 207) similarly noted the detrimental effect that multiple and inconsistent privacy laws have on sharing and the need to simplify data sharing:

> Research shows most citizens and business see different levels of government as a single, amorphous body which they expect to be sharing data. Private organisations are able to share data with other organisations when providing services, or monitoring risks, but governments often only do so when they are forced to do so via legislation. For governments to deliver services in the 21st century, they need to see themselves as the 'government sector' and not isolated eco-systems … Many pieces of legislation are drafted with explicit barriers to data sharing due to perceived risks, or have contradictory or overlapping data sharing restrictions. For public sector data sharing to be improved, efforts at all levels of government need to be made to significantly simplify the legislative framework with respect to privacy (Privacy Act 1988 (Cth) and/or Information Privacy Act 2009 (Qld)) as the cornerstones on which all data sharing is built upon (p. 8)

Moreover, the Bureau of Meteorology (sub. 198) states:

> The list of legislation, international treaties and agreements, national strategies and policy statements that applies to the Bureau's activities is extensive. Policy and regulation development is struggling to keep pace with government administrative changes, let alone the consequences of the development of digital technologies, increased user expectations, and increasing volumes of data. Navigating between differing interpretations and applications can lead to suboptimal outcomes and impose considerable overhead and often inertia… The Bureau considers that its users would benefit from a lighter touch and more consistency across applicable policies and regulation relevant to data. (pp. 10–11)

Only New South Wales has legislated to overcome this complexity in a somewhat comprehensive fashion, by creating a whole-of-government system for inter-agency data sharing (the *Data Sharing (Government Sector) Act* was passed in late 2015) (box 3.12). More recently, South Australia has also made promising progress, with the introduction of the SA Public Sector Data Sharing Bill 2016. This aims to legislate the 'five safes' model (discussed above and in appendix B) to enable data sharing across trusted entities.

Jurisdictional and sectoral barriers are a further issue

Despite these recent moves, the common practice — particularly at Commonwealth level — is still to impose barriers to data sharing and release, both within sectors and across jurisdictions. The SA-NT DataLink (2015) has argued that there is significant scope to improve state access to Commonwealth data for the purposes of linking population health data. Regarding education, they have further noted:

> Education is primarily a State/Territory responsibility but also Commonwealth (and privately) funded. These two jurisdictional areas have differing priorities and accountabilities…incentives between government and private school sectors to participate in more open data collection would vary … differences in the legal and policy governance and authorising environments between jurisdictions present particular challenges. … While it would be ideal to create a nationally consistent approach in the provision of data, and its governance, Australia's federated model works against this (SA NT DataLink 2016b, p. 5).

The adverse impacts of such jurisdictional barriers were described by the NSW Government as manifest in the:

- the additional effort needed by agencies or sectors to create one-off arrangements for the flow of information (for example, current work to establish a national picture on education outcomes)
- difficulties faced by organisations trying to grapple with national issues but having only partial, fragmented data (for example, child protection care arrangements and payments)
- the challenges when service provision involves non-government providers which may lie outside or in differing formal jurisdictions of public access and privacy regimes
- the varying approaches between regimes on the extent to which they focus on the agency-public axis alone, neglecting agency-agency sharing opportunities. (sub. 80, p. 10)

As pointed out in that submission, and discussed in greater detail below, data sharing is even more complicated when non-government entities, such as those created via public-private partnerships, are involved.

Failure to adequately consider data access in machinery of government changes

The complexities of data storage and management are exacerbated by government agency restructuring (often after a change of government). Machinery of government (MOG) changes sometimes require agencies to transfer responsibilities for data collection, storage and custodianship. This can disrupt projects that are underway, including sharing arrangements (which sometimes take years to progress even in the absence of MOG changes) (DPMC 2015). Moreover, where all data functions and responsibilities are not passed on, MOG changes can result in single datasets having differing collectors and custodians, or the data custodian being lost altogether.

Whilst policy to govern these transitions is in place (in which the National Archives is responsible for providing guidance on policy, mechanisms and standards for the transfer of

information), MOG changes are nonetheless disruptive. They have, for example, arisen during the late stages of integration projects (ANDS 2016), as well as during efforts to create new data collections (NSS 2013). The OAIC (2013) reports that MOG changes increase the complexity of information storage, in some cases resulting in systems that are within the same agencies but not compatible with one another. As the Department of Employment (sub. 18) states:

> Machinery of government changes also present challenges in terms of data management over the longer term. Data access may be lost when a function transfers from the department and often complex access agreements require lengthy negotiation thereafter to reinstate access. When functions are supported by IT infrastructure this also requires complex arrangements to ensure business continuity. The implications of this is that significant resources are invested in re-establishing basic data management practices, such as putting in place new procedures and protocols, as well as technical and service delivery solutions, rather than on value-add projects to improve our use of data. (p. 5)

At the Commonwealth level, the frequency of MOG changes in recent years has proven significantly challenging for the management of some datasets, and the jurisdictional complexities outlined above are magnified when MOG changes are overlaid.

As described in greater detail in later chapters, the realities and complications created by multiple jurisdictions are nevertheless manageable with a more effective approach in place.

Other countries have taken steps to streamline their data sharing arrangements

Whole-of-government approaches to data release are gaining support in some nations that are exemplars of leading practice. In the United Kingdom, draft legislation covering the better use of data in government was being developed across 2015-16, prior to the referendum on Britain's membership of the European Union (Cabinet Office (UK) 2016). The intended legislation is instructive as it provides an example, albeit in draft form, of the main elements that could form part of a legislative instrument in this area.

Based on the consultation processes that were undertaken for this legislation, it would have likely included several elements, such as proposals to:

- improve public services, by allowing public authorities to share personal data in specific contexts to improve the welfare of a specific person, and to access civil registration data (births, deaths and marriages)
- improve use of data for research and official statistics. This included giving the UK Office for National Statistics access to detailed administrative government data to improve their statistics; and proposals enabling the use of de-identified data in secure facilities to carry out research for the public benefit.

At the time of writing, the legislation was yet to be enacted.

New Zealand's Integrated Data Infrastructure (IDI) has been in place for around six years. It was introduced without new legislation, but has instead relied on several important amendments to existing legislation — in particular, amendments to the Privacy Act in 2013 to enable Approved Information Sharing Agreements between government agencies (Yarnell 2016, p. 1).

There has been some limited commentary around the lack of a dedicated legislation to underpin the IDI. The New Zealand Law Commission (2012), for example, considered in 2012 the possibility of introducing a specific legislative requirement on agencies to proactively release material.

The New Zealand Law Commission's view was that such a legislative requirement could steer a middle course between a voluntary approach and a more prescriptive, mandatory approach (such as via the stipulation of specific datasets or categories of data for which release was required) (box 3.12).

Box 3.12 The New Zealand Law Commission's consideration of 'umbrella' legislation

The new provision we recommend is a flexible one, so that each agency will be able to develop its own reasonable standard of release, rather than having to meet an externally imposed requirement. This means that some larger central agencies will be required to take meaningful steps to implement a proactive release strategy, in light of the government directive, while some smaller agencies such as school boards may not have to do much at all in present circumstances.

An umbrella provision would essentially provide a statement of principle as to the desirability of proactive release, which agencies could implement as their particular circumstances allow. It would send a clear signal to all public agencies about the expectation that proactive release will be increasingly used as a channel to release official information and encourage agencies to do what they can to improve the availability of official information in this way.

In legislative terms, an umbrella provision may be less complex than differentiating between agencies under a staged approach, as it would be clear that the legislation applies to all agencies without carve-outs, although the strength of the obligation may vary depending on the circumstances of the particular agency and the information they hold. An umbrella provision would also ensure that there are no gaps in coverage between the official information legislation as it relates to proactive release and the Declaration on Open and Transparent Government or NZGOAL. An additional factor that supports an umbrella provision is ensuring that New Zealand does not fall behind comparable overseas jurisdictions that have opted to introduce mandatory proactive release requirements.

Source: New Zealand Law Commission (2012, p. 268).

The New Zealand Government did not implement the Law Commission's recommendation. Instead it amended the NZ *Privacy Act 1993* to include an 'Approved Information Sharing Agreement' that *authorises*, by Order in Council, departures from the privacy principles (except those giving consumers rights to access and correction) if there is a clear public policy justification and the privacy risks of doing so are managed appropriately.

The New Zealand Data Futures Forum also discussed legislative aspects in its 2014 report (New Zealand Data Futures Forum 2014, pp. 27–29). Their view was that both review of existing legislation, and legislative change, were required. On the former, they argued that a broad review include consideration of:

> … the Copyright Act and other intellectual property legislation, the Official Information Act, Privacy Act, Public Records Act, Statistics Act and consumer law, with the purpose of achieving better, faster, trusted and more collaborative use of data and a more coherent and responsive data-use ecosystem. (New Zealand Data Futures Forum 2014, p. 29)

The Forum endorsed the earlier recommendation of the Law Commission regarding the introduction of a statutory release requirement.

In sum, the lack of a whole-of-government approach appears to impede data sharing in Australia's public sector, and sharing between sectors and jurisdictions remains particularly challenging due to legislative complexity and inconsistencies.

Intellectual property

Intellectual property can limit the sharing of public sector data. In some cases, governments cannot share or release data due to a third party holding intellectual property over it. In other instances, governments hold intellectual property rights over data and use licenses that limit the extent to which it can be redistributed and re-used.

Public sector licensing arrangements can restrict data sharing and reuse

Australian intellectual property law provides for copyright in datasets where compiling information involves sufficient 'authorship'. This applies to some public sector databases, but not all (appendix C). However, where copyright exists, it can inhibit access. Indeed, the Tasmanian Government (sub. 205) notes that common practice in the past has been to release data under crown copyright and in non-machine-readable formats. In such instances, data holders appear then to possess an exclusive right to reproduce, publish, adapt and communicate the copyrighted content (albeit with some important exceptions — see appendix C) in the absence of explicit permission or licence.

Even where copyright does not exist, uncertainty around the copyright status of a dataset can restrict its use. As such, licensing of datasets can be crucial to allowing the use of copyrighted public sector data.

Copyright limitations on the use of government data have been recognised for some time, and there have been a number of initiatives aimed at encouraging the application of open licensing in recent years. Internationally, open licensing has been a key concern of the open data movement since its very early stages. In 2007, key open government data advocates and pioneers listed the absence of copyright restrictions as one of eight Open Government Data principles (Tauberer 2014).

In Australia, several policy documents have been released to encourage open licensing, including:

- the *Guidelines on Licensing Public Sector Information for Australian Government Agencies* (Attorney-General's Department 2011), which specified a default position that public sector information (defined as material with the essential purpose of providing Government information to the public, including forms of data) should be released under the Creative Commons Attribution Licence
- the *Australian Government Public Data Policy* (Turnbull 2015), which specified that all non-sensitive public data (defined as data collected by government entities for any purposes including government administration, research or service delivery) should *inter alia* be published under a Creative Commons Attribution Licence unless a clear case is made for another open licence.

Additionally, the Australian Government's Open Access Licensing (AUSGOAL) framework provides guidance on licensing decisions (box 3.13), with endorsement of this framework in most state and territories, including New South Wales, Tasmania, Queensland and Western Australia.

Box 3.13 Copyright licensing of datasets

The Australian Government's Open Access Licensing (AUSGOAL) framework provides guidance for the licensing of government-held information. It endorses eight licensing options, including the six Creative Commons 4.0 International Licenses, the Restrictive Licence Template (RLT) and the BSD 3-Clause Software Licence. The Creative Commons Attribution licence is the most common, allowing others to use, and distribute content as long as they credit the copyright holder. Other Creative Commons licences involve combinations of user requirements, including whether users must apply the same licensing in their use of the work, are able to change the work, or can use it for commercial purposes.

Source: AUSGOAL (2011).

Some evidence suggests that the use of open licensing has been low, although it is not known how much this has changed in recent years. In 2013, an OAIC (2013) survey found that only 28% of agencies reported using the CC Attribution licence by default, with another 5% reporting that they had adopted a default position of other open licensing terms. Some government agencies have been reluctant to adopt standard licensing on the grounds that it is not suitable for their needs, particularly for sensitive datasets. The Victorian Government (2009) Open Data Review estimated about 85% of Victorian public sector datasets were suitable for open access licensing — it was mostly confidentiality considerations that prevented open access licensing of the remainder (appendix C). Similarly, the Bureau of Meteorology (sub. 198) noted that open access is not always appropriate for the datasets it licenses, either because of the license restrictions involved or because of the cost of provision (which necessitates provision of the data on a cost recovery rather than open access basis).

Intellectual property can limit sharing when third parties are involved

Public sector agencies frequently use data generated by other organisations, including those from the private sector. The sharing and release of this data can be restricted by laws relating to copyright, patent, confidential information, trade secret and trademark (see box 3.14 for some examples). In many cases, this is unavoidable. For private sector data providers, losing control over the distribution of data would make many data exchanges non-viable. However, where possible, government should endeavour to access information on the basis that it can be shared, released and re-used.

A separate issue arises where governments contract private sector organisations to provide services — such as public transport — and are unable access data relating to the operation of those services. This should be avoided through the establishment of contractual obligations to provide data, particularly where such data is required to monitor performance. At present there appears to be no consistent approach in the public sector for achieving this. This is discussed further in chapter 4.

Box 3.14 Instances of restricted sharing due to third-party copyright

Local councils are custodians of a large amount of data used in land planning and emergency response planning. In some circumstances, councils are offered subsidised land data if the privately contracted collectors retain intellectual property rights (ABRDRSC 2014). Local governments face litigation risks for publishing such data or sharing it with other levels of government.

In a similar case, ABRDRSC explained how the vast majority of the sophisticated Light Detection and Ranging (LiDAR) elevation data held by Geoscience Australia is unable to be used by parties outside of the Australian Government.

This is due to the intellectual property ownership residing with third-party data suppliers. A key barrier to opening up access to the data holdings is the cost of implementing new intellectual property arrangements for approximately 200 previous LiDAR acquisitions (Select Committee on Health 2016, p. 18).

Lack of discoverability is a major barrier to public sector data use

Before data can be used by those other than the original collector, it must first be discoverable. Data registries are crucial to data discoverability, but they are far from common across government agencies. In 2013, the OAIC (2013) reported that only a third of Australian Government agencies had a register of their data holdings, and in most cases this register was incomplete. More recently, the Senate Select Committee on Health reported that 'during the course of this Inquiry [into health policy] it became obvious that some departments were uncertain about what datasets they held' (Select Committee on Health 2016, p. 18).

The DPMC (2015) recommended that all agencies' details of all non-sensitive major datasets (through metadata) on data.gov.au in its report, *Public Sector Data Management*. In the course of consultations, the DPMC developed a register of high value datasets with input from a wide range of agencies. The register included reasons for withholding data, which ranged from lack of resources to the need to delay release until a reporting authority had done so. Encouragingly, several datasets listed as unavailable within the register have since been made available — for example, data on visa grants and lodgements, temporary entrants to Australia and permanent migration outcomes (Hemans, Department of Immigration and Border Protection, Canberra, pers. comm., 27 September 2016).

State government agencies have only made limited progress in setting up registers — for example, in 2015, the Victorian Auditor-General reported that major agencies in the Victorian Government did not publish their information registers, and in one case, such a register did not exist. The auditor found that '[c]onsequently, the public has no way to understand the agencies' full [data] holdings, limiting — and in some cases, even removing — their ability to access the [data] that they hold' (VAGO 2015, p. 11).

Data contained in registries also requires metadata to allow potential users to understand specific datasets. However, a substantial part of governments' data holdings is not accompanied by metadata. In fact, the OAIC (2013) found that over 40% of agencies do not regularly include accompanying metadata for their information holdings, and to date, progress in making metadata registries available remains very poor (appendix B). Additionally, lack of standardisation of metadata means the registries that are available are not reciprocally searchable.

There have been steps taken recently to improve the use of data registries and metadata in the Australian public sector. The Australian Government (2015) committed Australian Government entities to publish (anonymised) government data, with descriptive metadata**,** through data.gov.au, with the aim of aiding discoverability. As discussed above, publication of datasets on data.gov.au has been concentrated among a small number of government agencies.

A more comprehensive approach is needed. All Australian Government agencies should publish information on their data holdings to a central registry. This will not only enhance the use of data as it become more discoverable, it will also reduce duplicated data collection, which burdens both government agencies and the public. Despite these benefits, there are some circumstances where data is too sensitive to have its metadata published. However, these cases are very limited, and given the existing culture, decisions relating to whether this information is too sensitive should not rest with the data custodian, but rather, a central agency with responsibility for the register.

DRAFT RECOMMENDATION 3.1

All Australian Government agencies should create comprehensive, easy to access data registers (listing both data that is available and that which is not) by 1 October 2017 and publish these registers on data.gov.au.

States and territories should create an equivalent model where one does not exist and in all cases should make registers comprehensive. These should in turn be linked to data.gov.au.

The central agencies responsible for data should:

- set measurable objectives, consistent with best practice, for ensuring that available data and metadata are catalogued and searchable, in a machine-readable format
- improve accessibility of data for potential data users.

Limited exceptions for high sensitivity datasets should apply. Where they do, a notice indicating certain unspecified datasets that have been assessed as Not Available should be published by the responsible department of state, on the relevant registry.

Additionally, there is little point improving the sharing and release of public sector data if all users need a degree in statistics to be able to use it. A primary purpose of data release or trusted access is to create the opportunity to confirm ideas. This should not be simply limited to those who have access to, or are themselves, statistical experts. There are a number of ways that we can make the benefits of this data more accessible to the average Australian. Once the data is found, it needs to be useable. Not only should data be machine readable and formatted and standardised in a way that allows for easy manipulation and application, but consideration should be given to ways to make open data more usable for the average consumer. This may include greater availability of data visualisation tools.

3.5 Research data reuse

Arrangements for reuse of research data

The public interest in reuse of research data has been recognised by policy in the academic sector. For instance, the Australian Code for the Responsible Conduct of Research (2007, section 2.1) notes that:

> The central aim is that sufficient materials and data are retained to justify the outcomes of the research and defend them if they are challenged. The potential value of the material for further research should also be considered, particularly where the research would be difficult or impossible to repeat. … If the work has community or heritage value, research data should be kept permanently at this stage, preferably within a national collection. (p. 12)

It is worth noting that access to researchers' data does not necessarily mean making these datasets public, given the privacy and confidentiality considerations of many social science studies — these studies often collect data under conditions of anonymity of participants

with mediated or restricted access to data a part of the ethics approval for the research project. But it does mean making them accessible, readily, to other equivalent researchers.

There have been a number of supporting initiatives aimed at providing access to, and increasing discoverability of, research datasets in Australia (appendix B). A key NCRIS-funded development in enabling researchers to discover research data has been the establishment of the Australian National Data Service (ANDS) in 2008. The central service of ANDS is its Research Data Australia discovery portal, which includes 73 453 records. ANDS does not have a curation role, and data is not directly accessible through the portal. Contact details of data providers are listed, along with information on accessibility. A similar tool is offered by CSIRO, with many datasets featured on both portals. Since 2007, the Australian Research Council has encouraged researchers to deposit data from publicly funded research into public repositories and since 2014 has required researchers to outline how they plan to manage data arising from their research.

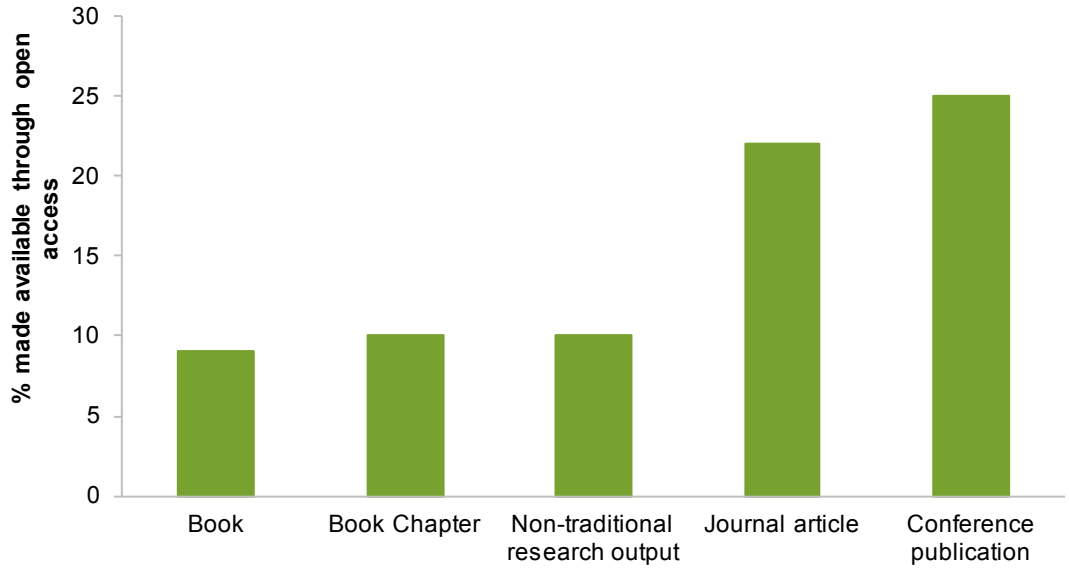
The Australian Data Archive (2016) has suggested that Australia's social science data research infrastructure is rapidly falling behind that currently operating in the United States and Europe — for instance, the Consortium of European Social Science Data Archives, which is a network of national social science data archives with a common core of technical infrastructure, policy and data and metadata standards. These models are designed to facilitate reuse of research data adopted by other countries (such as the United Kingdom, Germany and Norway) and are based around country-specific data archives with the capacity to engage in international collaborative networks to operate as an international facility, including with in-built trusted user capability. Australia's existing arrangements do not allow interoperability with international resources (Australian Data Archive 2016).

Arrangements for sharing and releasing research data in Australia are under review. NCRIS is currently running a consultation on Australia's 2016 National Research Infrastructure Capability Roadmap. One capability of focus is *Data for research and discoverability* — the Issues Paper proposed that research data be Findable, Accessible, Interoperable and Reusable. The Open Access Working Group (involving the Department of Education and Training, the Department of Industry, Innovation and Science, the Department of Health, the Australian Research Council, and the National Health and Medical Research Council) has also been established to look at issues around open access to research data later in 2016. The ability of this group to translate its views into action must depend in substantial part on political and bureaucratic leadership.

Given the limitations in Australia's current arrangements for reuse of research data, it is difficult to evaluate progress. Attempts to monitor data availability are impeded by the lack of a registry of datasets that result from Australian Research Council (ARC) funding. This is despite the fact that research overseen by the ARC is funded publicly and the principle of access is routinely endorsed. One metric that is available is the portion of research outputs, which are very rarely sensitive in nature, but are nevertheless made openly available in only a minority of cases (figure 3.5). Given the additional complexities

involved with making data available, it is likely that the availability of data is lower than this measure.

Figure 3.5 **Open access to output of publicly funded research**[a]
2015



[a] 'Open access' for the purposes of the data in this figure refers to output that is available at no charge.

Source: Australian Research Council (2016).

Current initiatives on research data do not go far enough

There would be significant potential benefits from increasing open access to data generated in Australia's research institutions. Not only is much of this data publicly funded, but as noted by the Australian Data Archive (Australian Data Archive 2016) in their recent submission to the National Collaborative Research Infrastructure Strategy Roadmap, there is a broader public interest in enabling the reuse of research data:

- The storage of data is a requirement of ethics committees and at present it is a challenge for both individuals and institutions to provide the necessary data repositories. It is important that data be available for the purposes of replication and extension of analysis in published papers.
- Reuse of data opens up opportunity for secondary analysis, particularly for PhD students and early career researchers.
- It is costly to collect high quality survey data and so it is important to avoid duplication and maximise the value of data that is collected.

- Access to ABS data can be highly constrained and access to public sector open data is limited — hence access to academic collections can be critical to research.

Greater access could help to build trust in the research sector and support implementation of the findings of the research via evidence based policy. A number of stakeholders have argued there is room to do more to support access to research data:

> Collectively, research data, techniques, algorithms and software (sometimes referred to as non-traditional research outputs or NTRO) are increasingly being considered as key enablers of innovation and especially in relation to improving the collaboration between researchers and business and industry. In the last few years the issue of scientific reproducibility is growing in importance, such that access to the research data is necessary and in many cases, mandatory. (ANDS 2016, p. 1)

> We would also add in principle support for open access to research data arising from government funded research after a reasonable period of exclusive use. Where open access to research data is not possible or appropriate due to privacy or other reasons, a record of the data should be made openly accessible and access to the dataset negotiated between parties. (The University of Sydney 2016, p. 1)

Moreover, increasing access to research data is consistent with a number of recent international academic developments, such as the Open Science Framework developed by the Centre for Open Science in the US, and several journals such as *Nature* requiring the underlying data to be made open as a condition of publication.

Just as data custodians and legislators should be adapting to the ubiquity and essentiality of data analytics opportunities through increased openness, so too should researchers. Researchers are an important group of data generators and users. Within research fields, increasing access to, and sharing of, data has the potential to generate significant gains and is clearly currently less than optimal.

Much research is publicly funded. But in the vast majority of cases, neither the public nor the bulk of any research community has access to datasets generated, despite there being a clear public interest in this occurring. Without more open access to research data, limited reproducibility will remain a major issue. Further, there can be significant duplication in data collection which is a waste of resources. Increasing the reuse of research data — which includes not destroying unique datasets at the end of research projects unless demonstrably preferable (chapter 5) — can improve the reliability of scientific findings through replication studies and increase the rate of scientific discovery through more thorough interrogation of results.

In the recent Intellectual Property Arrangements Inquiry draft report, the Commission recommended that publications from publicly-funded research be made open access after one year. This should be extended to underlying data as well. This will most often be achieved by allowing a wider field of trusted researchers to seek to use the data and replicate results or supplement them. Completely open release is not always optimal. Research data can be commercially sensitive, particularly where there has been a

substantial private sector contribution in funding research; and some such cases will occur. Yet, while it is said that releasing the data might even jeopardise commercialisation of the research, this excuse is over-used at present, as demonstrated in the Commission's Intellectual Property Inquiry. And the Commonwealth Government itself has expressed frustration at the poor conversion rate of intellectual property to commercial development. Timing of making a dataset more widely available is also much more manageable than is asserted at times in support of retaining or destroying the source data.

None of these challenges negate the need to increase access to research data. Rather, they are considerations that should be taken into account in reform design. The public paid for this research, the public should have access to the benefits of this research.

Research data should be made as open as possible and only as closed as demonstrably necessary. We consider that in line with the Australian Government Public Data Policy Statement, there should be a presumption that research data be made open by default (Research Data Services 2016) within a reasonable timeframe of funding or project conclusion, and always after publication of results. Where this does not occur, a public explanation should be provided by the research funder where public funds are involved — 'if not, why not'. Where privacy or confidentiality considerations tell against the public release of data, arrangements should be made to provide the data through secure sharing environments (trusted access models). Specific restrictions in legislation will still need to be observed, and the efforts of governments to address these will determine how effective policies of openness are.

DRAFT FINDING 3.4

There is a clear public interest in having research-oriented data widely available to trusted researchers in a timely manner. A corresponding presumption that it be released needs to be balanced against a number of potentially competing interests, including:

- the need for the researcher to benefit from their own research
- interests in commercialisation of research — for example, if the research was partly privately funded
- specific legislative or ethics approval restrictions
- privacy or confidentiality considerations
- capacity to provide access through secure sharing environments, where privacy or confidentiality considerations cannot be managed to enable the release of data.

Barriers to making research sector data available

As the extent of data collection within the research and university sector has increased, issues around access and reuse have become increasingly important. There are large

potential benefits from the increased sharing and release of research data, particularly in allowing for new questions to be asked using the same data, and for findings from research to be replicated. However, there are several barriers preventing access to research data.

Lack of discoverability of research data

It is impossible to re-use research data if people do not know what data exists. Failing to re-use publicly funded data means significant duplication is likely to occur if the same data is funded to be collected over and over again.

Arrangements ensuring that *all* researchers in a field know of the existence of such uniquely developed data sets could also be substantially improved. The Australian Research Council could, in the Commission's view, be doing more to communicate up-to-date detail on research data and metadata that it has funded and which, under a policy of open by default, should be available for access. Public funding entities are in a unique position to influence a shift in culture, and at the same time strengthen the case for others, such as administrative data holders, to equally act in a more open fashion. Publication of indicators on the proportion of projects involving datasets not available for release would provide greater transparency, as well as the opportunity to ask what is being lost if knowledge of data is limited. Moreover, the publication of research datasets would likely reduce duplication in data collection across research projects.

DRAFT RECOMMENDATION 3.2

Publicly funded entities, including the Australian Research Council, should publish up-to-date registers of data holdings, including metadata, that they fund or hold.

Publication of summary descriptions of datasets held by funded researchers but not released, and an explanation of why these datasets are not available, are also essential and would provide far greater transparency about what is being funded by taxpayers but withheld.

Incentives and countervailing considerations

A key issue is the incentives faced by the data holders in research institutions and the countervailing considerations that may tell against research data being made open. In cases where researchers are the sole data custodians, they hold a monopoly over the ability to derive publications using the data, and often this data is part an ongoing research program. As publications and citations are key drivers of career progression among academic researchers, the loss of sole control can be a strong barrier to the release of research data, as it may jeopardise an institution's control of ongoing research programs.

In other cases, sharing data can affect the commercial viability of research. The insights contained within the data of research projects, once made public, may render privately

funded projects non-commercial. Finally, as noted by Australian Policy Online (2016), copyright can be a barrier to making research open, as it can be the norm for researchers to sign away their rights as a condition of publication.

A factor that cuts both ways is that the release of data can increase scrutiny over research results. In parts of academia, there has been increasing concern regarding replicability (finding the same result following separate data collection) and reproducibility (reusing data to verify findings). While some academic journals now require researchers to submit data (in the absence of a strong reason not to do so), sharing data does not occur in all cases. In a recent study, Chang and Le (2015) attempted to replicate findings from 67 economics papers but were only able to access code and data for 40 of these. The need to re-examine research findings was further highlighted in the field of psychology, in which an attempt to replicate peer-reviewed research findings was successful in only 39 of 100 cases (Baker 2015).

Finally, where research data may be sensitive and have significant privacy and/or confidentiality concerns, there may be good reasons against making the data publicly available — nevertheless registries should surely indicate its existence

Management of this data poses additional challenges (PHRN 2016). Arrangements for secure sharing of research data would be beneficial — no such capability appears to exist currently (Australian Data Archive 2016).

These challenges do not, however, negate the potential public benefit of making research data more open.

Practical challenges for research data reuse

Even where researchers or research institutions want to make the data available for reuse, practical and resourcing challenges can stymie this.

Quality of the data available is a significant issue. No entirely centralised storage facilities are currently available and the costs of storing and curating data can be prohibitive for researchers and research institutions, particularly for very large datasets. Failure to curate and update data reduces its usefulness, and is often the result of insufficient resourcing and lack of skills and technical capability of researchers (Australian Data Archive 2016). At present there is a lack of coordination in Australia's approach rather than whole-of-system coherence (Research Data Services 2016).

Moreover, a lack of common standards for data curation even within disciplines can also pose a significant barrier — making data and metadata be retrievable by their identifier (using a standardised communications protocol) requires the use of a standard that enables at least readability and actionability by machines or by humans, or ideally both (Australian Data Archive 2016). Producing sufficient metadata for even small datasets can be resource intensive. A lack of standardisation of data and metadata can reduce findability and

usability. Further, there is no capacity to disseminate and access this data, particularly in a machine-to-machine way, although the Australian Data Archive is currently working with the Australian Urban Research Infrastructure Network to develop a machine to machine capability for the delivery of data into its environment (Australian Data Archive 2016). Issues of standardisation, storage and data quality are discussed further in chapter 6, and we understand that NCRIS is also considering these issues as part of the development of the Australia's 2016 National Research Infrastructure Capability Roadmap.

4 Private sector data collection and access

Key points

- With the advent of remarkable improvements in the capability of digital data collection, storage and analysis, private sector organisations now rival (and in some areas, such as employment services, clearly exceed) governments for their capacity to collect and use information.
  - This threshold change, in less than two decades, has important implications for the future of data management, not least that much of this data collection occurs in an unequal exchange between well-informed data collectors and less well-informed consumers/businesses.
- While individuals can currently ask to see their own data, there is no obligation on any data collector to provide this in a form useful to consumers/businesses.
  - Consequently, access to privately held data (and the desirability of having more access) varies considerably between industries. Several organisations and reviews have given support to the principle of expanding consumer access to their data.
- There is a good case in principle that businesses which collect data as a result of a regulation or public sector funding should have greater obligations to make data available to consumers in machine-readable form and to governments for the generation of community-wide benefits.
- In the United Kingdom and the United States, models have been established for improved consumer data access in areas such as banking, health insurance and electricity.
- When entering into contracts with the private sector, governments should consider whether there is a public interest from requiring greater access to data that is privately generated but as a result of contractual engagement on behalf of the public sector (for example, public transport).
- In credit reporting, low levels of voluntary sharing support a continuing debate about the merits or otherwise of *mandating* the shared provision of data through the comprehensive credit reporting (CCR) system.
  - Participation in Australia's voluntary CCR by holders of at least 40% of accounts has been proposed as the critical mass needed for the benefits of CCR to be delivered.
    - This benchmark should now be set by the Commonwealth as a target to be delivered by 2018.
  - In the interim, exposure drafts of legislation to deliver mandatory reporting should be prepared and circulated, such that mandated provision of data is in place by the end of 2018 if the benchmark is not achieved.
  - Greater clarity on how the hardship provisions should interact with CCR could help pave the way for broader industry participation in the scheme.

Private entities have always collected data on their customers, including personal details and data relating to transactions. As noted in chapter 1, the expansion in the Internet and the evolution of digital technologies have accelerated the generation and collection of data. With the privatisation of many government services (telecommunications and energy providers for example), the expansion in private sector Internet–based businesses, and intelligent products, much of the growth in what once were large holdings of data in the public sector now occurs within the private sector.

For example, ANZLIC — the Spatial Information Council noted:

> … the private sector has traditionally built on government spatial data and relies on it for their revenue and products. This is changing rapidly, however, as the private sector is increasingly generating this data for itself through technological innovation (e.g. Google Street View) and data volunteered by individuals through apps (e.g. smartphone locations aggregated into a live traffic feed). (sub. 164, p. 4)

This chapter draws a distinction between:

- commercial entities subject to regulation that permits or requires the collection of a range of data in fulfilling public interest obligations and
- those entities that do not acquire data by dint of either regulatory requirement or public funding.

The first group operates in sectors that are typically concentrated and larger scale; where competition can be inhibited by regulatory barriers; and there is often the need for significant upfront capital. Entities in these sectors typically provide complex products and services that are important to individuals and the wider community — hence the public interest rationale for regulation, but also importantly, underscoring the importance of data and information in enabling better decisions — by the business itself, consumers and regulators.

Data collected under such regulation includes that used for forecasting infrastructure investment, maintaining community protection, ensuring the stability of the financial system, and managing risk over generations. Businesses collecting data to meet regulatory, licensing or other public interest requirements have the capacity to also use this data to enable vertical integration and broader diversification in support of competitive advantage. Importantly, because much of this data collection is mandated via regulation or guideline, customers cannot make decisions based on their preferences about what data is collected on them and their activities.

The second group of entities, in contrast, necessarily must depend on the continuing quality of their relationships with customers, including as regards the collection of data, in order to compete and exploit data to innovate. They consequently have generally stronger incentives to respond to customer data needs and interests, and to assist a customer in using their data.

The distinction will not apply perfectly. It is a useful starting point though, when considering the question of whether there is any public policy need to be addressed in

balancing the interests of the subjects of data collection and those of data collectors in the private sector — particularly in light of burgeoning sources of data and exceptional new techniques for data analysis and use. In principle, collection of data for public interest requirements may mean higher obligations on private sector data collectors to make this data available.

4.1 Commercial entities in regulated sectors

A number of commercial entities in Australia are subject to specific industry regulations and requirements that allow or oblige the collection and utilisation of a range of customer, transactional and other data – that is, regulatory requirements may have allowed private 'data monopolies' to emerge and flourish. Banks, health insurance funds, and energy providers are possible examples.

Other circumstances where governments may specify data requirements for businesses can include:

- where there are information asymmetries facing consumers — for example, ingredients of food products (food labelling requirements)
- where orderly markets need to be maintained — for example, real estate markets (for example, standard contracts and clear title)
- where regulatory compliance requires creating or lodging records
- where fitness to undertake an activity is relevant, such as working with children (police checks)
- where government purchasing is a primary source of market activity, such as health services
- where governments act on behalf of the community as a steward — for example, businesses are required to lodge mineral and energy exploration data.

In the context of highly regulated commercial entities, some researchers, such as Joskow and Schmalensee (1986), have noted the concept of a 'regulatory contract' between customers (represented by regulators) and regulated entities. In return for the benefits that are conferred by regulation (such as a monopoly right to sell electricity in a particular geographic area), the regulated commercial entities has obligations, such as to provide a reliable supply. There is an argument that access to data obtained under public fiat could be implied by the concept of the regulatory contract, with obligations on regulated commercial entities to release or share data in some form where there is a net public benefit from doing so. The costs to the entity of making data available would nevertheless be a relevant factor in determining whether there is a net public benefit from wider release or sharing.

Many regulated businesses are already required to make some of their data available to other bodies, such as their regulator. But digitisation of processes and transactions means

that considerably more information is now collected from customers than when current regulatory frameworks and data reporting obligations were established. And customers are becoming aware of how important their data is, if they are to exercise choice or otherwise act in a manner expected of a competitive market, to enhance their own welfare and that of society.

The incentives to share data for businesses protected from a high degree of competition by licensing and yet able to gather data due to government fiat will be less than in circumstances where these advantages are not available.

In such markets, new entrant firms may seek access to customer data held by incumbent businesses where it offers them the chance to compete more effectively. It may be a public policy matter to address this where the absence of access to data itself is a — *or the* — key barrier to entry and hence greater competition. That matter will have to be decided case-by-case, as a competition policy issue and in the first instance by the Australian Competition and Consumer Commission (ACCC). For this Inquiry, the matter of competitors is only relevant in a narrow and defined area — for considering credit reporting — as it is a specific issue cited in the Inquiry terms of reference.

On the other hand, the question of access by individuals to their data *is* central to this Inquiry. As datanomics considered:

> … the development of citizen/consumer side market infrastructures, that enable consumers to make better and informed consumption decisions offers the greatest potential economic and social value creation, however is also the least mature or developed. (sub. 129, p. 14)

Elsewhere in this Report, the Commission proposes a mechanism to allow consumers access to their information, which may, if implemented, address this question across *all industries*. It is usually preferable, if the subject of intervention by government is a common matter across all consumers, for the intervention to also be common across all markets. This tends to incur least compliance cost and greatest understanding and awareness of a new social standard.

Should the proposal made elsewhere in this Report not be addressed, however, it is worth considering whether there are sufficient differences between regulated and unregulated markets *in relation to data collection* to make it preferable as a partial solution to data needs of consumers to address the matter either:

- on an industry-specific basis in future, or
- across all regulated industries.

The Inquiry terms of reference encourages us to look at the finance sector.

Financial institutions

Australia's financial institutions operate under a range of legal principles derived from common law, legislation (including the *Banking Act 1959* (Cth) and the *Corporations Act 2001* (Cth)). Those institutions that meet regulatory requirements (aimed largely at maintaining stability of the financial system), receive the advantage of having customer deposits (up to $250 000) guaranteed by the Australian Government. The Financial System Inquiry (Murray et al. 2014) also noted that there are perceptions of an 'implicit guarantee' in Australia, such that creditors believe that if a bank were to fail, the Australian Government would rescue the institution, which has the effect of reducing banks' funding costs.

In providing financial products and services to customers, financial institutions such as banks and credit unions, collect data related to the customer's identity (name, address, telephone number, tax file number, date of birth, and e-mail address) and data on the customer's financial position (such as their income, expenditure, savings, and credit history).

Where possible, financial institutions collect data directly from the individual. However, some (such as credit providers) also collect information about their customers from third parties such as: credit reporting bodies; other credit providers; organisations with which the financial institution has arrangements to jointly offer products and/or has an alliance with to share information for marketing purposes; marketing businesses; and brokers and other parties who may have introduced a customer to the institution (ANZ 2016). Similarly, personal information may be collected from a customer's referees, their employer, public registers, social media, or other information made publicly available by third parties (PCU nd; CBA 2014).

Financial institutions may also make use of cookies — small files placed on a browser or device by the website, app, or advertisement a person is using — to collect data on people's web browsing habits (appendix F). Each time a person visits Westpac's website, information is collected about their use of the website, such as which pages are viewed, the date and time of visits, location information, IP address, and information about the device used to visit the site (Westpac 2014). (The data collected by financial institutions is discussed in greater detail in appendix E.)

How financial institutions collect, hold, and share personal information is governed by the *Privacy Act 1988* (Cth) obligations in relation to consumer credit reporting, and other specific obligations (appendix C). Under common law, financial institutions have a duty to not disclose to a third party confidential information related to a customer's accounts including ' … information obtained as a consequence of the relationship between the customer and the bank' (McCoach and Landy 2014, p. 89). This duty is excepted only when the use and disclosure is:

- made with the customer's express or implied consent
- mandatory (or compulsory) under law

- necessary for the fulfilment of a public duty (such as in a time of war or emergency)
- in the interests of the institution, which occurs where disclosure is necessary to protect the legal rights of the financial institution — for example, when suing a customer to recover a debt, in which case prevention of disclosure would affect the institution's ability to enforce its rights (Chaikin 2011).

Use and sharing of data in the financial sector

Basic data on customer identity is used by financial institutions to verify identity and establish accounts, to communicate with customers, to market products and services, and to design and price products.

Some data collected by financial institutions fulfils legislative requirements (appendix E). For instance, financial institutions (along with the gambling sector, bullion dealers and remittance services) are required under the *Anti-Money Laundering and Counter-Terrorism Financing Act 2006* (Cth) to verify the identity of clients before services are provided, which may require the collection of information from sources such as passports and drivers' licences, and to perform ongoing customer due diligence (appendix E).

There is potential for data collected under regulatory requirements to be used in the commercial interests of the institution. At least one major financial institution in the United States used its anti-money laundering analysis of accounts to gain insights into customer travel patterns and foreign transactions, enabling the institution to market products such as travel insurance and targeted credit card offers (PwC 2015).

The *National Consumer Credit Protection Act 2009* (Cth) requires credit providers to enquire into the financial situation of individuals who apply for consumer credit. This is intended to ensure that the credit provider has sufficient information to make an informed determination about whether individuals can afford the credit for which they are applying.

As part of their business operations, many financial institutions share data (such as customer personal information) with third parties, including:

- superannuation and managed funds organisations, and their advisers
- valuers, insurers, re-insurers, claim assessors, and investigators
- real estate agents
- loyalty program partners
- fraud reporting agencies
- organisations that assist with product planning, research and development
- mailing houses and telemarketing agencies (NAB nd).

In some instances, when financial institutions share data with third parties, those parties may not necessarily be located in Australia. The Commonwealth Bank of Australia specifies a list of countries that customers' data and information may be sent to, and states that it ensures appropriate data handling and security arrangements are in place for these transfers (CBA nd). The bank also notes however, that Australian law may not apply to some of these entities (CBA 2014).

An important avenue for data sharing by financial institutions is via credit reporting and credit checks. Financial institutions obtain credit reports to get a clearer picture of whether a customer applying for credit or proposing to be a guarantor is likely to be able to meet payment obligations. Being able to obtain these credit reports hinges on financial institutions having shared data with credit reporting bodies (discussed further later in this chapter).

As also discussed in other sectors below, there is likely to be benefit in permitting consumers to have greater access to financial data that pertains to their transactions and accounts. For example, access to such data could help individuals better understand their finances and budget more easily. It may also assist individuals in determining the likelihood of whether applications for credit will be accepted or rejected — because it is presently not customary for financial institutions to explain the reasons for the rejection of a credit card application.

Telecommunications

Australia's telecommunications businesses are regulated primarily under the *Telecommunications Act 1997* (Cth) and the *Telecommunications (Interception and Access) Act 1979* (Cth). Those supplying telecommunications services to the public must have a telecommunications carrier licence issued by the Australian Communications and Media Authority and comply with associated regulations. There are no restrictions on the number of carrier licences that may be issued (ACMA 2016).

Telecommunications businesses routinely collect personal information including identifying and financial information, some of which may be used, for instance, when providing support to customers. Additional personal information may also be collected — for example, Telstra (2015) notes that it may collect some health information from customers to provide priority assistance services or a Centrelink customer reference number to provide eligible customers with a pensioner discount. Optus states that it generally does not collect 'sensitive' information (information such as race, religion, and political beliefs) about individuals, however:

> There may be times when you choose to tell us about your health, and we might collect biometric information for use with new technologies like voice or fingerprint recognition. This could happen as technology changes over time. (Singtel Optus 2016, p. 1)

Like financial institutions, several telecommunications businesses make use of cookies and other digital identifiers to collect information from individuals' online activity. Vodafone states that it uses cookies to measure website traffic patterns and web beacons (a small, transparent picture file used to monitor the navigation activities of a person browsing a particular website). Among the cookies that the company may use is a persistent cookie, which stays linked to a person's browser and records their visits, allowing the company to keep track of the products and services viewed. This is not, of course, a unique factor in telecommunications — many web sites use similar capabilities. Vodafone also states that it may log Internet Protocol (IP) addresses to track user movement and gather 'broad' demographic information (VHA 2016, p. 1).

Telecommunications businesses also collect information from third parties, including associated entities and credit reporting bodies. Telstra (2015) advises that it may collect information from publicly available sources of information, while Optus may buy or obtain information from 'trusted sources' to identify people to whom they may market their products (Singtel Optus 2016, p. 1). Again, it should today be expected by subscribers that almost any service provider may take such steps.

If a business allows payment for services, in full or part, to be deferred by at least seven days, then it can collect credit data and information on individuals' eligibility for credit seek credit reports. The use and disclosure of credit information means that telecommunications businesses have responsibilities under Part IIIA of the Privacy Act and under the *Privacy (Credit Reporting) Code 2014 (Version 1.2)*.

Although outside the scope of our approach to this Inquiry, for completeness, it is worth recording that there are legislative requirements that compel telecommunications co businesses to collect and retain certain types of data. In March 2015, the *Telecommunications (Interception and Access) Amendment (Data Retention) Act 2015* (Cth) was passed by the Australian Parliament. The Act requires telecommunication businesses to retain and secure metadata (information about the circumstance of communications, such as the phone numbers of people involved in a phone conversation) for a period of up to two years (Attorney-General's Department 2015). Furthermore, while the *Telecommunications Act 1997* (Cth) and the *Telecommunications (Interception and Access) Act 1979* (Cth) generally prohibit the disclosure of information obtained during the provision of telecommunications services, there are exceptions that allow information to be disclosed in certain circumstances. Where telecommunications providers disclose information, they are required to keep a record of the disclosure (OAIC 2016c).

Apart from using data for credit checking, and to provide and administer services, telecommunications businesses also use data to: monitor network use and performance; develop and upgrade infrastructure; develop products and services; conduct research, analysis and advertising (Telstra 2015).

In the hands of a customer, their own telecommunications data has the potential to be quite valuable. This data could be used, for example, to solicit competitive offers for service delivery from other providers based on actual consumption patterns and needs. As it

stands, the complexity (and bundling) of phone and data plans makes it difficult for customers to determine and compare genuine alternatives; and yet the burden is squarely on them to do just that. A different approach to data would allow the opportunity for providers to offer customers a better plan, based on actual use data.

Growing demand for connectivity and data usage in mobile data, in particular, underscores the potential significance of customers having more information and control in decision making.

Factors such as complex tariff structures and discount schemes, technological advances, and changing consumer experiences can make it difficult to make value judgements about which plan or plans best suit a consumer. Low levels of consumer switching between providers may be indicative of complexities involved in telecommunication service agreements (Harrison, Hill and Gray 2016). Granting consumers greater access to information relevant to their usage may assist them to understand which services are best suited to their needs and preferences, or facilitate the ability of agents to analyse a consumer's data on their behalf, thereby enabling them to make more informed choices. Indeed, the Australian Communications Consumer Action Network submitted:

> … there are currently barriers to prevent third parties accessing consumer data directly. Ways in which third parties can access this data securely and safely and in which consumers give genuinely informed consent must be investigated, and existing barriers addressed accordingly. (sub. 54, p. 4)

Telecommunications businesses share data with a variety of third parties, including:

- technicians, to resolve faults with service
- market research, telemarketing and marketing businesses
- debt-collection agencies.

As noted in appendix C, the Australian Privacy Principles provide guidance and limits on the extent of information disclosure to other entities, as do industry-specific legislation such as the *Telecommunications Act 1997* (Cth). Further, commercial considerations also play a role in deciding whether data sharing should take place — as stated by Telstra:

> … private sector datasets tend to reflect the business operations of the relevant data holder or compiler, and while the number of these is increasing due to digitisation and innovation, actual availability is a separate issue with access typically dependent on some form of commercial negotiation … we would typically view questions of access to private sector datasets — including any access to Telstra's datasets — as being in the domain of commercial negotiations, reflecting their proprietary nature. (sub. 88, pp. 5–8)

Mobile phone location data could be utilised for a variety of business and social applications (although the disclosure of mobile phone location data is generally restricted under telecommunications law). As noted by Cavoukian and Castro (2014), while an individual's mobility data is unique, re-identification from anonymous mobility traces is difficult, and would require significant additional information from public sources.

Accordingly, a number of alternative approaches have been used by businesses to obtain location data — for example, many apps collect location data, even where that data is not related to the primary purpose of the app, and can combine this data with other customer information.

Energy and other utilities

Retail utility markets in Australia typically have a small number of large firms, or even a single firm, serving a regional market. Many utilities were formerly government trading enterprises, prior to widespread privatisation in the 1990s. Energy and other utilities, such as water and sewerage, collect much of the same basic data and personal information as financial institutions and telecommunications businesses in order to provide services. Utilities may also collect other 'sensitive information' about individuals, such as whether a person uses life support equipment at their home, to determine whether restrictions on disconnecting the premises will apply (EA 2015).

Utilities increasingly have the ability to collect detailed data on consumer use via sophisticated metering technology, commonly referred to as 'smart meters'. Smart meters can enable consumers to view their detailed usage data and better manage their use to reduce utility costs. In Victoria, where smart meters were made compulsory, AGL, Australian Power and Gas, Energy Australia, Lumo, Origin Energy, Powershop, Jemena, United Energy and AusNet services have launched smart meter compatible web portals. Typical uses of these portals for consumers include accessing electricity use data, setting budgets and tracking progress, comparing electricity use to similar households, and projecting future usage (DEDJTR 2016). Smart meters are part of machine-to-machine data and the Internet of Things (IoT).

While smart meters are, in theory, supposed to provide consumers with data that can be used to inform usage decisions, there is some question as to whether this intended outcome has been achieved in practice. In an assessment of the Victorian adoption of smart meters, the Victorian Auditor-General noted that while $9.19 million in benefits from innovative tariffs and demand management were originally anticipated to be realised by 2014, only $0.23 million worth of benefits was actually realised. Consumers' ability to access relevant data appears limited by the nature of the meters (box 4.7).

The Auditor-General also found that the uptake of flexible pricing offers remained low, and recommended that future communication strategies focus on opportunities to use smart meters to reduce energy consumption and take up flexible pricing offers (VAGO 2015b). The ability for consumers to access their data in machine readable format, and to make it available to other electricity providers could improve consumer knowledge of their own usage — and allow expert advice to be obtained more conveniently — and so enhance competition and increase the take up of more efficient and cost effective pricing options.

Billing statements do provide consumers with data on usage, and many utilities also operate online portals where consumers can view data about their usage and see estimates

of their next bill. However, at present, even where consumers use such tools to manage their consumption, understanding how to compare different providers and services may not, like telecommunications, be easy. Privately-operated, third-party price comparison sites provide one option for using to data to try and improve market outcomes, although few actually compare the entire market and a number of sites earn revenue from service providers whose products are compared (ACCC 2014b; ASIC 2012) and so have conflicted objectives.

Because utilities have the option to provide services to customers on credit, they can meet the definition of a credit provider under the Privacy Act. Consequently, utilities may collect and hold credit data related to their customers. They may collect this from the individual customer, from other credit providers, and from credit reporting bodies. Origin Energy (OEL 2016) provides a number of examples of such permitted use and disclosure, including disclosing credit information to third parties such as debt collectors and credit management agencies and disclosing credit information to other energy providers that provide, or are considering providing, credit to an individual.

Insurance

Australia's insurance industry is governed by a number of laws and regulations, such as the *Insurance Act 1973* (Cth) and the *Insurance Contracts Act 1984* (Cth) and the voluntary *General Insurance Code of Practice*. Insurers (along with others providing financial services) are required to be licenced by the Australian Securities and Investment Commission (ASIC) and are regulated by both ASIC and the Australian Prudential Regulation Authority.

The range of data that insurance companies collect on their customers varies with the type of insurance offered. While they collect basic personal data, they also collect details of the product or individual being covered, claims made by customers, details of products used, changes in cover, and any suspensions or cancellations of policies. Many insurance companies also collect data on website usage (Allianz Australia 2016; nib health funds 2016; Suncorp Group nd).

Health insurers collect data about an individual's health, medical history, and associated services that have been provided to them, in addition to pension and health care card numbers, Medicare numbers, income tiers for rebate purposes, and employment details (especially for those participating in schemes such as corporate health plans). Some health insurers also seek information on sporting and lifestyle interests (Medibank Private 2015).

In its submission to this Inquiry, the Australian Dental Association noted that there are some private health insurers that are vertically integrated and operate practices, giving them access to granular data on the pricing, clinical practices of competitors, specific procedures performed, and the identity of patients receiving treatment through HICAPS. The Association argued that this access to data had a 'materially detrimental effect on competition' (sub. 8, p. 2). This may however be more about the impact on competitors,

but adverse impacts on consumers — the primary focus of competition policy — cannot be ruled out without detailed examination by the relevant authority. For the purposes of this Inquiry, it exemplifies the advantage that access to vast quantities of data could offer by way of market power.

Beyond information obtained directly from individuals, insurance providers also procure and combine data on individuals from a variety of third parties. These may include partner companies, insurance brokers and insurance agents (including comparison websites), marketing organisations, industry databases, statutory and government bodies, and service providers (for example, hospitals and medical professionals in the case of health insurance). Such data can provide insurers with a very detailed and insightful picture of an individual's habits, preferences, state of health, and ownership of assets (box 4.1). This has the potential to result in lower premiums for low-risk groups, but may also raise insurance premiums for higher-risk groups, including those who suffer from chronic conditions through no fault of their own.

Box 4.1 Using customer data to estimate risk and set insurance prices

One of the reasons insurance companies endeavour to obtain quite detailed information on the characteristics of individual consumers is to fine tune the pricing of their products. The more information an insurance company has on the person or items covered, the greater their ability to charge prices commensurate with the customer's risk.

Although not subject to the same degree of regulation as private health insurance, car insurance provides an illustration of this practice — a company offering car insurance may use information on a person's age, claims history, driving record (such as penalties for speeding and drink driving), and address when determining what premium they should charge a customer (ICA 2013). Young and inexperienced drivers will typically face higher premiums than those in middle-age. Address can provide information on the likelihood of criminal activity such as car theft and vandalism. Indeed, the Insurance Council of Australia observed:

> The use of geocoding techniques in locating precise geographic coordinates is another example of improving sophistication in data capture and analysis; the pricing of insurance for a motor vehicle can now factor in not only the address where the vehicle is garaged, but also the impact of any nearby traffic black spots. Many insurers are gradually expanding the range and precision of location based pricing as new datasets come to hand. (sub. 66, p. 2)

Taken to its extreme, insurance could be paid based on vehicle use, and as a driver entered more risky driving locations or heavy traffic, insurance paid per kilometre or per hour on the road would increase.

Some insurers use additional information on the behaviour of their customers to tailor insurance policies. Insurance Australia Group, for example, has access to Coles FlyBuys data, which it uses to assist in product design (Williams 2013), while QBE uses 'Insurance Box', a device that plugs in underneath a car dashboard and transmits data such as speed and distance travelled to QBE. The data are used to determine the likelihood of the customer having a collision, and the customer's insurance is priced based on the data (QBE Insurance 2016).

The Insurance Council of Australia recognised the effect of technological advancements on both the volume of data available and the capacity of insurers to collect it:

> The proliferation of data and the technological advances enabling its capture and analysis have already had a profound impact on the industry. Advances in scientific research and other digital modelling has been particularly important in understanding natural hazards and other catastrophic risks. This has enabled more accurate risk-based pricing. For example, while flood and cyclone risk was previously underwritten at the postcode level, increased granularity of data has enabled most insurers to price at the individual address level. (sub. 66, p. 2)

Apart from pricing purposes, insurance companies use the data they collect to administer their services, undertake research, and conduct marketing. Some insurers also use the information they collect to identify and market related services to customers — for example, health insurer Bupa Australia (2015) uses data it collects to determine whether a person is a suitable candidate for participation in a health management program, as well as to advise them of other services that may improve health and wellbeing.

In the case of insurance, it may be that the availability of claims data could help individuals make better decisions about their choice of insurer and coverage with respect to health insurance. However, given the likely confidentiality and commercial sensitivity of the relationship between individual data and insurance outcomes, the rationale for greater accessibility may not be as strong in the case of general insurance.

In relation to health insurance, Medibank Private stated:

> The health sector is very good at generating and storing data. It is less effective at translating this data into useful information. It is poor at linking and sharing information between health professionals, where it could be used to improve health outcomes and system efficiency. Worst of all is the health sector's ability and willingness to share data and information with consumers, whom it fundamentally mistrusts to be able to use health data … (sub. 98, p. 2)

Medibank Private further highlighted that:

> Many Australians have encountered difficulties understanding their cover and it is important insurers help minimise such issues by providing clear, digestible information about cover to consumers. (sub. 98, p. 7)

4.2 Entities in less regulated sectors

The prevalence of corporate entities in data collection in Australia is a remarkable shift from the data environment as little as 20 years ago, when data had yet to become 'big' and social media was far less pervasive in the community. Notably — and in contrast to those commercial entities discussed in the previous section — these organisations and others considered below, undertake their data collection activity without either regulatory fiat or public funding. Social media organisations such as Facebook and Twitter did not exist in 2000, and Google would only be formally incorporated in 1998. These entities are now some of the most active corporate entities in data collection in Australia, and many are able

to combine observations about consumer characteristics to build accurate profiles of individuals. Reflecting the importance of data analytics in the modern era, US author Geoffrey Moore stated that 'without big data analytics, companies are blind and deaf, wandering out onto the web like deer on a freeway' (Moore 2012, p. 1).

Indeed, in many fields, it is probable that private entities know much more about market trends and current economic activity than governments.

- LinkedIn is a repository of detailed data on individuals' employment characteristics, including professional qualifications and employment experience (LinkedIn included details of 3.4 million Australians as at August 2016 (Cowling 2016)). Services such as Seek collate comprehensive data on job listings, including such features as pay range and location. By contrast, employment data collected by government agencies has traditionally focused on surveys and the collation of broad data at industry level.
- With regard to trends in youth suicide and in the spread of flu (or even in possible pandemics, a matter of major if unlikely import to public policy), Facebook and Google may be better information sources for public authorities than traditional data sources. Facebook, for example, has developed tools to assist suicidal people and asks users to report suicidal content to it (Facebook 2016b), while Google has tools to estimate the spread of flu and dengue fever (Google nd), and other researchers, such as the Centre for Disease Control and Prevention (CDC 2016), are now out-performing the original efforts with updated analytics.

Regulation has not been a factor in the growth of these sources of greater knowledge; indeed, it is highly probable that the absence of regulation has helped the rapidity of innovation. This may matter little in public policy terms, or eventually it may matter a lot. Either way, it is an illustration of how far data collection and analytics have moved.

Supermarkets and other large retailers

Supermarkets collect large quantities of data that relate to the shopping habits of customers. Although customers generally have the option of not identifying themselves in supermarkets if they pay with cash, those who use other payment technologies and/or a customer loyalty and reward program will have personal information collected about them. Woolworths (2016) — and Coles similarly — collects:

- personal details such as name, address, telephone numbers, age and gender
- customer reference number or loyalty card number
- whether a customer has taken up other offerings, such as membership of clubs and loyalty programs, financial services products, and mobile applications
- rewards and redemption details applicable to membership and loyalty programs
- whether a customer has a connection with other people whose personal information the business collects or holds, such as family members linked to loyalty program membership

- what, how, and when a customer buys from one of the businesses' stores, or what they have expressed an interest in buying
- a person's stated or inferred preferences.

Other large retailers, such as Myer and David Jones, collect similar information, including information collected electronically, such as device identification information, apps used, and webpages visited. Surveillance cameras in stores are another source of data collection by retailers. Many retailers offer free instore wi-fi for customer use; this is generally unsecured, and hence, traffic sent on the network on an unencrypted basis can be relatively easily monitored.

In the context of retail markets, consumer group CHOICE submitted:

> CHOICE believes the best way to drive greater efficiency in complex retail markets is by giving consumers access to the data collected about them. By allowing consumers to access their transaction and consumption data, and making it sharable in a secure digital format, we can create opportunities for third party innovators to provide services that help consumers. We can also create pressure for product innovation and price-based competition by the businesses that hold this information. (sub. 167, p. 4)

Uses of data collected through retailer programs

The data collected by large retailers, primarily through debit, credit and loyalty cards, can provide powerful insights into consumer shopping behaviours, even at a de-identified level. Loyalty cards are widely used by Australians — in 2015, 84% of Australians were enrolled in at least one loyalty program, and on average, four memberships were held. Some 59% of people were active in all of the loyalty programs in which they were enrolled (Directivity et al. 2015).

Data collected via reward and loyalty cards gives retailers access to very detailed demographic and spending information on their customers. It enables marketing to be targeted at the individual level. However, insights can also be gleaned from electronic payments technology, even if the identity of the purchaser is not known. Credit and debit cards allow retailers to track the purchasing habits of individual cards, providing information on which products are selling well and which items are frequently purchased together. This, in turn, assists businesses to refine store layout and management choices (Graham 2016). The comprehensive data held by supermarkets facilitates a deeper understanding of customer attributes (box 4.2), and, in combination with changes in technology, offers the *potential* to change the way market participants interact.

While these private sector data collections are extensive in scope and application, they were recently assessed by the Office of the Australian Information Commissioner (OAIC) as largely being compliant with privacy requirements (OAIC 2016a, 2016b).

Box 4.2 Secondary uses of data collected by supermarkets

Australia's two major supermarket chains, Coles and Woolworths, have collected significant quantities of data on customer behaviours and attributes, and are exploring new ways to use this data. The vast data accumulated by large retailers has been used to identify the most profitable sites at which to build new stores (Technology Transactions 2013). Coles stated:

> We're always looking for new ways of delivering better value, but today we actually use a very traditional mechanism, looking at types of car, where people live, to calculate their insurance pricing … as technology changes, we will reassess that … (quoted from (Rubinsztein-Dunlop 2014))

Woolworths advised:

> What we've been able to do is take our insurer's car crash database and overlay it with our Woolworth's Rewards database …
>
> Customers who drink lots of milk and eat lots of red meat are very, very, very good car insurance risks versus those who eat lots of pasta and rice, fill up their petrol at night, and drink spirits. What that means is we're able to tailor an insurance offer that targets those really good insurance risk customers and give them a good deal via direct channels … And it helps to avoid the bad insurance risks. (quoted in Ma 2013, p. 1)

More generally, Woolworths' 50% non-controlling stake in data analytics firm Quantium has potentially allowed Woolworths to improve its capacity to derive insights about consumers. It has also provided Quantium with access to Woolworths' de-identified customer data, enabling it to improve its product offering to its own clients, such as eBay, IAG, Suncorp and Qantas (Technology Transactions 2013). Other major clients and partners of Quantium include Coca-Cola, Facebook, Foxtel, Google, NAB and NewsCorp (Quantium 2016).

The depth and breadth of data collected by large retailers, often at a de-identified level, appears to have aided the movement of Australia's two largest supermarkets into other markets that are not associated with their traditional area of business. Coles and Woolworths have opened up lines of business in insurance and credit card provision, and have used their data and research methods to discover insights that can be applied to these areas. Arguably, this has been to consumers' benefit.

Privacy is one lens through which to view data collection and use — that is, in terms of potential risks and costs. But on the other side of the ledger, is the potential opportunity for customers to better understand their own consumption patterns and choices through access to and the ability to share data held about them. It seems highly likely that affording consumers the right to obtain their data and to share it with others would quickly spawn a range of value-adding applications to facilitate better decision making and increased competition from retailers in the face of more informed consumers.

It is worth considering the potential for customer access and portability of loyalty data to enhance competition in Australian retail markets. In principle at least, there seems to be no reason why a consumer's right to access their data should be restricted to just one set of industries, as occurs for example with the UK midata web site. While there is a clear public interest rationale in industries where data is collected under regulation, the benefits to consumers of being able to use their data to obtain a better deal or gain third party advice is likely to apply more broadly.

The case for a wider right to data access for consumers would need to be made on competition grounds. One countervailing consideration is that apart from the potential to increase competition, data portability will also likely result in additional costs to commercial entities (that they may, in turn, pass on to consumers).

The likely costs imposed on businesses should figure in the determination of the public benefit from data portability, and be weighed against positive outcomes such as enhanced competition. Data portability (for which we prefer the term 'data transfer', to avoid confusion with EU standards) is discussed in further detail in chapter 8 of this Report.

Social media organisations

Social media allows its users to create, publish and share information with each other in a ready fashion in web-based environments. There are numerous social media platforms in operation — prominent examples include Facebook, Twitter, Google+, LinkedIn, Snapchat, Instagram, Pinterest and WhatsApp.

Many social media platforms are provided free of charge to the user. However, to be able to use a platform, the user is ordinarily required to provide data to the organisation in question. In a sense, the provision of data is the 'price' a user pays to access the platform.

The type of data provided, generated and shared differs between platforms. Facebook — the most prominent social media company today — collects very detailed information about people from the postings that they make on the platform. Facebook (2015) states:

> We collect the content and other information you provide when you use our Services, including when you sign up for an account, create or share, and message or communicate with others. This can include information in or about the content you provide, such as the location of a photo or the date a file was created. We also collect information about how you use our Services, such as the types of content you view or engage with or the frequency and duration of your activities. (p. 1)

Facebook collects information on the people and groups its users are connected to, and how users interact with those groups. Contact information (such as from address books) is collected if it is uploaded, synced or imported from a device. Facebook also collects device information from its users — that is, attributes such as operating system, hardware version, file and software names and types, specific geographic device locations, and connection information, such as browser type and IP address. The company may receive information from any third parties with which it is associated (Facebook 2015). It has been reported that Facebook uses 98 data points to target advertising to individuals (Dewey 2016). The company has launched a portal explaining to users why they see the advertisements they do (Facebook 2016a).

One of the means by which Facebook obtains data on its users has caused considerable consternation overseas, chiefly in Belgium and France. Specifically, studies in these countries found that Facebook makes use of cookies and plug-ins (appendix F) to track the

Internet browsing activities of its users, whether they are logged in or logged out of the service. The Belgian and French studies suggested that non-Facebook users who visit a page in the facebook.com domain may also end up with cookies set on their device. This form of Facebook data collection, and the activities of privacy offices overseas, is discussed in detail in appendix F.

In Australia, Quantium has partnered with Facebook to measure the effect of Facebook advertising on consumer purchases. A representative from Quantium stated:

> We have been testing this capability with a range of our clients … There have been a number of interesting learnings, including how past-purchase based targeting has led to twice the sales uplift compared with demographic targeting. (Macoun, quoted in Canning 2016, p. 1)

To give another example of the data collected by social media organisations, Pinterest works by having users pin images and videos to a 'board' and browsing what other users have pinned on their own boards. Functionalities include liking and commenting on user's pins, and sharing pins using other social media platforms like blogs, Facebook and Twitter.

Standard data that Pinterest collects are a person's name, profile photo, pins, comments, likes, email address, phone number, and in the case of a mobile device, potentially location data. Users who purchase items via Pinterest also provide Pinterest with their payment information. If a user links their Facebook or Twitter account to their Pinterest account, Pinterest can obtain information from those accounts, such as the user's friends and contact list. The company also uses cookies to collect information, and obtains information on the devices used to access Pinterest (Pinterest 2016).

None of this activity should be a particular surprise to well-informed users of these sites. They are cited more to illustrate the value that data can create than to raise any eyebrows.

The particular business models employed by social media companies differ, but typically, advertising plays a key role. Information collected about site users is, in turn, used to sell advertising placements that targets particular customers or customer segments, including providing suggestions on alternative sites the customer might be interested in (more detail on social media advertising practices can be found in appendix F).

Numerous other websites and apps adopt some of these strategies. This Report cites the trends in order to demonstrate the broad-based nature of information-gathering capabilities that Australians have accepted in large numbers. At August 2016, for example, some 15 million Australians were using Facebook, over 14 million were using YouTube, and 5 million were on Instagram (Cowling 2016).

For the large corporate data collectors, their reputation among consumers matters a great deal. Thus, in 2014 (and again in 2016), Facebook launched and then adapted its News Feed service to broaden user involvement with the site and at the same time in an effort to reduce the amount of 'click-bait' headlines (Al-Erini and Tang 2014). Most news sites have to manage the choice between damage to subscribers through advertiser-driven click bait and the need to continuously upgrade the novelty of a site.

For the innumerable smaller entities, the internal pressure for good behaviour may be lower, although both Apple and Android device providers monitor the behaviour of app producers to a degree. Developers with apps on Apple's App Store, for instance, are required to agree to specific guidelines, and if an app violates those guidelines, the violation must be addressed or the app will be removed from the store. Once an app is installed on a device, a user is prompted for permission the first time the app tries to access information such as location or photos (Apple 2016).

But consumer awareness is otherwise the primary protective mechanism. Governments have very limited ability to intervene effectively, although many are understandably loathe to admit this. Internet filters generally fail, as the technology behind them is easily defeated if (as is the perceived need) they are to be widely and simply adopted. Such interventions can also create a false sense of confidence.

The ultimate consumer choice (where data collection is concerned) is to shut the site's access to their devices. This can be easier said than done, as the capabilities of technology and software are not always transparent in any meaningful way to an individual once initial access has been granted. For these circumstances, greater clarity in the law concerning consumers' ability to cease being tracked or monitored may be needed.

Wearable and mobile application providers

Wearables — principally smart watches and fitness trackers — have emerged as a significant means by which data is generated and captured by individuals (box 4.3). Market intelligence firm International Data Corporation has estimated that worldwide shipments of wearable devices will reach 110 million by the end of 2016, representing growth of nearly 40% over the previous year (IDC 2016). The most popular category of wearables, fitness trackers, are capable of collecting an array of data related to an individual's fitness and health outcomes. Information collected may include the number of steps taken in a day, distance covered, heart rate, food consumed and activity location. Additional personal information is provided to the business that created the device via installation or registration.

Internationally, Australians are big adopters of wearable technologies. In its 2015 Mobile Consumer Survey of 2000 Australians aged 18–75, Deloitte found that 13% of respondents possessed a fitness band, while 3% also possessed a smartwatch. Of the other countries in which Deloitte conducts the survey, only China, where 18% of respondents owned a fitness band, recorded a higher level of fitness band ownership.

Box 4.3 Fitness tracker data collection — an example

Jawbone, a prominent brand of fitness tracker, works in conjunction with an app, and its privacy policy states :

> When you download our UP App, register for an account, connect your Device to your account, use the UP Service or send us requests, we may ask you to provide your first and last name, email address, postal addresses, account name, password, photo, gender, height, weight, and date of birth. You can also choose to upload your address book and Facebook contacts to our servers, or through email address lookup, so we can help you find friends using UP. Other information you may choose to input includes what you eat and drink, your mood, and other activities. We use information on what you eat and drink to provide you with calorie and nutritional information.
>
> … When you use or synch your Device, it automatically transmits activity and physical information to us including, but not limited to, detailed physical information based on monitoring your micro movements, including when you are asleep, when you are awake, when you are idle, and your activity intensity and duration. Some Jawbone devices also capture heart rate and other biometric data. This data is translated into information such as your sleep patterns, calories you burn, activities undertaken and your trends and progress. This data can also provide information on certain conditions you may have. (Jawbone 2014, p. 1)

When an individual uses the app, Jawbone also collects information on the device used by that person, such as manufacturer, model, and device ID. Jawbone uses the information it collects to refine it services, create statistics, and for promotion purposes. The company shares aggregated usage data on its blog and in the media, and shares some data with service providers, such as third party data analytics platforms. Jawbone also facilitates the sharing of information with a person's contacts (Jawbone 2014).

Given the very personal health information recorded by many wearables, they are of increasing interest to health insurance companies, with a number now offering free wearable devices and policy discounts, in exchange for use of the data generated.

Apps (too many to cite) have provided yet another major avenue for data generation, collection, transmission and storage. The most popular apps in Australia are social media, maps, news and weather, and game related (appendix F).

The Office of the Information Commissioner in Queensland highlighted that apps can collect significant amounts of personal information about users, often without them being aware of what information is being collected (OIC (Qld) 2014). This information can include calendar data, Internet usage logs, the user's address book and contact lists, photographs, location data, and information about how the user uses the app (box 4.4). The data is typically collected by app owners, and may be passed on to third parties, such as marketing businesses.

Wearables and apps are discussed in further detail in appendix F of this Report.

Box 4.4 Examples of popular apps and data collected

Among the most popular apps in Australia and overseas are the mobile versions of social media sites, such as Facebook, YouTube, and Twitter. Some other popular social media apps include:

- WhatsApp — a cross-platform mobile messaging app that allows people to exchange messages without having to pay for SMS. Data collected includes mobile phone number, numbers from contact list/address book, but does not collect names, emails and addresses. Users may choose to share location information (WhatsApp 2012).
- Snapchat — provides users with the ability to take a photograph, add text and art if they wish, and share with recipients for a set period of time, after which the photograph will delete itself and be removed from company servers. The app may collect a user's email address, phone number, date of birth, as well as usage, device, content, location and log information. Data may also be collected via the use of cookies, and from third parties (Snapchat 2016).
- Instagram — an app designed for sharing smart phone photos and videos up to 15 seconds in length. Users create a profile, and can view the postings made by those they follow, and vice-versa. Data collected by Instagram includes name, email address, user content, contact list (if a user elects to employ a 'find friends' feature on the service), device identifiers, metadata, log information, and other data collected by cookies (Instagram 2016).

Intelligent products

Advances in technology, especially the development of remote Internet access and cloud storage, have facilitated greater data creation and collection by product manufacturers, and those using appliances and systems made by those manufacturers. Machine-to-machine (M2M) applications and the Internet of Things (IoT) have been particularly significant in this regard. According to Heydon and Zeichner (2015), while the basic technology underpinning the IoT has existed for some time, three key factors have facilitated its expansion:

- decreased cost of intelligent sensors and actuators
- availability of near-ubiquitous connectivity at a progressively decreasing cost
- increased sophistication in handling large volumes of data from disparate sources.

M2M communication refers to the exchange of data between machines, without human interfacing or interaction (Link Labs 2015), and is typified by the use of process specific sensors and devices (Heydon and Zeichner 2015). The exchange of data between machines or devices can occur over the Internet or via other means such as radio frequency identifiers (RFID), near field communication (NFC), or Bluetooth. The networks on which M2M communications take place allow sensors, controls, and other machines to communicate with each other (though not all devices necessarily communicate with each other), enabling them to complete a task (Link Labs 2015).

Examples of M2M applications and the broader group of IoT include smart meters for utilities, smart city architecture, monitoring of manufacturing processes, and the tracking

of freight (PC 2016). A growing number of home appliances also have IoT capabilities, such as refrigerators with Internet connections and sensors which allow users to look at the refrigerator's contents remotely, and ovens that permit remote monitoring (Samsung Electronics 2016). In agriculture, John Deere now includes sensors on much of its farm machinery to communicate information on current activity to an iPad in the cab of the vehicle. For large tractors, data on GPS location, diagnostic readings and current operations (such as tilling and planting) is tracked by the John Deere operations centre.

Related to the IoT is what has become known as the 'Industrial Internet' or the Industrial IoT, which combines the IoT in manufacturing with big data analytics (Accenture and General Electric 2014). A leading proponent of the Industrial Internet is General Electric, which traditionally sold industrial hardware and repair services to earn the majority of its revenue. Faced with increased competition, the company refocused to place a greater emphasis on software analytics. It included digital sensors and microprocessors in its machines, and partnered with Intel for sensor technology, Cisco for network hardware, Accenture for service delivery, and Amazon Web Services for cloud operations. Since adopting this approach, General Electric has embedded software in applications as diverse as oil rigs, power plants, jet engines and rail infrastructure (Iansiti and Lakhani 2014). This has generated data that has enabled efficiencies to be realised, such as after-sales service shifting from reactive, to predictive and preventative, allowing products to be repaired before they fail and to be repaired remotely in some cases (Porter 2016).

The approach taken by General Electric demonstrates the commercial justification for businesses to innovate using data, which in turn benefits customers. In doing so, businesses may form partnerships and share data with each other as they develop data networks and specialise in different parts of manufacturing and service delivery.

4.3 Are current private sector arrangements sound?

The sheer diversity and growth in private sector data collection is a relatively recent phenomenon. Regulatory structures addressing access to data could hardly have anticipated it and are adapting in a reactive and piecemeal fashion. Lack of clarity around access rights — arising for example, when data is generated through the commercial interactions of multiple parties — further muddies the water and can hinder data exchange. Businesses are as much affected by queries over who exactly controls their data as are individuals.

Innovative use of data collection, storage and analysis have afforded significant service improvements to some businesses — indicated by their willingness to pay other businesses for data (and offer discounts and rewards in return for customer data). The case for a general shift in policy towards intervening to ensure one business has the same access to data as another business — which has featured in some submissions — however seems weak. There may be significant shifts in strategic advantage that arise from time to time, but innovation in market places has long involved just this.

There is no evidence of resource misallocation; and competition regulators do not appear to lack the ability to examine conduct that may substantially lessen competition, should such an eventuality arise. The case for intervention must show some form of public interest and although this might be plausible where public regulation has offered an information advantage to incumbents, the case for a general shift is hard to discern on these grounds. We would, however, be pleased to consider further advice on this point.

In the case of consumer to business relationships — and this includes the ability for a consumer to shift allegiance from one supplier to another — the question is substantially more open. Submissions to this Inquiry point to control of data as being a potential impediment to improved consumer choice.

And the terms of reference requires that one particular point of contention in consumer to business relationships — the ability to have positive behaviour in support of a credit rating considered by lenders — be specifically examined.

Some of the data collected by the private sector may have the potential to deliver significant public benefits if shared more widely *across the community* than the incentives of the private owners may dictate. For instance, research into the causes and treatment of cancer is assisted by the collection and collation of data on all cancer diagnoses from the variety of private (and public) sources of such data, including hospitals, pathology laboratories and radiotherapy centres. However, even in this area, any government intervention in private sector data would have to be carefully considered to ensure it maintains private sector incentives to collect, curate and exploit the value of their data holdings.

Voluntary participation of businesses in data sharing

Comprehensive credit reporting

The terms of reference for this Inquiry require the Commission to 'provide an update on existing data sharing initiatives in Australia, including the uptake of the credit reporting framework', and consider recommendations to improve participation in such initiatives.

The credit reporting regime in Australia historically limited the information that could be shared by lenders to so-called 'negative' information about an individual's credit delinquency (Veda nd). The main kinds of information permitted were related to:

- a credit provider having sought a credit report (from a credit reporting bureau (CRB)) in relation to an application for credit by an individual, and the amount of the credit sought
- an individual's current credit providers
- any credit defaults (in the previous 5 years)
- a credit provider's opinion that an individual had committed a serious credit infringement (such as credit fraud).

In March 2014, the provisions of the *Privacy Amendment (Enhancing Privacy Protection) Act 2012* (Cth) allowed credit providers to collect and share positive information (through formal arrangements with CRBs such as Veda). Participation is voluntary, with information being shared on a reciprocal basis (participants have access only to the types of information that they are willing to share) (ARCA nd). Comprehensive credit reporting (CCR) therefore aims to achieve a fuller profile of consumer credit behaviour than what is available under systems where only negative information may be shared.

The efficient and effective operation of credit markets relies on credit providers being able to access sufficient (and reliable) information about borrowers to form a basis for making decisions about who to lend to, and at what price. Where lenders face limitations in the information available to them, they risk lending to borrowers who are unable to meet their repayment obligations; or conversely not lending to parties who are a better risk than negative reporting suggests (all of which affects allocative efficiency). Further, relying on applying *average* risk premiums across broad groups of borrowers results in some borrowers paying a higher (lower) risk premium than warranted (and can lead to adverse selection) (Australian Government 2012).

This is less of a problem for large credit providers who have large customer bases of detailed information and therefore potentially, a competitive advantage in assessing credit risk (ALRC 2008). In addition to improving efficiency in how credit is allocated and priced, CCR might also lead to lower costs of assessing applicant credit worthiness, and help lenders better meet their responsible lending obligations (Australian Government 2012). The Australian Retail Credit Association summarised the benefits of CCR as:

- improved access to credit for minorities that would otherwise struggle to access credit
- more accurate pricing of credit
- promotion of responsible lending
- lower rates of over-indebtedness and default
- greater competition in the credit market
- fraud mitigation to assist in controlling the cost of credit to consumers
- increased financial stability and efficiency across the economy (sub. 87).

Additionally, Dun & Bradstreet noted greater efficiency and effectiveness of data gathering, improved financial literacy, and potentially lower levels of cross subsidisation as further benefits of CCR (sub. 135). [14]

[14] The Financial Rights Legal Centre submitted that those 'advocating for mandatory credit reporting have asserted there would be a number of benefits including better access to finance (at a better or more competitive price) and better lending decisions. There is no independent evidence — taking into account the current regulatory environment in Australia — to support these claims' (sub. 107, p. 7).

Comprehensive credit reporting was introduced in New Zealand in April 2012. The scheme is voluntary, and is not legally binding (ACCC 2015). Initial uptake was slow, with early adopters mainly comprising second-tier financial institutions and energy companies. The system is still evolving, and some institutions have yet to adopt the new regime (Landgraf 2016). Indeed, the Australian Retail Credit Association submitted to the ACCC in 2015 that the New Zealand exchange was missing roughly half of the comprehensive data potentially available to the system (ACCC 2015).

In the United States, the Fair Credit Reporting Act 1970 (US) and the Fair and Accurate Credit Transactions Act 2003 (US) permit the exchange of positive and negative credit information. Data is exchanged on a voluntary basis, and while reciprocity and consistency in the contribution of data has been subject to some regulatory oversight and comment, the United States has experienced instances of strategic incomplete data provision by credit providers (ACCC 2015).

There is little doubt that CCR is a desirable reform. Key questions are how strong are the incentives in Australia to make it work while it is a voluntary system; and what level of participation is necessary in order to ensure that data collected is reliable, contemporary, and gives a full picture of a customer's credit history.

Does the comprehensive credit reporting system need to be mandated?

According to Veda, at present, just over 25% of all retail credit accounts contain some amount of comprehensive credit reporting information — this amounts to 8.41 million account holders (Veda, sub. 163). In its submission to this Inquiry, the Customer Owned Banking Association stated:

> The majority of lenders that are contributing data are doing so in 'private' mode, with very few lenders exchanging data. (sub. 132, p. 3)

Further, Dun & Bradstreet commented:

> The largest two credit providers to begin supplying data remain in this 'private mode'[15] — after more than 12 months. (sub. 135, p. 11)

The Financial System Inquiry (Murray et al. 2014) assessed that participation in CCR by an individual credit provider would depend on the perceived net benefits to that provider. A major institution with a relatively large customer base would provide relatively larger benefits to other, smaller institutions if it participated in CCR early and fully. In the Financial System Inquiry's view, as participation and system-wide data grew, net benefits for all CCR participants would increase. The slow uptake of voluntary CCR in Australia therefore raises the question of whether government should mandate participation, or take some other policy action to encourage participation, so that system-wide net benefits can

[15] In 'private mode', data is shared with credit reporting bureaus, primarily for analytical purposes, but is not mutually accessible to other institutions.

be realised. A number of participants to the data Inquiry also had suggestions in this regard (box 4.5).

Box 4.5 Participant suggestions on comprehensive credit reporting

Several Inquiry participants expressed concern about mandating Comprehensive Credit Reporting (CCR) — for example, the Financial Rights Legal Centre stated that they had 'grave concerns' about the unintended consequences that might arise from mandatory CCR (sub. 107, p. 3). The Customer Owned Banking Association (sub. 132) stated that while it was sympathetic to calls by the fintech lobby to mandate participation in CCR, it was concerned that this will come at too high a cost compared to the benefits. They considered that if CCR were to be mandated, there would be no case for mandating participation by smaller lenders.

The Australian Bankers' Association (sub. 93) submitted that it would be premature for government to mandate CCR, citing investment already undertaken by industry which should be allowed to come to fruition, and the presence of a number of issues affecting the rollout of CCR, particularly hardship reporting in repayment history information (discussed below).

The Australian Retail Credit Association — notwithstanding varying views within its own membership — expressed a belief that industry is best placed to manage and develop frameworks around how CCR is best implemented. It further noted that a mandated approach may not be economically viable for some organisations and may affect their competitive position (sub. 87).

Veda (sub. 163) suggested that a Treasury-led reporting mechanism on CCR be established, in which CEOs of leading credit organisations reported to the Treasury on their contribution of CCR information. Advice would be required on the intention of the organisation to start CCR contributions, providing a start date.

Other submissions suggested a stronger approach for accelerating participation in CCR. For example, Financial Institutions & Management Advisory (sub. 73) suggested mandating partial reporting of CCR by all licensed credit providers by the end of 2017. This would involve data on the date an account was opened and closed, as well as the type of credit and credit amount. The Advisory also recommended that the reporting of consumer credit defaults be mandated for all licensed credit providers by mid-2017.

In the opinion of Tyro Payments:

> … banks seem to think that locking in the customer rather than competing is in their interest. Thus the voluntary approach of industry self-regulation has again wasted time and has failed. That is not new. (sub. 7, p. 5)

InFact Decisions and Verifier Australia (sub. 27) recommended that full CCR be enabled either by mandatory means, or by supporting voluntary CCR by including telecommunications and utility data, and by mandating partial CCR to enable consumer opt-in.

The incentives for an institution to participate in voluntary credit reporting are not strong. As the Financial System Inquiry (Murray et al. 2014) noted, for credit providers, participation depends on the perceived net benefits, which will differ between different classes of credit provider.

For a major institution with significant data holdings, early and full participation may provide, at least initially, relatively low benefits and will certainly incur costs. These costs

(making data machine-readable in a consistent format) may be encountered anyway by an innovative institution, and as such could be considered sunk. It is doubtful however that shareholders of such institutions would agree. For small or new participants, participation would provide access to a greater pool of data with potentially significant benefits.

The case for CCR rests on the public interest that is likely to be served if a CCR system is more complete, enabling both more accurate rating of a consumer's credit status, and that resources are allocated more efficiently. The case for simply assisting an entrant over an incumbent is not strong: many markets have aspiring entrants.

Several Inquiry participants indicated that uncertainty around the way in which CCR interacts with the hardship provisions of the National Credit Code (in Schedule 1 to the National Consumer Credit Protection Act 2009 was discouraging participation in the scheme.

The Australian Retail Credit Association (sub. 87) and Dun & Bradstreet (sub. 135) suggested that an April 2016 Financial Ombudsman Service determination (on the reporting of repayment history information (RHI) for a consumer who failed to meet monthly repayment obligations) means that credit providers should not report consumers who are permitted to meet repayment obligations at a later date as having failed to meet their repayment obligations. Dun & Bradstreet (sub. 135) submitted that under this determination, the repayment history of groups with different risk profiles would be represented in the same manner, making it difficult to assess differences in risk.[16] The Australian Retail Credit Association stated:

> We have been advised by a significant number of credit providers that the issue around reporting RHI [repayment history information] during arrangements is a road block to their transition to CCR. Those credit providers that have already transitioned to CCR have indicated they are in the process of returning to 'negative mode'. (sub. 87, p. 14)

Greater clarity on how the hardship provisions should interact with CCR could help pave the way for greater industry participation in the scheme. Alternatively, the inclusion of a hardship flag in credit reports could address the concerns expressed by participants to this Inquiry. Evidence from submissions indicates that this issue should be resolved to enable the benefits of CCR to be realised.

A way forward

We note that the comprehensive credit reporting system is relatively new, with the necessary legal, regulatory, and industry frameworks required to facilitate CCR only completed in December 2015. Until then, industry had no certainty that a workable model

[16] Dun & Bradstreet stated that those making repayments in full as per their original contract, those who had been granted formal temporary hardship contract variation and were meeting the relevant terms, and those who had promised to make a payment to a credit provider, though not agreed to a formal temporary hardship contract variation with the provider, would all be represented in the same manner on the basis of their repayment history (sub. 135).

for CCR had been achieved (Australian Retail Credit Association, sub. 87). The Commission considers that it is too early to be certain that the scheme in voluntary form will fail. Mandating participation may align incentives to meet the public interest and efficiency in capital markets better than voluntary participation but the jury is still out on how far the major financial institutions are prepared to risk corporate reputation by seeing the current scheme fail.

It would be highly advisable for the finance industry and its regulators to meet and determine a collective position on the RHI issue noted above. As this is a draft Report, the industry has time to show its willingness to solve the matter prior to our final Report.

The Commission considers that a further year is needed to see if indeed major institutions are willing to see a voluntary scheme fail. The Commission has previously recommended a review of the degree of participation in 2017, to determine whether participation should be mandated (PC 2015). If by the end of that year, there is no progress towards a critical mass of 'open' participation by major parties, draft legislation should be circulated early in 2018 to impose mandatory participation.

To judge success or failure in 2017 may be assisted by establishing a target for critical mass. There has been much consideration of the potential 'tipping point' at which those financial institutions who had hitherto not participated in CCR find it in their interest to participate. The Customer Owned Banking Association suggested that international experience shows that the tipping point is reached when there is more than 40% participation in CCR (sub. 132). Similarly, Veda submitted that stability of credit scores does not require a majority of data, and that modelling showed that critical accuracy and stability could be achieved with 40% participation (sub. 163). The Commission recommends the Government adopt such a target in developing its strategy for wide participation in CCR.

DRAFT RECOMMENDATION 4.1

The Australian Government should adopt a minimum target for voluntary participation in Comprehensive Credit Reporting of 40% of accounts. If this target is not achieved by 30 June 2017, the Government should circulate draft legislation to impose mandatory reporting by 31 December 2017.

Sharing of vehicle data

Reservations about the sharing of vehicle information with vehicle repairers have been expressed for at least the past six years. A review by the Commonwealth Consumer Affairs Advisory Council (CCAAC) in 2011 found limited and conflicting evidence that access to data and information was a barrier for independent repairers. But with uncertainty around whether the cost and timeliness of data access were issues, the Council recommended expedition of an outcome such as a voluntary industry code of conduct, and that the

Australian Government canvass regulatory options if the industry was unable to arrive at an effective outcome of its own (CCAAC 2012).

A voluntary agreement for the sharing of service and repair information — between the Australian Automotive Aftermarket Association, the Australian Automotive Dealer Association, the Australian Automobile Association, the Federal Chamber of Automotive Industries (FCAI), and the Motor Trades Association of Australia — has been in place since December 2014. The principles of the agreement specify that consumers are entitled to full information regarding the maintenance and repair of their vehicle, and that repairers should be able to access all information required for diagnosis, body repair, servicing, inspection, periodic monitoring, and reinitialising of a vehicle (FCAI et al. 2014).

The Australian Automobile Association submitted to this Inquiry its concern that (despite the voluntary agreement) manufacturers could restrict access to, and control of, data produced by vehicles to advance their own interests:

> Allowing only one service provider (i.e. the vehicle manufacturer) to access vehicle data about an accident, service schedules, breakdown or other safety features of a vehicle, limits consumer choice, competition, and may create monopolies in the automated repair and service market. (sub. 157, p. 17)

Further, according to the Australian Automotive Aftermarket Association, as at October 2015, only nine of the 68 car brands available in Australia had made information available under the agreement. The availability of repair and service information and data has been raised as an area of interest by the ACCC in its study into the new car retailing industry (ACCC 2016).

As with retailing, it is plausible to consider a broader consumer right to access their data as having competition benefits this industry, too.

Commercial competition as an impediment to data release

The Australian Computing Society submitted that voluntary data sharing is only likely to occur in instances where 'there are clear business benefits in the market and/or appropriate incentives are provided' (sub. 134, p. 24). While there are incentives for commercial entities to share data with each other under some circumstances, data on some markets may be undersupplied, as the Financial System Inquiry found:

> In many circumstances, private sector organisations have strong incentives to restrict access to the data they hold, as it serves as a competitive advantage. However, this may create inefficiencies where the benefits to the economy of releasing data are greater than the benefits to individual institutions of restricting access. (Murray et al. 2014, p. 185)

Some of the data collected by the private sector may have the potential to deliver significant public benefits if shared more widely than when guided by the incentives of the private organisations. For example, the Centre for International Finance and Regulation commented:

> Intervention might be appropriate if the data-gathering of private firms … has some clear public interest purpose. Airliners carry flight recorders, by law, to better determine the cause of accidents. There may come a time when similar requirements are expanded from trucking fleets to private motor vehicles. (sub. 9, p. 18)

Resource exploration data may be an example of where the amount of information released privately could be less than socially optimal, in the absence of regulatory requirements. The Commission's 2014 report on mineral and energy exploration recommended that exploration companies should be required under legislation to publicly disclose information about resource discoveries in Australia, on the same basis as the current requirements for those exploration companies listed on the Australian Stock Exchange (PC 2013).

Commercial entities subject to a significant degree of regulation, such as banking and insurance as discussed above, may in some instances face weaker incentives to meet the interests of their customers than entities operating in other markets. This may be caused by high barriers to entry and the presence of relatively few competitors. Some indication of the weaker incentives to meet consumer interests may be gauged by complaints received by industry ombudsmen. For example, the Financial Ombudsman Service (FOS) reported over 34 000 disputes in 2015-16 between consumers and financial service providers, an increase of 7% compared to the previous year (FOS 2016). The Telecommunications Industry Ombudsman (TIO) received more than 31 000 new complaints in the March quarter of 2016 alone (TIO 2016).

Advancements in technology have facilitated the ability to collect and disseminate data more easily. In the future, consumers may have a greater interest in viewing data relating to their consumption and use this to inform their decisions. Data could be used by consumers in order to allow them to obtain a better deal, and thereby enhance the competitiveness of markets. Overseas, policy initiatives such as midata in the United Kingdom (discussed below) have already sought to improve the competitiveness of markets in some industries and their ability to respond to consumers' interests by increasing consumer access to data.

The potential for increased levels and new sources of commercial competition within an industry is a powerful deterrent against incumbent businesses, including businesses that obtain their data due to regulated advantage or under public funding, openly sharing data collected in the course of their business activities (Manyika et al. 2013). Often, the initial firm to provide a new product or service enjoys a first mover advantage and retaining data may be seen by the firm as necessary to retain this advantage. For some first mover businesses, there may conversely be benefits to data release — if, for example, release enabled a business to better draw on external research or development expertise.[17]

[17] The first mover advantage may be thought of by some firms as a first mover disadvantage (Boulding and Christen 2001), if data collected by first movers is required to be released and gives potential new business entrants information on consumer demand or preferences. For some businesses though, there

Some commercial entities see data's commercial-in-confidence nature as being an effective rationale for non-release of all data. The Australian Private Hospitals Association points to the potential impact of data release on the insurance costs of private hospitals:

> Any data that gives one hospital competitive advantage over another, or data that would place one hospital in a less beneficial position when negotiating contracts with health insurers, would be considered commercially-in-confidence by the private hospital sector. This would include data in respect to activity, costs, volume, price and quality. This may be data that is publishable in the public sector (such as the number of separations a hospital provides annually) which in the private sector should only be published as a sector-wide aggregate. (sub. 183, p. 9)

An alternative view might be that, while such data undoubtedly is a source of commercial advantage, withholding it at a net cost to the public interest (say, community health) would be unjustifiable.

The commercial-in-confidence classification is also invoked by public sector agencies as a factor preventing the release or sharing of information obtained from the private sector. This has also been a perpetual and often misleading defence used by governments in the course of making commitments to infrastructure projects (PC 2014). Usually, the data applied in business cases is commonly exchanged amongst all the advisory firms working for all the bidders or contractors. Its sensitivity is much less, if any, if kept at level that is not about a specific purchase or individual contract. Much the same might be said of hospitals.

More generally, industry-wide data sharing arrangements can see a movement of demand towards smaller firms with less marketing and reputational power (Murray et al. 2014, p. 188), and may conversely act as a further disincentive for larger firms to share their data. In sharing its data, a larger firm would be contributing a proportionately larger amount of data to the pool, but accessing a proportionately smaller benefit in doing so. Dun & Bradstreet (sub. 135) and ANZ (sub. 64) testify to data's role as a business asset and note the disincentive of a claimed 'free rider' risk, whereby small participants obtain large amounts of market data without contributing an equal amount.

The extent to which this is a problem in practice is questionable however — not least because it is far more likely to be large consulting firms and market analysts (rather than small market participants) who analyse and report on such data. And improving the quality of such analysis would appear to be to the benefit of all market participants and governments.

In practice, there are examples of commercial entities voluntarily entering into partnerships to share data, primarily in large markets. One example is the relationship between technology company Palantir and chocolate manufacturer The Hershey Company to form an industry-wide data exchange hub. Palantir and Hershey's have reportedly used customer

may be potential benefits from early data release — if, for example, release enabled a business to better draw on external research or development expertise.

transaction and store data to derive insights on issues such as how sales respond to the placement of confectionary in different locations within stores (Rao 2015).

Similarly, aircraft suppliers and contractors share information on safety and engineering risks — the public interest (and confidence in aviation) demands this.

Further analysis of the interaction between the public interest and data sharing is likely to be conducted for the final Report.

Where industry-wide sharing of data would benefit consumers

There can be benefits to the general public, in terms of greater consumer choice, from an industry-wide sharing of market information. The Financial System Inquiry report (2014) gives the hypothetical example of a reciprocal, industry-wide data-sharing agreement for food truck businesses, whereby a trader would be able to access the aggregated market information and location data of competitors. Consumer choice would be maximised by the trader's consequent ability to target under-served locations, refine promotional offers, and adjust food and beverage offerings to match consumer demand (Murray et al. 2014, p. 183).

Other examples of industry-wide data sharing arrangements that could benefit consumers, but may come up against the barrier of businesses' inclination to maintain a competitive advantage, include:

- superannuation funds and self-managed super fund advisory firms could pool aggregate data to create a clearer picture of average superannuation balances, instead of these figures relying on reports to the Australian Prudential Regulation Authority (APRA) or the Association of Super Funds of Australia (ASFA). Analysis of averages by gender or geographic area could be extremely useful for policy development
- retail banks could share aggregate credit card data to create a more accurate picture of per capita debt and credit card usage. Access to these figures could help individuals make decisions about their own finances as well as contributing to the Australian Bureau of Statistic's studies of household debts in their Australian Social Trend reports.

In sum, there are situations where the private sector lacks incentives to disclose information that is necessary for markets to function properly, or where there are significant positive public benefits from this disclosure. In such cases there may be a role for government action to ensure broader access to data in the public interest. The benefits and costs of any such action would need to be rigorously assessed beforehand.

Consumer access to data

There has been a movement internationally towards increasing consumers' access to the information that businesses hold about them and their transactions.

In the United Kingdom, the 'midata' program was initiated by the UK Government with the objective of giving consumers access to their data in a portable, machine-readable

electronic format (box 4.6). In the United States, 'Smart Disclosure' aims to expand access to data in machine-readable formats to help consumers make better choices for services such as health care, energy, and communications. In the US health sector, the Health Information Exchange gives doctors, nurses, pharmacists, and other health care providers to access and share a patient's medical information via electronic means (DHHS (US) 2014). In France, the MesInfos pilot project was launched in 2012, in which six large companies (including a bank, an insurance company, and a telecommunications company) participated with 300 testers to enable the latter to make improved consumer decisions by sharing customer data.

Other personal information management systems are in the process of being developed. The Hub-of-All-Things was created by researchers at six universities in the United Kingdom. Individuals using the system can acquire their data from Internet connected devices, which is then transformed by the individual's personal data platform, enabling the contextualisation of data and use for decision-making. In the social media space, digi.me is a tool that allows people to download and unite their social networks and transfer their social media information from one social media platform to another.

In Australia, there is no equivalent industry-wide platform or economy-wide approach to consumer access to data. Despite this, some larger businesses have seen commercial advantage in improving the (selected) access of customers to data about them. ANZ, for example, stated:

> ANZ makes financial data available to customers to assist them with managing their finances. For example, customers can download transaction data in common formats to their computers. Business customers are able to register so that automatic, direct bank feeds of transaction data are sent to customers' compatible accounting software packages. ANZ has set up direct bank feeds with a number of accounting software providers to make reconciling business accounts easier … (sub. 64, p. 6)

The movement to increase consumer access to data has occurred in a context in which control of data has traditionally centred on the commercial entities that have gone to the effort of collecting it. Efforts to boost consumer access to data may partly be seen as an effort to loosen that control. Further, while data advocates may refer to consumers accessing 'their' data, Australian law does not actually recognise 'ownership' of data, by either a consumer or business (chapter 1).

Box 4.6 Operation of midata in the United Kingdom

midata was launched in the United Kingdom in November 2011. The program focuses on consumer data in the banking, energy, and mobile phone sectors. These are largely sectors where consumers have long term and frequent interactions with service suppliers, and where it is difficult to compare the cost of services. Under the midata program, consumers who have access to data in a portable format are able to use that data to participate in third party price comparison sites.

In a 2014 review, the UK's Department for Business, Innovation and Skills found that most current account providers allowed customers to download PDF statements and data for a period of several months to three years. PDFs have not traditionally been machine readable and are therefore limited in their ability to allow consumers to use data. The Department noted that there was no consistency between providers with respect to data fields and the length of time for which data is archived.

In the energy sector, the six largest energy companies in the United Kingdom all provided midata downloads within a consistent set of fields. However, comparison websites did not provide facilities to upload midata energy files. The next phase of midata implementation for energy would allow consumers to be able to give a third party, such as a price comparison website, the ability to download their midata in a 30-minute window. This phase is expected to be implemented in late 2016.

Although all of the major telecommunications companies in the United Kingdom provided customers with online accounts and the ability to download PDF bills, most did not provide the option to download customer data in a machine-readable format. Nevertheless, consumer engagement with the industry was regarded as high by consumer groups and comparison sites, facilitated by defined contract ends and consumer desires for handset deals.

A benefit of the midata model is that it relies on a carrot and stick approach — the emphasis is on voluntary engagement of business, but with the threat of mandatory action. It also recognises the importance of industry standards in achieving data portability. And it is confined to a small number of high value sectors — energy, banking, mobile phone providers. But the midata example also shows the necessity of a high level of political will and stakeholder engagement in achieving meaningful reform.

A recent investigation into retail banking by the UK Competition and Markets Authority concluded that while midata is a positive development, it is not straightforward to use, its current application is not fully effective, and usage remains low. Problems encountered using midata included that midata files did not take into consideration reward and service quality dimensions, limiting the ability of consumers to fully compare all aspects of a service, and that the customer experience was not a seamless one, given that customers are required to download a CSV file and upload that file to a third-party price comparison website (in addition to the fact that limited compatibility with some tablet and mobile devices has also been a problem for some users).

Further issues encountered using midata were difficulties relating to security risks, in addition to the fact that in practice, some transaction data downloaded is redacted to address confidentiality concerns. In turn, this prevents price comparison website systems from identifying payments which would qualify for rewards from some providers of current accounts.

In sum, while the UK Competition and Markets Authority considered midata to be a positive development, they concluded that there were many limitations associated with its current implementation. Consequently, it has not been used by many customers or third parties, and had limited effectiveness enhancing price transparency.

Sources: CMA (2016a, 2016b); DBIS (UK) (2014).

Broad schemes for consumer access to data, such as midata, are promoted on the basis that they will enable consumers to better understand their own consumption behaviours and patterns, helping them to become more informed. Businesses have the opportunity to improve their relationships with their customers, and develop trust-based relationships with them. Greater consumer knowledge may result in greater competition, as consumers 'shop around' to find businesses with the goods and services that best meet their needs and preferences and offer a mix of price and quality (DBIS (UK) 2011). As stated by the Financial System Inquiry in reference to the private sector generally (that is, more broadly defined than consumer access to data alone):

> Increasing access to private sector data would have large efficiency benefits across the economy. It would support innovation and competition, as new entrants and smaller businesses with smaller datasets could better compete with larger incumbents. Online comparator sites and other advice services could better serve their clients, and consumers could make more informed choices. (Murray et al. 2014, p. 188)

There have been some efforts in Australia towards increasing consumers' access to data, specifically in relation to energy data (box 4.7).

Box 4.7 Consumer access to energy data in Australia

In 2012, the Australian Energy Market Commission released a report setting out its recommendations for market conditions to facilitate efficient demand side participation — that is, greater and more active participation by consumers in the electricity market. Among the Commission's recommendations were that a framework be developed for smart meters, and that competition be introduced for metering services. Further, the Commission recommended that the National Electricity Rules be amended to provide a framework for consumers to request and receive energy metering data from their retailer (AEMC 2012).

By late 2014, there had been some progress in enabling greater access by consumers to energy data. In its final rule determination, the Australian Energy Market Commission changed the National Electricity Rules to make it easier for customers to obtain their electricity consumption data from their distributor in addition to their retailer, and for customers to authorise third parties to obtain their electricity data from retailers and distributors. The new rules also required retailers and distributors to adhere to minimum standards regarding the format, time frame and cost by which usage data are delivered to customers (or parties authorised by that customer).

As a consequence, customers and parties authorised by them have been able to access electricity consumption data from 1 December 2014, although the minimum requirements relating to the provision of metering data did not come into force until 1 March 2016 (AEMC 2014). In September 2015, the Australian Energy Market Commission produced a final report and determination in relation to metering data provision procedures, outlining in detail the requirements of retailers and distributors (AEMC 2015).

A number of energy companies also provide online portals that enable consumers to manage their details and pay their electricity bills, as well as monitor their usage. For instance, AGL offers AGL Energy Online, and Origin Energy provides a portal called 'My Account'. In addition to these tools, consumers can view usage data on their billing statements, and some energy companies have phone services that may provide further account information.

Besides the Murray report, a number of reviews and organisations have expressed support for the principle of giving consumers access to greater data. For example, the ACCC submitted to the Competition Policy Review (the 'Harper Review'):

> The ACCC recognises the potential for consumer access to data to improve competition and consumer outcomes. The ACCC encourages further consideration of initiatives in this area. (ACCC 2014a, p. 27)

The ACCC stated that the ability for individuals to request access to their personal information from an entity covered by the Privacy Act under Australian Privacy Principle 12 provides a foundation for consumers to access their transaction and consumption data to improve competition (ACCC 2014a). However, as acknowledged by the ACCC (2014a) and the Financial System Inquiry (Murray et al. 2014), the Privacy Act does not provide guidance on the format in which personal information is to be provided to consumers.

In addition, the Financial System Inquiry (Murray et al. 2014) also noted that in most cases, consumers are unable to authorise third parties to access their personal information directly from a service provider, limiting the ability for consumers to make more informed consumption choices.

The presence of significant regulation in areas such as banking, insurance, and utilities is a prima facie indication of the presence of a public interest that requires safeguarding. For instance, prudential regulations aim to protect the public by dealing with potential threats to financial stability, mitigating the risk of highly adverse events — such as the collapse of financial institutions — occurring.

More concretely, any organisation wishing to operate as an authorised deposit-taking institution (ADI) in Australia is required to be authorised to do by APRA under the Banking Act. ADIs are subject to supervision by APRA, which includes adhering to requirements contained in prudential standards, as well as providing data to APRA in compliance with reporting standards (APRA nd). Hence, the collection of data from regulated commercial entities plays some role in ensuring system stability.

The concept of a 'regulatory contract' (discussed above) may raise the question as to whether highly regulated entities are meeting the obligations of this 'contract' in a data-rich world. That is, in today's regulatory and technological environment, are today's regulated entities meeting the public interest obligations that lie behind their licensing?

Further consideration applied to regulated commercial entities would therefore consider whether ensuring better data flows and consumer empowerment would better serve the public purposes under which regulated businesses operate.

Other entities that receive less or no such regulatory attention, such as technology businesses, might argue that no such regulatory contract exists between them and the public.

The Commission's preferred approach to improve consumer access to data and data portability is discussed in detail in chapter 9 of this Report. The Commission considers that data access could provide benefits to consumers by allowing them to be more informed about their purchasing decisions. It may also help to counter what may be unequal exchanges between well-informed commercial entities with significant data capabilities on the one hand, and individual consumers on the other. Individuals generally have less technical capabilities and are less well-informed regarding the precise uses to which data may be put than large commercial entities such as banks and utilities. The latter have sophisticated internal networks and capabilities to analyse data, as well as possessing legal and contract experience.

Further, giving consumers greater access to data may increase their confidence in data collection by commercial entities, thereby enabling subsequent new uses and data innovations. For these reasons, the question of whether regulated entities owe greater obligations to their customers, is for the purposes of this draft Report, unresolved.

4.4 Consumers beware

Consent, and community perspectives on providing data to private firms

As illustrated above, the evolution of technology has enabled an increase in the collection, sharing, and use of data by the private sector. There has been an increase in voluntary data sharing mechanisms, most notably social media, as well as the creation of apps, the use of cookies to record Internet browsing, and partnerships between large commercial entities and organisations that specialise in data analytics. Consumers are often granted access to certain benefits or services that are 'free' of charge in notional terms but come with a requirement to provide data to the firm providing the benefit or service. Therefore, consumers 'pay' in the form of providing data, rather than (or in addition to) paying in monetary terms.

Given these developments, questions that arise include the extent to which consumers are aware that they are sharing personal information and the nature of that information, and are comfortable with its various uses by a growing number of businesses. In general it appears that people are aware that their data is being collected, but they are less aware of just how extensive that can be. A study examining transparency around customer data looking at the United States, United Kingdom, Germany, China and India, found that only 25% of people understood that there data footprint included information on their location, and just 14% realised that they were sharing their web surfing history (Morey, Forbath and Schoop 2015). Entities covered by the Privacy Act are required to have a privacy policy that sets out how the entity collects and manages personal information. Genuine consent requires that people are able to understand their options and make a meaningful choice.

One study examining whether people read privacy policies found that 95% of participants agreed to a 'gotcha' clause in the terms and conditions they were given that signed away rights to their first born child (Obar and Oeldorf-Hirsch 2016). Earlier empirical research also supports the notion that most people do not ready privacy policies on a regular basis (Solove 2013). Some researchers have also questioned the extent to which information is understood by consumers, and, even if understood, the extent to which the information is subsequently acted upon (Ben-Shahar and Schneider 2011).

Automatically ticking a box is not informed consent. But genuine consent requires that people are informed, are able to understand what they are agreeing to, and be given a meaningful choice. Genuine consent is an important component of building trust — as recognised by Shaw (2014, p. 2):

> The change in how organisations use our personal data is happening whether we like it or not and we risk destroying trust if consumers are harmed or even surprised, by how their personal data is used. We need consumers to trust how their data is used or they will be slower to engage by sharing their data. This will delay the benefits of a Big Data society and leave the UK to be potentially overtaken by other countries with a different view of the importance of consumer trust.

A further reason why consent may not always be genuinely given is due to the 'take it or leave it' approach embodied in the terms and conditions of data use. Individual consumers may lack the bargaining power and organisation necessary to result in terms of data access that are the result of negotiation between commercial entities and consumers (Braucher 2006).

In a survey undertaken for the Australian Communications and Media Authority (ACMA) in 2013, it was found that just over half of survey respondents said that they would never give inaccurate information about themselves online. Indeed, one-third of survey respondents replied that they would not use a service at all, rather than provide inaccurate information. Of those who had given inaccurate information online, the most common motivation for doing so was relevance, followed by concerns about the trustworthiness of a site. About three-quarters of survey respondents indicated that they would stop using a site if it mishandled their health information, email address, phone number, credit card details, or photograph (ACMA 2013).

In a 2013 OAIC survey on community attitudes to privacy, when asked what they believed the biggest privacy risks facing people were today, roughly half of all respondents nominated online services and social media sites. Nearly all respondents considered that activities such as an organisation revealing a customer's information to other customers, supplying information to an organisation for a specific purpose only for it to be used for another purpose, and organisations asking for personal information that does not appear relevant to the transaction ought to be considered misuses of information (OAIC 2013).

Particularly relevant given the global reach of many entities that collect data is the matter of data being sent offshore. In the OAIC's survey, only 10% of respondents replied that they were not concerned about their personal information being sent overseas, while 62%

replied that they were 'very concerned' with the prospect of their personal information being sent overseas (OAIC 2013).

When asked how willing they would be to give personal information in exchange for a benefit, the majority of respondents indicated that they were not willing to exchange personal information for benefits in the form of a discount, prize or improved service. However, a sizeable minority indicated that they would — 28% indicated they would give personal information in exchange for a discount, 14% for a prize, and roughly one-third would give information for better service (OAIC 2013).

Despite stated concerns about privacy and the use of personal data, there is a comparatively high use in Australia of product offerings such as loyalty programs and the adoption of new technologies like wearables (highlighted earlier in this chapter), raising what Acquisti, Taylor and Wagman (2016) refer to as the 'privacy paradox'. One explanation for this apparent paradox is the existence of decision-making biases in consumer behaviour — for example, if individuals place a high value on the short-term benefit they receive from making certain data available in exchange for a service, while discounting the future reduction in privacy or not being able to assess this reduction (as a result of not reading privacy policies or being unable to determine how data is used). Other reasons for the apparent paradox include that knowledge of possible solutions, such as privacy-enhancing software, is limited, and that people mentally assess privacy trade-offs and make different decisions in different situations (Acquisti, Taylor and Wagman 2016).

While there may be several reasons for individuals expressing concern about privacy on the one hand, and trading elements of privacy in exchange for goods and services on the other, there is scope to improve outcomes for consumers. Doing so would be desirable, particularly as it seems neither businesses nor consumers want to erode the benefits delivered by the creation and use of date. Chapter 8 further discusses issues around consumer trust in data collection and use.

Privacy requirements and private sector data access and use

Commercial entities will in most cases have a clear financial incentive to ensure that the privacy of their customers is protected and that data protections are adequate, particularly as security of data and privacy rate high on consumers list of concerns. Even for the most well intentioned businesses, however, human error and malicious efforts mean risks cannot entirely be eliminated and considerable diligence is required to keep protections up to date. Consumers and those who represent them need to keep in touch — and be enabled to keep in touch — with how data security is being managed on their behalf. For those firms that have been involved in prominent data breaches, sustained damage to their reputation and lost trust of consumers is a significant risk (breaches are further discussed in chapter 5). Conversely and not surprisingly, studies highlight the benefits of transparency around data collection and protection in building and retaining consumer trust (Morey, Forbath and Schoop 2015).

The Australian Privacy Principles (APP) create the current framework within which entities can disclose data to third parties and manage personal information. The implications of these for data collected on individuals are discussed further in chapter 5. Data Republic indicates that their use is a matter of significant judgment — noting that the APPs require interpretation by specialists, often resulting in review before, during, and after solutions have been implemented to ensure that processes adopted meet expectations established by the APPs (sub. 176).

That the regulatory framework for protecting privacy affects the ability of businesses to carry out certain operations using data is desirable, but its objective must be clear, defensible and implementable at a cost that is the minimum necessary to achieve the regulatory objective.

Private sector businesses that extract and exchange data

An increasingly significant part of the data landscape is businesses that collect, collate, sell and organise data to other organisations for profit.[18] Data brokers, web scrapers and data facilitators and intermediaries are the most prominent categories of businesses in this field.

In the context of data, there are two key reasons underpinning the emergence of specialist data businesses. The first is network externalities — the insights drawn from a particular dataset could be more valuable if combined with other datasets. While concerns about competitors and other parties having access to what is seen as proprietary data can impede data exchange, Westpac (sub. 197) noted that bilateral data sharing is occurring based on commercially negotiated prices. The second reason underpinning the emergence of specialist data businesses relates to specialisation. Analysing data to derive insights is not straightforward, and many businesses would make commercial decisions to not invest in such a capability. This creates a niche for firms who can add value to data and derive valuable insights, even where such data is publicly available.

As discussed below, some concerns have been raised about the practices of data brokers and web scrapers overseas. Such concerns are not yet widely held in Australia — the sector is new and emerging and has not undergone the degree of scrutiny that has been applied to equivalent activities overseas.

Data brokers

Data brokers, also referred to as ‘information brokers’, ‘data vendors’, and ‘information resellers’, collect information to create detailed profiles of individuals, which is then sold

[18] There are also not-for-profit organisations that collect, collate organise, and sell data, such as the UK-based organisation Community Insight, which sells subscriptions to a Geographic Information System tool based on open data that provides community mapping and reporting for housing providers (Community Insight nd).

to other entities. They may also provide other services, such as data management and consulting.

There are several data brokers with operations in Australia, often operating as subsidiaries of organisations that are headquartered overseas. One such business is Acxiom, which offers services such as consulting, analytics, and database services. The business offers a number of data packages among its services, including a 'Christmas Spenders' database, which Acxiom markets as helping clients find consumers with a high propensity to spend in the lead up to Christmas Day and Boxing Day. Similarly, Acxiom offers another data package for 'Easter Segments', and advertises this product as helping organisations maximise the effectiveness of their Facebook advertising by targeting families and consumers who spend the most over the Easter long weekend (Acxiom 2016).

Similarly, Experian promotes itself as a marketing services company, and offers products that span credit risk and fraud management, decision analytics, and customer-centric marketing. It is another data business that has operations in Australia, and has worked with clients such as Fire and Rescue New South Wales and the State Library of Victoria. In one example of its business activities, Experian was requested by an established, US based luxury retailer to maximise returns from email and social marketing and reengage with customers that the retailer had lost contact with. Experian used Facebook's Custom Audiences product (appendix F), which provides for customer emails, mobile numbers, and Facebook IDs to be used to identify customers that a business would like to target. The retailer's email database was uploaded and checked against Facebook's customer database, matching customers against Facebook profiles. No personally identifiable information was disclosed during the process and the tool did not identify contact details. This approach allowed the retailer to target advertising to those customers who receive emails, but do not make a follow-up visit to the website, as well as those customers who no longer received the company's emails because they may have switched to another email account (Experian nd).

There has been some concern in technological media publications about the activities of data brokers in particular. For example, trepidation has been expressed about data brokers in the United States knowing, as well as buying and selling detailed (and in some cases, sensitive) personal information (Beckett (2013) and Maus (2015)). In the United States, the data broker industry was the subject of a US Senate Committee report and a report by the Office of the Privacy Commissioner of Canada (box 4.8).

Web scrapers

Web scraping, also referred to as screen scraping and web harvesting, is a computer software technique used to extract data from websites. The objective of web scraping is to gather data and compile it into a more useful format for the user's purposes. Web scraping software interacts with websites in the same way as a browser does, and saves data required from a webpage into a local file or database (SysNucleus nd). This happens despite many websites containing copyright and licensing provisions that do not permit use of data for commercial purposes.

Box 4.8 North American reviews of the data broker industry

In December 2013, the US Senate Committee on Commerce, Science and Transportation published a review of the data broker industry in the United States. The Committee argued that advances in technology had fuelled the growth of an industry that operated in a manner largely hidden from the view of consumers. It was found that data brokers' customer base encompassed virtually all major industry sectors of the economy, and collected and sold information for purposes such as credit risk assessment, fraud prevention, and marketing.

The Committee's major findings were that:

- Data brokers collected a large volume of detailed information on hundreds of millions of consumers — this included personal characteristics, preferences, and financial and health information.
- Data brokers sold products that identified financially vulnerable consumers — some companies were found to compile and sell consumer profiles without the permission of those consumers, with some products focusing on the financial vulnerability of those consumers, carrying titles such as 'rural and barely making it' and 'ethnic second city strugglers'.
- Data broker products provided information about consumer offline behaviour to tailor online outreach by marketers — increasingly, the information data brokers sold marketers was provided digitally, based on dossiers of offline data collected about consumers.
- Data brokers operate behind a veil of secrecy — data brokers were found to typically amass data without direct interaction with consumers. Three of the largest data brokers in the United States — Acxiom, Experian and Epsilon — were, the Committee argued, secretive with respect to their practices, refusing to identify the specific sources of their data or the customers who purchased it.

The Office of the Privacy Commissioner of Canada observed that there was a lack of comprehensive oversight of the data broker industry in the United States, partly due to the fragmented nature of US privacy laws and regulations. By contrast, in Canada, a single act applies to all organisations that collect, use, and disclose personal information in the course of commercial activity (except in provinces with substantially similar legislation). While data brokers operated in Canada, the Office noted that data brokers in Canada tended to use fewer information sets than their counterparts in the United States, and that the scale and scope of information available in Canada was less than in the United States.

Sources: OPCC (2014); Senate Committee On Commerce, Science and Transportation (US) (2013).

There are a number of businesses across the world that offer web scraping services to clients. In Australia, for example, The Data Scraping Group provides web scraping services to businesses and lists its website scraping and harvesting abilities as including:

- monitoring competitors' prices, locations and service offerings
- harvesting directory and list data from the Internet and transferring it to Excel or CSV
- social media scraping for customer and trend analysis
- regular (and even real-time) updates of exchange rates, insurance rates, interest rates, real estate and stock prices
- updating a business' website with extracted details and product prices from suppliers
- monitoring webpages and providing alerts for changes (The Data Scraping Group nd).

Web scraping typically endeavours to emulate human browsing, so that the scraping activity blends in with 'legitimate' website traffic. Generally, content that is presented on a website can be accessed by web scrapers (Brody 2013). Preventing web scraping requires conscious effort by website owners, and there are a number of anti-scraping techniques — such as employing captchas and blocking or redirecting requests from particular IP addresses. Even so, many of these techniques can still be circumvented by determined scrapers (Brody 2013; Daityari 2014).

'Scraped' data might be put to use for such purposes as optimising search engines, monitoring business social networking pages, and for the purposes of providing price comparisons across products offered by different suppliers. As an example of the latter, the ACCC noted that web scraping was one of the techniques used by comparator website operators to extract pricing information from websites, often without the approval or knowledge of the business that owned the website (ACCC 2014b).

It appears that web scraping can be both supportive of consumer interests and yet possibly also exploitative. Considered from the point of view of this Inquiry and its interest in access to data, the case for public policy action (and its likelihood of effectiveness) seems low — but we would welcome more information.

Government purchases and contracting of private sector data

In certain situations, public sector organisations use data generated by, or held by the private sector. One example is the use of property data by the ABS. To calculate its Residential Property Price Indexes for Australia's eight capital cities, the ABS (2016) uses data supplied by the private firm CoreLogic RP Data (a wholly-owned subsidiary of the property information group CoreLogic). The Reserve Bank of Australia also uses data from CoreLogic RP Data (CoreLogic RP Data 2014). In the United States, Uber has recently agreed to sharing ride pattern data with Boston officials to enable better transport planning and the prioritisation of road maintenance (Morey, Forbath and Schoop 2015). The use of such data by governments, either by itself or in combination with public sector data sources, is increasingly seen as having great potential in providing richer statistical information and potentially improving economic and societal outcomes.

Governments also require some private sector bodies to collect certain types of data, many of which are industry-specific. This may occur as a condition of regulation or of receiving public funding. Governments may also enter into contracts with the private sector for the provision of goods and services, with data collected by firms in the course of fulfilling these contracts. Not all contracts governments have currently entered into with the private sector would necessarily address the issue of access and use of data in a manner that provides net benefits to the public.

In Victoria, the contract for the myki ticketing system for public transport requires the operator to prepare a monthly performance report across 27 indicators. The information is used by Public Transport Victoria (PTV) to assess the performance of the contractor. The

Victorian Auditor-General reviewed the operational effectiveness of myki, and found that PTV's verification of the contractor's monthly reports was limited and did not provide sufficient assurance that the source data used by the contractor to compile reports was accurate. The Auditor-General noted that under the myki contract, PTV had the power to conduct an audit, which would give it access to the performance data and systems maintained by the contractor, although PTV had not done so. The Auditor-General also found that on most occasions, PTV was unable to verify the accuracy of 18 performance measures, nine of which directly related to incentive payments (VAGO 2015a).

The Opal card ticketing system in New South Wales operates on train, bus, ferry, and light rail services in Sydney, and surrounding regions, with management of Opal data outsourced to the private sector. In a 2015 report, the NSW Auditor-General judged that Sydney Trains and NSW Trains (both state-owned organisations) were not able to obtain required information on passenger revenue and trips to assist in operating an efficient and effective business, based on existing Opal data. The Auditor-General noted that service operators would like to have boarding and exiting data by station 24 hours a day, seven days a week, to appropriately manage demand (Audit Office of New South Wales 2015).

The ability of the public sector to access data generated by private organisations in the context of public-private contracts can have important implications for the administration of public services and other services. It is imperative for governments to ensure that the contracts to which they are party provide a degree of access rights that is consistent with the interest of the wider public, as this may have consequences for service quality and government expenditure (for example, through contractor performance payments). For some applications, it may also be beneficial for government to ensure that it can make at least some of the data accessed via public-private contracts open and freely available to the public, for the purposes of promoting transparency and accountability. Conditions could also be placed on the timeliness, format and delivery method of data to ensure usability.

The Commission considers that all governments should, in future contract negotiations, consider the potential public interest of any data collected, including the likelihood of its effective use.

The ferry system contract for the provision of ferry services in Sydney Harbour stipulates under section 10.9 (f) that:

> The Operator acknowledges that all information and data collected by the Electronic Ticketing System:
>
> (i) is the property of the Director-General [of the NSW Department of Transport] or the Director-General's Associate; and
>
> (ii) may be used by the Director-General as he sees fit in administering public transport in the State and this Contract, including in preparation for the appointment of a Successor Operator on, or in anticipation of, expiry or termination of the Term. (MinterEllison nd, p. 53)

This is one example of a contract with the private provider of a good or service (in this case, Harbour City Ferries) that explicitly builds in an ownership right for government over the data collected by the provider. It also gives the public sector rights concerning the usage of that data in the administration of the service. Access to this data may be used to improve the operation of related public services, or in the planning of infrastructure developments, thus benefitting the public.

DRAFT RECOMMENDATION 4.2

All Australian governments entering into contracts with the private sector which involve the creation of datasets in the course of delivering public services should assess the strategic significance and public interest value of the data prior to contracting. Where data is assessed to be valuable, governments should retain the right to access or purchase that data in machine readable form and apply any analysis that is within the public interest.

5 It's all about you: the challenges of using identifiable information

Key points

- Identifiable information includes any data about an individual or a business from which their identity can be directly established or easily deduced. Knowing an identity, and yet protecting it, is important to both the public and private sectors.
  - In the public sector, sophisticated use of separate personal identifiers to link data improves policy yet protects individuals. Countries such as New Zealand are leaders in this field.
  - In the private sector, identifying a user enables advertisers to tailor services to individuals. Corporate regulation nevertheless encourages firms to take care with identifiable data.
  - These trends come at a potential risk of personal exposure. The *Privacy Act 1988* (Cth) and many other legislative instruments are *intended to protect* individuals' identifiable information, but also *enable its release* where research may benefit the community.
- In the public sector, the primary impediment to more effective use of identifiable data is typically *not* the Privacy Act, but legislation specific to the field in which the data is collected; plus conservative interpretations of these obligations; and a consequent incentive for data custodians to limit access out of an abundance of caution.
  - Some of the legislation restricting access to data was formulated more than a century ago, and may no longer be fit for purpose; 506 separate restrictions in 176 pieces of legislation have been identified by the Australian Law Reform Commission.
- To ensure all legislative requirements are met, data custodians and research institutions have created complex approval processes for data access that are inefficient. In one ongoing case, researchers anticipate that they will receive the linked data they have requested eight years after their initial application.
- Linkages of identifiable information can substantially increase the usefulness of data. However, the Australian Government imposes unique restrictions on data linkages, and is particularly severe in requiring the destruction of linked datasets after use, despite the often significant cost involved in establishing these linkages.
- The perceived risk inherent in increasing access to data is unauthorised release of identifiable information. Across public and private sectors, examples of this are very rare. The evidence shows that the greater risk comes from poor practice by data collectors or individuals, often exploited by hackers with malicious intent.
- The incremental personal risk to privacy from increasing access to identifiable data for policy and research purposes is likely to be very small, but data managers cannot afford to be sanguine — privacy is not a trivial matter and reputational risks for institutions are increasing.
- There are numerous ways to increase access to identifiable data while minimising risks — such as using a trusted user model and implementing strong de-identification techniques.
- Legislative reform, leadership in addressing a culture of risk aversion in the public sector, and dedicated funding, are needed to support increased access to identifiable information and better management of risk.

5.1 What is identifiable information?

Identifiable information is information, or its underlying data, that can lead to the direct identification of an individual person or business. While the importance of using identifiable information is well understood (chapter 2 outlined a range of potential uses), the practical applications of such data in Australia are still very limited, and some other developed nations do much better. This chapter details the legal barriers to increasing availability and use of identifiable data, discusses the potential risks involved and the community perceptions of such risks, and presents issues that will need to be considered in designing policies for improved access and use of such data.

What is identifiable information about people?

Identifiable information about an individual person (as opposed to a business) is referred to as 'personal information'. The *Privacy Act 1988* (Cth) specifically defines personal information as:

> … information or an opinion, whether true or not, and whether recorded in a material form or not, about an identified individual, or an individual who is reasonably identifiable.

Such information can be generated in many ways, through direct or indirect collection. There are three broad categories of identifiable data on individuals (WEF 2011):

- Volunteered data — created and explicitly shared by individuals, or collected from them directly. Names, addresses and contact details are common examples.
- Observed data — captured by recording the actions of individuals. This includes administrative datasets that record details about individuals' interaction with government agencies (such as Medicare or Centrelink), or datasets generated by companies on their customers' behaviour and preferences (such as Internet browsing history collected by social media networks and other websites, or location data collected when using mobile phones, and through frequent flyer programs). In some cases, public and private data observations coincide, for example, where companies are required by law to collect and retain data on their clients (for example, anti-money laundering legislation and metadata retention laws).
- Inferred data — based on the analysis of other types of data. For example, retailers can observe changes in shopping patterns, and use these to draw inferences on a range of topics, potentially including people's health concerns or other personal information (in the US, retail giant Target identified customers' pregnancies from changes in their shopping habits (Hill 2012)). Further, information about one person's location or activities can be used to infer information about other people as well.

Increasingly, data custodians are using statistical techniques to de-identify data (for example, by removing personal identifiers – see box 5.2), in order to enable increased use. Depending on the sophistication of techniques applied, re-identification remains possible. This occurs, for example, when a sufficient number of de-identified datasets are linked

together, and the chance of re-identification, or the creation of identifiable data, increases (section 5.5).[19]

Very large volumes of personal information are collected daily, much of it without the knowledge of the people it is collected about (chapter 4):

> A consumer may be familiar and comfortable with data about the date, time and location of a credit card purchase being recorded as part of a transaction. They may be less familiar with technology and practices that recorded their movement through a store and associated purchasing behaviour and uses this data to target marketing material to them … citizens are concerned about the collection of this data, how it might be used and who has access to it. (ACMA 2013b, p. 12)

Along with offering vital insights for policy makers, this information is being used to generate increasing economic value (WEF 2011). A range of new businesses have emerged (particularly in the United States) that specialise in the indirect collection of personal information, which is then sold to third parties. Data brokers collect demographic information, financial and health details and records of social media activities. The information is highly detailed — in 2013, one data broker in the US was reported to hold up to 75 000 separate data elements about each individual consumer, which it aggregates and uses in a wide range of products (Senate Committee on Commerce, Science, and Transportation (US) 2013). Such practices are very likely to affect Australians, as their details are recorded on various websites. Although in Australia the *Spam Act 2003* (Cth) prohibits the use of address-harvesting software to identify individuals' email addresses, obtaining unsecured personal information may occur via web scraping (subject to certain provisions contained in the Privacy Act) (chapter 4).

The boundaries of personal information are increasingly unclear, particularly since people have different perceptions of privacy and personal information — for example, there is an ongoing debate whether IP addresses[20] constitute personal information. According to the Office of the Information Commissioner in Queensland (2012), while IP addresses in isolation are not personal information, when they are linked to information that can allow an individual to be reasonably identifiable, they can become personal information. The Federal Court is currently considering these issues as part of the Grubb v Telstra case, which raises the question of whether mobile phone location is part of personal information and should be accessible to individuals (Telstra, sub. 88). Another aspect of personal information that may require further consideration is the handling of information about deceased persons, which is currently not covered by the Privacy Act.

---

19 For example, if one dataset only contains a person's age and gender (say, male aged over 95), and another only contains the town they live in (a small rural town), combining them together could lead to re-identification based on unique characteristics that only a fairly small number of people possess (Menzies Foundation 2013).

20 An IP address is a code that identifies each device accessing the internet, and is routinely recorded by web site operators and internet service providers (OIC (Qld) 2012).

Elements of the legal definition of personal information give rise to ambiguity. The Australian Law Reform Commission (ALRC) has concluded that 'while much information will fall clearly inside or outside the definition, there will be a need for ongoing practical guidance in relation to areas of uncertainty' (ALRC 2008, p. 309).

This issue will become increasingly important as public and private sector entities collect more and more information about people, while grappling with community expectations of privacy and control over their data. Expanding consumers' access to data collected about them, similar to changes adopted in the UK (chapter 4), raises further questions about the boundaries of personal information. The current approach of the Privacy Act may not suffice in dealing with new challenges raised by big data collection and advanced analytics.

DRAFT FINDING 5.1

The boundaries of personal information are constantly shifting, in response to technological advances and community expectations. The legal definition of personal information, contained in the *Privacy Act 1988* (Cth), gives rise to uncertainty. This uncertainty will only increase in future, as new technology continues to emerge.

What is identifiable information about businesses?

While there are clear legal and community expectations that identifiable information about people is protected, the approach to identifiable business information differs, depending on the type of information.

In their dealings with regulators, all businesses supply an array of identifiable information, and some of this information is publicly available. For example, the Australian Securities and Investment Commission (ASIC) allows individuals to search a wide range of information about companies and their directors — basic information is freely available, and more detailed extracts can be purchased.[21] The Australian Government is currently in the process of outsourcing the management of the ASIC registry, while retaining ownership of the data (Department of Finance 2016).

Specific types of identifiable information about businesses are protected by law, for example:

- Trade secrets and commercially valuable information (defined as information having a commercial value that would be, or could reasonably be expected to be, destroyed or diminished if the information were disclosed) are exempt from the right of public access to information created under the *Freedom of Information Act 1982* (Cth).

[21] The specific types of information and access are defined in the *Corporations Act 2001* (Cth).

- The *Taxation Administration Act 1953* (Cth) imposes a two-year imprisonment term on taxation officers who disclose protected information, which was obtained under the purposes of taxation law, and can be reasonably used to identify a specific entity.
- Companies operating in specific industries can apply for their information to be declared confidential. For example, medical research companies can apply to the Gene Technology Regulator to declare certain information as Confidential Commercial Information. Such information is not released for public consultation or included in public registers (OGTR 2014).

Similarly to the difficulties raised by the definition of personal information, the way identifiable business information is defined in various legislative instruments can give rise to confusion. For example, past reviews have recommended clarifying the definition of commercially valuable information contained in the Freedom of Information Act (Belcher 2015). Commercial-in-confidence data is also affected by this lack of legislative clarity, as there is no legislation describing how it can be shared. This creates confusion for government agencies and may unnecessarily restrict data use (Department of Industry, Innovation and Science, sub. 69, NSW Government, sub. 80).

More complex issues may arise over access to identifiable information, where the distinction between personal and business information is not clear. For example, the Governance Institute has raised concerns that the release of personal details of company directors and office holders may expose them to increased risk of identity theft (Governance Institute of Australia 2015).

For both individuals and businesses, maintaining confidentiality is an important factor when considering increasing the availability and use of data. However, this must be balanced against the need to allow access to data for purposes that will meaningfully improve community wellbeing.

What can we do better with identifiable data?

Many submissions to this Inquiry, as well as past reviews, have illustrated the benefits that could arise from increased access and use of Australia's health data (see, for example, Telethon Kids Institute, sub. 5; AITHM, sub. 52; Joint Council of Social Service Network, sub. 170; OECD 2015b; PC 2015; SSCH 2016). The health sector exemplifies many of the issues impeding improved access to and use of identifiable data, and while progress has been made, substantial challenges remain (box 5.1).

Box 5.1 Australia's health data — an underutilised resource that could be saving lives

Due to a multitude of legal, institutional and technical reasons, Australia stands out among other developed countries as one where health information is poorly used (OECD 2015b):

> The health sector is very good at generating and storing data. It is less effective at translating this data into useful information. It is poor at linking and sharing information between health professionals, where it could be used to improve health outcomes and system efficiency. Worst of all is the health sector's ability and willingness to share data with consumers (Medibank Private, sub. 98, p. 2).

The implications of this situation are significant. At the individual level, patients are required in many cases to act as information conduits between the various health care providers they see. Inadequate information can lead to errors in treating patients (Joint Council of Social Service Network, sub. 170). At the system level, inefficient collection and sharing leads to data gaps and unnecessary expenditure, while the lengthy approval process for researchers requesting access to health data limits their ability to make potentially life-saving discoveries. For example:

- Nearly **five years** after requesting the data, researchers at the University of Melbourne received de-identified information about CT scans and cancer notifications. Their work showed there was an increased cancer risk for young people undergoing CT scans, and led to changes in medical guidelines for the use of scans. 'Had [the] study been approved sooner, and been able to proceed at an earlier date…, we would have had results sooner, with potential benefits in terms of improved guidelines for CT usage, lesser exposures and fewer cancers' (John D Mathews, sub. 36, p. 13).
- Since 2008, the Australian Research Council and other government bodies have been providing funding to the Vaccine Assessment Using Linked Data Safety Study. Among other objectives, this study examines whether there is a relationship between vaccination and admission to hospital or death. The study requires data from both the Australian and State Governments. Obtaining data from the Australian Government has taken **six and a half years**; state data has not yet been linked. According to Research Australia (sub. 117), linkage is expected to occur in late 2016, **eight years** after the project commenced.

Despite this challenging backdrop, progress in using personal health data is occurring (see appendix D for a detailed discussion on health data). For example, once fully rolled out, the Australian Government's eHealth initiative, My Health Record, may have the potential to allow healthcare providers to share information about patients relatively easily, as well as enabling patients to access their own records for the first time. More broadly, a recent Senate Inquiry has made numerous recommendations on ways to improve the use of health data (SSCH 2016).

As illustrated in chapter 2, increased access to data can have benefits across many sectors. The use of identifiable information can support better service delivery from government and improve social outcomes; and offer consumers better choices in many markets. The extent of personally identifiable information needed to achieve these potential benefits differs depending on the purposes for which the data is used. For example, financial service companies tailoring products to individual customer needs may require much more detailed information, compared with social science researchers, who can use probabilistic matching and inference techniques to minimise the need to access identifiable data.

The potential benefits of increased access and use of data have long been recognised by the private and public sectors. However, the legislative environment (section 5.2), complex

approval processes (section 5.3), existing incentive structures (section 5.4), risk averse culture and lack of capability (chapters 3 and 6) all serve to restrict access to and use of identifiable data.

Much though we might prefer to think otherwise, Australia stands out among OECD countries for its relatively poor use of identifiable data (box 5.1).

5.2 The legal environment aims for flexibility, but risk aversion results in paralysis

Extensive legislation — at the Commonwealth and State level — restricts access to identifiable data. This legislation centres on the protection of privacy and government secrecy; intellectual property and copyright legislation can also restrict access to data, but its effect on identifiable information is limited (chapter 3).

On the flip side, the authority to release information is far more limited. At the Commonwealth level, such authority is only included in the Freedom of Information Act (appendix C). There are also cases where some personal information is made publicly available (for example, through land title registries or electoral rolls), or where government entities have the authority to release information if they believe this will benefit the community.[22] Overall, however, even where legislation includes provisions that allow for data use, access is often impeded (for example, sharing of information for child protection purposes fails to occur even when it is possible, as described in chapter 3).

Commonwealth privacy legislation

The Privacy Act aims to 'promote the protection of the privacy of individuals', while at the same time ensuring that such protection 'is balanced with the interests of entities in carrying out their functions or activities'. The Office of the Australian Information Commissioner (OAIC) oversees the operations of the Act (OAIC 2016c).

The Act applies to most Australian Government agencies, private sector organisations with turnover of over $3 million, and organisations deemed to be health services.[23] It applies to activities both within and outside Australia, if the organisation has some representation or activity within Australia — therefore, the information provided by Australian citizens to many foreign entities (for example, when making purchases online from websites overseas) is not necessarily protected by the Act (see appendix C for a detailed discussion of the Act).

---

22 For example, the Australian Prudential Regulation Authority can release documents submitted to it if it believes that 'the benefit to the public from the disclosure … outweighs any detriment to commercial interests that the disclosure may cause' (Australian Prudential Regulation Authority Act 1998, s57).

23 Apart from hospitals and doctors' clinics, this also includes pharmacies, aged care and child care services.

All relevant entities covered by the Act must comply with the Australian Privacy Principles (the APPs), which set out the framework for the collection, management and disclosure of personal information. While generally the disclosure of personal information held by an entity is only permitted with the individual's consent, the Act acknowledges that in some situations, disclosure without consent may be necessary. Therefore, it permits the disclosure of personal information in situations where it is required to lessen or prevent serious threat to life, health or safety of an individual or the community, to locate missing persons and to take appropriate action in relation to unlawful activities. It also includes special provisions that allow access to identifiable information required to conduct medical research (see below).

An issue of interpretation

Information sharing is permitted under the APPs in some circumstances. Under APP 6, information sharing for research purposes can be considered a secondary use of information, which may occur if 'the individual would reasonably expect the APP entity to use or disclose their personal information for the secondary purpose'. For the purposes of sharing health information, the disclosing entity should take 'such steps as are reasonable in the circumstances to ensure that the information is de-identified' (OAIC 2014a).

These conditions put the onus on the entity releasing the information to ensure that this is done in accordance to 'reasonable expectations', and if required, 'reasonable steps' are taken to ensure de-identification (box 5.2). As a result, in the public sector, '[d]ata custodians … may not always feel confident about decision making in such a complicated legal environment and may act cautiously' (PHRN 2016, p. 4). And, in effect, access to data is more difficult than it needs to be:

> Information-handling policies that do not clearly identify the circumstances in which information can be disclosed to Australian Government agencies or others may result in an officer not disclosing information that could properly be shared — or a lengthy delay before such sharing occurs. (ALRC 2010, p. 525)

These legal restrictions also affect the ability of private sector entities to share identifiable information. For example, private health insurers hold data about medical practitioners' billing patterns and care outcomes, which cannot be released without their consent. If some of this information was made publicly available, as is already occurring overseas,[24] it would assist individuals in choosing a medical practitioner and reduce their transaction costs. Additional data about billing and other performance indicators collected by private health insurers would also give GPs more information when choosing a specialist to whom they refer their patients (Medibank Private, sub. 98).

[24] In the United Kingdom, patients can access data on care outcomes of GP practices. Data on billing is available to US consumers.

Public and private organisations must also comply with APP 12, which requires entities to allow individuals to access the data held about them, subject to a range of restrictions.[25] The Financial Services Inquiry (Murray et al. 2014, pp. 184–185) found that:

> … a number of impediments are still preventing consumers from being able to use their data effectively:
>
> - Little guidance is available on how personal information should be provided, including delivery method, timelines and standards for representing data.
> - In most cases, consumers are unable to authorise trusted third parties to access their personal information directly from their service provider. This reduces the ability of competitors to offer consumers better value or tailored services, or develop advice services to better inform consumer decision making.
> - Confusion exists over what constitutes personal information, which may limit individuals' access to data.

The Information Commissioner argued that being a principles-based law enables the Privacy Act to 'apply to many different Australian Government agencies and industry sectors, and to the myriad of ways personal information is handled in Australia' (OAIC 2016e, p. 4). And while this is certainly the case, principles have their drawbacks.

By being open to the interpretation of organisations that collect and manage data, the outcome will reflect the culture of the entity — more adventurous cultures will push boundaries and the less adventurous will find a case for inactivity. The result is a lack of clarity as to the rights of individuals and the responsibilities of data custodians, which impedes access to data. This is not a matter of attributing fault, but rather a description of an unfortunate reality.

De-identified data can be shared, but is relatively rare

While the Privacy Act protects personal information, de-identified information that is no longer about a reasonably identifiable individual can be shared and published freely. The Act requires information to be de-identified if it is no longer needed for the purpose for which it was collected, or another specified secondary purpose. Organisations also de-identify data before sharing or publishing it (box 5.2) (OAIC 2016b).

If data is not appropriately de-identified, the risk of re-identification increases significantly (El Emam et al. 2011). Therefore, data custodians must be satisfied that the data has been successfully de-identified before releasing it. This is usually done by completing a risk assessment, which can look at different aspects of re-identification, such as:

- the motivated intruder test, which assesses whether a 'reasonably competent motivated person with no specialist skills' would be able to use the data to identify an individual

[25] Individuals can also request access to government records about themselves under the provisions of the Freedom of Information Act. These provisions are often used in some areas; for example, the Department of Immigration and Border Protection (sub. 168) handled over 20 000 requests in 2014-15.

- re-identification 'in the round', which attempts to establish whether the data being disclosed could be used to identify an individual, either on its own or in combination with other information that may be available (OAIC 2016b).

Box 5.2 How de-identification works

De-identification refers to a process that alters data such that it can no longer be used to identify a specific individual. This is a two-step process, including:

- removing personal identifiers such as names, addresses or dates of birth, and
- removing or altering information related to a rare characteristic (such as people with unique occupations or small populations).

Beyond removing or modifying personal identifiers, there are a range of statistical techniques that can be used to de-identify data. Examples include combining information into categories (such as expressing a person's age as a range rather than single years); swapping identifying information between individuals; creating a synthetic dataset, that mimics the trends present in the real data but does not reflect the information of any real individual; and supressing data.

In addition to statistical techniques applied to data, data custodians can include clauses protecting the data in the contract signed with data users, and limit data users' ability to access the information directly (for example, by providing remote access or analysis services, so that the data remains with the custodian but analysis can still occur) (OAIC 2016b).

The Australian Taxation Office's individual sample files are an example of a de-identified dataset that is available online. The files contain data extracted from 2% sample of individual tax returns lodged each year. The ATO de-identifies the information by removing all personal identifiers and combining the data into categories (such as age and occupation groups). Further, the data is altered if it is deemed to increase the risk of identification (ATO 2016).

De-identification (and the more complex step of confidentialisation, which requires more changes to data) can also occur for specific limited purposes. The ABS's TableBuilder is an online tool that allows users to create customised tables using data from the 2006 and 2011 Censuses. To minimise the risk of identification, TableBuilder applies confidentiality techniques dynamically, so that data is randomly adjusted as it is retrieved in response to the user's request (ABS 2015).

Assessing the risk of re-identification is far from precise. First, the tests used are limited in their ability to address the most common reasons for data breaches — malicious intent and human error. De-identified datasets that satisfy the 'motivated intruder' test could still be breached by criminal organisations, using specialist skills and advanced equipment.

Second, the privacy risk posed by de-identified data is 'not just unknown, but unknowable' (Narayanan, Huey and Felten 2015, p. 357). It is very difficult to establish what other information or technology is or will be available in future that could be used to re-identify previously de-identified datasets (University of Sydney, sub. 35).

Realistically, hackers are likely to see the original datasets as a more attractive target, rather than attempting to re-identify individuals from de-identified datasets. Unless de-identification is particularly weak, re-identification requires more effort than attempting

to hack directly into a dataset, since every de-identified dataset has at least one more transformed layer than the original data for a hacker to penetrate. To further deter malicious re-identification, the Australian Government introduced amendments to the Privacy Act, to create new criminal offences of re-identifying government data that has been de-identified and disclosing such data (AGD 2016a).

However, in the face of uncertainty, and given the high agency reputation and individual costs involved if data is re-identified, data custodians tend to approach de-identification cautiously.

Guidance to agencies looking to publish de-identified data is available from both the OAIC and the National Statistical Service, but appears deeply underutilised given how few de-identified datasets exist. The guidance offered is generic; however, researchers have shown that a blanket approach to de-identification is likely to be unsuccessful, and each type of data requires a different combination of legal and technical measures to ensure the risk of re-identification is minimised (Narayanan, Huey and Felten 2015).

Nevertheless, a number of government agencies, such as the Australian Taxation Office (ATO, sub. 204), have successfully published de-identified data. Further, state-based linkage units such as the Data Linkage Branch in Western Australia (sub. 13), have successfully linked personal information for three decades, while maintaining individuals' privacy, as have statistical agencies in other countries, such as New Zealand, Denmark, Sweden and Finland (SSCH 2016).

In most cases, however, government agencies are likely to consider the worst case scenario, and avoid releasing the data altogether (Ritchie and Welpton 2014). In some cases where data is confidentialised, stakeholders have argued that agencies adopt standards exceeding the requirements of the legislation. The Australian Institute of Health and Welfare (AIHW, sub. 162, p. 18) states that 'frequently, these standards are applied without any real attempt at balancing the levels of risk against the research benefits of releasing finer-grained data'.

The Information Commissioner (2016b, p. 12) has acknowledged this issue in his recent submission to the Commission's Education Evidence Base Inquiry:

> Many organisations and agencies are unsure about how to de-identify data appropriately, and in some cases end up releasing personally identifiable information. I consider that this is an area of regulation where agreed industry terms and standards are important — not only to the actual efficacy of de-identification, but also to provide public confidence in it as a solution. To this end, I will be commencing a national conversation about de-identification and opening up consultation on renewed guidance later this year.

The need for better guidance on robust de-identification has become increasingly pressing following the reported re-identification of personal information from data released by the Department of Health. Researchers at the University of Melbourne found that it was possible to use de-identified MBS and PBS data, which was published on data.gov.au, to uncover provider numbers issued to doctors by Medicare. The Department and the OAIC

are investigating this event (Department of Health 2016; OAIC 2016a); any findings should be used to inform future guidance on de-identification. The significance of this event appears to be in the capacity to draw on advice from researchers about de-identification.

Robust and practical guidance will be a welcome step towards supporting agencies and businesses to use best practice de-identification processes. In conjunction with other agencies that have relevant expertise, such as the Australian Bureau of Statistics (ABS), the OAIC should issue detailed technical guidance on what constitutes best practice de-identification processes, including a risk-based approach to manage the risk of re-identification. This guidance should be regularly updated. However, it is important to realise that while guidance should be issued, neither the OAIC nor any other entity can guarantee risk-free data management.

Further, the OAIC should have the power to certify when best practice de-identification processes are being used by an organisation in accordance with this guidance, where it sees value in doing so (for example, to promote better practice). Having the experience and guidance of the OAIC will remove much of the uncertainty that stops data custodians from releasing de-identified data, as the OAIC certification will confirm that they have taken all appropriate steps to minimise risks. This will give confidence to individuals and organisations and expand access and use of data.

DRAFT RECOMMENDATION 5.1

In conjunction with the Australian Bureau of Statistics and other agencies with data de-identification expertise, the Office of the Australian Information Commissioner should develop and publish practical guidance on best practice de-identification processes.

To increase confidence in data de-identification, the Office of the Australian Information Commissioner should be afforded the power to certify, at its discretion, when entities are using best practice de-identification processes.

Privacy Act exceptions apply only for health research

The Privacy Act also recognises that in some cases (for example, some data linkage projects and research involving rare diseases), the use of de-identified data or obtaining individual consent for data use are impractical. Under section 95A of the Act, the National Medical and Research Council has created guidelines for human research ethics committees to approve the use of identifiable information in medical research without consent, where appropriate (NHMRC 2014).

These guidelines are limited only to health research — meaning that any other human research, such as social science, has no means to request access to identifiable information, regardless of the potential community benefit that such research could have. Both New

Zealand and the United Kingdom allow the use of identifiable information in all types of research, subject to maintaining privacy and data security (ADLS 2012; OPC (NZ) 2013).

In Australia, the ALRC (2008) recommended amending the law to extend the special arrangements for data access to all human research (more recently, the Productivity Commission (2016) recommended extending these arrangements to public interest research).

The ALRC recommendation has not yet been adopted by the Government; however, the Information Commissioner (sub. 200, pp. 30–1) flagged a potential change in this area:

> Given technological advancements and shifting community attitudes since the publication of the ALRC's For Your Information Report in 2008, I am of the view that it may be timely to re-evaluate the provisions, and consider whether it is still reasonable to limit the existing exceptions to health and medical research.
>
> Questions around the secondary use and disclosure of personal information have often proven to be problematic, particularly where an entity is unclear about whether or not a collection, use or disclosure for a secondary purpose would fall within an exception to the APPs. This uncertainty may contribute to a reluctance to make information available, even where this would be permissible under the framework. A review of the framework for research under the Privacy Act would therefore enable other mechanisms to be explored, alleviating this uncertainty, and could thus improve the availability of data for research.

Amending the Privacy Act exceptions, such that they would support improved access to data for all fields of research that are in the public interest, would be a positive development in improving data availability and use.

The final Report will outline a structure for the public interest definition, to support the OAIC in developing its guidance.

DRAFT RECOMMENDATION 5.2

The *Privacy Act 1988* (Cth) exceptions that allow access to identifiable information for the purposes of health and medical research without seeking individuals' agreement, should be expanded to apply to all research that is determined to be in the public interest.

The Office of the Australian Information Commissioner should develop and publish guidance on the inputs required to establish a public interest case.

Is privacy to blame?

The Privacy Act has often been singled out as a primary reason for the limited extent of information sharing between government agencies. The Australian Information Commissioner (sub. 200, p. 2) strongly opposed this view:

> Privacy is often named as one of the primary barriers that prevents the sharing or accessing of personal information from and between government agencies – that is not correct. … one of the main impediments to information sharing is rather a general reluctance to disclose personal information, due to a number of misunderstandings about obligations under privacy and other laws. Rather than preventing the sharing of personal information, privacy law places important limitations around the circumstances under which it can be collected, used and disclosed, consistent with the community's expectations.

Submissions to this Inquiry support the Information Commissioner's statement. The evidence suggests there is a level of misinterpretation — or at least a very risk averse interpretation — of the Privacy Act and its accompanying Privacy Principles by public sector data collectors and custodians, which in turn leads to an overly cautious and risk averse approach to data management (see, for example, AIHW, sub. 162; Cancer Australia, sub. 104; Department of Social Services, sub. 10; Grattan Institute, sub. 12; Office of the Information Commissioner – QLD, sub. 42). Similarly, confusion about the interpretation of privacy law limits the ability of consumers to access data about themselves that is held by data custodians in the private sector, such as telecommunications and utility companies (Murray et al. 2014).

In effect, the legal framework, which aims to create the flexibility to release data where appropriate, results in a more restrictive environment, as data custodians are uncertain about community expectations or de-identification techniques. Custodians of government datasets (usually mid-level public servants) carry the legal risk — it is unsurprising that they are risk averse, in the absence of clear and persistent guidance and direction from senior leadership on appropriate circumstances to release data.

In the private sector, the lack of clarity about what constitutes personal information, coupled with sometimes strong commercial disincentives to allow access to data, also result in restricted access, even when it is allowed by law. There may also be occasional genuine misunderstandings of the privacy legislation, lack of guidance from regulators or concerns regarding the potential consequences from making the wrong decision.

To remedy this situation, the Department of Prime Minister and Cabinet (2015, p. 36) has suggested that:

> Significant gains can be made in the short-term by educating staff on how to interpret legislation to share and make better use of data. This requires a change in mindset for staff to look for ways to make data available within the law.… In the medium-term, legislation should be reviewed to identify whether privacy and secrecy laws can be streamlined and modernised.

Our work indicates more is required. There is a need for structural reform. Frameworks that could be put in place to give data custodians more guidance on data release are discussed in chapters 8 and 9 of this Report.

Secrecy protections in specific Acts

Beyond the general protection afforded to personal information in the Privacy Act, a wide range of other Acts contain provisions preventing disclosure of information about people and businesses (appendix C). Such secrecy provisions are included in the Acts governing the activities of many Australian Government bodies. The ALRC (2010) identified 506 secrecy provisions in 176 different pieces of legislation. Unlike the Privacy Act, more than 350 of these provisions create criminal offences and are punishable by imprisonment.

Secrecy provisions have been restricting access to Australian Government data for over a century (ALRC 2010). For example, when it was first introduced, the *Crimes Act 1914* (Cth) prohibited Commonwealth officers from publishing or communicating any fact or document that it was their duty to keep secret. Except for broadening the scope of this prohibition to include former Commonwealth officers, which was introduced in 1960, this section of the Crimes Act has not been substantively changed in over 100 years.

These secrecy provisions can impose substantial limitations on access to and use of identifiable data. For example, section 135A of the *National Health Act 1953* (Cth) prohibits divulging any information about individuals collected through the operation of the Act, unless authorised by the Minister.[26] The penalty for divulging such information without authorisation is either a $5000 fine, two years imprisonment, or both. Inquiry participants have voiced similar concerns about the effect of secrecy legislation on access to identifiable data in other policy areas:

- For [the Department of Social Services], there are numerous separate pieces of legislation for social services across the three key areas of social security, family assistance and child support which limit the sharing of debt-related data. Legislative limitations in portfolio specific legislation also prevent the effective sharing of debt-related data between agencies. For example, the *Taxation Administration Act 1953* restricts sharing of tax information to specified purposes, and limits the use of Tax File Numbers as a de facto client identifier. Collectively, these legislative limitations prevent effective debt-related data exchange within and between agencies and limit a coordinated approach to debt management. (Department of Social Services, sub. 10, p. 10)
- Access to the EABLD [Expanded Analytical Business Longitudinal Database] is significantly constrained by the legislation governing ABS and ATO data. Currently, mechanisms have been devised to facilitate access by Australian Public Service staff, although these are less than satisfactory. **(**Department of Industry, Innovation and Science, sub 69. p. 3)
- The protection of privacy and security within legislative frameworks can, at times, limit the department's ability to use and share data. For example, there are specific protected information provisions contained in the Social Security Law, which go beyond the Privacy

[26] Numerous other Acts impose secrecy provisions on health data, including the *National Health Act 1953* (Cth), the *Health Insurance Act 1976* (Cth), and the *Aged Care Act 1997* (Cth). Each Act uses different terminology in its secrecy provisions, and this can result in situations where an agency restricts access to information that other agencies can release (ALRC 2010).

Legislation and Privacy Principles, which set a high threshold for allowing the use and reuse of protected information. (Department of Employment, sub. 18, p. 5)

- Until 2012 the ABS provided the Commonwealth Grants Commission with Government Finance Statistics (GFS) unit record data, under a return to source protocol, as the State and Commonwealth treasuries had endorsed this. Since then, however, the ABS has reinterpreted its Act and now considers that it requires the permission of *all* individual agencies described in the data, not merely the agency providing data to the ABS. This new interpretation of the Act has created significant barriers to [the] ability to analyse data without any improvement in the privacy offered to providers. (Commonwealth Grants Commission, sub. 58, p. 3, emphasis added)
- A report released in March 2016, made use of linked student data from Victoria recorded across four NAPLAN test years … Having student records that were linked enabled the analysis to focus on student progress rather than simply outcomes. Unfortunately, the same linked NAPLAN data were not available for other states – in some cases student identifiers were not properly recorded, and in other cases the education departments were not allowed to share the data with us. (Grattan Institute, sub. 12, p. 4)

Given the potential penalties that they carry, these secrecy provisions have a substantial effect on data custodians' willingness to share data. The concerns around compliance with secrecy provisions result at times in 'cultures of secrecy within some agencies [that] pose a greater barrier to information sharing than legislative restrictions' (Attorney-General's Department, quoted in ALRC 2010, p. 537).

DRAFT FINDING 5.2

A wide range of more than 500 secrecy and privacy provisions in Commonwealth legislation plus other policies and guidelines impose considerable limitations on the availability and use of identifiable data. While some may remain valid, they are rarely reviewed or modified. Many will no longer be fit for purpose.

Incremental change to data management frameworks is unlikely to be either effective or timely, given the proliferation of these restrictions.

Legal limitations on data linkages are a destructive policy

Data linkages, where different datasets containing information about the same individuals are brought together, add substantial value to data collections, enabling more insights to be derived from already collected information. By painting a more complete picture of individuals, data linkage supports the development of academic research and government policies (WA Data Linkage Branch, sub. 13).

Particularly in the health space, data linkages are conducted in a highly restrictive regulatory environment. For example, section 135AA of the National Health Act explicitly prohibits the linkage of data from the Medicare Benefits Program (MBS) and the Pharmaceutical Benefits Program (PBS), unless conducted under guidelines issued by the

Privacy Commissioner. These linkages have occurred, but guidelines only allow them in a limited set of circumstances, and require that any linked datasets are destroyed as soon as a specific project is completed (OPC 2008).

The legal limitations around access and use of MBS and PBS data have had substantial implications for medical research and policy evaluations. According to evidence submitted to the Senate Select Committee on Health (2016), if MBS-PBS linkages were routinely allowed, they could provide insights into clinical outcomes, access to services and cost-effectiveness of health policies that are currently not available. The requirement to destroy the datasets after projects are completed leads to duplication of effort and potential waste of public funds. The Information Commissioner acknowledged the need for a review of the guidelines, in consultation with the Department of Health (OAIC, sub. 200).

The requirement to destroy datasets is not limited to health data. The High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes, adopted by the Australian Government in 2010, mandate that *all* datasets resulting from linkages that use Commonwealth data are to be destroyed at the completion of projects, unless specific exceptions are put in place (CPSIC 2010). Linkage keys must also be destroyed when used to link Commonwealth data (NSS 2016c). These linkage keys comprise a code or set of indices that enable two or more records belonging to the same individual to be brought together from separate datasets (ABS 2012; Data Linkage WA 2016).

The Commonwealth's policy is anachronistic. Linkage keys are successfully re-used in projects that are based on data at the State level — for example, the Centre for Health Record Linkage (CHeReL) maintains a master linkage key system, which can create links between the health records of over 11 million people in NSW and the ACT (CHeReL 2016). These linkages meet the requirements of privacy legislation (see appendix B for further information on data linkage systems).

There are also numerous examples of successful enduring data linkages undertaken overseas. Since 2011, Statistics NZ has been managing New Zealand's Integrated Data Infrastructure (IDI), which holds and links a wide range of data about individuals, including health and safety, education, travel history, tax paid and benefits received (Statistics New Zealand 2016). Access to the datasets is managed through a risk-based approach, as well as a series of statistical processes to ensure confidentiality. Since its inception, the IDI has supported a number of policies, ranging from research and development subsidies to businesses to crime prevention and early intervention for disadvantaged households (Statistics NZ, sub. 62).

There has been some ad hoc progress in overcoming the restrictions imposed on users of Commonwealth data, and specifically data linkages. For example:

- There are a number of projects working to create enduring linked datasets from Commonwealth data, such as the Multi Agency Data Integration Project undertaken by the ABS along with four major Commonwealth agencies (the Department of Health,

the Department of Social Services, the Department of Human Services and the Australian Taxation Office) and the Business Longitudinal Analytical Data Environment, using data from the Australian Taxation Office and ABS surveys, in partnership with the Department of Industry, Innovation and Science (ABS, sub. 94).

- The AIHW (sub. 162) is also conducting a trial to establish arrangements for ongoing linkage keys involving the Commonwealth, New South Wales and Victorian health departments. The trial aims to establish arrangements for an ongoing national data linkage.

In its submission to this Inquiry, the Department of Prime Minister and Cabinet (sub. 20) indicated that the principles governing data integration projects may need to be reviewed. Such a review is indeed required.

DRAFT RECOMMENDATION 5.3

The Australian Government should abolish its requirement to destroy linked datasets and statistical linkage keys at the completion of researchers' data integration projects.

Data custodians should use a risk-based approach to determine how to enable ongoing use of linked datasets. The value added to original datasets by researchers should be retained and available to other dataset users.

Single identification numbers for better linkages?

Successful data linkages rely on the ability to identify the same individual in different datasets. This task may be simplified if different datasets use the same unique identifier for each individual.

Many data collections in Australia do not include a consistent unique identifier for individuals. APP 9 restricts the use of government related identifiers (such as Medicare or Centrelink reference numbers, or passport numbers) to specific circumstances, in order to prevent them from becoming universal identifiers. Such universal identifiers are seen as a potential risk to privacy, as they may enable data linkages that people do not consent to (OAIC 2014a).

Several participants in this Inquiry supported reform to introduce a single individual identification number in Australia, in order to improve the accuracy of data matching. For example, Professor Fiona Stanley (sub. 5, p. 12) stated:

> Linkage would be facilitated if Australia had a system, as in Scandinavia, of unique ID numbers for every individual in the country. This could actually be done very efficiently by using the Medicare Number in Australia; having such unique identifiers actually protects privacy and enhances confidentiality and results in more accurate linkage.

Evidence presented to us on probabilistic linking techniques (appendix B) indicated that these new techniques can improve the accuracy of data linkage without the need for a single identifier — some have claimed they can result in greater than 90% accuracy in some cases and, when done by a body with sufficient expertise, can also significantly reduce privacy risks (Curtin University, sub. 41). However, other stakeholders argue that such techniques are not fit for all purposes, as they may skew the data, and a personal identifier will be required for some datasets to be linked (AIHW, sub. 162).

For countries with a single identifier, advantages are apparent regarding the more uniform collection and compilation of data.[27] But for Australia, the benefits of moving to a single identification number, beyond what is currently used, are unlikely to outweigh the costs, or overcome community concerns (see, for example, Australian Privacy Foundation, sub. 142).

State privacy legislation

All states and territories, except for Western Australia and South Australia, have separate privacy legislation. This legislation applies to the state and territory public sector,[28] but not the private sector in most cases (unlike the Commonwealth Act) (appendix C discusses the privacy legislation in the states and territories in detail).

State legislation is based on similar principles to the Commonwealth Privacy Act. Despite these similarities, the existence of multiple regulatory schemes restricts data sharing across jurisdictions (Data Linkage Branch, sub. 13, NSW Government, sub. 80). All agencies involved in data sharing must ensure that the applicable legislation is adhered to. This necessitates multiple approval processes, conducted by agencies that adopt different approaches to data sharing (OAIC 2016b). As a result, access to cross-jurisdictional data can be severely impeded. For example, in the context of educational research, the Tasmanian Government (2016, p. 5) noted that:

> The lack of uniformity of privacy legislation across states has certainly impacted upon the willingness of jurisdictions to participate in cross-jurisdictional research and projects which have the capacity to enhance educational outcomes. …. It should be noted that these barriers are often perceived rather than real.

In some instances, sharing identifiable information between jurisdictions occurs successfully (Department of Prime Minister and Cabinet, sub. 20) — for example, between law enforcement agencies. CrimTrac (which since July 2016 has been amalgamated into the Australian Criminal Intelligence Commission) is a collaborative partnership between

[27] More than 100 countries have compulsory identification cards with a unique identification number, and others have optional schemes.

[28] The definition of the 'public sector' in this context varies significantly between jurisdictions (for instance, in their application to government-owned corporations, contracted service providers, and universities). Additionally, some states but not all of them have separate regimes for health privacy (appendix C).

state, territory and federal police agencies and the Attorney-General's Department, which provides tools for police to share information across jurisdictional borders (CrimTrac 2015). At the same time, States and Territories do not share aggregate data on criminal activity in a consistent manner. According to the NSW Bureau of Crime Statistics and Research (sub. 23), the lack of appropriate cross-jurisdictional data precludes analysis that could point to more effective policies to reduce crime rates.

There are, however, examples where de-identified and aggregate datasets are effectively shared between jurisdictions. For instance, the Department of Health (sub. 99) reported that it has released de-identified MBS and PBS data to state and territory health departments. All states and territories supply aggregate health data to the AIHW (appendix D).

In their dealings with the private sector, governments have the ability to compel entities to share identifiable information in some cases. For example, in Queensland, Victoria and New South Wales, police officers can use data collected by private operators of public transport billing systems to assist in their investigations (Mickelburough 2015; Munro 2015b). In many cases, authorised officers can request this information without the need for further approvals (see, for example, PTV 2014).

Privacy legislation affects data sharing between the public and private sector in other circumstances, such as emergency alerts. For example, private telecommunications companies work with governments to provide alerts to people in areas that are threatened by natural disasters. Alerts are delivered without emergency services having access to individuals' phone numbers or information about phone locations (Telstra, sub. 88). In other instances, private sector entities face difficulties in accessing data held by governments, and have to contend with overly bureaucratic application processes or high fees for data access. Archerfish Consulting (sub. 30, p. 5) submitted an example of this:

> On behalf of NSW Treasury, [Archerfish Consulting] applied to the Australian Institute of Health and Welfare (AIHW) for access to a data extract from State and National admitted patient datasets … By the time individual States had approved the release of the information, more than 12 months had elapsed and the total cost of the extraction was $2582. Access to the AIHW data was not unencumbered; access was contingent on Treasury's agreement that the data was solely for use on the nominated project.
>
> In 2015, when Archerfish's contract with NSW Treasury had expired, Archerfish accepted a pro bono economic evaluation engagement with a Sydney not-for-profit organisation … In a formal request to AIHW [Archerfish Consulting] asked for an extension of the user license for the admitted patient datasets …
>
> Rejecting [the] appeal for reconsideration, the Director of AIHW reaffirmed that AIHW intended to charge $2200 to reissue the data set and alter the data user license. The Director of AIHW claimed that it was an administrative requirement for AIHW to seek the clearance of each affected State for even trivial extensions of de-identified data and the internal costs charged were justified.
>
> The economic evaluation of the charity's work did not proceed.

Overall, successful collaboration is more likely where there is a vital need to share data to maintain the safety of the community, and where privacy concerns play a minor role. In many cases, however, complex approval processes and concerns about intellectual property can limit the efficacy of data sharing (PC 2014). Concerns about the way data will be used — for example, to make new comparisons of state and territory governments' performance in delivering services to the public — can also limit the extent to which data is made available. Access to data across all domains of government activity has substantial potential to generate important insights that will improve policy making and community wellbeing — and should occur routinely.

5.3 Lengthy approval processes waste time and money

Access to datasets containing identifiable information is subject to complex approval arrangements. Depending on the data requested, this can involve multiple data owners, custodians and stewards, integration units, ethics committees and other advisory bodies (chapter 3).

Data users (including researchers within and outside government) looking to access identifiable data must obtain consent from the individuals about whom the information was collected. Where this is impractical, researchers must have the approval of a human research ethics committee before using identifiable information (NHMRC 2014). Data custodians must also approve applications for access. If the dataset contains information from multiple sources, approval usually needs to be sought from each data custodian separately and often sequentially (see, for example, Archerfish Consulting, sub. 30, Centre for Big Data Research in Health, sub. 21).

Each policy and approval step is intended to ensure privacy and confidentiality are maintained and corporate reputation protected; however, the fact that approvals must be sought separately from numerous organisations that are governed by many different policies and use different application processes creates a major obstacle for data access.

Stakeholders have reported many instances, in particular in regards to health data, where governance processes were lengthy and inefficient (box 5.3). This has substantial implications for conducting health and other research:

> While the infrastructure is in place and there is impetus to conduct multi-jurisdictional research that involved record linkage in Australia, the administrative effort required and the lengthy time frames for approvals do not allow research involving national record linkage to be conducted in a timely manner. The current process is costly to funding bodies, the data custodians' institutions and also for the public, who ultimately support the research. (Mitchell et al. 2015, p. 325)

These difficulties do not affect only large scale projects, involving data from multiple jurisdictions. Accessing single datasets can also be onerous, due to the difficulties in identifying the data custodians, and the time taken for requests to be approved (chapter 3).

Even large government agencies that were tasked with analysing data and supplying information face challenges when trying to gain access to datasets in different jurisdictions. The AIHW (sub. 162) submitted that in creating the Enhanced Mortality Database, which linked four datasets, it encountered cumbersome and lengthy negotiations with multiple data custodians and ethics committees. Some approvals were only granted for a short time, and had to be renegotiated before the project could start, due to the delays in obtaining permission from other organisations.

The issues highlighted by these examples refer to two aspects of data governance — the role of the data custodian, and that of the human research ethics committee.

Box 5.3 Navigating the maze of data access: the example of research into Indigenous women's health outcomes

Compared with the rest of the population, Indigenous women are twice as likely to get cervical cancer and four times more likely to die from it. Yet there is limited information about the participation of Indigenous women in the National Cervical Screening program. To overcome this, in 2011, a group of researchers commenced a project to link data from cervical screening registries to other health datasets that contain information about Indigenous status (Garvey et al. 2016). Five years later, the researchers have only been able to analyse data from one jurisdiction, due to ongoing delays in approval processes.

A time line of delay

> The time from initiation to completion of the ethics committee approval process ranged from two to 32 months, and final approval to link and access all datasets took five years. In one jurisdiction, a data custodian provided conditional approval pending ethics committee approval; by the time the HREC [human research ethics committee] approval was received, a new employee held the data custodian position and the conditional approval was deemed to be invalid, requiring the approval process to start afresh. The first set of data was obtained in December 2013 and one dataset remains outstanding as of April 2016. While the professionalism, support and thoroughness of almost all individuals involved has been exemplary, the process has been fraught with duplication, ineffective regulation and delay.
>
> Over 400 days of person time, at a cost in excess of $200,000, were spent obtaining HREC and DLU [data linkage unit] approval. Datasets contained different variables or the same variables with different naming conventions; the researchers worked with the DLU to obtain variables that were necessary to answer the project's research questions. In some jurisdictions, a request for each variable had to be justified and negotiated and, where changes to the original request were necessary (either researcher or DLU driven), an amendment was required by the relevant ethics committees and/or data custodians.
>
> The first results, based on population data (1,334,795 women aged 20-69 years) from one jurisdiction, show that Indigenous women have a 20-point lower screening participation rate than other Australian women, with no improvement over time, and a higher rate of high-grade cervical abnormalities. **Had the process been more efficient and less protracted, results for the whole country would have been available by now, information which could have underpinned interventions to reduce cervical cancer occurrence in indigenous women.** (Garvey et al. 2016, pp. 95–96, emphasis added)

Data custodians: between a rock and a hard place

Data custodians (often mid-level public servants) have substantial, and at times conflicting, responsibilities in regards to the data they oversee, in particular when it is integrated with other data (NSS 2015).

They are expected to maximise the value of data holdings (which can only occur if the data is used). It can often be difficult to assess whether the value of the data is indeed maximised, and government departments do not always have clear ways to measure this.[29] At the same time, custodians must also ensure that data is only released in accordance with the relevant legislation. In many cases, if this legislation is not followed correctly, the data custodian may risk imprisonment or substantial fines, which creates strong disincentives for custodians (section 5.4). Even more intimidating can be internal procedures, which in some cases have evolved to take on a more substantial role than that originally intended in the legislation (box 5.4).

The result is overly bureaucratic approval measures, which are inconsistent across jurisdictions and in some cases lack clarity and transparency. Researchers have reported that data custodians are often under-resourced, and this results in delays in processing applications (SSCH 2016).

Human research ethics committees have an important role, but create further duplication of effort

Human research ethics committees (HRECs)[30] hold an important role in approving requests for data access. This role is particularly important in health research, where ethics committees are empowered by the Privacy Act to allow access to identifiable information, where obtaining consent or using de-identified information are impractical. In doing so, HRECs attempt to balance the individual's right for privacy with the potential benefit to society from conducting medical research using identifiable information (NHMRC 2014).

[29] For example, when asked by the Senate Select Committee on Health what was the Key Performance Indicator used to evaluate data usage, the Department of Health stated that 'it facilitates a Data Governance Council that includes representation from the Department, the AIHW, the Australian Bureau of Statistics, Department of Human Services and other Health portfolio agencies. The Council is responsible for ensuring effective policies and governance for the Department's approaches to data collection, management, interrogation, sharing, access and release' (SSCH 2016, p. 45).

[30] In addition to human research ethics committees, universities have ethics committees that oversee research that is conducted on animals (University of Melbourne, sub. 148).

Box 5.4 Not all approvals are made equal – the interactions between data custodians and human research ethics committees

While there is no consistent approach to the approval processes for data access, in many cases data users first obtain in-principle approval from data custodians before submitting an application to the relevant human ethics committee (HREC). Data custodians give their final approval only after the HREC signs-off on the application (see, for example, Population Health Research Network 2011).

From a legal perspective, once the HREC has approved the application, there are no further requirements that the data user needs to fulfil. However, in practice, data custodians refuse to allow access to data even after HREC approvals. Data custodians are not required to provide potential users with an explanation as to why their application was refused, and there is limited transparency in the process (Adams and Allen 2013).

In one recent case, the Australian Electoral Commission (AEC) rejected the approval given by the Monash University HREC to a research project funded by the National Health and Medical Research Council (NHMRC) (Loff et al. 2013, Monash University, sub. 133).

The *Electoral Act 1918* (Cth) allows the AEC to provide sample data from the electoral roll to medical research that has been approved by a HREC and adheres to NHMRC guidelines. However, despite this, the AEC states that its 'first obligation is to the elector. Applications [for data] will be rejected if there is a risk that medical research will breach elector security and privacy, is politically biased or has the potential to discourage electoral participation' (AEC 2016).

In the Monash University case, the AEC refused to supply data to the project despite it having obtained the approvals required by legislation. According to the researchers, this was because in the AEC's view, the HREC members did not have sufficient qualifications and expertise to deal with privacy issues, and the research topic (to establish the views of Australians on privacy and participation in epidemiological research) did not constitute medical research (Loff et al. 2013, Monash University, sub. 133).

While the intent of the legislation was to increase access to identifiable information that will be used in medical research, the complexities of the HREC application and approval process act as an additional barrier to access. This is particularly the case in projects requiring data linkages, where approvals from multiple HRECs are required, and each committee needs to consider numerous separately-generated guidelines for the review of each project (Judy Allen and Carolyn Adams, sub. 106). Apart from using a national application form, there is little similarity in the way the committees operate. Not only do they require vastly different documentation to be submitted, but there is limited mutual recognition between committees. Therefore, researchers must re-submit their application and negotiate with each committee separately (Mitchell et al. 2015).

A number of participants to this Inquiry supported introducing a more harmonised system of ethics committee approval. For example, the Population Health Research Network (sub. 110, p. 2) suggested that:

> A system of national mutual recognition of ethics review of applications for research using linked data should be implemented. It may be possible to adapt the current system for review of clinical trials applications for this purpose.

Various organisations have been working towards mutual recognition between HRECs; however, progress has been slow (SSCH 2016). The National Health and Medical Research Council (NHMRC), which oversees the operation of HRECs in Australia, is promoting a National Mutual Acceptance Scheme, intended to streamline the approval of research conducted in public hospitals (NHMRC 2016). It also runs a National Certification Scheme for HRECs. Approval from committees that are certified by the NHMRC is accepted nationally. However, only a small minority of committees are certified (NHMRC 2012, 2016).

Some academic institutions are implementing other approaches to streamlining ethics approvals — for example, the University of Melbourne (sub. 148) signed a memorandum of understanding with participants in the Melbourne Academic Centre for Health to implement mutual recognition of ethics reviews. The value of such memorandums, discussed in chapter 3, can become a negative rather than positive bureaucratic force.

While these issues are particularly prevalent in the health sector (due to the sensitive nature of the data, but also due to the special role played by HRECs in medical research), governance arrangements present a challenge in all cases where access to identifiable information is required. The Department of Social Services (sub. 10, p. 13) identified 'aligning multiple ethics committees and data custodian approval processes' as one of the significant operational challenges agencies face in sharing data.

Streamlining governance arrangements — by simplifying application processes, clarifying the role of data custodians and promoting mutual recognition between ethics committees — would go a long way towards supporting increased use of identifiable data. It is unlikely that this change can be achieved incrementally, given the extent of accreted layers of approvals. However, governments can support this process by issuing clear instructions to data custodians, particularly in regards to their interaction with HRECs. Further, since much of academic research uses public funding, institutions that promote HREC mutual recognition should be given funding priority. Such an approach would create incentives for academic institutions to implement mutual recognition, and increase efficiency in access to data.

DRAFT RECOMMENDATION 5.4

To streamline approval processes for data access, the Australian Government should:

- issue clear guidance to data custodians on their rights and responsibilities, ensuring that requests for data access are dealt with in a timely and efficient manner;
- require that data custodians report annually on their handling of requests for data access;
- prioritise funding to academic institutions that implement mutual recognition of approvals issued by accredited human research ethics committees.

State and territory governments should mirror these approaches to enable use of data for jurisdictional comparisons and cross-jurisdiction research.

Current data linkage capacity unlikely to be sufficient into the future

Once all approvals are obtained, data linkage for research projects is carried out by accredited bodies — six state-based linkage 'nodes', which can link state data, and three accredited integrating authorities (the ABS, AIHW and the Australian Institute of Family Studies) which can link Commonwealth data (AIHW, sub. 162). Australian Government agencies are able to link data without an accredited authority in some circumstances.[31]

Researchers have raised concerns about the limited number of organisations accredited to link Commonwealth data, suggesting that the increasing demand for linkages creates capacity bottlenecks and delays research projects (Centre for Big Data Research in Health, sub. 21). The AIHW, one of the three integrating authorities, argues that while demand for linkages has increased, any delays are the result of the time invested in preparing and standardising the data, rather than the linkage process itself (sub. 162).[32]

It is likely that both the number of accredited integrating authorities and the time required to prepare data for linkage have a role to play in any delays experienced by researchers. Regardless, it seems that the capabilities and resources to undertake these tasks exist in additional institutions, beyond those currently allowed to link Commonwealth data. It

[31] Using an accredited linkage authority is only required for projects that are conducted for statistical purposes, involve two or more data custodians, where at least one holds Commonwealth data, and where the users are not the original data custodians (for example, where the linkage is undertaken for use in academic research). Some data integration projects, such as those undertaken for regulatory purposes and compliance monitoring, do not need to be undertaken by an accredited integrating authority. In these cases, data custodians can choose another integrating body or outsource part of the project to a linkage unit (NSS 2016d).

[32] This view is also echoed by the Data Linkage Branch in the Department of Health WA (sub. 13). In addition to the need to help prepare the data for linkage, the Data Linkage Branch has argued that the time to complete data linkage projects has increased as the projects have become more complex.

seems unnecessarily restrictive and is certainly not in the wider public interest not to accredit more such institutions.

Current guidelines for accrediting existing linkage units so that they are allowed to integrate Commonwealth data for research projects state that '[d]uring the early implementation phase, only Commonwealth agencies will be approved for accreditation so that the system can be fully tested and evaluated before it is extended beyond the Commonwealth' (NSS 2016b). The guidelines have been in place for over four years, and the Department of Prime Minister and Cabinet (sub. 20) has stated that it is reviewing the governance arrangements for accrediting additional integrating authorities.

Stakeholders have suggested that the state-based linkage units should also be accredited to work with Commonwealth data (Centre for Big Data in Health, sub. 21, Population Health Research Network, sub. 110). These units have many years of experience in linking data and, in some cases, have already linked Commonwealth data. Most notably, the WA Data Linkage Branch linked Commonwealth health data for a decade until the Department of Health defunded the project, apparently due to other funding priorities (appendix D). The recent Senate Inquiry into health policy has recommended that the Australian Government consider extending accreditation to link Commonwealth data to State linkage units (SSCH 2016).

DRAFT RECOMMENDATION 5.5

In light of the Australian Government's commitment to open data, additional qualified entities should be accredited to undertake data linkage.

State-based data linkage units should be able to apply for accreditation by the National Data Custodian (Draft Recommendation 9.5) to allow them to link Australian Government data, and the intention of 'open by default' should apply to these exchanges.

5.4 Data custodians have limited incentives to release identifiable data — and plenty of reasons not to

Incentive structures faced by data custodians have the effect of limiting access to identifiable data. The act of increasing access to identifiable information creates personal and agency costs for data custodians, both directly and indirectly. Risks also increase, and capabilities required will be different. If passed on to data users, these costs may be prohibitive and result in limited demand for data, and no progress in either innovation or discovery, both of which are vital to the development of data use.

The cost of increasing access

Custodians face a range of costs when deciding to increase access to data. These costs can be incurred either directly, in the course of preparing the data and managing access, or indirectly, as a result of reputational risk. The reputation of the data custodian can be compromised as a result of data misuse, which can lead to privacy breaches. There is also a risk that the data reveals unfavourable information about the agency's operations and performance (DPMC 2015; Ritchie and Welpton 2011).

Both public and private sector organisations grapple with the costs incurred in the process of increasing access to identifiable and other data (chapter 7). Specific costs to manage the risks around identifiable data include the time taken to ensure all legal requirements are complied with, the costs (both in terms of staff time and technology) required to de-identify the data and ensure the risk of re-identification is minimised, and set up appropriate ways to access the data securely. For example, the Department of Foreign Affairs and Trade (sub. 202, p. 9) stated:

> The vast majority of passport data held is personal information that cannot be used or disclosed for secondary purposes unless one of the exceptions to the general prohibition against secondary use or disclosure under the Privacy Act applies. To depersonalise the data in order to share it more widely would be time consuming and costly, given the architecture underpinning how the data is currently stored.

Access costs can limit the development of new research proposals and new services using identifiable data (CIFR sub. 9; SA NT DataLink sub. 123; Western Sydney University sub. 119). SA NT DataLink submitted examples of this to the Senate Select Committee on Health (2016, p. 51):

> For one project with a cohort of about 10,000 individuals and linking 4 datasets, the SA NT DataLink cost was estimated at $10,000. The researchers also wished to link to a Commonwealth dataset for which they quoted approximately $160,000. Because of the high Commonwealth costs, the researchers could not include this data.
>
> Another project with a cohort of about 240 individuals and linking 4 datasets, the SA NT DataLink cost was estimated at $8,500. The researchers also wished to link to another Commonwealth dataset for which they quoted approximately $40,000. Again, because of the high Commonwealth data costs, the researchers could not include this data.

Incentive structures

Similarly to other types of data, the incentives to release identifiable information are limited. Data custodians often do not benefit from the value generated by the data, and therefore have little incentive to incur the costs required to prepare it for release (chapter 7).

On the other hand, legislation creates clear disincentives to release identifiable data for public sector employees. The various Acts under which Australian Government agencies operate — including the ATO, Department of Social Services, Department of Health, Department of Immigration and Border Protection and many others — state that

unauthorised disclosure of protected information, including identifiable data, may result in a two year imprisonment term, a financial penalty or both (see, for example, Department of Immigration and Border Protection, sub. 168). Each Act defines the type of disclosure that is allowed, often focusing on sharing data for operational reasons.[33]

For private sector entities, identifiable data about their customers is often a substantial commercial asset and they have strong incentives to restrict access (see, for example, ANZ, sub. 64; Telstra, sub. 88). This data is used to develop new products and services, and may also contribute to revenue, if parts of the dataset are sold to third parties.

Private companies are unlikely to share identifiable information about their operations with competitors.[34] Attitudes may be more nuanced than first appears; for example, Westpac (sub. 197) agrees that sharing de-identified information may have benefits to all market participants. Companies such as Data Republic (sub. 176) are developing solutions that allow corporations to securely share de-identified information.

The Financial System Inquiry suggested that if small businesses were able to share de-identified data with a third party aggregator, such as a payments service provider, this could greatly improve competition and target products and services to customer demand (Murray et al. 2014). Such data sharing is already occurring on a small scale. For example, the ANZ Business Insights initiative offers small business customers the ability to benchmark their operations against aggregated transactional data from similar businesses (ANZ nd). Similarly, using de-identified data from NAB and UBank in conjunction with ABS census data, Quantium (sub. 187) developed a website where individuals can compare their detailed spending patterns with other people across Australia.

5.5 There are risks of improved access, but most breaches happen in data collection and storage

As identifiable information is collected and generated in increasing volumes across disparate entities with widely differing commitment to individual or customer welfare, this increases the risk of a data breach. A breach occurs when 'personal information held by an agency or organisation is lost or subjected to unauthorised access, modification, disclosure, or other misuse or interference' (OAIC 2014b). Data breaches can be a result of human error or systems failure; however, in most cases, they occur due to malicious or criminal

[33] For example, the *Aged Care Act 1997* (Cth) permits the disclosure of personal information to government agencies, such as Medicare, Centrelink and the Department of Veterans' Affairs to enable them to carry out their duties, but specifically prohibits using the information for any other purpose. Information can be disclosed if the Secretary of the Department administering the Act believes it is in the public interest to do so.

[34] Some markets have unique characteristics that lead to identifiable data being shared. For example, dentists provide information to private health insurers when treating patients. According to the Australian Dental Association (sub. 8), vertically integrated health insurance companies can then use this information to increase the competitiveness of their own dental practises.

activity, such as data theft or hacking (Ponemon Institute 2016). By comparison, breaches due to sharing or release are far fewer in number and reach.

Over time, Australians have become increasingly concerned about data breaches and the risk of identity theft. However, concerns about loss of privacy online do not always translate to proactive measures to protect personal information (OAIC 2013).

Community attitudes to privacy and data access

Community surveys indicate that the level of concern about the privacy and security of personal information has been increasing, as more individuals use the internet to access a range of products and services, and post more information online.[35] In 2013, three in four Australians reported that they were more concerned about the privacy of their personal information while using the internet, than they had been five years earlier (ABS 2016a; OAIC 2013).

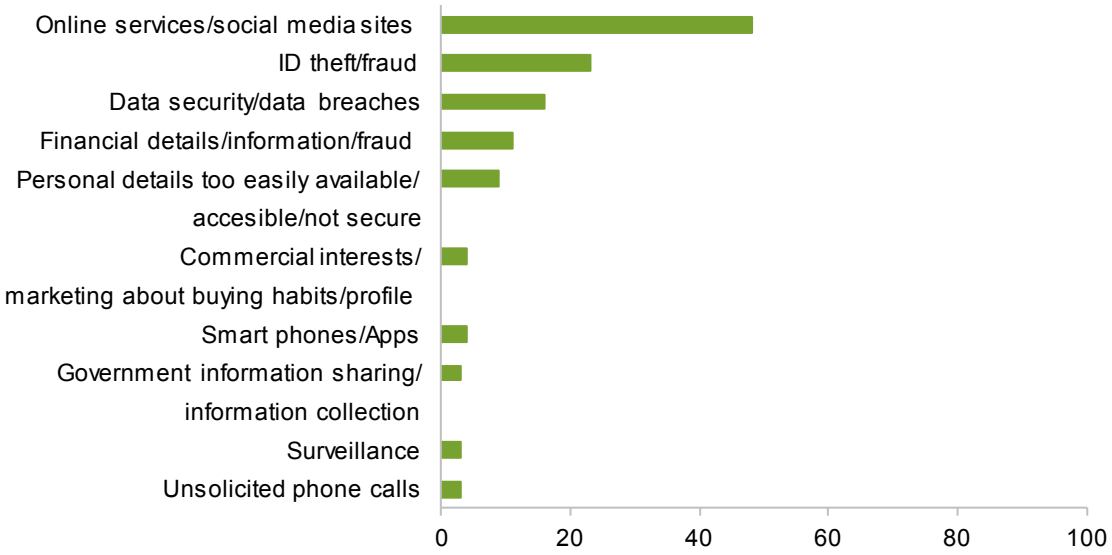
Much of this concern relates to online services, and in particular social media. Online services and social media sites are seen as the biggest risk to privacy, and people have limited trust in the ability of social media providers to keep their personal information secure (figure 5.1 and 5.2). Nearly all users of online services take some measures to ensure their information is secure (OAIC 2013). But they also expect government to play a role in keeping data safe — research has shown that there is an expectation that the responsibility for the protection of personal information is shared between users, service providers and government (ACMA 2013).

Despite the growing concern about the safety of information provided online, this is not always reflected in daily behaviours. While Australians believe it is very important that they are informed of how public and private organisations handle their information, only half of them read websites' privacy policies, which usually contain details on the way data is collected and handled (OAIC 2013).

Many people are willing to forgo some of their privacy and provide personal information online in order to access improved services, discounts or tailored services (Acquisti, Taylor and Wagman 2016). For example, many Australians are not concerned about the privacy implications of participating in brand loyalty programs, which collect their personal information, and choose to provide this information in exchange for benefits and discounts. In 2015, 84% of Australians participated in brand loyalty programs, such as those operated by Coles, Woolworths and Qantas. Only 9% of people who stopped participating cited privacy as a concern (Directivity et al. 2015). People can also choose to provide their location information online, which can be used for targeted advertising (chapter 4), but also to assist them with contacting family and friends in an emergency via social media (Facebook, sub. 172).

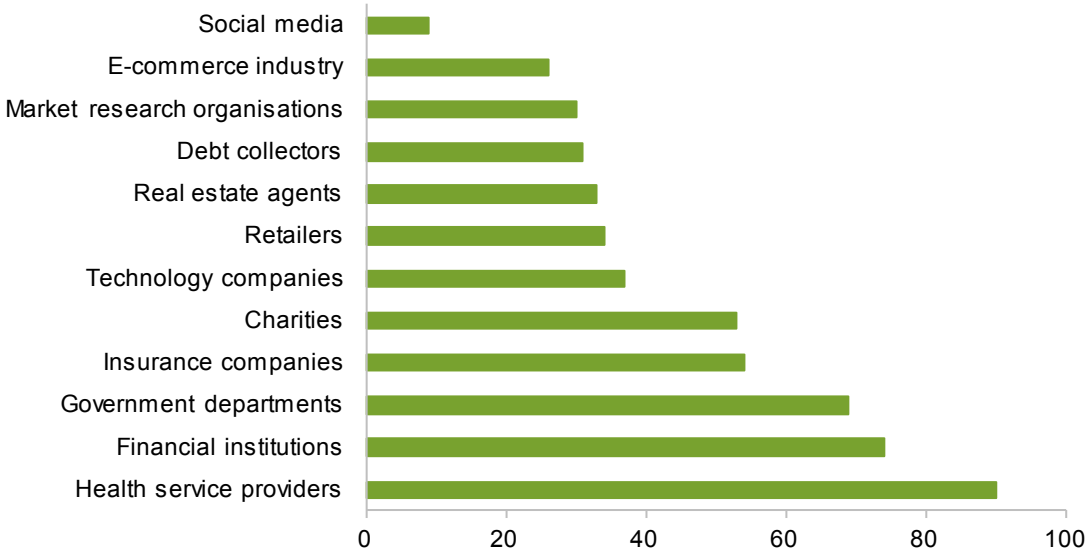
35 In 2014-15, over 85% of Australian households had internet access at home, and over 70% accessed the internet for social networking (ABS 2016a).

Figure 5.1 **Biggest privacy risks perceived by people**
Per cent of respondents[a]



[a] Sample size = 1000. This chart only shows the most common responses indicated by participants.
Source: OAIC (2013).

Figure 5.2 **Trust in organisations to handle personal information**
Per cent of respondents trusting organisations listed[a]



[a] Sample size = 1000.
Source: OAIC (2013).

DRAFT FINDING 5.3

Although parts of the government view community expectations as a factor that limits the use of data, reliable surveys have shown that most individuals believe sharing personal information between government departments can be beneficial, and indeed is occurring without damage.

However, individuals expect to remain in control of who data on them is shared with.

In dealing with government, people tend to trust that their information is held securely, but there are concerns about how the data is used. The public sector's perceptions of these community views tend to limit the use of data (DPMC 2015). However, survey respondents do not generally view information sharing between departments as major threat to privacy (OAIC 2013). In fact, according to the Queensland Government (sub. 207), most people expect the different parts of government to share data. Overseas studies have found that people overestimate the extent of information sharing that is already occurring within government (Bickers et al. 2015); in Australia, there is 'anecdotal evidence that suggests the community already believes there is widespread sharing of data across government' (ATO, sub. 204, p. 2).

Most people believe sharing personal information between government departments would be a good thing, as long as individuals had the choice whether to allow sharing to occur (Bruce and Bruce 2015). They see benefits in sharing, as they believe it will enable more efficient and accessible government services. But at the same time, individuals expect governments to share information with their consent, only when strictly necessary, and be transparent about their data handling processes (Bickers et al. 2015). Overall, it seems the community may be far more accepting of data sharing than government agencies believe, provided individuals have a degree of control over their data and the benefits of sharing are evident. The onus is on government to communicate these benefits effectively.

Similarly, despite the highly sensitive nature of medical information, individuals are willing to share this data in order to promote health research. They are willing to do so, however, if they are well informed about the research and the organisation conducting it, and if they give their permission for the data to be used (King, Brankovic and Gillard 2012). A recent survey found over 90% of Australians were willing to share their de-identified health data to advance medical research and improve patient care (Research Australia 2016).

These findings are a reflection of community attitudes to increasing access to data more broadly — individuals may be willing to share their information where they see a personal benefit in doing so. They may also allow third parties to share such information where there is a clear benefit (such as improved government services or advances in medical research), as long as they are well informed and trust that their information is handled appropriately, and they feel that they have a certain level of control over their data. For governments seeking to increase access and use of data, the level of trust in the community

and the effective communication of benefits derived from data use are important considerations. Chapter 10 describes how the Commission proposes to maintain trust while increasing access to data.

How real are these risks involved in increased data access?

A data breach causes identifiable information, which should be protected, to become publicly available. This can happen either when identifiable information is compromised directly, or when linkages of de-identified information provide a sufficiently detailed dataset to allow re-identification of individuals. Re-identification technologies are developing rapidly, and there is a heightened fear of inappropriate disclosure (Data Linkage Branch, sub. 13, attachment 5). However, breaches that are caused by re-identification are far rarer and smaller in their scale, compared with direct unauthorised access to datasets as a result of poor data security by collectors, individuals or both (box 5.5).

For the organisations affected by a data breach, there are substantial costs involved in addressing the security flaws that enabled the breach to occur, notifying the clients involved and potentially paying damages and fines. Beyond the direct costs, the biggest consequence for private sector organisations is loss of trust from their customers, and as a result, lost business opportunities. While some industry sectors, such as financial services, are more vulnerable to data breaches than others, any organisation that holds identifiable information can be at risk (Huq 2015; Ponemon Institute 2016).

When government agencies are affected by data breaches, this can have very substantial implications for their operations, and the community trust in the public sector handling of personal information. For example, in 2007, the UK Government lost two CDs containing the personal details of about 25 million people. The review into the data loss uncovered systemic problems in the Government's information security policies and data governance. A wide range of measures were put in place in order to move government data management policies 'away from [the] current operating model where it typically takes responsibility for collecting and maintaining data on its customers, to one where its customers, be they individuals or businesses, entrust their information to HMRC on the understanding that [the Government] will keep it secure' (Poynter 2008, p. 44).

The consequences for individuals whose details are disclosed as a result of a data breach can be far reaching. Such disclosure of data can:

- cause humiliation, embarrassment or anxiety for the individual
- impact on the employment or relationships of individuals
- affect decisions made about an individual or their ability to access services, such as their ability to obtain insurance
- result in financial loss or detriment

- pose a risk to safety, such as identifying a victim of violence or a witness to a crime (Privacy Committee Of South Australia 2015, p. 4).[36]

Data breaches can result in identity theft, which affected about 126 000 Australians in 2014-15 (ABS 2016b).[37] Earlier surveys suggest that most personal information used in cases of identity theft was obtained online, either through theft, hacking or from information sent by email or placed on a website. Some victims suffered financial losses, and others reported being refused credit or being accused of a crime (Smith and Hutchings 2014).

Data breaches are receiving a lot of media attention, and the number of reported breaches has increased substantially in the past decade (Office of the Privacy Commissioner (NSW), sub. 173). Globally, the majority of these breaches are due to hacking activities. While evolving technologies are creating new opportunities for hackers, the most common ways to gain unauthorised access to data rely on exposing long-standing user vulnerabilities — such as individuals using weak passwords or opening malicious attachments sent by email (Verizon 2016).

There are no official statistics on the number of data breaches occurring in Australia. The limited number of cases that are reported to the OAIC are likely to represent a very small proportion of the overall number of incidents involving misuse of personal information.[38] Examples of the data breaches handled by the OAIC included fraudulent requests for personal information held by large organisations and personal data being mistakenly released on a government website (OAIC 2015).

Research has shown that nearly half of reported data breaches in Australia are the result of malicious activity. Human error or systems issues each account for about a quarter of the remaining data breaches (Ponemon Institute 2016). While neither of these factors can be completely eliminated, organisations can use risk management processes to mitigate risks while increasing access and use to data (section 5.6).

Is re-identification a major risk to identifiable information?

In addition to data breaches resulting in direct access to personal information, an increase in data access and availability may raise the prospect of re-identification. This can occur in

[36] On the other hand, for some individuals, increased access to data may result in criminal liabilities, such as the discovery of welfare fraud.

[37] Overall, 1.6 million Australians experienced a form of personal fraud (including card fraud, identity theft or scams) in 2014-15 (ABS 2016b).

[38] Notification of data breaches is mostly voluntary, although a data breach notification bill is likely to be introduced by the Australian Government in 2016 (AGD 2016b). Currently, only eHealth records require mandatory notifications of data breaches. In 2014-15, seven such notifications were made by Medicare, each affecting one individual and resulting from a system error in the Medicare database. The number of voluntary data breach notifications received during 2014-15 increased to 110, 64% more than the number of notifications received in the previous year (OAIC 2015).

two ways: either the data can still be used to identify a specific individual even though identifiers (such as names, addresses and dates of birth) have been removed, or the combined dataset resulting from linkage of different de-identified data allows an individual to be identified (NHMRC, ARC and AVCC 2007).

The technology that can be used to re-identify individuals is rapidly evolving, and the risk this poses to privacy may increase as more datasets become available (Data Linkage Branch, sub. 13, attachment 5). However, the risk of re-identification is mostly the result of a failure to adopt the most effective de-identification techniques (El Emam et al. 2011). In the recent past, there have been a number of celebrated breaches of identification overseas, caused by linking apparently de-identified data sets. But the number is small, vanishing in comparison to the range and volume of data breaches caused by hackers (box 5.5).

Box 5.5 Health data identification risks in the US — data breaches are a thousand times more risky than re-identification

In 2015, the US health sector was affected by some of the largest data breaches in its history. In February, health insurance provider Anthem announced that its servers had been hacked, leading to the theft of the personal records of nearly 79 million people. In March, Premera, another health insurance company, revealed that its servers had also been hacked and 11 million personal records were stolen. Overall, in 2015, more than 111 million individual healthcare records were stolen by hackers (DHHS (US) 2016; Munro 2015a).

Most breach incidents occurred as a result of human error, coupled with insufficient data security measures used by companies (Yaraghi 2016). In both the Anthem and Premera breaches, hackers sent employees emails that seemed legitimate, and led them to log into fake websites with their company username and password. These credentials were then used by the hackers to break into to the companies' servers (bitglass 2016).

Hackers have been using these strategies, known as phishing, for a number of years, and many such attacks occur each day (Verizon 2016). Re-identification attacks seem far rarer — usually, reports of successful re-identification come from researchers trying to expose weaknesses in data security measures, rather than steal information.

In previous years, there have been a number of prominent cases where individual health records were re-identified. For example, in 1997, researchers successfully re-identified the health records of Massachusetts Governor William Weld, using health insurance and voter registration data. The health insurance records included individuals' full post codes, dates of birth and gender. By linking this information to the details included in the voter roll, the researchers were able to successfully identify the Governor (Barth-Jones 2012).

Since then, however, de-identification protocols have evolved significantly, including aggregating data about individuals in larger geographical areas, rather than specific post codes. Under these protocols, there is only a 1 in 3500 chance of an individual being successfully identified (Barth-Jones 2012) — in comparison, Americans currently face a 1 in 3 chance of having their complete health record stolen by hackers.

Data linkages have been conducted in Australia for many years, and until recently, there has been no evidence of re-identification occurring (Menzies Foundation 2013; Stanley 2010). This has been attributed to the contractual and legal obligations imposed on

researchers using linked data.[39] Researchers have little incentive to breach the trust of data custodians by intentionally re-identifying data, as this will be detrimental to the entire academic community (Menzies Foundation 2013; Ritchie and Welpton 2014). In the recent case of re-identification using data released by the Department of Health, it was the researchers themselves who alerted the Department to the vulnerabilities in the data (Department of Health 2016). Similarly, there are limited incentives for data linkage units to breach confidentiality and misuse data. Data linkage authorities use a range of safeguards to minimise the risks to privacy (see, for example, ABS, sub. 94).

Undoubtedly, increasing access to identifiable information requires more investment in risk management and constant attention to statistical techniques and learnings. However, enormous volumes of identifiable information are being generated, accessed and used every day. Individuals and governments have limited control over how this data is generated and used in the private sector. At the same time, new frameworks (such as Privacy by Design, discussed below) enable increased access to data without compromising personal information (Office of the Information Commissioner – QLD, sub. 42).

In this context, the incremental risk of allowing increased access and use of identifiable information, using security protocols and trusted user models, is very small. At the same time, the potential benefits from increasing access to data are substantial.

DRAFT FINDING 5.4

Large volumes of identifiable information are already published online by individuals or collected by various organisations, with or without explicit consent.

In this context, the incremental risk of allowing increased access to formerly identifiable data by public and private sector organisations, using security protocols and trusted user models, is likely very small.

Breaches of personal data, often enabled by individuals' unwary approach to offering data, are largely dominated by malicious or criminal activity. By comparison, breaches due to sharing or release are far fewer in number and reach.

[39] For example, to access data from the WA Department of Health, researchers sign a declaration of confidentiality, where they agree to use the data only in accordance to legislation. Researchers must declare that the data will only be used for the purposes approved within their specific project, and destroyed after the project is completed (Department of Health (WA) 2013). Consequences for breaching confidentiality range from increased monitoring by the human ethics committees, cancellation of the project, reporting to the researcher's employer and funding institution, and in extreme cases, reporting criminal conduct to police (Department of Health (WA) 2012).

5.6 Data access protocols are already in place — but progress has been slow

Despite the numerous barriers discussed above, public and private sector organisations have succeeded in using data without breach for many years. Data linkages have also been conducted — in the case of Western Australia, for over three decades — and used in academic research and policy evaluations (appendix D). This shows that some of the factors that impede data availability and use can be overcome in certain circumstances, by using contemporary frameworks and technologies, as well as persistent effort.

Addressing security and privacy concerns

Organisations in the public and private sectors can manage security and privacy risks while increasing access to identifiable data and complying with the legislation. This can be done in various ways, for example: designing data management systems that are based on privacy and security; implementing trusted user models, which only allow certain individuals to access the data; and using advanced techniques that allow data to be analysed without compromising privacy. And where datasets are not capable of strong re-design (for example, very small samples), data may not be accessed. There is no case for universal release, but rather, for comprehensive review. A number of these models for secure access have already been implemented in Australia, and consideration should be given to expanding their operation.

Data management systems that promote privacy and security

As protecting individuals' privacy has become a priority for data custodians, new privacy-oriented data management practices have emerged. Privacy be Design is one of the most common of these practices, and it has been adopted by many public and private sector organisations, including the Victorian Government (CPDC (Vic) 2014).

Privacy by Design is based on a proactive and preventative approach to privacy, aiming to prevent privacy breaches from occurring rather than taking remedial steps after a breach. This approach is embedded into the design of information systems as well as business practices and processes, so that privacy and security are maintained from the point of data collection, until the data is archived or destroyed (CPDC (Vic) 2014). In effect:

> Privacy by design enables public sector policy makers, information technology professionals and those responsible for delivering services to the community to approach privacy as a 'design feature' of public sector processes and activities rather than as a compliance burden to be endured or to which lip-service is given (CPDC (Vic) 2014, p. 1).

The OAIC has issued guidance supporting Privacy by Design, as a way for entities to comply with their legal requirements under the Privacy Act (OAIC 2016d). Additional

support for entities may be required in order to promote broader implementation of this approach.

Trusted user models

Much of the legal responsibility for keeping data safe is placed on the data custodian; however, data users also have a role to play in maintaining privacy and security of identifiable information. This can be achieved by using trusted user or trusted access models that:

> … would enable approved researchers access to sensitive and linked public datasets using safe and agreed platforms, enabled by mutual commitment and support from participating institutions and public agencies.
>
> … A trusted access model would require shared principles and clear responsibilities and accountabilities on the parts of government bodies and researchers. Its design should aim to achieve a balance of flexible, dynamic and efficient accessibility for data users along with adequate protections for data owners and data subjects. (University of Melbourne, sub. 148, p. 23)

Trusted user models are already used by a number of organisations in Australia. For example, the ABS uses a principles-based approach to improve access to data, while maintaining security and privacy. The Five Safes Principles framework, which has already been implemented in ten government agencies using ABS data, examines different aspects of access in order to determine the safest way to release the data (chapter 3).

Technological developments support the development of various safe settings, where researchers can analyse identifiable information. Funded by governments and academic institutions, the Sax Institute operates SURE, the Secure Unified Research Environment. SURE facilitates remote secure access and analysis of large and sensitive datasets (Sax Institute, sub. 56). The use of SURE can overcome many risks to unauthorised release of identifiable data.

However, it is the only facility of its kind in Australia, and given increased demand for data access, its capacity may be exceeded (SSCH 2015). To give effect to the open by default intent, augmentation of capacity is surely needed.

Privacy preserving techniques

Many data linkage projects use the 'separation principle', where no individual working with the data can see both the identifying information and the underlying data contained in the dataset (for example, medical or other private information) (NSS 2016a). Researchers have stated that many aspects of the data linkage process can help to preserve privacy, and limit the risk to individuals (Data Linkage Branch, sub. 13, SSCH 2016).

Evolving technology has introduced a range of new possibilities for identifiable data to be used for analysis, without compromising privacy and security. For example, the CSIRO (sub. 161) has done extensive work in the area of confidential computing, which allows for sensitive data to be analysed without the underlying information being revealed.

Investing in further data analysis and linkages that use innovative privacy preserving techniques may enhance the protection of individuals' privacy, while allowing better use of data.

Strengthening transparency and individual control over data

As the collection and reuse of data and analytics increases, '[i]ndividuals are becoming more transparent to organisations, but it is not clear that there is a parallel advance in the transparency of the data-processing practices of organisations to individuals' (OECD 2015a, p. 220). Data breach notifications are one way in which transparency around privacy protection can be increased (OAIC 2014b). As noted earlier, such breaches are generally not created by increased access such as release of de-identified data sets. But there is no doubt that breaches are persisting.

Australia has no system of mandatory notifications in case of data breaches. The Australian Government is currently preparing to introduce to Parliament legislation that will require organisations to notify the OAIC, as well as affected individuals, when they experience a 'serious data breach'. Such a breach would involve personal information being subject to unauthorised access that puts the individual at real risk of serious harm. The purpose of the notification is to allow the individual to take any possible steps to avoid harm, such as changing their passwords or cancelling credit cards (AGD 2015).

In the EU, mandatory data breach notifications have been part of a broader reform to data protection rules. The reform also formalised a 'right to be forgotten', which gives individuals the right to request that references to historical information about them be deleted (European Commission 2016b). In practice, EU citizens are now able to request that internet search engines such as Google remove links to information about them that is inaccurate or irrelevant; the search engine must consider the request in light of the public interest in continuing to access the information (European Commission 2016a). Since the first court ruling introduced the right to be forgotten in May 2014, Google has received over 566 000 requests to remove links, and accepted 43% of them (Google 2016). From 2018, individuals will also be able to request information collected about them be deleted by data custodians, although some exceptions apply, such as data used for research purposes (chapter 8).

In 2014, the ALRC considered the option of providing Australians with an enforceable right to remove certain information about themselves. It argued that the 'right to be forgotten' that was introduced in the EU was not the appropriate approach in Australia, given that the Australian Privacy Principles (APPs) require entities to destroy or de-identify information in some cases. However, it flagged the possible need for a specific

APP to deal with this issue, as 'existing APPs do not require an entity to provide a simple mechanism allowing an individual to request the destruction or de-identification of personal information' (ALRC 2014a, p. 321).[40] Chapter 8 further discusses the right to be forgotten and its possible application in Australia.

5.7 A comprehensive policy approach would be better to tackle the challenges

Governments and other organisations in many countries are realising, and harnessing, the power of the identifiable data that they hold to make better decisions and offer better products. In Australia, both the public and private sector accept that data should be more accessible and used — but a range of barriers have caused progress in data use to be relatively slow. Nonetheless, progress has been made. Over the years, governments have put in place a range of policies that have enabled some use of their identifiable datasets. Guidelines have been developed to allow secure access to data, in ways that minimise privacy risks.

The main obstacle organisations continue to face in increasing the use of identifiable data is collection-specific legislation (e.g. in health or education), and in particular the interpretation of clauses in the legislation that were intended to allow data sharing — but are very rarely used in that way, by both the public and private sectors. To overcome the disincentives created by the legislation, Australian Government agencies are now encouraged to adopt a 'reasonable interpretation' of their legislation (DPMC 2016). While this may support progress, impediments spread across literally hundreds of Acts also need to be addressed. In chapters 8 and 9, proposals on how to do so effectively will be examined. A structural approach is clearly needed.

A further challenge for the Australian Government, stemming in part from the overly complex legislative framework that applies to data, is the lack of a coordinated approach to policy development. In effect:

> Commonwealth data sharing initiatives are currently undertaken on a largely ad hoc basis between individual agencies. This can be unnecessarily complicated and time consuming (Department of Social Services, sub. 10, p. 6).

While in recent times there have been a number of declarations and initiatives from the Department of Prime Minister and Cabinet, encouraging all government agencies to

[40] In its discussion paper on *Serious Invasions of Privacy in the Digital Era*, the ALRC (2014a) recommended the introduction of a new APP that will allow individuals to ask organisations that information about them is deleted or de-identified. However, this recommendation was not included in the final report from the ALRC, as it was opposed by the OAIC based on the fact that the APPs already require personal information to be destroyed or de-identified when no longer required. The ALRC accepted that more consideration needs to be given to this topic (ALRC 2014b).

improve the use of their data, these are yet to be fully implemented (see, for example, Department of Industry, Innovation and Science, sub. 69).

In the course of this Inquiry, we have come across numerous examples of such initiatives. Two of them are the Australian Government Linked Data Working Group (sub. 46), which is creating tools for better data integration, and AusGoal, which is developing licensing frameworks for open access to government data (AusGOAL 2011) — both are very small groups of motivated individuals that are making progress within their own, self-determined area of work. The Queensland Government (sub 207, p. 9) warns that:

> There are pockets of sharing excellence embedded within agencies, however without a mechanism to make these more visible agencies will potentially develop solutions that already exist elsewhere.

The Department of Prime Minister and Cabinet is working towards greater coordination between the various organisations working in this space (DPMC 2015). However, in order to achieve large scale reform, governments must implement broad, systemic changes, coupled with strong leadership and dedicated funding. Some of these changes are already occurring in some states — for example, the NSW Data Analytics Centre and Victoria's forthcoming data agency (Andrews 2016; DFSI (NSW) 2016). Chapter 9 presents the Commission's recommended approach to improving data availability and use across jurisdictions, with particular emphasis on the role of the Australian Government.

Community involvement and communication will become increasingly vital in maintaining the trust of individuals in governments and other institutions as access to data increases. Chapter 10 discusses ways to maintain trust and incentives, as part of the implementation of the Commission's recommended data framework.

6 Making data useful

Key points

- Data availability is only one half of the equation — the ease with and extent to which data can be used, transformed, stored, and reused to achieve specified goals is also a challenge.
- Barriers to data access and sharing are compounded by poor quality data, and even some widely available or shared data is of very low usability.
- This can be the case for both public and private sector data, and irrespective of whether the data is generated for administrative, research or other purposes.
- Points of inefficiency exist throughout the life cycle of many datasets, adversely affecting quality and usability, and limiting the uses to which the data can be put.
  - The initial collection of the data can involve duplication and fragmentation between government departments, levels of government, and the research sector, which affects data quality and creates barriers to sharing the data between entities.
  - Insufficient or inconsistent adherence to standards for data storage and curation, including the application of metadata, mean that it is difficult to compare and/or integrate datasets, especially when data is not stored in a machine readable format.
  - The continued use of incompatible or legacy IT systems limits the range of available data formats and hinders data transfers and the application of standards.
- These issues are costly to resolve, often requiring the purchase of new IT systems and/or bandwidth, considerable time and labour in standardising data to be fit-for-purpose, and ongoing training for staff.
  - Entities will not be able to make use of even the most well-curated and standardised datasets unless users have adequate data management and analysis capabilities.
  - The extent to which these skills are sought and used is affected by organisational culture.
- Strengthening standards under which consumers can access their data and redirect it to others (whether held in public or private collections) is likely to contribute to the efficiency of markets. In the private sector, some limited government intervention may be necessary to ensure the interoperability of such standards across competing firms.

Data in Australia is collected, stored and managed in systems of vastly different quality. It can be accessible without being usable or useful, irrespective of what type of data it is or who holds it. That said, private sector businesses tend to resolve data quality issues when motivated to share data with their counterparts, and most of the unresolved quality problems relate primarily to:

- public sector (government and research body) data; or
- private sector data available to consumers or governments.

In these areas, inconsistent data management practices make it difficult to find and use data, and limit the value that can be extracted from data collections (box 6.1).

Box 6.1 **Participants' views on public sector data quality**

Numerous submissions to this Inquiry, from a wide range of stakeholders, presented examples of the technical difficulties they encountered in finding and using public sector data:

- It is still common for government datasets to require heavy management by human processes to locate them, evaluate the reason for accessing them, select and then transfer the data. Once received by another party, the data typically is stored again, and reformatted to suit individual software tools — with this process often occurring multiple times over. Further, if data is managed by a narrowly purposed community, then the data availability is narrowed (captured) by the needs of that particular community. And lastly, even if such data is provided online, the data is commonly provided either through human interaction via portals which are designed for a specific set of use-cases, or an old-style "data download" and end-user data wrangling. (National Computational Infrastructure, sub. 189, p. 9)
- [T]he quality of the public data itself is variable and datasets cannot always be used to help build an evidence base … Without data dictionaries and an independent assessment of the quality of the data itself, requests may be submitted for data that then proves to be worthless because the data is incomplete or incorrect. (Brotherhood of St Laurence, sub. 186, p. 5)
- There is no national agreement regarding the definition of disability, meaning that key state-based administrative data sets are routinely not accurate or nationally comparable. National reports such as the annual Child Protection [report] continue to fail to report on the overrepresentation of children with disability in these systems, despite being required to do so by the National Framework for Protecting Australia's Children. The Australian Institute of Health and Welfare has suggested that this was because each state functioned with a different definition of disability, and so the data was not commensurate. (People with Disability, sub. 203, p. 5)
- Prior to 2016, data on school attendance for both Indigenous and non-Indigenous students was published for each jurisdiction and a national figure was not available given differences in definitions across jurisdictions. … A large range of regional data is already published by state and territory and Commonwealth entities; however, there would be value in making this data more accessible and findable. In addition, ideally regional data would be available against agreed geographies such as the Australian Bureau of Statistics SA1 and SA2. Postcodes have not been designed to facilitate sound regional data analysis. (Department of Prime Minister and Cabinet, sub. 20, p. 31)

Making data useful and usable — for both its initial purposes, and for further use by other parties — can raise the value of data, increase efficiency and facilitate collaboration between entities. This chapter discusses the factors that affect data usability, from the way it is collected, managed and stored, to the technological and capability challenges that affect data custodians and users.

6.1 What makes data useful?

Usability refers to the extent to which a dataset can be employed to achieve specified goals and the ease with which data can be transformed and reused. To improve usability, data

should be machine readable and managed using open standards (or commonly accepted definitions and methodologies — see below) (DFSI (NSW) 2016).

Usability is heavily dependent on data quality, but the measurement of data quality is a complex proposition. The ABS's Data Quality Framework, for example, factors in seven different characteristics of a high-quality dataset: institutional environment, relevance, timeliness, accuracy, coherence, interpretability and accessibility. Some of these characteristics are considered in more detail below.

Accuracy can be influenced by many factors

Accuracy refers to the degree to which a dataset correctly describes the phenomenon it was designed to measure (ABS 2009), and can be affected by factors such as measurement, transcription, sampling, rounding, and empty cells caused by lost or missing data.

The level of accuracy required may differ depending on the way data is used, and the intended purpose of a dataset can affect accuracy. For example, NBN Co said that it had encountered several hundred inaccuracies in the G-NAF address dataset repeatedly over the past four years. According to NBN Co, inconsistencies between the geocoded addresses in G-NAF and the actual geographical locations of houses and apartments contributed to considerable delays and cost blowouts in the National Broadband Network schedule (Colley 2016; Hutchinson 2012). However, PSMA Australia, the government-owned company that produces G-NAF, argued that the dataset was not constructed for the purpose of detailed public service delivery and that it was not sold to NBN Co on that basis (Sharwood 2012).

Timeliness is important, but relative

The accuracy of information about people, businesses, infrastructure and the environment (and, consequently, a dataset's usefulness) is contingent upon timeframes. However, 'timeliness' is a relative quality, affected by factors such as the rate of change of the phenomenon being measured, the frequency of measurement, and the immediacy of the response that users want to make based on the information (ABS 2009).

Poor timeliness in public sector data affects the extent to which an agency can use data operationally (that is, for carrying out policy and regulation), as well as in evaluation and policy development. For some aggregate datasets, the amount of labour required to ensure the data is cleaned and standardised may present a major barrier to achieving rapid release. As discussed in more detail below, the amount of work required by the Australian Institute of Health and Welfare (AIHW) to standardise health data can cause considerable delay in the release of aggregate figures. For datasets that contain personal information, the time taken to de-identify data in accordance with the Privacy Act can also add to delays in data sharing. Dynamic de-identification systems, such as that used by ABS's TableBuilder, can help overcome such delays.

Coherence and interpretability affect data reuse and comparisons

Coherence refers to the internal consistency of a statistical collection, product or release, as well as its comparability with other sources of information, within a broad analytical framework and over time. Therefore, standard definitions and units of measurement are necessary to achieve coherence. *Interpretability* refers to the availability of information to help provide insight into a dataset. Information that could assist interpretation may include the variables used, the availability of metadata, concepts, classifications, and measures of accuracy (ABS 2009). Here, too, adherence to standards can maximise data usability and the potential for dataset linkage.

Consistent use of data standards supports coherence and interpretability, and, in turn, enables accurate analysis of data (section 6.3). However, much of Australia's public sector data and metadata does not use clear, consistent standards within or between jurisdictions (CSIRO, sub. 161). This shortcoming has also been observed in much of Australia's research data output (Australian Research Council, sub. 22). The importance of standards and machine readability are well-observed by data custodians such as the Department of Social Services (sub. 10), the Australian Government Environmental Information Advisory Group (sub. 32) and the Department of Prime Minister and Cabinet (sub. 20). And as discussed later in this chapter, a range of government initiatives designed to achieve greater data standardisation have been put in place over the past few years (section 6.3).

6.2 Data collection: a fragmented picture with too much overlap

Duplication and inconsistencies in data collection

Data collection in Australia is highly fragmented. A lack of data sharing between the public, private and academic sectors, and across and within jurisdictions, results in duplication and inconsistency in data collection, not only for surveys but also for the compliance reporting responsibilities of individuals, businesses and nonprofit organisations (see, for example, Victorian Alcohol and Drug Association (sub. 91)). This results in inconsistent data collections or multiple collections of the same data.

The Australian Statistics Advisory Council (ASAC) cites duplication as a major cause of suboptimal use of data:

> ASAC has long noted a tendency for government agencies to address gaps in data for decision-making by introducing new, ad hoc collections for a single purpose without reference to existing data or data collection processes, or the potential to frame the response to enhance broader statistical analysis. …
>
> [This] demonstrates a lack of coordination and collaboration across governments resulting in inefficiency. In 2009, the Australian Government established the Statistical Clearing House to be the mandatory central clearance point for business surveys that are run, funded, or conducted

> on behalf of the Australian Government, to reduce the load on business and improve government coordination. … However there is still no equivalent authority over-seeing household collections, nor are state and territory governments covered under this mandate. …
>
> Disaggregated efforts to develop new data limit potential for broader statistical purposes such as statistical data integration, and lead to the suboptimal use of scarce technical expertise that exists … for developing statistical solutions to inform complex policy issues. (sub. 25, p. 2)

For example, Leading Age Services Australia has indicated that significant duplicated data collection takes place in the regulation of the aged care sector.

> Feedback received by LASA members suggests that there are a number of inefficiencies in the reporting requirements of providers, including significant duplication of requested data. … LASA is also concerned that instead of seeking to utilise existing data, government departments may introduce new reporting requirements for service providers. (sub. 47, pp. 1–2)

These concerns are not limited to any one sector. The lack of coordination of data collection, management and publication standards across jurisdictions can lead to different measurements, making it difficult to aggregate data at a national level, or to share and link data across jurisdictions. Similarly, comparisons between jurisdictions become less effective — or even impossible — when datasets are incongruent (SCRGSP 2016).

Low data discoverability is an important factor behind duplication in data collection. When agencies are unaware of their own, or others', data holdings, the likelihood of inefficient data collections increases (chapter 3).

Survey fatigue

Poor coordination of data collection, leading to the repetition of surveys and compliance-based reporting, places a larger regulatory burden on individuals, services, suppliers and businesses (box 6.2). In some cases, this may result in those parties submitting poor-quality survey data to save time and effort (ABS 2016b) and can also waste public funds:

> [Duplication of surveys] places unnecessary burden on business, households and the broader community by duplicating effort in collection and wasting government resources. There are numerous examples of agencies allocating large sums to the conduct of household surveys when ABS collects similar information already, or in discussion could modify future collections to meet any perceived need. (Australian Statistics Advisory Council, sub. 25, p. 2)

Box 6.2 Case study: data collection in Australian agriculture

Agricultural statistics are important for policy planning and delivery. However, current public sector agricultural data collection and management systems in Australia appear problematic. CSIRO pointed to the variation in measurements that occurs as a result of overlap:

> There is considerable duplication in the collection of data from industry on farm performance, practices and attitudes. The ABS, ABARES, Rural Research and Development Corporations (RDCs) and other organisations … collect similar information at varying degrees of spatial and temporal granularity. There is an obvious opportunity for harmonisation, efficiencies, reduction in respondent burden and consistency in data collection. (CSIRO, sub. 161, p. 37)

The Commission's (2016c) Inquiry into agriculture also received submissions to this effect:

- 'The [National Agricultural Statistics Review] found that survey burden on agricultural businesses has been adversely affecting the agricultural statistical system through its impact on data quality. Farmers who are disengaged with the system due to survey burden are less likely to return survey forms or provide incomplete or less accurate responses, reducing the quality of information provided. This, in turn, requires greater effort by agencies such as the ABS to collect, check and validate survey data, affecting the timeliness and usefulness of the survey outputs. This can then lead to an increase in survey activity when organisations initiate additional surveys due to timeliness and quality concerns with the existing data sources, further exacerbating the survey burden on agricultural businesses. In this way the problem of respondent burden can become a self-perpetuating cycle.' (ABS 2016b, p. 2)
- 'A common complaint for farmers is the inability of governments to share information internally, and across jurisdictional boundaries. Even within agencies, farmers and industry representative bodies have to provide the same data numerous times to various bureaucratic silos. Farmers are always looking to ensure the data they collect in their business is done in an efficient manner and only collected when it serves a valuable purpose, and they expect government agencies to do the same.' (GrainGrowers 2016, p. 11)

These issues are not new, and they have been highlighted in previous reports (Murray Irrigation 2014; Retailer and Supplier Roundtable Ltd 2014). Almost a decade ago, the Commission recommended that improved coordination between Government agencies in collecting farm data could reduce the time spent by farmers completing surveys (PC 2007). Limited, if any, reform followed. The National Agricultural Statistics Review carried out over 2013 and 2014 by the ABS and the Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES) concluded that fragmentation, duplication and a lack of coordination and clear governance arrangements continue to be the key problems to be addressed in improving Australia's agricultural statistics system (ABS 2015).

Minimising the duplication of data collection is an important step in improving data quality. Much greater collaboration between governments at all levels is needed. ASAC points to greater consultation with the ABS as a method of achieving this:

> … before any major new survey or administrative data developments are approved, all levels of government should be encouraged to consult with the ABS in its role as Australia's central statistical authority. Advice should be sought on what data may already exist; what data is needed; how that data may be collected; and how the data may be developed with an eye to broader statistical and research purposes. ABS advice should also be sought on the technical and quality aspects of the data under development, and the burden that could be avoided by leveraging existing data or existing infrastructure … (sub. 25, pp. 9–10)

While the ABS has made attempts to promote coordination and consistency in data collections across government, only limited progress has been made. Stakeholders were reportedly frustrated not only at this lack of progress, but also with the significant limitations that the ABS places on access to its data:

> [M]any stakeholders experience frustration at their inability to gain access to microdata, to enable further statistical analysis. Access is denied appropriately to ensure confidentiality, but the potential value of further analysis is sometimes lost. The ABS has been seeking ways to protect confidentiality while still enabling the access stakeholders require. Until this dilemma is resolved, customer frustration will continue, and opportunities will be missed …
>
> The appetite for cohesion within the broader system has been reenergised … however, the ABS recognises that historically little progress has been made. Similarly, stakeholders report frustration around this lack of progress. The problems the [National Statistical Service, where the ABS facilitates collaborative use of statistics among government agencies] faces today are not dissimilar to those outlined in the original review undertaken some 30 years ago. (APSC 2013, pp. 25–26)

Overall, inefficiencies in data collection result in substantial duplication of effort and wasted funds, across both the public and private sectors.

DRAFT FINDING 6.1

The lack of public release and data sharing between government entities has contributed to fragmentation and duplication of data collection activities. This not only wastes public and private sector resources but also places a larger than necessary reporting burden on individuals and organisations.

6.3 Data management: standards turn data into a useful asset

Data standardisation involves presenting and managing data in a consistent, documented manner. Data standards can be broadly categorised as governing:

- *data* — standards are applied to the content of a dataset
- *metadata* — standards are applied to the context of the data — that is, the information describing a dataset's characteristics (National Statistical Service nd). Structural metadata describes the definitions used in a dataset (for example, through a data dictionary), while content metadata includes the context of the data (such as date, time and place of collection), its provenance, and how it was generated, including any analysis or calculations performed (appendix B).

The use of standards improves the curation and usability of data, and can facilitate data sharing and linkage. Standards not only make the linkage process technically easier, but also provide contextual information crucial to data analysis (AIHW 2007). How much

context needs to be captured will depend on the intended use of the data. The more contextual information is available, the more questions the data can answer, but the cost of standardisation will also rise as more information is captured.

Apart from supporting data usability, appropriate standards and data management foster trust in data and may increase its value (National Computational Infrastructure, sub. 189). For example, the Cooperative Research Centre for Spatial Information argues that understanding data provenance allows potential data users to make better decisions:

> Provenance understanding provides a measure of trust to the data being accessed in providing evidence in decision making processes. This is critical to ensure that policies put into practice are not just supported by data, but the understanding of how that data was generated is also known. (Cooperative Research Centre for Spatial Information, sub. 43, p. 2)

The work of the Australian Institute of Health and Welfare (AIHW) provides an example of how standards can be developed and implemented, as well as the costs and benefits involved. The standards developed by the AIHW are used across the health system, and enable data to be shared efficiently (box 6.3).

Box 6.3 Putting standards into practice

The Australian Institute of Health and Welfare (AIHW) develops and maintains the National Health Data Dictionary and the Metadata Online Registry (METeOR) in order to ensure the Australia's health data is managed with consistent definitions and metadata (AIHW 2007). The AIHW uses international standards (such as ISO/IEC 11179, which defines data dictionaries and is widely used in major data collection in other countries) to underpin its work. In essence, the AIHW's standards require that:

- each data element (variable) must describe one concept only
- each data element must have a unique name and definition, and the use of abbreviations or acronyms other than those widely accepted must be avoided
- the accompanying metadata is also subject to standards and must include the variable name, definition, data type and context (AIHW 2007, 2010).

These requirements are applied to datasets from nine different jurisdictions that include information about numerous medical conditions and patient characteristics. Changes and updates to the agreed standards involve consultations with all relevant stakeholders (AIHW 2007). AIHW also encounters issues of data formats that affect dataset interoperability, and has developed a system to address this by enabling custodians and source agencies to develop data submissions in consistent, interoperable, machine readable formats (sub. 162).

Implementing and maintaining data standards requires ongoing investment. However, the AIHW has warned that:

> There is a cost associated with creating data standards, but the cost of not creating data standards is likely to be even higher. This includes the loss of information that occurs due to staff changes, data redundancy, data conflicts, liability, misapplications and decisions based upon poorly documented data. These costs should be factored into the data development budget. (2007, p. 17)

Having a central institution that develops and maintains standards makes the health sector unique in Australia's data landscape.[41] Most other areas — whether public, private or research sector — do not have similar bodies, and therefore each data custodian can choose to employ different definitions and measurements, as well as different ways to manage data. This is a recipe for incoherence.

The consequences of substandard adoption of standards

Inconsistencies in data standards and management have significant implications for usability because they limit users' ability to compare, aggregate and link data.

For example, in the case of education data, there are wide variations in reporting protocols for both the initial capture and final reporting of data (Department of Education (NT) 2016). Substantial variations in aspects of education delivery (such as school starting ages, which differ from jurisdiction to jurisdiction, or the definition of commonly used terms such as 'disadvantage') further contribute to data comparability issues (NSW Government 2016; Tasmanian Government 2016). This lack of coordination between states and territories makes it difficult to create national aggregate datasets — in fact, aggregations of identical datasets by different public sector bodies can result in different final figures depending on the business rules surrounding both the initial data capture and final reporting of data (Department of Education (NT) 2016). The Commission's draft report on the national education evidence base also saw jurisdictional measurement consistency issues raised in the areas of school attendance data, teacher education data and teaching workforce data (PC 2016b).

Submissions to this Inquiry highlighted similar experiences across different types of data and sectors, from property titles to natural hazards (CoreLogic Asia Pacific, sub. 102; CSIRO, sub. 161). The Department of Infrastructure and Regional Development's experience in developing its regional yearbook provides a case in point (box 6.4).

Inconsistent data standards, or a lack of adherence to standards, also exacerbate problems in the construction of longitudinal datasets (achieved by linking multiple collections of the same data over different points in time) since measurement practices can change over time, even for the same data collection, in the absence of standardisation. Given the high demand for, and potential value of, longitudinal datasets (chapter 2; DPMC 2015) this can be highly detrimental to the outcomes of data use. Inconsistent use of standards is also an issue with private sector data. Whether voluntarily agreed-upon, or mandated as for Standard Business Reporting (SBR)[42], varied use of standards contributes to difficulties

41 See appendix D for a more detailed treatment of nationwide health data management.

42 SBR is a compulsory program for business-to-government compliance reporting that attaches to existing accounting software. One of its major functions is to provide a standardised data dictionary for the terms commonly used by businesses in their various financial reports to government. The Australian Reporting Dictionary substantially reduced the duplication and variation of data collections across different forms and reports. Overall, more than 33 500 data elements were consolidated into about 6600 terms (ABR 2014).

with comparing or aggregating data. These difficulties may be encountered both by private sector parties wishing to share data among themselves, and by public sector agencies receiving data from the private sector.

Box 6.4 Case study: standard geography classification

The Department of Infrastructure and Regional Development (DIRD) publishes the Progress in Australian Regions Yearbook to help benchmark how our regions are progressing against social, economic, environmental and governance indicators. In its submission to this Inquiry, DIRD detailed some of the data standardisation issues it has encountered:

> The development of this publication highlighted limitations in the availability of regional data. Some indicators included in the publication were only available with a limited level of geographic detail, or were available with various customised geographic classifications that did not allow for easy comparison.
>
> In developing the Yearbook, regional researchers were faced with the challenge of trying to integrate data sets that are based on different geographic classifications. While the ABS tends to apply a consistent geographic classification (the Australian Statistical Geography Standard), other government agencies and research bodies use a variety of geographies that vary according to the way that information is collected. This means that different data sets are not directly comparable, and poses challenges for making inferences about how different statistical indicators relate to one another. Examples include:
>
> - NAPLAN data on the literacy and numeracy of Australian school students is published on a geographic scale that classifies school locations into Metropolitan, Provincial, Remote and Very Remote areas. This is similar, but not directly comparable to the ABS Remoteness Areas.
> - Tourism Research Australia provides data on tourism demand and supply according to the boundaries of tourism regions. This is a custom geography designed to reflect tourism markets, but which is not directly comparable to the more widely used ABS classification system.
> - Health system data is published according to Medicare Local Service Delivery Areas, another custom geography based on the catchment areas of local Medicare providers.
> - The Australian Electoral Commission publishes electorate statistics, including enrolment and voting distribution, according to electoral division boundaries.
>
> Datasets are usually compiled with a geography that is well suited to their immediate purpose or the way source data is arranged, but this approach is often poorly suited to wider comparison and analysis. With access to greater detail (such as street address), it may be possible to reverse-engineer datasets to a different geographical classification; however a more efficient solution would be for agencies to collaborate on a single classification to be used across the board.

Source: Department of Infrastructure and Regional Development (sub. 201, p. 6).

Why are standards often not widely adopted?

For some entities, implementing consistent data and metadata standards comes at a high cost, not only in terms of labour and technology but also the amount of expertise required (Department of Employment, sub. 18; Bureau of Meteorology, sub. 198). In the public sector, departments and agencies typically have limited resources to devote to improving their data curation and it is often not highly prioritised. The AIHW (sub. 162) argued that costs could be mitigated if data integrators and research users were not required to destroy datasets at the completion of projects, as is currently required. Datasets are standardised before linkages are performed, but this investment in data curation is routinely lost.

Enforcing uniform standards can also be a costly and complicated exercise in the private sector, given the levels of specialisation and variation between individual businesses. At worst, mandating standards without a process of consultation and inclusion across sectors may not achieve the desired outcome and can waste time, skills and funds (CSIRO, sub. 161). However, the operation of the SBR initiative indicates such standards *can* be implemented across an industry in some cases (ABR 2015).

Standardisation and open data

When deciding to openly release data, agencies make decisions about whether to invest in processing data so that it complies with standards. If it is decided that data will be processed, agencies must also decide what level of standardisation they aim to achieve.

The key argument in favour of standardising data before publication is that well-formatted data is much easier to use, and therefore it is more likely that its release will generate benefits in the community from reuse. On the other hand, there are substantial time and cost savings possible from the release of raw data.[43] The Department of Employment noted:

> Resources involved in preparing our data for reuse and sharing is a challenge for the department. The challenge is in the time it takes staff to prepare, clean and de-identify the data for external use. For example, one of our key data holdings, employment services data, is contained in a transactional database, which was not constructed for research purposes. We are currently developing a prototype data dictionary for this data, to be implemented going forward. However this particular solution will not be retrospectively applied to the system due to prohibitive costs. Over time, as the system is refreshed, we can apply the data dictionary, and we estimate this project may take around five years. (sub. 18, p. 5)

From the users' perspective, there are benefits in receiving the original dataset, rather than one that has been reformatted or transformed, as processing may compromise the validity of the data (Janssen and Kronenburg 2012).

Typically, if data is of sufficient value, private companies and not-for-profit organisations will process it to suit their purposes. Therefore, researchers have suggested that public sector data should be published 'as is', with others to 'play a role in cleaning up the data and making it re-usable. This community can then in its turn also 'give back' and play a role in the standardisation of the open data' (Janssen and Kronenburg 2012, p. 16). These issues are discussed further in chapter 7.

The adoption of open standards in the initial collection and management of government data can remove the need to process datasets before publication (as the standards are freely available to users who interpret the data — see below). The Open Government Partnership,

[43] The timeframes to standardise existing data can be very long — for example, the European Commission's INSPIRE project, which will create an integrated data set of all European spatial information, will take over 12 years to complete (European Commission nd).

which Australia joined in late 2015, works to promote the use of open standards by governments involved in open data initiatives.[44] Progress in this area seems slow. In 2015, the Partnership's open data working group found many inconsistencies in the adoption of open standards by government agencies globally (McKinney, Guidoin and Marczak 2015).

Too much of a good thing?

Stakeholders have suggested that, while generally beneficial for open data, standardisation can also have unintended effects. For example, it can further delay access to data:

> The experience of Landgate … over thirty years is that whilst standardised approaches to the collection of data have proven important in linking data from various sources, an overly heavy focus on standardisation in the sharing and release of public sector data can delay the release of potentially useful data. For example, the use of complex metadata standards has often meant that data publishers wait until the data is 'perfect' before agreeing to release the information. (SPUR powered by Landgate, sub. 67, p. 3)

The AIHW notes that different levels of standardisation require different investments of time, skills and funds, and suggests that the costs of *some* standardisation may outweigh the potential benefits:

> … [M]any aspects of collection, sharing and release of data should be standardised; however, to do this fully for all public sector data is neither cost effective nor necessary. The costs involved with standardising all collections to the highest level (for example, linkage enabled, geospatially enabled, complete and standardised metadata, sharing enabled for a data system, transfer enabled and cleaned, standardised data items) would be prohibitive.
>
> … Standardising data to enable the highest quality linkage and analysis should be reserved for the highest value datasets as this will incur the highest costs. On a fit-for-purpose basis, lower value datasets may not need this level of standardisation and could, for example, be auto-standardised using semantic and machine learning techniques. (sub. 162, p. 18)

Mandating particular standards for data also has the risk of reducing its usefulness. The Bureau of Meteorology states that it is important not to apply a one-size-fits-all methodology, for several reasons — including sector specialisation, technological requirements, storage space and cost.

> The Bureau is acutely aware of the current format for ubiquitous data exchange being XML as the content is self-described with delivery. [XML] also provides the ability to dynamically link between data sets. The freedom that XML promises is a trade-off with storage and transfer cost through the extremely verbose nature of the language. The verbosity problem [increases] with the addition of complex scientific content. The Bureau recommends that a single format to

[44] The Open Government Partnership was established in 2011 by Brazil, Indonesia, the Philippines, Mexico, Norway, South Africa, the United Kingdom and the United States. It currently has 70 member countries, which each create a national action plan to advance transparency, accountability, participation and innovation. Open data initiatives are significant in these plans. The Australian Government is in the consultation phase of developing its first national action plan (Australian Government 2015; OGP 2015).

share data should not be mandated: a same size fits all approach would be counterproductive. Instead, a suite of accepted data formats should be espoused. (sub. 198, pp. 20–21)

The Commission considers that the need for standards to be relevant to their specific subject area is a legitimate challenge. A one-size-fits-all solution can be difficult to implement and may be counterproductive to data usability.

Work on standards is progressing … slowly

Several whole-of-government policies have been developed to encourage the adoption and implementation of data standards across the public sector, including:

- the 2006 Australian Government Information Interoperability Framework, which promoted common data standards to enable data sharing across government agencies (AGIMO 2006);
- the 2009 *National Standards Framework for Government*, a guide for endorsing and managing standards;
- the 2011 updated Australian Government Architecture Reference Models, which contain a detailed guide to standardisation; and
- the Digital Continuity 2020 Policy, implemented by the National Archives of Australia (NAA 2015; sub. 114) which promotes consistent information governance across the Australian Government, including minimum data and metadata standards.

While the implementation of the Digital Continuity Policy is ongoing, other initiatives have made little progress and standards of public sector data appear poor, particularly in aspects such as metadata application and data cataloguing (chapter 3). The Queensland Government suggests that open standards 'seem to be struggling for acceptance in some areas of government' (sub. 209, p. 9). According to the Department of Social Services, the financial cost is a major reason behind some government entities' minimal investments into developing capabilities for sophisticated data management and release (sub. 10).[45]

If Australia is to improve the usability of its public sector data, the application of consistent data and metadata standards will become increasingly important. However, it must be noted that if most or all of this standardisation happens before the data is ever shared or released, this imposes more upfront costs on the initial data custodian, against the backdrop of unlikely, or significantly delayed, returns on that investment (chapter 7).

Some parts of the private sector are also working to voluntarily promote data collection and management standardisation. For instance, the Australian Farm Institute (2016) recently recommended industry-wide commitment to a nationwide open data alliance, the adoption of standards for data collection, metadata and privacy, and interoperability of farm data management software to maximise data transferability.

[45] Chapter 7 provides a more detailed discussion of approaches to cost recovery.

Options to improve the implementation of standards

When seeking to standardise its data and metadata, an organisation can use one of a number of approaches:

- define its own standard;
- choose to use an accepted standard already developed by one of the many standard bodies, such as Standards Australia; or
- participate in a community of users (both data custodians and data users) to reach agreement on the best way to define, structure and manage data.
  - The work of the AIHW, which brings together a range of stakeholders in the health sector to define common terminology for the datasets it manages, is an example of this process. While this approach helps in creating data that is comparable across jurisdictions, it is also complex and time consuming.

Technological advances can make a substantial contribution to standardisation. For example, modern techniques such as machine learning provide opportunities for computers to assist the process of understanding data holdings and how they relate to the existing standards. Semantic and machine learning can be used to distil common meaning from different data definitions, and generate standards (AIHW, sub. 162). Such standards will still need to be validated by experts in the relevant field.

Public sector

The policy push to increase data sharing and release relies on the assumption that the data held by the public sector is of sufficient quality to deliver insights. For much data, this requires some level of adherence to data standards.

The Australian Government's open data portal, data.gov.au, contains guidance for agencies publishing open data, and various state governments (NSW for example) have developed guidelines for the implementation of standards. But more needs to be done; according to the Department of Industry, Innovation and Science (sub. 69, p. 2), the Australian Government should put in place a strategy to 'establish enforceable standards for data use in the public sector, preferably across all levels of government'.

The application and enforcement of standards to datasets could take place:

- In a centralised manner within particular fields of collection (as jurisdictional health agencies do via AIHW). A central release authority would receive datasets from various public sector agencies in their initial formats and classifications, and would then transform and harmonise those datasets for integrating, comparing, linking and release. This approach is used by Integrated Data Infrastructure in New Zealand.
- In a more devolved manner, whereby agencies would be expected to transform their own datasets before sharing or release. A central agency of some kind would still be

needed to facilitate communication between agencies and advise on or adjudicate differences in standards application. This is similar to the approach developing in NSW, where the Data Analytics Centre (DAC, discussed in more detail in chapter 3) works to coordinate consistent data management definitions and standards in collaboration within and across NSW public sector agencies. At the same time, the DAC has the authority to compel agencies to release datasets to the DAC for its own transformation and analysis (NSW Government 2015; Pearce 2016).

In its Public Data Policy Statement, the Australian Government declared that it will make high-value datasets available using high quality standards (Turnbull 2015). Progress towards this goal has been limited so far. Currently, the National Archives of Australia (sub. 114, p. 2) provides 'digital information management policy, standards and products to improve interoperability and enable discoverability of information within government'. The Commission considers that the implementation of standards should be made a matter of priority, to support increased availability and use of data. In this context, high-level approaches by central-agency entities such as the National Archives may not be sufficient.

DRAFT RECOMMENDATION 6.1

Government agencies should adopt and implement data management standards to support increased data availability and use as part of their implementation of the Australian Government's Public Data Policy Statement.

These standards should:

- be published on agency websites
- be adopted in consultation with data users and draw on existing standards where feasible
- recognise sector-specific differences in data collection and use
- support the sharing of data across Australian governments and agencies
- enable all digitally collected data and metadata to be available in commonly used machine readable formats (that are relevant to the function or field in which the data was collected or will likely be most commonly used), including where relevant and authorised, for machine to machine interaction.

Policy documents outlining the standards and how they will be implemented should be available in draft form for consultation by the end of 2017, with standards implemented by the end of 2020.

Agencies that do not adopt agreed sector-specific standards would be noted as not fully implementing the Australian Government's Public Data Policy and would be required to work under a nominated Accredited Release Authority (Draft Recommendation 9.6) to improve the quality of their data holdings.

Private sector

As in the public sector, standards can play an important role in enabling the productive sharing and transfer of data between parties in the private sector. There are several means to encourage the expanded use of standards in the private sector, involving different levels of government intervention.

Mandated standards

One option is for governments to mandate data-sharing standards (FinTech Australia, sub. 182). At the broadest level, the Commission has previously noted that this approach can be too heavy-handed:

> As many standards are developed by non-government organisations, or evolve as de facto standards, the role for government often requires a light touch aimed at facilitating new standard setting and adoption. … [S]tandards should be the minimum necessary to achieve the regulatory objectives, should be outcome focused and not overly complex or prescriptive. … [S]tandards should be subject to regular review to ensure that they remain relevant and are not unnecessarily impeding the adoption of new technologies. (PC 2016a, pp. 104–105)

Governments may struggle to build or retain the expertise necessary to mandate standards that are commercially relevant to different businesses and sectors. They also need to be mindful of the risks that mandating standards can have on competition. The Commission has previously pointed out these risks:

> Instead of facilitating market exchange, standards … can unduly inhibit competition. There are incentives for dominant enterprises to restrict competition by working to have standards favour their product against other producers. This is particularly a risk with highly prescriptive standards, as they can favour or exclude particular technologies and thus favour or disadvantage particular companies. … That said, where a dominant firm imposes its own standard, and uses it anti-competitively, the government may need to implement policies aimed at increasing competition. (PC 2006, pp. 14–17)

Thus the argument is not against standards per se, but about their public policy purpose. Standards that ensure market participants can overcome information asymmetries and behave efficiently in sharing data are becoming a matter of public policy interest. There may be a potential role for governments in mandating data standards for the purpose of improving consumers' access to data collected about themselves (this issue is discussed in more detail in chapter 8).

The midata experience in the United Kingdom points to a potential role for government in this area (chapter 4). An impact assessment for the midata program identified several factors that could disincentivise firms from voluntarily developing standards that assist consumers in accessing their own data:

- *Competitive disadvantage*: an individual firm is unlikely to make consumers' transactional data available if it believes it will create a competitive disadvantage that its competitors can exploit, even if firms in a sector may generally be willing to do so.
- *Positive spillovers*: a midata-style standard may facilitate competition and new entrants with new business models, and the individual incumbent firms are unlikely to value these positive spillovers.
- *Inefficiency/cost*: The application of matching standards may be costly where firms have to accommodate large numbers of different data formats and insufficient progress is made to generate a critical mass of any one format.

The impact assessment concluded that, while governments could have a role in mandating the *existence* of standards, the extent of its involvement in determining specific details should be limited:

> It is important that any Government intervention does the minimum required to ensure a consistent standard of data being released while not preventing innovation around the exact format the data is available in. For this reason the Government is proposing to set a standard that data should be released in an electronic machine readable format and not to specify any other standards. (DBIS (UK) 2012)

Improving the accessibility and usability of consumers' own data is a desirable objective. The question is how best to develop standards that may support such an objective, without creating disincentives to collect data or other unintended consequences.

Private sector determined standards

An alternative to mandated standards, which is consistent with the midata experience discussed above, is to require that standards be developed, but to leave the actual development work to the parties directly involved — consumers and data collectors, and their advisers. Private participants in markets will often, over time, settle on a particular standard, particularly once use of a standard has achieved a critical mass. Such developments are often underpinned by voluntary, cooperative efforts on the part of industry participants. An example is the World Wide Web Consortium (W3C), in which an international community of member organisations, full-time staff and the public work together to develop the standards that enable the operation of the Internet.

The development of private sector standards may take considerable time. For example, three years after midata commenced, a review of the program found that no agreed standards had yet been developed for the data made available to consumers. This inhibited the ability of consumers to determine which financial product would be most beneficial for their needs (DBIS (UK) 2014).

As outlined above, there are several reasons why this may occur. Chief among them are the costs associated with upgrading systems and modifying existing practices, and the belief

that the development of standards would provide an advantage to commercial competitors (such that there is no benefit from being the first to adopt a standard).

Government facilitates development of private sector standards

A model that may be considered a template for government facilitation of the development of private sector standards is the Australian Communications and Media Authority's (ACMA) role in the development and enforcement of industry codes, set out in the *Telecommunications Act 1997* (Cth).

In this process, industry bodies and associations are able to develop codes on matters relating to telecommunications activities, and those codes may be registered by ACMA. If registered, ACMA can direct an industry member to comply with the code in question, thereby ensuring the adoption of codes that have been developed on a voluntary basis (ACMA 2015b). ACMA may also create industry standards itself in certain situations, such as: when a request to develop an industry code is not complied with; where no industry body or association exists to address the issues in question; or where an industry code has been developed but has been deemed to be deficient (ACMA 2015a).

If this model were adopted and circumstances arose where an industry regulator was required to develop the industry standards, that regulator should take steps to ensure a consultative process of standard development, such as:

- convening industry expert groups for input
- circulating proposed standards to stakeholders and interested parties for feedback
- consulting with the relevant industry Ombudsman or the Commonwealth Ombudsman.

The advantage of this approach is that it would provide the private sector with incentives to develop standards suitable for their industries, with scope for customisation to a fine degree. However, it would be necessary to ensure that standards were not designed by incumbent industry members as de facto protection, with a view to restricting the future entry of competitor organisations.

DRAFT RECOMMENDATION 6.2

The private sector is likely to be best placed to determine sector-specific standards for its data sharing between firms, where required by reforms proposed under the new data Framework.

In the event that voluntary approaches to determining standards and data quality do not emerge or adequately enable data access and transfer (including where sought by consumers), governments should facilitate this, when deemed to be in the public interest to do so.

The benefits of providing data via Application Programming Interfaces (APIs)

An API is a structured set of functions and procedures that allows applications to access the features or data of an operating system, application, or other program (Christensson 2016). APIs enable two or more pieces of software to communicate with each other without the need for manual data transfer, or the need to align the structure of the data used internally by the organisations sharing the data. It has been suggested to the Commission that requiring data to be provided in an API format could substantially improve data portability (see also appendix B for more detail on how APIs operate, and appendix E for discussion of the role of APIs in financial data and open banking).

Through APIs, information can be transferred automatically (for example, moving from a consumer phone bill to a comparison website) following permission from the individual/consumer, rather than requiring the consumer to obtain the data themselves. Similar methods of providing data already exist in the Australian banking industry, where 'feeds' of transaction data are provided to users (such as finance brokers) for a fee.

The Centre for International Finance and Regulation (sub. 9, p. 23) supported regulatory intervention in favour of APIs:

> … the use of standardized APIs for consumer access and download of transaction data would facilitate the development of entire new categories of consumer advisory services. … Previous attempts to introduce APIs such as the Open Financial Exchange (OFX) standard failed due to active resistance from industry incumbents. In our view, such behaviour is short-sighted since the incumbents are well-placed to develop their own advisory services, and have been for some time. In view of this behaviour, it seems likely that a regulatory enforced solution will prove necessary.

In the United Kingdom, an investigation into retail banking by the Competition and Markets Authority (CMA) noted a lack of automation in midata file transfers — such that users had to manually share their consumption data with third parties — and concluded that this limited the program's effectiveness and contributed to sluggish uptake (2016). Consequently the CMA ordered the largest UK banks to adopt common API standards so consumers could more easily share their transaction data with other financial institutions.

While a business-to-business API implementation of data transfer would allow for a smoother consumer experience, it would also impose greater costs on businesses, particularly for data security and governance. The compliance costs of an API implementation may be justified where the transfer of data offers significant benefits to consumers and is used by many. The difficulty of introducing an API will vary significantly across data holders — many potential costs have been considered, but realistic costs are difficult to pinpoint (CMA 2016).

Ultimately, decisions regarding the implementation of APIs for data transfers may be best assessed on an industry basis and by industry standard-setting groups with the presumption favouring the use of APIs. It appears that provision of data in an API format most closely

achieves the objective of allowing consumers to derive value from their data, and most closely satisfies the criteria of improving the availability and usefulness of consumer data.

INFORMATION REQUEST

The Commission seeks more information on the benefits and costs of a legislative presumption in favour of providing data in an application programming interface (API) format, specifically:

- *In which sectors would consumers benefit from being able to access data in an API format?*
- *What are the main costs and barriers to implementing APIs?*

Standardisation and curation in the research sector

At present, unless research datasets are published on an open access portal (for instance, under the requirements of a particular journal publication), they remain held by the researcher and indexed by the Australian National Data Service (ANDS) under the Research Data Australia portal (appendix B). Participants in this Inquiry have noted a number of issues with this approach (Centre for Policy Development, sub. 11; Council of Australian University Librarians, sub. 97) and have suggested options to improve reuse of these datasets, including:

- regular updating and curation of research data;
- development of standards for research data management; and
- principles that decide what to store and how it should be funded.

Chapters 3 and 8 discuss the public interest in making research data open. Chapter 3 also recommends that the Australian Research Council (ARC) should publish up-to-date registers of its data holdings and summary descriptions of the datasets held. Research data curation and standardisation methods can help make such registers more accurate and quicker to construct, and make open research data more discoverable and usable.

How should research datasets be curated?

Along the lines of the centralised ARC register for publicly funded research data, one model could involve a central body being made responsible for curating research data to make it easier to reuse. This could involve standardising, indexing and updating it on an ongoing basis. Differing degrees of centralisation could see this model based around a single data curation authority, analogous to the ANDS's role as the central research data index, or a more dispersed model with broadly discipline-specific authorities. This model could see central data custodians also made responsible for the curation of the datasets they hold, as suggested by the National Aboriginal Community Controlled Health Organisation for the discipline of Indigenous health (sub. 192, p. 4).

Such a model would have all the advantages of a central repository (such as ease of dataset identification, simple management of storage, and consistency in indexing). It seems impossible, however, that a single such body could cover the arrays of datasets that might be accumulated across the research community, particularly if active curation is involved. Moreover, there is the danger that a body such as this might lack the specialised expertise in the data that the original researcher has and, as a result, may apply inappropriate transformations to the data. While the concept of centralisation may be attractive for ease of use, at this point the primary needs are discoverability and wider availability.

Funding researchers and/or universities directly to curate their own data, meanwhile, would have the benefit of building on existing arrangements, but may encounter a lack of incentives for researchers to keep their datasets or metadata up-to-date and standardised.

Requiring (or at least enabling) that research datasets 'cleaned' or otherwise improved through better curation or transformation, be provided back to the original data holder once a project is completed, would seem the most practical option. The original data holder, presented with potentially improved versions of its dataset (or code that allows improved versions), could then decide whether to retain all versions in full or to adopt only a single complete version. Retention in full might require ongoing curation, but the originating collector is probably best placed to determine the costs and benefits of such a decision.

Standardisation for research data

Standardisation of the format of research data is undoubtedly desirable — particularly if reuse of the data is to be encouraged — but difficult to address on an aggregate basis. Even more so than for public sector data, the specialisation and complexity of some research datasets means that it is not practicable to impose a one-size-fits-all approach.

If the aforementioned central data curation body were to be created, one option would be to make that body also responsible for developing standards for storage. Another option would be to mandate that universities store their research data in a particular format. As discussed above, these options would need to be accompanied by funding to address associated data storage and curation costs.

An alternative option would be to promote greater use of open access platforms for research data storage by academics. Several types of these platforms exist, including:

- institution-specific (such as the University of Western Sydney's Research Data Repository Project);
- cross-institutional (such as eResearch South Australia, a collaboration between the University of Adelaide, Flinders University and the University of South Australia);
- discipline-specific — typically managed by a consortium of institutional members (such as the Australian Data Archive, for social science data, or the Aboriginal and Torres Strait Islander Data Archive);

- national, publicly funded (such as the National Computational Infrastructure's research data storage service, or the ongoing Research Data Storage project); and
- international (such as the Open Science Data Cloud, a resource of the Open Commons Consortium, which is funded by both public and private sector members). (ANDS nd)

Such platforms generally have their own standards for storage; for example, eResearch South Australia's data storage service carries some (unspecified) formatting requirements, as well as requirements for the creation of collection-level metadata records.

This approach would have the added benefit of not requiring academics to bear the cost of data storage. However, as discussed above, work would need to be done to coordinate interoperability between the various publishing platforms' standards.

A final option is to convene government and academic working groups to develop sector-specific standards. While this approach would have the benefit of sector-specific knowledge, it would not reduce the plethora of standards that already exist.

What to keep and what not to keep

The enormous datasets generated by some research — such as the Australian Square Kilometre Array Pathfinder telescope, which currently generates around 100 terabytes of raw data per second (Poloni 2014) — can be very expensive to store. Because of the costs of storage, these very large datasets are often used and deleted relatively quickly. Deciding what to store means assessing the possible value of data before it has been used. Since this may be a difficult task for an individual research team, bodies such as NCRIS and the Open Access Working Group are likely to provide the most useful guidance in this area (see also chapter 8 for more discussion of the decisions around which research data should be made open).

One type of data that may not need to be completely opened up, or shared in full, is data generated by modelling simulations. The University of New South Wales (UNSW) notes that given the vast amounts of simulation data that exist in certain research areas, and the fact that the data can be recreated using the source data and the model or code, archiving of this data can be a waste of storage space (and funding). UNSW proposes that for such data, it might be preferable to archive, or publish as open data, the means to recreate the data (the model/code and source data) rather than the dataset itself (sub. 50).

More generally, participants in this Inquiry expressed concerns about the effect that reforms to research data information systems and retention requirements would have on university budgets (University of Tasmania, sub. 196). We recognise that curation and storage of research data is costly, and chapter 7 discusses options for funding and cost recovery.

6.4 Technological challenges and opportunities

Technological aspects of data management are crucial in improving data usability. Issues such as the system and format in which data is stored can heavily influence the extent to which data can be reused and the degree to which metadata is retained. The costs of data storage can also affect data retention decisions. Though these issues occur to some extent in both the public and private sectors, they typically have a greater impact in the public sector, due to difficulties in attracting skilled staff and the limited scope for investment in new IT systems (Department of Agriculture and Water Resources, sub. 37; Dun and Bradstreet, sub. 135).

Legacy IT systems and incompatibility impede data transfer and integration

For data to be most usable, it must be transferrable between different users and combinable with existing datasets held by those users if necessary. Legacy IT systems — essentially outdated or superseded systems that remain in use — and non-interoperable software can hinder transfers of this kind. The Australian National Audit Office (ANAO) describes the example of the Australian Childhood Immunisation Register (ACIR), where significant limitations on system interoperability were delaying data processing and necessitating manual input:

> Limited interoperability between [the Department of] Human Services' ICT systems (ACIR, MCD and ISIS) and external providers' practice management software makes it necessary for the department to supplement automated data exchange processes with daily manual data cleansing and matching activities. For instance, departmental operational reports of transactions between ACIR and MCD indicate that some 4900 records required manual resolution over a two month period.
>
> … [A] number of persistent data synchronisation errors were identified, arising primarily from limitations in the interoperability of the relevant departmental ICT systems. Further data synchronisation errors arise from known system-to-system issues, such as misalignment of the business rules between systems. [These errors] are managed through daily manual reviews undertaken by departmental staff — an essential but resource-intensive exercise which does not incorporate a quality control process.
>
> … [T]he process [of manual data validation] involves a significant additional workload for departmental staff … delays in this information being accepted into the Register may affect a child's immunisation status and a parent's eligibility for family assistance payments, where the delay coincides with key milestone and payment timeframes. (ANAO 2015a, pp. 20, 23, 73)

There are many other examples of continued use of decades-old 'legacy' software and hardware systems, and of system incompatibility, across the entire public sector (see, for example, ANAO 2015b; Department of Agriculture and Water Resources (sub. 37)). A review of the Australian Government's use of ICT suggested that public sector agencies may be spending too much of their ICT budgets on maintaining legacy systems (Gershon 2008). More recently, an audit of Victoria's financial systems IT assets found

that many Victorian public sector agencies had made limited or no progress in upgrading from end-of-life IT systems (VAGO 2015). And the Department of Industry, Innovation and Science (DIIS) has acknowledged that the age of some of its systems presents barriers to optimal data access and use:

> There is also the challenge of dealing with data that is contained in legacy systems. Even though these systems can be modernised, there are a number of technical challenges associated with the effective sharing of such data. These challenges include: an inability to link data tables; standardisation issues between data systems; gaps in metadata availability; and inconsistent storage formats. All of these can lead to data quality issues and difficulties in automating data provision. (sub. 69, p. 9)

However, with current Australian Government Public Data Policy requiring government entities to ensure all *new* systems support data discoverability, interoperability, and cost-effective access (Department of Agriculture and Water Resources, sub. 37), the replacement of systems over time should see technological issues posing less of a barrier to data access and use. For example, in 2015 the Commonwealth Government launched the Welfare Payment Infrastructure Transformation program, which plans for the Department of Human Services' (DHS) 30-year-old digital payments platform to be replaced over seven years with a system that can support greater automation and linkage between government entities (Cowan 2016; Nott 2016; Tudge 2016). An external capability review of DHS in 2012 found that existing legacy systems had significant costs in terms of customer compromise, elevated costs and reduced flexibility (APSC 2012).

Data management software procurement issues can also affect data usability in the private sector. For instance, along with a surge of interest in agricultural data access and analysis, there has been rapid growth in farm data capture and storage programs. Attempts to capture market share in this field have seen issues of program interoperability and proprietary data formats arise, where data generated by one type of software cannot be viewed in another type of software and cannot be aggregated automatically (Beef Central 2016a, 2016b; Pawsey 2015). Not only does this render data pooling or sharing between individual farmers difficult, it also presents complications for public sector bodies wishing to analyse the data for policy purposes. Appendix D contains details of how IT interoperability can present issues in both public and private healthcare.

Demand for data storage is increasing

Public sector agencies are collecting and processing increasingly large volumes of administrative information, enabled by digitisation. This has resulted in the need to store vast amounts of data. The 2014-15 Australian Government ICT Trends Report notes that the total amount of data stored by Australian Government agencies has increased by 259% in the five years from 2009-10 to 2014-15. This has created new challenges for data storage, and managing storage costs. While data storage is becoming cheaper, this is not consistently offsetting the expense associated with rapid growth in data holdings. The total

cost of storage in large and medium sized Australian public sector agencies increased 25% from 2009-10 to 2014-15 (figure 6.1).

Increasing use of cloud storage

Cloud storage involves data being stored and backed up on remote servers, which users access via the Internet, with physical server infrastructure typically owned and managed by a hosting company. Since cloud storage does not require physical proximity, it has the potential to lower costs and increase the effectiveness of data storage in the public sector.

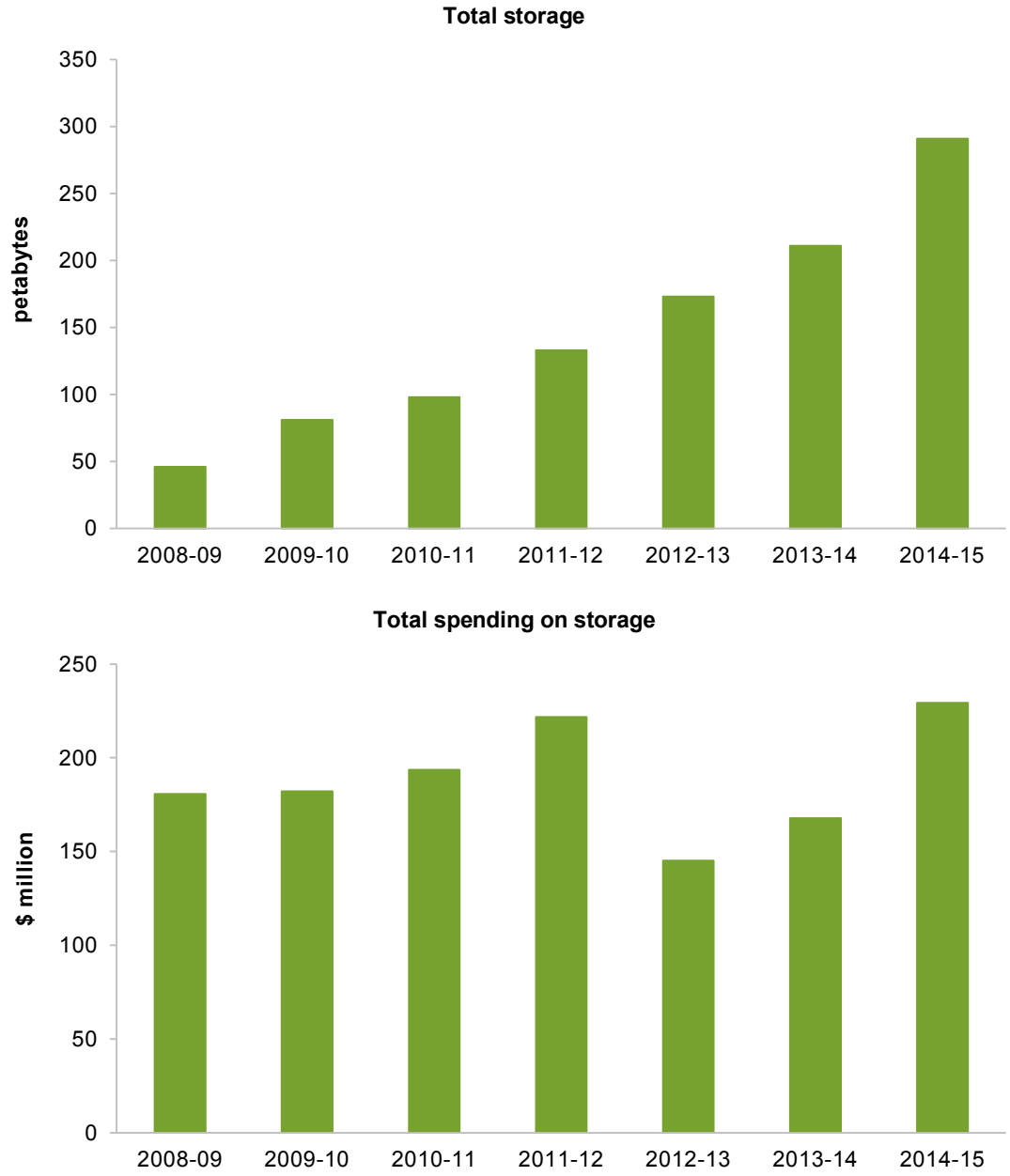
The Australian Government Cloud Computing Policy requires non-corporate government entities to use cloud services wherever they are fit for purpose, offer best value for money and provide adequate management of risk to information and ICT assets (Department of Finance 2016a). The Department of Finance noted in 2014 that the take up of cloud services had been sluggish. Since then, spending on cloud storage has accelerated, with quarterly spending in late 2015 more than three times that of mid-2014 (Department of Finance 2016b).

This accelerated use could be beneficial for public sector entities in light of the various costs that data management changes may impose, but cloud storage providers' security protocols need to be monitored closely, particularly given that many large providers have servers located in several different countries.

Moreover, agencies are still faced with considerable uncertainty about cloud storage, particularly with regard to the storage of personal information on the cloud:

> There will need to be increased support for, and advocacy of, standard permission schemes, such as the Australian Signals Directorate (ASD) Certification which awards certification to a range of cloud service providers for specified cloud services. This could help departments understand and mitigate security risks, and create a more flexible and technology adaptive data sharing environment for sensitive data. (DIIS, sub 69, p. 8)

Figure 6.1 **Public sector data storage**



Source: The 2014-15 Australian Government ICT Trends Report (Department of Finance 2016a).

Uncertainty around storage security requirements can impede data use

The need to maintain the security of data, and prevent unauthorised access to identifiable information, is an important consideration for both public and private datasets. The community expects personal information to be handled securely, and surveys have shown that many individuals will avoid dealing with organisations that do not emphasise data security or have been affected by data breaches (OAIC 2013).

While data security measures can pose a barrier to increased use of information, protecting data — particularly identifiable information — is vital to maintain trust in data collections and the organisations that manage them. And costs in other forms can also be avoided by keeping data secure: in 2015, Australian companies that had to deal with a data breach incurred average costs of $2.64 million (Ponemon Institute 2016).[46]

There is no overall policy framework for data security in Australia. Australian Government agencies are required to comply with the information security measures included in the Protective Security Policy Framework and the Information Security Manual (ASD 2016; Attorney-General's Department 2016). Separate requirements apply to state and territory government agencies (appendix C).

Entities covered by the Privacy Act are required to take 'active measures' to protect personal information from misuse, loss and unauthorised access (OAIC 2014). However, this requirement does not apply to parts of the private sector (for example, some small businesses are not covered by the Privacy Act). Moreover, given that holes in security measures can be very difficult to detect and preventative audits are not required by legislation in any part of the private sector, inadequate data security is generally only discovered after a breach, even when the most advanced data security technology is used.

The substantial direct and indirect costs of breaches — including a loss of trust — and community expectations regarding data security create a strong incentive for organisations to protect the personal information they hold. However, as far as the public sector is concerned, there is limited guidance from government leadership on implementing an approach to data security that will balance the need for protection with appropriate access to data. The Department of Industry, Innovation and Science (sub. 69) has called for additional support for government agencies to help them mitigate security risks, in particular in the case of sharing identifiable information about businesses, where the Privacy Act does not apply and de-identification is not required.

6.5 Capability and resource constraints

Data skills are lacking

Maintaining and maximising a dataset's usefulness require specialised skills and an understanding of data's potential uses. However, these skills are in relatively short supply, and are more thinly spread across the public sector than the private sector (Dun and Bradstreet, sub. 135). Several participants to this Inquiry noted a lack of data capabilities in the public sector (box 6.5), which can have substantial consequences for both the quality and the beneficial use of data.

[46] This reflects the costs incurred by 26 Australian companies, from 11 industry sectors, that experienced loss or theft of personal information in 2015 (Ponemon Institute 2016).

Box 6.5 Participants' views on data capability in the public sector

- Decision makers want "tools" rather than raw data services, but often lack the specialist capabilities or staff to build these tools. In addition, the software industry has not been successfully incorporated into the open data journey, meaning that it is difficult to acquire the right skills to utilise complex data services. … [T]here is often a skills and knowledge gap between data release and data use which [limits the benefits] of releasing the data in the first place. (CSIRO, sub. 161, p. 15).
- [A]cross all levels of government there is a lack of capability to analyse data which results in not making full use of data that already exists. This includes both the technical data analytical skills of employees and the use of systems and processes located within agencies. Suggested strategies to enhance capability might include … deliberate recruitment of employees into government with data skills; agency examination of data stores and the analytical skill and resource required to maximise the use of data for a range of purposes; and a deliberate move by governments to work in partnership with others to utilise and develop skills … (Australian Statistical Advisory Council, sub. 25, pp. 3–4)
- The Australian Public Service … faces a shortage in employees with the requisite data skills and capability. The Public Sector Data Management Report identified that ready for work graduates with data capabilities are in short supply in the public service, and during consultations, most entities expressed a need for more data capabilities. (Department of Prime Minister and Cabinet, sub. 20, p. 26)
- The data resources in government agencies are often directed to reporting rather than analysis. As a result, data science and analytics are treated as a second-order priority. The knowledge and mechanisms required to extract, interrogate, manipulate, analyse, communicate and interpret data are not often present in public agencies. … Staff with relevant skills … may not be involved in the design of processes for data collection and release. These staff also may not have a strong understanding of key technologies such as APIs and web services that can automate the application of policies and processes for ongoing data release. (NSW Government, sub. 80, p. 9)

Skills development has been identified as a priority by governments seeking to increase the use of their data (see, for example, Department of Finance, Services and Innovation (NSW) 2015; Victorian Government 2016), and several governments have introduced initiatives to address this skills shortage. For example, the Tasmanian Government's 'Stats Matter' strategy, in operation since 2013, includes as a key facet the development of a statistical capability framework (sub. 205).

Similarly, in August 2016 the Australian Government released a strategy document for data skills and capability in the Australian Public Service that proposes both foundational data literacy training for all APS employees and specialised 'data fellowships' with Data61 (DPMC 2016). The Department of Prime Minister and Cabinet has also suggested that data analysts could become a shared resource across multiple government agencies (DPMC 2015).

These developments are promising. Coupled with an increasing focus on hiring employees with data skills and recognition of the need to promote 'a culture of valuing data' (Tasmanian Government, sub. 205), greater emphasis on development of data skills should

enhance the efficient and effective use and reuse of data in the public sector. Similarly, an increased level of community knowledge of the utility and importance of data science skills (Centre for International Finance and Regulation, sub. 9) should mean that necessary upskilling also continues in the private sector.

Improving data quality and usability can be costly

Transforming data to a standardised, structured, machine-readable state can come with high costs. For example, the ABS warns that 'the cost and effort required to develop new technology, and [required to perform] such mundane tasks as digitising or transferring existing information to a new common digital format, should not be underestimated' (ABS 2016a).

While acknowledging potential cost pressures, the Department of Prime Minister and Cabinet (2015) concluded that the upfront costs of migration to new standards 'should be outweighed by reduced development costs over time'. However, considering that most of the benefits of standardisation and other improvements to enhance data usability are unlikely to accrue to the government agencies responsible for them, the promise of future cost offsets may be underwhelming relative to the disincentive of immediate expenditures on hardware, software and labour.

Some level of dedicated funding to incentivise the wide and timely adoption of standards, and the provision of guidance around how agencies can introduce and implement standards efficiently, may need consideration. CSIRO (sub. 161) recommends that long-term funding models be developed for managing and distributing public data, including support for government agencies in making available standards-conformant data services and APIs, as well as ongoing funding for data.gov.au as a major component of public data infrastructure. Funding issues are discussed in more detail in chapter 7.

Everyday citizens need to be able to use data too

The benefits of improved sharing and release of public sector data to consumers (chapter 2) will not be realised if they need a degree in statistics to be able to use the data. CSIRO (sub. 161) notes that a gap exists in the ability of end users to consume many of the complex data services that are produced by data providers.

There are different ways that we can make access to data more beneficial for all Australians.

First, people should be able to easily establish what data is available, and where they can obtain it. To achieve this, available data must be catalogued and searchable in a central location. This does not necessarily require that all data be stored in a single database, but rather that the databases are interoperable (chapter 3).

Second, once the data is found, it needs to be usable. Not only should data be machine readable, it should be formatted and standardised in a way that enables it to be read by most domestic computing systems. Greater use of data visualisation tools could be useful. A central repository of practical experiences, similar to usability.gov in the United States (box 6.6), would provide a rich resource to help data custodians and analysts maximise the accessibility and usability of open data for consumers.

Box 6.6 **usability.gov**

The United States Government operates usability.gov as a resource for parties looking to develop or expand a digital 'user experience', improving a consumer's engagement with a website, system or application. It publishes guidelines, methods, tools and examples of best practices for providers of digital material in both the public and private sectors.

Much of the material on usability.gov relates to the 'user-centred design process'. In effect, this is a framework in which the needs, wants and limitations of the digital material's end users are given extensive attention throughout the design process. The aim is to allow users to find information quickly by presenting it in an accessible layout and design and using simple and consistent language, and to submit information equally quickly where it is required.

Making digital interactions more user-centred can have significant benefits for service delivery, information accuracy, and productivity — and can potentially save costs. Government agencies often realise immense returns on investment from usability improvements because they operate on a large scale — some government websites serve hundreds of thousands or even millions of users, and agency intranets may also be used by thousands. Usability improvements can have similarly large impacts on commercial and nonprofit organisations (Nielsen 2007).

Source: Department of Health and Human Services (US) (2013).

The Commission considers that central agencies should be responsible for promoting the discoverability and accessibility of data and developing standards to promote consistency and interoperability across jurisdictions, sectors and users — including consumers. Input from the private sector and from the states and territories may be helpful in achieving this goal (chapter 9).

Where open data relates to individuals, it is important that privacy and data security safeguards, including best practice de-identification techniques, are in place. How the Commission envisages management of these risks is discussed in chapter 5.

DRAFT FINDING 6.2

Data standards should aim to ensure that the content produced is usable by those who seek access to their own data. To achieve this, available data needs to be published in machine readable and commonly used formats that are relevant to the function or field in which the data was originally collected or will likely be most commonly used.

7 Value adding and pricing decisions

Key points

- Market opportunities for harvesting data and creating new business ventures are important determinants of the extent to which businesses, industry organisations and not-for-profits choose to add value to data, and/or share with other parties. Both of these activities have grown rapidly in the private sector over the last decade.
- Public sector datasets can also offer attractive opportunities for new and innovative private sector services, but government agencies have limited ability to determine the value of data they might release. Such valuations are important for assessing the likely level of returns (or value added) from additional dataset processing, such as tailoring datasets for specific uses.
- Processing of data that adds value and takes a dataset beyond the standard required internally and for other regulatory requirements, should only be undertaken by government agencies when:
  - there is a previously-unaddressed public interest purpose clearly identified by the agency, and accepted by the government, for the agency to undertake additional value adding and make the value-added data available; or
  - the agency can perform the value adding more efficiently than either users of the data or any private sector intermediaries; and potential users of the data have a demonstrable willingness to pay; and agencies have the capability in-house or under contract with a third party.
- Beyond this, government agencies should refrain from additional value adding to datasets because the delay incurred by agencies in doing so prior to release (or sharing) can be substantial, and data users in general appear to have a preference to access data 'as is' and in a timely manner.
- There are various approaches for pricing public sector data, ranging from free provision and marginal cost pricing to commercial pricing. The preferred approach will vary according to user demand and agency capability to act commercially.
  - For a given level of data quality, making data freely available will maximise use and hence deliver the highest level of social benefits (before taking costs into account). But it will increase the net cost to government of data release.
  - Where agencies undertake substantial value adding because it meets the principles above, there are strong grounds for passing these additional costs on to data users.
  - An exception may be where data is used for research in the public interest. Pricing of data for the publicly-funded research community should be the subject of a review.
- Maintaining and increasing the availability of public sector data is costly. Sharing or release of minimally processed datasets should be funded by agencies from existing budgets. For datasets that have significant and demonstrable public interest, additional funds should be provided to agencies to ensure the quality of dataset curation and release.

One of the vital enduring characteristics of data is that many parties can access it simultaneously and yet its value to all remains undiminished. The Internet creates opportunities to add value not only to newly generated data but also to existing data. For example, Uber's creative use of data has re-shaped urban transport, Airbnb's data analytics is transforming the accommodation market, and IBM Watson Health's cognitive computing is providing access to health data that was previously 'hidden'. These examples, and many more like them, reveal the extent to which the private sector has awoken to the value of data. Entrepreneurial forces, supported by price signals and the potential for profits, are driving innovative collection and use of data by businesses. The private sector will identify a purpose, value it strategically or financially (or both) and determine how best to utilise it. This valuation will guide business decisions on value adding (and pricing if a business decides to market some of its data holdings).

By contrast, the value of data tends to be less clear when it resides in government hands. Previous chapters have dealt with the value to public policy development and service delivery that can come from data sharing and integration (chapters 2, 3 and 5). This chapter focuses on the known and unknown value that can come from data release. Compared with the private sector, government agencies understandably find it hard to value the many disparate and potential purposes to which the data they hold, if released, might be put.

For public sector agencies, the purpose of release often reflects pressures on government to be more open, a desire to meet the demands of potential users to *assess* the dataset for potential use, or the advancement of social, environmental or growth objectives. However, any financial returns from these purposes are unlikely to end up with the agency incurring the cost of releasing the dataset. Thus 'the market' is not likely to offer much guidance to valuation, other than in very specialised circumstances. These valuation difficulties in turn confound value-adding decisions.

7.1 Value adding and sale of private sector data

Market opportunities for harvesting data and the possibilities of creating new business ventures are the substantial determinants of the quantity and quality of data that private sector organisations (businesses, industry organisations and not-for-profits) choose to collect, add value to, and share with other parties (either by sale or other mutually amenable arrangements).

Where data is shared with other parties, the value placed on it will be determined by: the existence of accessible alternatives (if it is easy to collect, there will be many substitutes and, other things being equal, its value will be low); the extent to which it has been or needs to be processed for use; potential uses to which the data can be put; and strategic leverage (the extent to which the business wants to maintain control over the data and its use). The constantly improving ability of smart technology to link to artificial intelligence (AI) and learning algorithms is likely to extend the value of even relatively well-analysed datasets in the future.

Value adding by businesses[47]

Value adding by businesses will be done for either internal purposes — to identify market trends, profile customers, and guide investment and pricing decisions (box 7.1) — or for external purposes — to share with or sell data to parties such as:

- suppliers — to enable suppliers meet specific input requirements, or promote goodwill and long term business relationships
- finance providers, advisors or managers — to enable business financing, accounts management, or legal or financial administration
- business partners — to maintain business partnerships or meet the requirements of a shared parent company
- customers — to build goodwill, for promotional purposes or to comply with regulatory requirements
- industry organisations — to enable industry development, benchmarking, promotion, or to inform and influence governments
- government agencies — such as the Australian Bureau of Statistics, the Australian Taxation Office, Australian Securities and Investments Commission, State Revenue Offices, or a local government — to meet regulatory or other reporting or licensing requirements.

Box 7.1 Examples of businesses adding value to data

Businesses add value to data to:

- make the data more attractive to potential purchasers — for example, the Australian Stock Exchange curates the transactions data it collects from the stockmarket
- improve the capacity to target customer preferences — for example, Amazon uses its customers' purchasing history to promote other products to them
- attract customers (promotional or goodwill purposes) — for example, Google incurred considerable expense in generating its extensive map data, which it provides free to most users
- meet standards specified for sharing with other parties — finance sector companies curate customer data in order to participate in the Comprehensive Credit Reporting scheme.

Some value adding may also be done with a view to selling data. This is a relatively recent development outside of specialised industries such as advertising. In future, by virtue of big data analytics, most firms will be able to engage in active management of their data assets. The question of value then becomes central: how much will additional value adding

[47] For convenience, this chapter will refer to private sector organisations as 'businesses'. The reader may assume that, at least with regard to data release, similar incentives and constraints apply to businesses and not-for-profit organisations and so the terms can be used interchangeably in this chapter.

cost and how much it is likely to return? While the public sector faces the same conundrum, private sector data holders will generally have stronger incentives and capacity to capture the value of data assets, and be less likely to be required to satisfy public interest objectives.

Role of government in value adding decisions of businesses

For most businesses, government will have little impact on decisions about the type and extent of value adding they perform on data collected, purchased or otherwise acquired.

However, businesses that collect data because of — or with the benefit of — regulated market access or receive significant amounts of public funding, are in a different position (chapter 4). In particular, the requirements on businesses to share or release data may vary over time with changes in public interest standards and regulatory requirements.

The costs of providing data to comply with regulatory requirements will often involve investment in data management (see, for example, ANZ, sub. 64). Such costs are either absorbed (that is, margins may shrink) or passed on to consumers in the form of higher prices, depending on the strategic considerations of the business. Any increase in the obligations of businesses to share their data holdings with other parties will likewise incur costs:

> The cost to industry implementing changes to its existing systems to facilitate greater access to data needs to be carefully considered. System changes to facilitate such increases in data availability will likely result in increased costs to consumers. As such, any changes considered need to result in an increase in the ability of consumers to make informed choices because of the availability of data, outweighing the costs of implementing its availability. (Red Energy and Lumo Energy, sub. 63, p. 2)

Release of private sector information collected by governments

Of the private sector information that governments collect, only a portion is released for access to potential users outside of government. Deciding what privately sourced data should be released can involve a difficult weighing up of a range of factors:

- Would releasing the data cause commercial detriment to the businesses or individuals the data is about? That is, is the data restricted under contract, commercially sensitive and provided to government in confidence or, in the case of individuals, does it impact on their privacy?
- Is there a public interest case for releasing the data (such as transparency of the activities of businesses and performance of government regulators, spillover benefits for the community, or better informed market participants)?

In the event that there is a public interest case for releasing such data, questions arise as to how much processing or value adding should ideally be undertaken, and what price, if any,

should be charged to access the data. These questions are the subject of sections 7.3 and 7.4.

Market power

Governments are already active in requiring data where market power is evident, such as price monitoring in essential services after they have been privatised (for example, energy utilities and telecommunications) and monitoring of service delivery in the transport sector (for example, ports, rail and aviation).

With the advent of the digital age, some data intensive businesses — such as Amazon, Facebook and Google — have gained substantial market share in the sectors that they operate. Market power often proves to be transitory, at least when it is not entrenched by regulation. Moreover, the interoperability of technology and linked services that comes with such market share can be to the considerable benefit of consumers; and, should adverse circumstances arise, the *Competition and Consumer Act 2010* provides the Australian Government with the authority to address any misuses of market power.

There are also other influences at work that tend to moderate the scope to misuse data-related market power by businesses. As noted by Acquisti (2010), there is a delicate balance of power between businesses and their customers:

> In choosing the balance between sharing or hiding one's personal information (and in choosing the balance between exploiting or protecting individuals' data), individuals and organizations face complex, sometimes intangible, and often ambiguous trade-offs. Individuals want to protect the security of their data and avoid the misuse of information they pass to other entities. However, they also benefit from sharing with peers and third parties information that makes mutually satisfactory interactions possible. Organizations want to know more about the parties they interact with, tracking them across transactions. Yet, they do not want to alienate those parties with policies that may be deemed too invasive. (p. 3)

In the case of data-related market power, any attempt by governments to force businesses to share data may have unintended consequences. The Financial Systems Inquiry (Murray et al. 2014), for example, noted that disincentives for businesses to collect data (or, indeed, collect it within the reach of Australian law) could be induced.

General government intervention to deal with real or perceived market power in a period of rapid change is likely to have other unintended consequences. The difficulty in foreseeing the next wave of innovation and change means that the imposition of regulations directed at curbing today's apparent market power may have the effect of constraining future innovation, at substantial economic and social cost (PC 2016). These are fine judgments, to be made carefully.

Data pricing by the private sector

The data that a business collects will be of varying interest (or value) to other parties. Valuation of a dataset will be subject to a range of influences including:

- the level of demand (willingness to pay) for such data
- the strategic value to the business of maintaining exclusive control over the data (or otherwise restricting with whom and on what terms it shares the data)
- the cost of any value-adding required to get the data to a marketable standard.

Role of governments in the pricing of private sector data

While governments directly set or approve prices in some parts of the private sector (such as for essential services), there appear to be limited instances where a public interest case could be made for government involvement in the pricing of specific private datasets.

A better case may be made for governments to mandate that businesses share certain data with their customers. Such rights already exist to some degree, but are limited by the absence of minimum standards (for example, that the data be machine readable) and provision to review approaches to pricing for such data sharing. Without these, it is possible for businesses to either provide customer data in an essentially worthless form, or — hypothetically, at least — charge large fees to deter customers from making requests. We note that the *Privacy Act 1988* (Australian Privacy Principle 12) seeks to set some boundaries around the charges an organisation may impose for giving an individual access to personal information they have requested.

In considering such standards and pricing frameworks, the policy focus should be on identifying the public interest to be served and limiting intervention to the minimum necessary to achieving it.

An alternative, or complementary, approach would be for governments to provide incentives for businesses to share data of their own volition. The Financial System Inquiry suggested that the Australian Government could compensate businesses that release or share their data. While noting that the most efficient approach may be to charge users of data in order to fund this compensation, that Inquiry suggested in other cases businesses could be compensated through greater access to others' data (the Comprehensive Credit Reporting scheme is an example of such an approach). Alternatively, if the data is viewed as a public good, the government may be justified in directly compensating the data holders and providing open access to the data (Murray et al. 2014).

Such activity is not unknown today. The ABS purchases fuel price data from a private sector data reporting service for use in the calculation of the Consumer Price Index, although it does not release the data to consumers. Late last year, the market intelligence company Informed Sources Australia agreed to make retail fuel price information available to consumers free and to third party service providers on reasonable commercial terms — at the same time as it provides such information to fuel retailers — as part of an agreement

reached between it, major fuel retailers and the Australian Competition and Consumer Commission (ACCC).

But these are rare and tangential examples. This Inquiry prefers the clearer and well-precedented policy path — that is, where a demonstrable public interest is capable only of being served by release of data held in the private sector, that release should occur. Regulation to achieve this should impose the least practical cost and seek to maximise the public benefit consistent with that. Compensation is an unlikely choice to make in such circumstances, but the case can be made only once the concept is confirmed.

7.2 Value adding and sale of public sector data

As noted earlier, the public sector data holder faces a conundrum when it comes to sharing or releasing data — how can it serve an identifiable public interest, in preparing data for sharing or release, where the value to be gained by doing so is unlikely to: accrue to the agency holding the data; and be readily valued (to enable clear justification for additional processing effort and cost)?

A key aspect of this conundrum is to determine the ideal level of value adding to undertake to data before making it available. A correlated issue is whether an attempt should be made to price the data that an agency shares or releases.

- Should public sector data be freely available to users (implying that its production and maintenance is funded out of general government revenues)?
- Or should users be charged, implying that the production and maintenance of the data should be paid for, at least in part, by those (or some of those) who access and use the data, with the government contributing to such costs only to the extent that it is also a user of the data or where there are broader public benefits associated with releasing or sharing the data?

These decisions will be influenced by a range of incentives, disincentives, obligations and constraints faced by government agencies (chapters 3, 5 and 6) and by the impact that charging for data might have on achieving economic, social and environmental objectives (chapter 2).

However, the overriding aim should be to ensure that there is timely release of data (and that is in a usable format) and, conversely, that time and resources are not wasted on value adding that is not valued by the ultimate users.

‘Basic data’ and ‘enhanced data products’

A distinction can be made between:

- ‘basic data’ — to which an agency has undertaken a minimal level of curating or processing (that is presumed to be value adding) to make the data fit for sharing or release

- 'enhanced data products' — to which an agency has undertaken more extensive curating or processing effort (that is, again, presumed to be value adding).

Agencies will usually have to undertake some processing of data to make it fit for release. This processing will generally have the objective of making the data:

- machine readable
- readily linkable to other datasets
- understandable — that is, what the data is and where it has come from
- de-identified (to protect the identities of the data subjects in the case of data on individuals or organisations).

Such processing was discussed further in chapters 5 and 6.

To produce an enhanced data product, an agency will usually have to undertake more extensive value adding to a dataset (except in cases where the agency undertook such value adding to the data for its own internal purposes). Where an agency undertakes more extensive value adding, it will usually be with the intention of targeting the data at a specific set of uses or users (rather than just general release).

To value add or not to value add — what are the incentives and disincentives facing agencies?

As governments move — desirably — towards making data more accessible, pressure will rise on some agencies to not just share or release data but to incur expense in undertaking additional value adding to some of their data holdings in order to meet user interest — that is, to produce enhanced data products. Users may have unrealistic expectations of the current capability of agencies and of the funding available to undertake such value adding.

Similar pressures may arise on agencies to devote resources to converting older records to digital form (with the advent of advanced digitisation techniques and as public data becomes increasingly attractive to external users). The costs and benefits of digitisation of old data should be examined closely — if public interest benefits do not warrant it, such efforts should only be made where cost recovery or other revenue possibilities compensate for the costs. The National Library of Australia, for example, has a program to digitise its analogue collections to make them accessible to everyone and to ensure their long-term preservation. Part of this program includes a digitisation on demand service, some of which is charged for on a fee-for-service basis or undertaken in partnership with (or with funding from) external parties.

Even in the absence of such pressures, some agencies may be motivated to add additional value to the datasets they intend to release or share:

- to bolster the agency's standing by increasing the attractiveness of data to end users (even if users' needs are often more prosaic)

- to improve the marketability of the data, with a view to raising revenue, either for cost recovery purposes or to generate a 'commercial' return
- as a result of cultural drivers and historic legacy.

In some circumstances agencies will be required to add value to data. As noted in chapter 6, some level of processing is often essential prior to using data for internal analysis or for sharing with external parties. Common examples of requirements for agencies to undertake some value adding are:

- for national consistency or benchmarking
- to meet statutory or administrative requirements — such as requirements to fund a certain percentage of an agency's expenditure with external revenue and cost recovery obligations
- to fulfil the reporting aspects of a specific program or legislation
- for program and policy evaluation and regulatory monitoring and enforcement purposes
- to meet international reporting obligations/standards.

There is usually a strong public interest element behind each of these requirements.

For most agencies, data sharing and release (and the associated processing of the data) are not a 'core' activity (motor vehicle and land title registries are two notable exceptions) and so these activities are accorded low priority and an over-riding wariness (chapter 3). Acknowledging data sharing and release as essential ingredients for effective delivery of core activities necessitates a substantial cultural shift and the acquisition of additional capabilities. Lack of staff skills and the costs of technology clearly impede agencies' value-adding capabilities, as noted by the Bureau of Meteorology (sub. 198) and the Department of Prime Minister and Cabinet (sub. 20).

The pressure now being felt to meet implied government objectives for datasets to be released wherever possible may also act as a disincentive to value add. A benchmark of the number of datasets released is a very rough and possibly misleading guide to the openness of data — for instance, the result may be the release of less usable or valuable datasets.

For government business enterprises (GBEs), the incentives to value add to their data are broadly similar to private companies — that is, they will value add for 'internal' purposes, such as planning, identification of market trends and customer profiling. However, they may also seek to market some of the data they hold. For corporatised GBEs, the separation of social objectives from commercial interests allows for a practical approach to value adding:

- To achieve social objectives, value adding by GBEs should be undertaken to the minimum standard — that is, to provide a basic curated data product that is released in a timely manner and fit for purpose (accurate, machine readable) — and shared or released free of charge or at marginal cost. Such value adding should be funded by the government.

- Value adding for internal processes or for commercial (data sales) purposes should be undertaken on the basis of commercial considerations, taking into account any restrictions on pricing (discussed below).

How much value adding is enough?

In determining how much value adding to undertake on data before it is released, the primary focus for all government agencies should be to establish first the public interest objective they would be serving by releasing the data in question. This will then guide the decision on value adding — namely, whether to simply process the data to achieve a basic data product (fit for sharing or release) or undertake additional value adding to produce an enhanced data product.

There are good reasons for agencies to, in general, limit the amount of value adding they undertake beyond the basic level. The principal reason is that the private sector is generally better placed than the public sector to:

- identify what value adding would be sought by potential users (businesses tend to be more 'market oriented')
- undertake such value adding
- distribute the outputs of that value adding (businesses tend to be more motivated to find potential users of their products and services).

Stiglitz, Orszag and Orszag (2000) note that as a generality:

> The government should exercise increasing caution as it adds more and more value to raw data or information, or as it provides a more and more specialised service. (p. 59)

Advising more specifically, Stiglitz, Orszag and Orszag (2000) echo the point that private sector businesses (and not-for-profits and researchers) are generally considerably better than governments at using and, even, 'cleaning' data:

> [T]he government should provide search engines and "ferret" tools to assemble data, but more specialized tasks – such as "cleaning" databases or linking official information to related academic articles – should generally be left to non-governmental entities (including academic institutions, non-profit organizations, and private-sector firms). Such case- or individual-specific tasks have less of a public good nature than the underlying data. (p. 59)

The report of the Victorian Parliament's *Inquiry into Improving Access to Victorian Public Sector Information and Data* (Parliament of Victoria Economic Development and Infrastructure Committee 2009) noted:

> Throughout the Inquiry, witnesses expressed concern about governments behaving in a business capacity and competing with the private sector by selling value-added information products. All of the submissions that addressed this matter were of the view that value-adding should typically be the role of non-government organisations. The Cyberspace Law and Policy

> Centre (CLPC) recommended that the Victorian Government avoid policies that allow commercial returns for value-added information. (p. 107)

The Australian Spatial Information Business Association stated that:

> … government agencies must re-focus on the management of good quality and current spatial data sets and further encourage the private sector to invest in the value add and deployment to the broader community. (Parliament of Victoria Economic Development and Infrastructure Committee 2009, p. 107)

The Victorian Parliament's Inquiry made the following finding:

> Finding 20: There is growing recognition that government should have a limited role in adding value to public sector information (PSI) for commercial purposes. The value of PSI should be enhanced through private sector activity for the creation of new products and services. (p. 109)

The Centre for Policy Development (sub. 11) provided an example of misplaced value adding by government agencies (box 7.2).

Box 7.2 Misplaced investment in value adding to public sector data

The Centre for Policy Development (sub. 11) gave an example of well-intentioned — but ultimately misplaced — investment in value adding by government agencies.

> As [their] digital resources accrue, there is a strong temptation for public institutions to package them up in value-added resources that promote the benefits of their ongoing collections management work and potentially to raise revenue through fees for value-adding services.
>
> A good example of this kind of initiative is the set of "field guides" to the fauna in each state that have been developed as mobile apps, through a collaboration between various Australian museums (e.g. the NSW fauna field guide released by the Australian Museum). They package up information and multimedia resources in mobile apps that are made available … free to the general public. The collaborative approach whereby the same code base was used for all apps has been an innovative way to mitigate development costs. The apps also provide exposure for the museums.
>
> However, the field guides have still involved significant cost and involve a number of fundamental drawbacks. These include:
>
> - A lack of ongoing funding to ensure that the code base remains usable on new releases of mobile device operating systems;
> - A very limited number of species in each type (birds/reptiles/frogs/fish/invertebrates etc.), preventing the app from reliably performing its core role of assisting users to identify species they encounter; and
> - No ability to capture and upload records of users' encounters with the fauna documented in the apps.
>
> With the limited species coverage and limited functionality, the app-store reviews of these "field guides" is very mixed, with a significant number of users being negative about their experiences. This public feedback undermines the very objective of creating and releasing the apps in the first place.
>
> Given these limitations, the apps are little more than electronic coffee table books, providing a restrictive window into the Museum collections. They are destined for obsolescence. This can be seen today because the code bases for the iOS and Android versions have been made available as open source by the Museum of Victoria. Neither of these public code bases have been updated since their original release three years ago. (p. 17)

Source: Centre for Policy Development (sub. 11).

In contestable markets, such as the dissemination of weather forecasts, the policy challenge is to determine the efficient boundary between public sector value-adding activities and private sector value-adding activities. For example, there are sound public interest arguments for the Bureau of Meteorology to collect and provide geographically disaggregated basic weather data on a consistent basis across Australia, and to be the 'official voice' in times of weather emergencies. It also provides a number of other services that, in some other countries with larger markets, are provided by the private sector (box 7.3). The extent to which these activities continue to be provided by a government agency as markets grow should be re-evaluated from time to time by the agency, to ensure it is not displacing value-added services that could be more efficiently provided by the private sector.

Timeliness of data release is paramount

A further consideration is the trade-off between timeliness and value adding. The Australian Property Institute, while making the case that users are best placed to assess the quality of data, also noted the importance of timeliness of public data release, even if it comes at the expense of the quality of the data:

> … [T]he Institute (and SIBA) consider that users are the assessors as to whether data quality levels are fit for a specific purpose. It is recommended that publishing data (with a quality statement) ought to take priority over improving data quality … (sub. 169, p. 4)

The Commission concurs that public sector data should be provided in a timely manner and notes that agencies should take into account that timeliness will, very often, be of more value to users than relatively minor issues with data quality. That is, users will generally prefer data 'as is', released promptly and accompanied by a clear statement of the data's quality.

A question of principles

A set of policy principles is needed to guide agencies' decisions on value adding. These principles should include only taking further measures to adapt or improve datasets beyond that required by internal agency and other regulatory requirements (such as privacy legislations) when:

- there is a previously-unaddressed public interest purpose clearly identified by the agency, and accepted by the government, for the agency to undertake such additional value adding and subsequently make the value-added data available free of charge or at marginal cost; or
- the agency is best placed to perform the relevant functions; and potential users have a demonstrable willingness to pay for the value added product; and the activity does not involve an unreasonable diversion of agency resources; and the agency has the capability in-house (or under an existing contract with a third party) to perform the value adding; and the IT risk is assessed and found to be small.

Box 7.3 Value adding in the weather forecasting industry — Australia, the United States and the European Union

In the United States, private-sector weather forecasting is a multi-billion dollar industry — a 2006 survey estimated annual revenues of US$1.6 billion (Mandel and Noyes 2013) — with a 24-hour cable channel, hundreds of private enterprises (Stiglitz, Orszag and Orszag 2000) and a substantial weather derivatives market. A survey of clients of private weather forecasting businesses revealed that the three main factors driving customer demand were the accuracy of their forecasts, assistance in operationalising the forecasts and the availability of one-on-one consultation(Mandel and Noyes 2013).

In Australia, the Bureau of Meteorology (sub. 198) provides a range of data, information and services. Aside from free public good products, it provides:

- tailored information for specific industries, where it has added value through analysis and processing, and for which it charges prices with an incremental cost recovery element.
- commercial products, that are bespoke, to which it has undertaken considerable value-adding, and for which the Bureau charges (p. 11).

Some private weather forecasting services have emerged in Australia — for example, Weatherwatch and Weatherzone — that make use of basic data collected by the Bureau of Meteorology. However, the private weather forecasting industry in Australia is small compared with that of the United States — most value adding is currently undertaken by the Bureau of Meteorology. This could reflect a range of factors, including differences in the size and structure of the Australian and United States agricultural sectors.

The development of the substantial private weather forecasting industry in the United States appears to have been influenced by the National Weather Service's policy statement of 1991 (since superseded) which stated that it:

> … will not compete with the private sector when a service is currently provided or can be provided by commercial enterprises, unless otherwise directed by applicable law. (National Weather Service 1991)

As noted by Stiglitz, Orszag and Orszag (2000), the National Weather Service's approach appeared to strike a sound balance between the public sector's role in providing basic information and concerns about displacing specialised, value-added private-sector services. Some researchers — for example, Weiss (2010) and Pettifer (2015) — have suggested that restrictions on data access and high data prices charged by European meteorological agencies may partly explain the significantly smaller size of the private weather forecasting industry in the European Union compared with that in the United States.

Sources: Bureau of Meteorology (sub. 198); Mandel and Noyes (2015); National Weather Service (1991); Stiglitz, Orszag and Orszag (1991); Weiss (2010); Pettifer (2015).

There will most likely be relatively few such circumstances, but a policy of open by default or similar sentiment should allow for this as needs that fully meet these principles arise.

Datasets should be fit for the purpose of release — meeting standards necessary for use in the field in which they are offered — but no greater than that unless they meet the criteria set out above. These standards will differ by field. Agencies might usefully consult with potential data users regarding their views on value adding.

In chapters 8 and 9, we examine the concept of release authorities in sectors where data offers the greatest potential to add to the community's welfare if released. These release authorities would be capability centres of excellence and setters of standards for effective use, and would provide guidance on the minimum (and maximum) level of value adding required.

DRAFT RECOMMENDATION 7.1

Beyond achieving a 'fit for release' standard (Draft Recommendation 6.1), government agencies should only value add to data if there is an identified public interest purpose for the agency to undertake additional value adding, or:

- the agency can perform the value adding more efficiently than either any private sector entities or end users of the data; and
- users have a demonstrable willingness to pay for the value added product; and
- the agency has the capability and capacity in-house or under existing contract; and
- the information technology upgrade risk is assessed and found to be small.

7.3 Pricing of public sector data

The cost incurred by a data holder to get a dataset to a state ready for release can be a key consideration in determining its pricing. But it is just one side of the coin. On the other side are:

- the intent of policy-makers to encourage wider use of the data
- the public interest purpose for release
- the likely scale and nature of benefits derived from data access and use, including whether and what benefits accrue to private users relative to wider public benefits
- the capacity and willingness of data users to pay, including consideration of the extent to which the dataset is unique or difficult to replicate, and how data use (and therefore potential benefits) may be impacted by pricing.

Further complicating all of this, the benefits of sharing and releasing public sector data often: depend on the uncertain, complex and dynamic contexts in which some data are used (for example, in research); take time to emerge; and can be significantly impacted by, or even dependent on, rapid advances in data analytics. These factors, and the potential for significant spillovers (box 7.4), make it almost impossible to estimate in advance the benefits of increased sharing and release of public sector data (OECD 2015).

Box 7.4 Spillover benefits from data use

Spillover benefits can arise from any users of data, including businesses and the publicly funded research sector. Some businesses use public sector data to produce products (informational or otherwise) to consumers, generating benefits to consumers that may substantially exceed the costs of purchase. Researchers that use public sector data may also generate (sometimes long term) benefits to the community.

The effect of various pricing approaches (such as free access, marginal cost pricing, cost recovery and commercial pricing) on the generation of direct benefits will tend to be very similar to the generation of spillover benefits — that is, the lower the price of data, the more likely it is to be used and the more likely that benefits will be generated from its use.

The Australian Government's cost recovery guidelines (Australian Government, Department of Finance 2014) are silent on whether market failure — in the case of data, the potential for spillovers associated with its sharing or release — should be taken into account in decisions on pricing. If there is potential for significant spillovers, the case for full or even partial cost recovery is weakened.

The Commission has previously noted that:

> … cost recovery is inappropriate where information products have a high degree of 'public good' characteristics or where there are significant positive spillovers. Information products that meet these tests would be budget funded as part of a basic product set. Other information products may nonetheless be included in the basic product set if the Government decides that there are explicit policy reasons for doing so. Additional information products [that is, enhanced data products] would be assessed for cost recovery. (PC 2001, p. XLII)

Sources: Australian Government, Department of Finance (2014); PC (2001).

Current pricing practices vary within the Australian government, although guidance from the Department of Finance (2015) notes that entities should 'only charge for specialised data services and, where possible, publish the resulting data open by default' (box 7.5).

In Australia and elsewhere, given the uncertain — and potential for large — benefits, many open data initiatives encourage the provision of data 'at the lowest possible cost, preferably at no more than the marginal cost'. For instance, in 2008 the OECD recommended that for public sector information:

> Where possible, costs charged to any user should not exceed marginal costs of maintenance and distribution, and in special cases extra costs for example of digitisation.(OECD 2008, p. 6)

The important point in all of this is to note that the approach taken to pricing public data should be consistent with achieving the objectives and purpose of data release. One pricing model will not meet all objectives or circumstances.

Box 7.5 Current data pricing practices within the Australian Government

On 24 December 2015, the Department of Finance released an information sheet on *Charging for Data Services* (Australian Government, Department of Finance 2015) as part of the Australian Government Charging Framework. The Information Sheet indicates that entities can consider charging for the following data services: specialised data collection; provision of specialised data and data analysis services; facilitating specialised access to data; and data support services.

The Information Sheet also advises that Australian Government entities should be aware of the Public Data Policy Statement's requirement that entities 'only charge for specialised data services, and, where possible, publish the resulting data open by default'.

It specified certain data services for which entities could consider charging, including:

- specialised data collection (for example, survey development and/or conducting a survey)
- provision of specialised data and data analysis services (for example, developing or tailoring existing data, provision of data in specific formats and provision of reports based on data analysis)
- facilitating specialised access to data (for example, provision of additional data infrastructure)
- data support services (for example, assistance with interpretation or data presentation of survey results, education, call centre assistance or maintenance of datasets).

How this is implemented in practice varies considerably across Australian Government agencies and across datasets, with a number of datasets released freely, while others are priced commercially. An example of the latter is ASIC's company database, for which the agency receives annual revenue estimated at about $700 million in company lodgement fees and another $60 million in revenue from company searches — the highest such fees in the world according to some sources (West 2016).

Sources: Australian Government, Department of Finance (2015); michaelwest.com (2016).

Cost recovery

Some governments have endorsed cost recovery; and certainly, the costs involved with standardising data and metadata to prepare the data for re-use can be significant.

Under a cost recovery approach, data is priced to recover the marginal costs and some or all of the sunk costs that an agency incurs — the costs associated with data collection, curation, maintenance, and storage and distribution infrastructure. The share of any costs that can be apportioned to internal use of the data by the agency are usually excluded, and any charges related to value adding should not attempt to claw back the costs of providing a basic dataset.

More broadly, the Australian Government's cost recovery guidelines (Australian Government, Department of Finance 2014) note that cost recovery pricing can improve the efficiency, productivity and responsiveness of government activities and accountability for

those activities. Indeed, this pricing approach can increase cost consciousness for all stakeholders by raising awareness of how much a government activity costs. In addition, the Australian Government's cost recovery guidelines also note that cost recovery pricing can promote equity, whereby the recipients of a government activity, rather than the general public, bear its costs.

However, cost recovery pricing may reduce the attractiveness of datasets to some users, including those with limited access to finance (the impact on small and medium enterprises is often raised), and those contemplating speculative or experimental uses of the data (where the chances of a 'pay-off' — commercial or otherwise — from the data are relatively low).

And it must also be acknowledged that cost recovery schemes themselves are not costless — there are necessary costs associated with administration (processing requests, payments and the like) as well as in maintaining the effectiveness of the scheme (through licensing and enforcement activities, for example).

Alternatives to cost recovery

It is reasonable to assume that if agencies are releasing data, and particularly if they are adding value before releasing it, that the question of what price to charge will arise. It would be desirable to develop pricing guidance capable of applying to all data release. But this is not achievable given the range of purposes for releasing data to external parties and the many and varied users and uses of different datasets. Nonetheless, it is important to consider the attributes of different pricing approaches, and their impact in the context of the purpose for releasing data.

Marginal cost pricing

Marginal cost pricing is a common approach, with prices based on the costs incurred in making data available to an additional user, while the agency funds all other sunk costs. In the case of data, the marginal cost of reproduction and release (distribution) is usually so low as to effectively be zero — that is, data is made available free of charge.

Providing data free of charge or at a price that covers only the marginal cost of distributing it promotes the greatest use of the dataset (at a given level of quality) and hence the greatest benefits to users. Of course, it may be possible to improve the quality of the dataset before its release and increase the range of users and uses to which it may be put.

Pollock's (2008) comprehensive study of the economics of public sector information concluded that there is a strong case for pricing at marginal cost or below:

> When it comes to charging 'users' of public sector information the case for pricing at marginal cost or below is very strong for a number of complementary reasons (note that, for most digital data, marginal cost will be approximately zero). First, the distortionary costs of average rather

than marginal cost pricing are likely to be high because: a) the mark-up to cover fixed costs is high, as marginal costs are such a low fraction of average costs; b) the demand for digital data as with other information services is likely to be high and growing; and c) there are likely to be large beneficial spill-overs in inducing users to innovate new services based on the data, as is evidently the case for other ICT services. (pp. 43–44)

Pollock (2008) also observed that for many datasets the government is already providing a large contribution to fixed costs and that marginal cost pricing or free access would allow external users better access to the dataset. Various studies have indicated that lowering prices from a cost recovery level to either zero or marginal cost can promote momentous increases in demand (box 7.6).

Box 7.6 Impacts on demand of making data freely available

In a study of 21 public sector agencies, the European Commission found that a change in pricing approaches to 'marginal and zero cost charging or cost-recovery that is limited to re-use facilitation costs only' increased the number of re-uses by between 1000% and 10 000% (European Commission 2011, p. 6). It also found that lowering the price attracted new types of users, in particular SMEs.

The OECD (2015) also noted that SMEs in particular increase their use of public sector data in response to lower prices:

> There is in particular cross-country evidence that significant firm-level benefits are to be had from free or marginal cost pricing, with small and medium-sized enterprises (SMEs) benefiting most from less expensive data and the switch to marginal cost pricing (Koski, 2011). For example, analysis of 14 000 firms in architectural and engineering activities and related technical consultancy services in 15 countries in the 2000–07 period shows that in countries where public sector agencies provide fundamental geographical information … free or at maximum marginal cost, firms grew about 15% more per annum compared with countries where public sector geographic data have cost-recovery pricing. Positive growth comes one year after switching to marginal cost pricing, but growth is higher with a two-year time lag. Apart from SMEs (once again) benefiting most from cheaper geographical information, switching to marginal cost pricing of PSI [public sector information] substantially lowers SME barriers to enter new product and service markets. (p. 413)

Reducing the cost of access has resulted in increased income for some government agencies:

> [T]he Austrian public sector body responsible for geographic information, Bundesamt für Eich- und Vermessungswesen (BEV), lowered charges by as much as 97%, resulting in a 7000% growth in demand for certain product groups. In essence, BEV was able to increase its geographic Open Data sales revenues by 46% in the four-year period after the pricing review. (Capgemini Consulting 2013, p. 9)

In Australia, reducing the price of access to public sector spatial information has had a significant impact on the use and re-use of such data, with a substantial increase in the volume of data sold (Spatial Information Industry Action Agenda 2001). In February 2016, the G-NAF database started being published under an open data licence at no cost to end users on data.gov.au (after years of users being charged for use) and, by August 2016, it had been downloaded more than 1500 times (Kantor and Bhunia 2016).

Sources: Capgemini Consulting (2013); European Commission (2011); OECD (2015); Spatial Information Industry Action Agenda (2001); OpenGov (2016).

This approach generates little or no revenue to offset the costs of collecting, producing and curating the data — which means the government must fund these costs. An agency that is mandated to price data at marginal cost (or provide data free of charge) may have reduced incentives to make data available or to add value to its data because it will not be able to reap any revenue benefits from doing so. In other words, this approach may reduce an agency's level of customer orientation — including, for example, how responsive it is to complaints (Pollock 2008). In cases where data revenue forms a significant portion of an agency's budget and cross-subsidises other activities, a reduction in prices may jeopardise the continuation of some of the agency's other activities if an alternative source of funding cannot be secured.

However, making data accessible free of charge eliminates the need to administer and enforce pricing schemes. In cases where data had been sold under licence (as noted above, licencing may be required to sustain cost recovery or commercial pricing), the costs relating to monitoring compliance with licencing arrangements would disappear (European Commission 2011). In the case of making data available free of charge, an agency's transactions costs will fall, sometimes significantly — for example, administrative costs, such as invoicing, will fall. Where free release leads to an increase in the number of data users, this can in turn sometimes have a positive impact on data quality if data deficiencies are reported back to the agency. This increase in quality can be beneficial for all data users, including the agency itself. Finally, any increased economic activity associated with data access could be expected to generate additional tax revenues in due course.

Commercial pricing

A more commercial approach to pricing would see agencies pricing data based on the price of similar products available in the market, or if no such product (that is, a substitute) or market exists, based on maximising returns (revenues) for the agency. At its extreme, the latter could involve the agency charging different prices for different users (price discrimination), and/or to reflect differing levels of value adding that is undertaken, including in response to user demand (a 'freemium' model — basic data is free, but value-added data products attract a premium).

The effect on demand of commercial pricing, and in particular, price discrimination approaches that charge non-commercial users (such as researchers) less than commercial users will depend on the nature of the dataset concerned.

Whether or not price discrimination is desirable will depend on a range of factors, including:

- the potential uses of the data
  - For instance, if non-commercial use is expected to dominate then the negative impact on overall data use from a price discrimination approach may be relatively small (there are not many potential users who will be required to pay), as may be the amount of revenue raised.

  - If, however, commercial use is expected to dominate, then a price discrimination approach would help to achieve any cost recovery objectives (compared with free data), but could also significantly dampen demand amongst some groups of users.
- how practical it is to discriminate between users
  - if expected revenues are small, it may not be worth charging for data.

Deloitte (2013) noted that pricing public sector data at a level above the marginal cost may help to 'protect' the dataset from any reductions in agency funding. That is, production of the dataset would, to some extent at least, be self-financing and hence more likely to be sustained regardless of fluctuations in an agency's budget. Deloitte also noted that prices paid are a signal of consumers' willingness to pay for a particular dataset, conversely they can also signal a commitment by an agency to maintain the dataset over time, and can provide an incentive for agencies to be more responsive to the needs and wants of their customers.

Commercial pricing, at least for commercial users of data, has the benefit of providing price signals to the agency supplying the data and to the broader market (prices would be fairly transparent and the annual revenue raised may be discernible from an agency's annual report). This may then:

- provide data managers and policy makers some guidance as to what *might* be worth investing in, in the future
- encourage private entrants into the market for supplying such data.

Commercial pricing could also be considered if the agency is competing with private sector data providers. However, the agency should not exploit the advantages it holds by virtue of public ownership by undercutting the prices of its competitors. In the latter instance, however, the agency should consider why it is competing with private operators — there may be a case for it to release its basic data and let the private sector assume responsibility for undertaking the value adding.

There is the prospect of commercial pricing creating new sources of revenue. In practice, however, commercial pricing will not usually provide much benefit to the agency supplying the data. Revenue streams may not be significant (see below) and, even if the significant revenue were to accrue to the agency, subsequent agency budgets are usually adjusted for such revenue streams.

In the context of these considerations, it is interesting to note that the OECD (2015) observed that sales of public sector information (including public sector data) tend to generate very little direct revenue for most governments compared with the costs involved in collecting, curating and distributing such data. The study outlined earlier, undertaken by the European Commission (2011), found that sales revenues recouped around 1% of the overall budgets of the agencies examined in the study. The OECD (2015) found that even in exceptional cases, sales revenue represented a maximum of about one-fifth of the total expenditures of the agency generating the information or data.

The uncertain nature and timing of the benefits flowing from public sector data use and the unique nature of public sector data can make it difficult to value in a commercial sense — particularly before widespread use and application has occurred — and this needs to be taken into account.

DRAFT FINDING 7.1

There is no single pricing approach that could act as a model for guiding public sector data release decisions. The identification by agencies of the grounds for undertaking each release will have a direct bearing on the choice of price approach. Cost recovery, long considered to be the default option in the public sector, is only one of a range of approaches and not necessarily to be preferred.

Price as a means to an end — maximising the use and benefits of data

As noted throughout this report, greater access to public sector data can offer attractive opportunities and benefits. The approach taken to pricing public data for release should seek to maximise those opportunities and benefits, while retaining the incentive and ability for agencies to collect and release data effectively and to remain responsive to user needs and interests.

In light of the many uncertainties involved, approaches to pricing should, all else equal, err on the side of maximising access to and use of data.

Maximising social benefit

In many instances, access to data will deliver strong and significant public benefits — chapter 2 for example, noted the many and varied health-related benefits from increased availability of administrative health data.

Where public benefit is likely, the clear aim of data release policies and pricing across all agencies should be to maximise access to and use of its data. There is broad consensus that making data freely available is likely to maximise use and hence deliver the highest level of social benefits (box 7.6). There will, however, be a net cost to government related to data release. These costs and the potential funding requirements need to be addressed if genuinely open data is to be achieved and maintained.

In view of the exceptional potential for the research sector to generate spillovers, the Commission is of the view that pricing of data to the publicly-funded research community should be the subject of a separate review.

Such a review should be undertaken independently of the Department of Finance. Key considerations should include maximising the productive use and re-use of data in research, including datasets created in the course of research projects. There should be no

expectation that one funding model will fit all public datasets of interest to the research community, given the varying degrees of potential spillovers and the different circumstances of individual agencies.

DRAFT RECOMMENDATION 7.2

The pricing of public-sector datasets to the research community for public interest purposes should be the subject of an independent review.

Enabling commercial interests and use

Some of the benefits created through access to public data will accrue largely to private individuals and enterprises that are able to create new products and services on the back of innovative applications and analysis of public 'big data'. In these circumstances, commercial interests may dictate that agencies are able to price data to deliver revenue, but also as a means of determining important information about users, including who users are and what is their willingness to pay for additional value-adding processing. Where agencies undertake substantial value adding (because it meets the principles outlined in section 7.3 above, including that there is a willingness and capacity to pay for this), there are strong grounds for passing these costs on to data users.

The example of the Bureau of Meteorology discussed above (box 7.3) is an important one. In addition to making some data and information available free, the Bureau of Meteorology provides tailored information for specific industries, where it has added value through analysis and processing, and for which it charges prices with an incremental cost recovery element. It also provides, for a fee, bespoke commercial products (sub. 198, p. 11).

In general, where data clearly has high commercial potential (for example, exploration data compiled by Geoscience Australia), there could be equity grounds for supporting at the least a cost recovery approach in line with the user-pays principle. It could be argued in such circumstances that free access or very low (marginal cost) pricing would amount to a subsidy to the data recipients, with the potential for efficiency losses that are commonly associated with that. Free access or marginal cost pricing should, in these cases, only be supported if there is a strong likelihood of sufficient spillover benefits being generated to offset the implicit subsidy or if charging users is impractical.

Conclusions on pricing

Despite uncertainty over the magnitude of the benefits of making public sector data more available, it is possible to draw some broad conclusions on pricing. A key point to note is that a single pricing approach for all datasets is not desirable. Rather, a distinction should be made between basic datasets (that have been minimally processed to a 'fit for release'

level) and those datasets that have undergone more substantial value adding by the data holder (to produce an enhanced data product) — but always also taking into account the likely public interests and spillover benefits.

Table 7.1 summarises the Commission's assessment of the various pricing approaches for a typical agency selling minimally processed data.

Table 7.1 **Pricing approaches for a typical agency selling minimally processed data[a,b]**

Pricing model	*Direct benefits*	*Spillover benefits*	*Net agency revenue*[a]	*Agency incentives for efficiency*[b]
Free / marginal cost	High	High	Low	Low
Cost recovery	Medium	Medium	Medium	Medium
Commercial	Low	Low	High	High

[a] Excludes adjustments to agency budgets made in response to revenue changes and ignores downstream impacts on central government taxation revenue generated by any increase in economic activity. [b] The risk of misuse of market power by agencies that have unique data holdings is not included.

Source: Commission estimates.

For minimally processed data, free access or marginal cost pricing is the preferred approach

For basic datasets (where only the minimum necessary processing has been performed) there is a strong case for free access or marginal cost pricing, given:

- uncertainty over the size of the benefits (direct and spillovers)
- evidence that demand for datasets is often highly responsive to price.

For datasets that are already available (at a price above zero or marginal cost), the main impact on agencies of adopting free access or marginal cost pricing is the revenue forgone. However, the reduction in agency revenue would be offset entirely by a reduction in costs for existing purchasers of such datasets — that is, it will be a transfer. Hence any resulting benefits from increased use of the dataset will largely represent a net benefit (from an economy-wide perspective), particularly if the marginal costs of supplying any induced additional users are very low.

For datasets that are not currently available, the cost of making them available would comprise the costs of curating the dataset to a sufficient minimum standard for release (these costs may be substantial) and the distribution costs (which are generally small).

For such datasets, an assessment would need to be made of the likely impacts of making them available to external users — that is, an assessment of the expected increase in use of a particular dataset and the size of the benefits likely to be derived from this additional use — versus the costs involved. On balance, if the benefits are expected to exceed the costs,

the dataset should be made available and, if it comprises basic data, made freely available or priced at marginal cost.

For value-added data, the case for cost recovery or commercial pricing is stronger

In the event that there are strong efficiency arguments for an agency undertaking value adding above the 'basic' standard, the case for adopting cost recovery or commercial pricing is strengthened, if there are users that are willing — and have the capacity — to pay and this is consistent with public interest case for release.

In such cases the agency could adopt either:

- a cost recovery model
- a price discrimination model whereby non-commercial users can access the data free of charge, while commercial users pay a higher price based on cost recovery or revenue maximisation (although it may be difficult to prevent commercial users from accessing free copies of the data)
- a freemium model, whereby all users can access the basic data free of charge but pay a higher price for the value-added data if they wish to access it.

On balance, for value added data, given the potential for abuses of market power, and the potential for unanticipated spillovers, cost recovery pricing is the approach that would be in the broad community interest.

However, agencies should consider experimenting with lower prices for value-added data to assess how demand (and revenue) responds. If demand for the value-added dataset is somewhat price sensitive — that is, if lower prices elicit a non-trivial increase in demand, then lower prices could be maintained. This would be in line with the experience of some European agencies that experienced stable or higher revenues when they lowered prices (European Commission 2011).

DRAFT RECOMMENDATION 7.3

Minimally processed public sector datasets should be made freely available or priced at marginal cost of release.

Where there is a demand and public interest rationale for value-added datasets, agencies should adopt a cost recovery pricing approach. Further, they should experiment with lower prices to gauge the price sensitivity of demand, with a view to sustaining lower prices if demand proves to be reasonably price sensitive.

7.4 Funding support for public sector data release

Pricing will not achieve full recovery of costs for agencies because of normal budget practice (where agencies often do not retain proceeds of sales). Further, public interest data release is an activity that can be expected to grow and cost recovery or full pricing is unlikely to be consistent with public benefit.

Costs can be substantial — a point noted, for example, by the Department of Employment, sub. 18). Releasing data may require agencies to acquire new skills, train employees, purchase technologies, and upgrade network infrastructure. There are also costs associated with ensuring timely updating of data as well as with organising and preparing data for release. The costs associated with managing sensitive personal data can be particularly significant (chapter 5).

These costs mean there is often reticence on the part of agencies to share or release data. The Department of Agriculture and Water Resources (sub. 37) stated that a lack of ongoing funding imposes barriers to access and use of data, and is likely to slow the rate of progress in achieving policy goals in this area. CSIRO (sub. 161) observed that data management activities are often funded through short-term initiatives or internal capital budgets.

More broadly, the costs of increasing public sector data availability need to be weighed against the public benefits of doing so. The objective of increasing data access should not simply be to increase the volume of available data.

Alternative funding approaches for maintaining and increasing data availability

There are several options for funding data collection and availability:

- agencies fund from existing budgets (re-prioritise current and future spending)
- government provides additional earmarked/tied funding to agencies
- a reward approach
  - under which agencies are rewarded for data releases that result in research outputs (Card et al 2010)
- a combination of any or all of the above.

Advantages and disadvantages of the various funding approaches

Agencies fund from existing budgets

International experience suggests that countries have not had particular difficulty in funding the switch to free and open data and information, and that this has not been the major barrier that was foreseen in the past (OECD 2015). Half of the respondents (12 of 20

countries plus the European Commission) did not have special funding or budgets for the switch to open and free public sector information strategies. The sources of finance were largely internal, or derived from reallocation of existing funds.

This approach instils in agencies a culture in which data sharing and release is core business. Further, it reinforces budget discipline in the process of making more data available — that is, agencies would focus on achieving outcomes in the most cost-effective way possible because funds were being diverted from competing uses. This approach also brings clarity to value-adding activities — they need to be justified for funds to be found.

Government provides additional earmarked funding to agencies

Additional funding could be provided through future annual budgets to agencies in line with each agency's expected costs of making its data publicly available. While intuitively appealing, this approach has a number of drawbacks. First, it may imply that releasing data is not core business for government agencies, to be pursued only as long as funds last. Second, the provision of additional funds could encourage inefficient processes and delivery because there were no opportunity costs for individual agencies. Finally, this approach places an additional burden on overall government expenditures, necessitating savings in other areas, tax increases or increases in debt.

On the other hand, funding does provide an inducement for data release. In New Zealand, for example, Land Information New Zealand was funded to help other agencies prepare their data for the Integrated Data Infrastructure (IDI).

In practice, supplementation will always be essential if agencies are genuinely not able to reallocate funds and costs are more than marginal.

A reward approach

A 'reward' approach, as a generic tool, is unlikely to motivate agencies in an efficient manner. It may substitute for market-related revenue generation in some cases, and may motivate agencies to develop data proposals for the sake of earning the reward. However, such motivations are not an efficient substitute for an intrinsic assessment of the direct and spillover benefits that should inform decisions regarding data release, value adding and pricing.

There are also inevitable lags between the incurring of costs by an agency and the delivery of 'reward' and so the reward would probably not provide a significant incentive.

Mandated non-budget revenue requirements

In some cases, agencies are required to achieve a substantial part of their funding from non-appropriation sources (for example, about 70% of the income of Australian Institute of Health and Welfare is from sources other than government budget appropriation). Such

requirements have implications for the type of data an organisation releases — for instance, they may focus on releasing data that is considered likely to generate revenue, rather than what may be in the broader public interest.

Budget policy is beyond the scope of this report but the implications of requiring agencies to maintain very high external revenue targets should be considered from time to time.

Mitigating the costs — outsourcing

Capital or recurrent expenditure?

Funding for data-related activities will usually comprise a mix of capital and recurrent expenditure. For instance, payments for server and storage hardware, associated infrastructure and software are typically capital expenditure. The main recurrent expense is often the labour needed to manage an in-house data system.

Lowering capital expenditures by outsourcing

While the Department of Prime Minister and Cabinet has suggested the upfront costs in standardising data will be 'outweighed by reduced development costs over time' (DPMC 2015, p. 34), agencies sometimes do not have the luxury of longer-term planning horizons where budgets are concerned.

One approach that government agencies can take to mitigate the costs of data management may be outsourcing. For example, with the advent of the cloud, storage can be outsourced and treated as recurrent expenditure. Where outsourcing is done for the right reasons — namely that a third party can provide better and/or cheaper services — it is to be encouraged. If done solely for the purposes of accounting in the context of near term budget constraints, the results may be less desirable.

A number of important government datasets are in the process of being outsourced, including the National Cancer Register (which will be managed by Telstra Health) and the ASIC registry (where a new operator is yet to be chosen). In both cases, the Australian Government will retain ownership of the data collected and stored (Department of Finance 2016; DoH 2016)

Contractors also have to be paid and managed, so outsourcing can become expensive. Some stakeholders have raised concerns about the effects of outsourcing on the accessibility of the data and potential imposition of new costs on users (CSIRO, sub. 161, SA-NT DataLink, sub. 123, Uniting Church in Australia: Synod of Victoria and Tasmania, sub. 137).

We also note that outsourcing the control of data without addressing the possible exercise of third party copyright by a contractor is a serious risk (chapter 3).

Is there a role for some centralised funding of some data-related activities?

There may be a case for centralised funding of some system-wide activities — for example, to promote the interoperability of datasets from different sources, or where there is clear potential for substantial spillovers from facilitating data release — such as spatial data — from a number of different jurisdictions.

A further factor to consider is that to the extent public data sharing and release boosts downstream private economic activity, it will contribute tax revenue. However, this can be a deceptive illusion — public funding is always likely to generate some tax revenue. Moreover, any such additional revenue will accrue to consolidated revenue rather than the agencies that made the data available and therefore have little direct impact on the relevant agencies' budget.

Conclusions on funding

As the Commission is recommending that for minimally processed datasets — which will potentially form quite a large proportion of future new releases — a free access or marginal cost pricing approach be adopted, the funding of increased access to data from data sales revenue is not a viable option.

Moreover, free access or marginal cost pricing will reinforce the primary advice of this chapter, which is for agencies not to value add beyond actions required for data collections to meet internal and intrinsic public interest requirements. This approach is a virtuous circle — agencies will incur the least additional cost, and external users will benefit from agency's data investment at no more than marginal cost.

For those agencies that can have a good understanding of the market capability to pay for data that is value added, current funding structures may be worthy of preservation. We have not been advised of alternatives, and nor can we design one on available information, because the field of potential users is simply too diverse.

In chapter 2, the Commission recommended that central government agencies with data responsibility should consult widely with stakeholders to identify demand for specific high value datasets — effectively, a crowd sourcing approach to estimating the relative value of public datasets. Should this recommendation be adopted, it should be accompanied by contingent additional funding for the agency or agencies that hold the datasets identified, through this process, as high value. The amount of such funding would be linked to the likely public value of releasing or sharing the data and the ability of the agency to pay the associated costs — and payable on release of the relevant data. It would be a limited form of funding supplementation, designed primarily as an incentive for external parties to spend their time assessing which datasets held now by agencies are of highest potential value for early release — that is, the availability of this supplementary funding for agencies would increase the likelihood of the crowd sourced advice being acted upon.

This incentive is not an unusual way to drive early adoption of change and determine early priorities for action — both areas where governments have limited information (although they will get more through this process) and face internal cultural resistance.

That aside, however, the Commission believes that normal government budgetary processes should determine funding for data-related activity in agencies.

DRAFT RECOMMENDATION 7.4

For datasets determined through the central data agency's public request process (Draft Recommendation 2.1) to be of high value and have a strong public interest case for their release, agencies should be funded for this purpose. Funding should be limited and supplemental in nature, payable only in the event that agencies make the datasets available through release or sharing.

Aside from this additional funding, normal budgetary processes should apply for all agencies' activities related to their data holdings.

8 Options for comprehensive reform

Key points

- Data management practices — including arrangements for collecting, analysing, linking and passing data between parties, and the obligations, rights and opportunities of the various parties involved — have developed in a reactive fashion to the rapid growth in digital data collection, storage and analysis.
- There is enormous scope for beneficial reform, if effectively risk-managed. System-wide change is required for the potential benefits to be fully realised.
- The Commission has compiled a set of criteria for assessing reform options. Reforms must: deliver a net benefit to the community; increase data availability and use; engender community trust; empower consumers; preserve productive commercial incentives; protect privacy; and be able to adapt to an evolving data landscape.
- A specific focus on improving the ability of individuals to better understand, access and use their own data is essential to gaining acceptance of a wider use by governments and the private sector of their data.
  - Rights must be genuinely exercisable to be of value to consumers. This requires legislative reform and follow up to create common standards for data exchange.
  - A new definition of consumer data may be required to enable data transfer. Simply amending the *Privacy Act 1988* (Cth) may not be the best way to proceed, given the myriad of restrictions on data access and use contained in other legislation.
  - In the face of data's exponential growth, there is a need to move away from seeing individuals' interest as being limited to privacy to a broader focus on effective control and choice and to accessing data's value.
- Public sector data reform requires comprehensive change including: legislative change, clear leadership and institutional reforms to drive culture change, measures to build trust, and a genuine and thorough risk-based approach to data management.
  - The Australian Institute of Health and Welfare provides a good institutional model for an entity, with sectoral expertise and cross-jurisdictional remit, that can curate, integrate and release sectoral data sets in a coordinated manner.
- It is essential to guide data custodians in how to improve data access for researchers. Current restrictions reflect a culture of distrust. Forcing the destruction of researchers' unique topic-specific datasets at the end of a project is unwise (why destroy knowledge?), a costly waste of effort, and poor practice as a risk-management tool.
  - Equally, though, researchers should themselves not prevent awareness and re-use of data generated in the course of publicly funded research. The pot and the kettle should be the same colour.
- Options for improving access to private sector data need to be justified by a clear public interest to be served by so doing, and consider the impact on incentives for data collection and value adding.

As made clear in previous chapters of this Report, there are significant opportunities to increase access to and use of public and private sector data in ways that will benefit the community. This chapter examines options for promoting data availability and use, including those based on models that have been tested internationally or within Australian states and territories.

8.1 What outcomes are we trying to achieve?

Australia's data policies have developed in a reactive fashion to the tremendous development of digital technology. While the detailed shape of future developments surrounding data is necessarily still unclear, the direction and strength of change to a society that is substantially data-dependent is now sufficiently apparent to make it necessary to assess whether existing arrangements are fit for purpose.

The potential benefits of increasing data availability and use are many and varied (chapter 2), including:

- better informed decision making by consumers, businesses and government
- improved public services arising from data-driven efficiencies and better targeting of government policies and programs
- more open and transparent government, generating greater confidence and empowered citizens
- better places to work, live and play — ranging from smart cities to improved use of natural resources
- boosting Australia's competitive advantage and business opportunities through innovation and a world-leading data environment
- transformation of everyday life through personalised products and services, and a greater variety of choices.

If Australia's data regulation fails to adapt to the rapid changes, we will be unable to realise the full value of our data and risk being left behind other countries that are adapting to these changes and taking the steps necessary to benefit from their data holdings.

Previous chapters have identified that individuals — as consumers and citizens — need greater clarity about what rights and control they do and do not have when it comes to data about them. Similarly, clarity about expectations and incentives for data collectors and users in both private and public sectors will drive better outcomes, and we look at ways to achieve this. Uncertainty in understandings of rights and opportunities is particularly dangerous in periods of rapid change.

Comprehensive reform to data availability and use will necessarily require a change to the culture around the sharing and release of data. Decisions about the sharing or release of public sector data are not typically based on an assessment that weighs the potential value

of that data against a realistic assessment of the likely risks involved. A combination of legislation specific to the field in which the data is collected (some of which was formulated over a century ago) and excessive caution in interpreting and complying with that legislation, results in risks factoring heavily in decision making (chapters 3 and 5). This risk aversion is impeding more effective use of data held by governments.

While there are risks associated with data use and release, there are also significant and growing costs associated with not using available data. The Commission believes Australia can have its cake and eat it too — we can maximise the benefits of data and minimise potential harm to Australians through robust risk assessment and management processes and a high level of transparency over how data is used. This is the overarching objective of the reforms analysed below.

We need to embed trust in Australia's data management framework

We recognise that establishing a social contract founded on trust is essential to facilitate data sharing and use. Trust is central to removing barriers to better data access and more productive use of data. From earlier work in this Report, we know:

- In the public sector, a lack of trust (or confidence) by data custodians in data users results in valuable public sector data assets being underutilised and the erosion of potential benefits.
- Maintaining community trust in the way data is handled and used by the public sector is crucial, but inaction in the face of researcher interest is unlikely to generate that. The community knows that data is being collected and expects that it is being used to advance their interests.
- In the private sector, a lack of trust in data management practices and consumer privacy can erode a business' social license to operate.

Lack of understanding of data collection and use can breed mistrust. A model of opaque information handling practices is not — and should not be — a sustainable business model. A loss of community trust has already been felt by some businesses (chapter 4). People appear to be more willing to share information if they trust how it is being used, and feel like they have control over its use (chapter 5).

Trust is built on several key pillars — a sense of shared control, sharing in the benefits of data collection and use (such as better service delivery), and a belief in the accountability and integrity of data collectors and users (box 8.1). Embedding these fundamental values into Australia's data framework, rather than simply asserting their importance, is central to realising the full value of Australia's data.

Box 8.1 **Trust requires shared value, control, and genuine accountability**

- *Shared value*: Data is most valuable when it is shared. The value derived from this data should also be shared among the private sector, public sector, researchers, not-for-profits, community groups, and individual consumers.
- *Control*: Individuals should know who holds their data and how it is used, and be able to exercise control over this.
- *Genuine accountability*: Data management in Australia should build trust and confidence in the system by being transparent, promoting responsible data stewardship, and appropriately safeguarding privacy and data security. This requires meaningful safeguards to be embedded into Australia's data framework to assure people their data is being used safely.

Source: Adapted from the NZ Data Future Forum (2014).

8.2 Criteria for assessing reform options

We are using nine criteria to assess options for improving data availability and use. Many of these criteria are explicitly encapsulated in the Inquiry terms of reference, but we have applied additional analysis from literature and submissions to arrive at a comprehensive approach.

1. Deliver net benefits to the community — Governments can increase data availability in many ways, but all involve potential trade-offs. The aim is to ensure greater data availability delivers net benefits to the community, taking account of potential impacts on privacy, compliance and administrative costs, and/or spillover impacts to the broader community. Few of these benefits and costs can be estimated with much certainty.
2. Increase the availability of data — We have identified a wide range of benefits that greater data availability in Australia is likely to deliver — indeed, the Inquiry is not about increasing the collection of data, but about making more efficient use of what already exists.
3. Increase the usefulness of data — The form in which data is provided can be just as important as its content — not least, its machine readability, interoperability with other datasets, and ease of transfer. Early benefits will be gained from prioritising release of higher value data.
4. Engender community trust and confidence in how data is used — A lack of confidence and trust in Australia's framework for data collection, availability and use will undermine realisation of the full value of the country's data holdings.
5. Enable individuals to understand, access, use and benefit from their data — Data can promote more informed decision making by individuals — for example, consumers could use their own transaction data to make better consumption decisions in the future.

Individuals should be able to establish what information is being collected about them by public and private organisations and correct errors in that information.

6. Preserve commercial incentives to collect and add value to data — Commercial incentives are driving growth in data collection and the productive uses of that data. The protection of commercial-in-confidence data is an essential underpinning of competitive markets. We recognise the importance of maintaining private incentives and protections in light of the benefits they can deliver in terms of innovation, competition and value to consumers.
7. Promote transparency and accountability of governments — A more open and less risk-averse culture towards public sector data availability will increase transparency of government performance (policies and decisions), improve the accountability of governments and encourage constructive engagement of individuals and organisations with governments.
8. Address potential risks to privacy — The unauthorised release or use of identifiable information can pose potential risks of embarrassment, financial losses, discrimination or identity theft. Transparent and robust processes for managing these risks contribute to greater confidence in data access and use. Individual privacy needs to be protected.
9. Establish adaptability in policy settings/processes to account for different data types, different data users and changes that innovation will bring — Frameworks for data sharing and release need to be capable of assessing and balancing the risks and benefits of data access and use across a diversity of circumstances, institutions and users. The benefits and risks will inevitably change over time as markets evolve and technology advances through innovation. Flexibility in the processes for how data is managed and used, and transparency of those processes, will help to maximise the benefits and contribute to greater community confidence in the approach adopted.

8.3 Policy options to improve outcomes for individuals

As discussed in previous chapters, there is considerable scope to improve outcomes for consumers (criterion 1).

At present, a wide range of information is collected about individuals by governments and the research community (chapter 3) and by private businesses and not-for-profits (chapter 4). Individuals report significant concerns about how their information is being collected and used, particularly by some online companies. At the same time many individuals continue to use these services, although they may not understand how their information is being used.

When individuals are able to access their information, there appears to be the potential for significant benefits — as recognised by the Harper Review and the Murray Inquiry, and reflected in criterion 5. For instance, the energy market in Australia has particular arrangements (smart meters and standards) that could make it easier for consumers to

request access to their consumption history than in other sectors due to standards and authorising third party access (chapter 4) — this allows innovative new businesses like Energy Tailors to emerge that promote consumer choice and competition between energy providers. Reforms to improve mobile phone number portability in Australia have had significant success — and unexpected benefits (for example, mobile phone numbers are replacing residential addresses as a more enduring way of contacting individuals). The emergence of Health& as a platform for individuals to manually input and manage their health information is an encouraging development, but there is still much progress to be made in the health sector.

Legislative frameworks have not been altered to reflect the realities of rapidly increasing digital data collection (University of Tasmania, sub. 196). Where changes have occurred they have mostly been directed defensively, towards restricting data use. Opportunity has been ignored, either for individuals or for governments. The private sector, meanwhile, has actively embraced business opportunities with individuals' data, but too often individuals lack understanding and control over this. And these divergent approaches create inconsistent expectations or understandings about how individuals' data can be used and what safeguards are in place.

Submissions to this Inquiry generally accept that people want personal information about them to be managed carefully and respectfully; as does this Inquiry (criteria 4 and 8). Effective governance confers confidence for data usage (for instance, Association for Data-Driven Marketing and Advertising, sub. 178; Australia Post, sub. 174; Centre for International Finance and Regulation, sub. 9; CSIRO, sub. 161; Datanomics, sub. 129; IoT Alliance Australia, sub. 188; NetApp Inc, sub. 166; Office of the Privacy Commissioner — NSW, sub. 173; Telstra, sub. 88).

DRAFT FINDING 8.1

It is important governments and businesses maintain a social licence for their collection and use of data. This can be built through enhancement of consumer rights, genuine safeguards, transparency, and effective management of risk. Community trust and acceptance will be vital for the implementation of any reforms to Australia's data infrastructure.

A framework to improve individual participation and trust

There is a range of options available to improve consumer outcomes and give individuals a sense of greater participation (or non-participation, at their choice) in the trading and analysis of personal data. Retaining the status quo is not considered one of them, as there is simply not sufficient transparency or scope to exercise personal preferences (criterion 9); and inaction is hardly likely to improve trust (criterion 4).

At one end of the reform spectrum, prescriptive restrictions on how businesses use personal data could be imposed — for instance, strengthening the default terms and conditions of a privacy policy, and leaving little room for businesses to deviate from them.

This approach would surely limit the scope for innovation and discovery (criterion 9), which are key benefits of digital data use seen to this point in the data revolution. And it would not recognise that consumers can gain from the active use of their data, nor that they have different preferences related to its use. While some consumers might find advertising that tracks their behaviour 'creepy', others might appreciate being told about the flash airfare sale on their frequently flown air routes, or not care at all. It is impossible to design a prescriptive solution that matches all preferences, yet allows new services to continue to proliferate — noting, however, that there are some wrongful practices that should not be permitted by law.

Moreover, taking such an approach also means adopting incremental regulatory shifts to specific failures as they become evident. In the light of the scale of data-induced change at the consumer level, it is preferable to first seek an adaptable solution that can evolve naturally *without* further regulatory intervention, to improve the ability of consumers to control their own outcomes.

Such an option is discussed in the remainder of this section. It seeks to give consumers meaningful choice, and greater control and transparency over how their personal data is collected and used. Control is a key element of community trust (criterion 4) (box 8.1). Control over how personal information is used can also help encourage information sharing (criterion 2) because it builds individuals' confidence that their personal information will be used in a way that reflects their preferences (Brandimarte, Acquisti and Loewenstein 2010).

Such demand for control is inherent in many consumers' attitudes to what is viewed as 'ownership' of their data (for instance, Gould, sub. 1). Codifying and strengthening consumer rights over data may thus not feel as transformational as it is. But consumers who believe they own their data are wrong, as demonstrated later in this chapter. (And in fact, ownership is likely to lead to perverse outcomes — discussed below).

Many participants to this Inquiry have supported this approach, arguing that consumers need greater control over how their information is used, for instance:

> Australia Post believe that the development of citizen/consumer side market infrastructures, that enables consumers to make better and informed consumption decisions offers the greatest potential economic and social value creation, however is also the least mature or developed. This view calls for a fundamental shift in thinking with regards to personal data. The World Economic Forum's multi-year Rethinking Personal Data initiative highlights the importance of moving from an institutionally oriented collection approach in which the individual is passive, to a shared data, shared governance model where trust and active engagement are possible. (Australia Post, sub. 174, p. 12)

> … the need for greater control over data accessibility in the context of data about individuals. Made available with appropriate user control, privacy, and trust, that data has great potential to generate social, economic and environmental benefits. (Centre for Policy Development, sub. 11, p. 4)
>
> We [Facebook] realise that we won't be able to achieve this goal without the trust of the people we serve if they don't have confidence that they can control the information they share on our platform. When people choose to communicate on Facebook, they're trusting us to treat their information with respect and to put them in control of who sees what they post. Our responsibility to uphold that trust is why privacy is at the core of everything we do at Facebook, and why we work every day to ensure that as we're building new ways to help people connect, we're also helping people stay in control of their information. (Facebook, sub. 172, p. 2)
>
> Confidence and trust is built through transparency. Organisations need to be clear, explicit and transparent about: what data they are collecting; what they are doing with the data and why; who will have access to the data and why; how the data will be protected from unauthorized access and use; how long the data will be maintained; how will the data be destroyed; what recourse an individual has if his/her personal data is misused or shared with unauthorized individuals; how an individual can opt out of having his/her data shared publicly; how an individual can have their data corrected or removed; and, the technology solutions implemented to manage the data. (NetApp Inc, sub. 166, p. 9)
>
> The most effective means to enable the expression of access, use and disclosure rights is to embed consumer data control mechanisms into data design protocols. … and thus build public confidence that data practices accord with community expectations towards privacy. (Office of the Privacy Commissioner NSW, sub. 173, p. 10)

In all of this it is important to be realistic about individuals' capacity to make decisions about their personal information, and design reforms accordingly.

The following sections evaluate options for designing a consumer data rights framework to improve outcomes for individuals (criterion 1), drawing on submissions to this Inquiry and overseas experience. Broadly speaking, these options are designed to:

- give consumers the ability to *derive greater benefit* from their own information (criterion 5)
- improve *consumer choice and control* over their personal information (criterion 4)
- improve *transparency* in how information is collected and used (criterion 7)
- maintain *safeguards* for consumers (criterion 8).

Regardless of any future reform, every opportunity should also be taken to reiterate to Australians how valuable their data is — by all means, trade it, but *caveat emptor* is a crucial piece of advice that currently only seems to appear when it is too late to help.

Consumer rights will give a better outcome than ownership

Many people think that they already 'own' their personal data in the same way that they own a pair of shoes (Gould, sub. 1), but this is not the case at law. In Australia, no one owns data (*Breen v Williams* (1996) 186 CLR 81), and this is generally the same overseas, although copyright and various other laws can ascribe various rights to parties — databases and medical records can be covered by copyright, for example. The Commission does not consider that this position at law should be changed — people should not be given ownership over their personal information, for the following reasons.

Thinking about data as personal property creates messy overlaps with copyright law — *Breen v Williams* held that doctors have copyright in medical records, and the patient could not access them. Untangling from data any copyright in how it is recorded might be possible, but simpler choices would be better. Existing privacy legislation has already modified the common law position and gives consumers a right to access personal information (medical records) even when another party (the doctor) has copyright over the document. Relying on ownership to achieve access might not achieve the same outcome — for instance, resolving competing interests in data if clear assignment of ownership was sought would be difficult where there are multiple owners (when there are multiple people in a photograph on Facebook, who owns that photograph?). Such a situation could render data unusable by any one party, frustrating criterion 3.

Thinking about personal information in the context of consumer 'rights', as many other countries do (United States, United Kingdom, European Union, New Zealand, and Canada), solves many of these problems more simply. And, a case can be made that the concept of *your* data always being *your* data suggests a more inalienable right than one of ownership. Rights may be balanced against other competing interests, but they cannot be contracted away or sold with no further recourse for the individual in the event of data misuse (Australian Bankers Association, sub. 93; ANZ, sub. 64; Dun & Bradstreet, sub. 135; Xamax, sub. 3;). This conclusion is consistent with that of the Australian Law Reform Commission (ALRC) (2003) when considering ownership of genetic information — that data rights give a more enduring and workable outcome for individuals.

Consumers may have some ability to influence privacy, but if data is viewed (as it should be) as an asset of increasing value, it is not an asset owned by the original source or subsequent holder of that data. To improve trust, reforms clarifying rights are addressed in this Report.

When talking about 'rights', the Commission considers that it is generally preferable to implement any new consumer rights in a way that is consistent with Australia's existing legal frameworks (for instance, as per the existing Australian Consumer Law framework).

Reform options to give consumers greater benefit from their data

Access and transfer of personal information

Access to personal information can provide consumers with significant benefits (chapter 2). In Australia, individuals have the ability to request access to information about themselves under privacy and freedom of information legislation (and in the energy market, under the energy market rules). Although under privacy legislation they can request information be provided in a particular format, there is no standard format for provision. A range of exceptions apply to this general ability to request access, including where access to the information would be prohibited by law. Additionally, individuals can request correction of their personal information — an entity must oblige if the correction would make the information more accurate, complete and up to date. But there is no simple, standard way for individuals to exercise this power — and in some industries they cannot authorise a third party to do so for them (chapter 4).

There have been problems even where initiatives to make it easier for consumers to access data have been introduced — lack of progress on common standards led to long delays in the implementation of midata in the United Kingdom, and there has been poor uptake by consumers (CMA 2016; DBIS (UK) 2014). The MyHealth Record in Australia has had some recent success, but implementation has been difficult due to poor incentives to participate and reluctance within the medical profession (appendix D). There appears to be significant unmet potential for more consumer-oriented data management in the health sector. And reforms to improve the access to and sharing of bank customer information in Australia and the United Kingdom have had limited success to date, although there are some recent moves on this front (appendix E).

For individuals to derive the most benefit they can from accessing their personal information, they should be able to use their data to move their custom to another preferred product or service provider or use their data to make more informed decisions about products and services of benefit to them, including being able to authorise a third party to do so on their behalf. Existing frameworks do not readily allow individuals to do this.

Information may not be provided in a machine-readable format, and even when it is, the format of provision may not be able to be read by another service provider and/or the data variables may be incompatible with product offerings of different providers. This is generally equally true of the public sector (for instance, under freedom of information requirements) as it is of the private sector. Thus the lack of formal standards is a serious potential impediment to the ready transfer of regained information, and thus to an individual's ability to benefit from it. Current Australian comparator sites are weakened by the limited supply of data from consumers. The United Kingdom is better, in part due to its midata reforms (chapter 4).

In considering reform options that would give individuals the capacity to readily transfer their personal information to a different provider, we have examined:

- how the ability to access/transfer should be expressed
- what data this ability should apply to
- implementation through standards.

How the ability to access and transfer is expressed

A minimal change option would be to amend *the existing privacy legislation* to specify that when a consumer requests access to their personal information, it must be provided in a machine-readable, safe, and electronically transferable format. Implementation of choice of format could perhaps even be left up to industry (for private data) or departments of state (for public data) — as the minimal change option, no incentives would be provided for performance.

However, this type of action seems likely to fail. Merely amending the privacy legislation to enable consumers to download their data will have limited benefits. Consumers are generally not skilled in determining exactly what part of their data is of commercial value, whereas third party intermediaries (such as higher quality comparison sites) or competitors to an existing supplier (in say banking or energy) might well be able to apply data analytics to just that task. Forcing consumers to do the downloading and uploading will moreover restrict the use of any new right to those who are strongly digitally literate. This would achieve very little progress against criteria 5 and 9.

Thus, to strengthen competition or user choice and demonstrate to individuals how they may benefit from control over their data — and in turn lift confidence in data analytics — a more powerful concept is to enable a right to data transfer *directly* to third parties, including allowing authorised third parties to exercise this right on behalf of the consumer. Two legislative models from overseas are helpful here.

The *Enterprise and Regulatory Reform Act 2013* (UK) enables the UK Information Commissioner to compel businesses to release consumer information if they do not do so voluntarily. Prior to introduction of this Act, progress on midata was slow. But a 2014 review of the midata program (DBIS (UK) 2014) found that the new Act had catalysed sectors into action, and progress had been sufficient to not require the legislative power of compulsion to be exercised; although even today the system is not fully effective (chapter 4).

Another model would be to introduce a right to data portability similar to that adopted in the new EU General Data Protection Regulation 2016/679 (EU GDPR), effective from 2018. This provides that the consumer has the right to transmit consumer data (which is a subset of personal data, discussed below) to another body without hindrance from the data holder, where technically feasible and the data is available. This right is required to be balanced against other competing rights such as necessary and proportionate restrictions to

safeguard important economic or financial interests of the country or to safeguard protection of the consumer.

The effectiveness of this form of portability has been questioned — consumers have to sue to enforce such a right, and it is often a poor financial decision to do so. And the requirement for the action to be technically feasible leaves in a data holder's hands a substantial ability to impede the transaction (Bapat 2013).

To address these deficiencies and improve the application of the right — namely, to be more comprehensive, we consider a right to data transfer in chapter 9.

What data should transfer apply to?

Consumer data sufficient to deliver an individual the ability to benefit from competitive offers of services amongst different market participants is likely to be a mix of some personal information as defined by the *Privacy Act 1988* (Cth)[48] and other transactional and technical standard data relevant to the service currently or previously received by the individual.

Examining the definition used where this right has been implemented elsewhere is instructive. The United Kingdom and European Union have separate definitions for 'consumer data' (data subject to a portability right) and 'personal information' (identifiable information subject to restrictions on collection and processing). This appears to be necessary for practical reasons — for instance, excluding copyrighted data analytics work, parts of proprietary statistical models, or data collected for law enforcement or national security purposes from transfer.

- In the United Kingdom, the *Data Protection Act 1998* (UK) applies to 'personal data', that is, data that relates to a living individual who can be identified from that data, or from a combination of that data and any other information likely to come into the possession of the data controller.
  - However, the Enterprise and Regulatory Reform Act only covers 'customer data' — information held in electronic form by or on behalf of the business, and related to transactions between the business and the consumer. Businesses it applies to are further limited to gas, electricity, mobile phone, and financial services providers.
- In the European Union, the new GDPR applies to 'personal data', which is any information concerning an identified or identifiable natural person (including someone able to be identified directly or indirectly, in particular by reference to an identification number or one or more specific factors (physical, physiological, mental, economic, social or cultural identity).

[48] Australia's privacy legislation generally applies to 'personal information' which is information or an opinion about an identified individual or an individual who is reasonably identifiable. Different states and territories have some variance in definitions, including how long after death the privacy legislation continues to apply (appendix C).

– However, the right to transfer only applies to personal data that was provided by consent, or where processing of the data was necessary for the performance of a contract. Thus the right to transfer does not apply where data processing was justified on another ground — for instance, where processing was necessary for compliance with a legal obligation or for carrying out a task in the public interest. Exercise of the right to transfer does not imply an erasure request (article 68, GDPR).

The desired outcome from defining consumer data should be that if the data point was received from an individual and subsequently remains substantially unaltered such that it is able to be linked within the systems of the firm back to that individual *then it is consumer data*. This approach would nevertheless allow entities who transform data and hold it in a way that does not identify or link back to an individual to continue to do so without impeding innovations that may flow from that. To apply more broadly than the Privacy Act, any new concept would need to be defined in existing legislation that has a broad coverage, as well as any new data-specific legislation.

In terms of what industries data transfer should apply to, it is possible to imagine the application of portability to only certain data collectors — in the United Kingdom, midata applies only to certain regulated sectors. But the approach adopted by this Inquiry is to seek to enable consumers to transfer information comprehensively across the economy until circumstances clearly demonstrate it is not tenable (that is, that criterion 1 is no longer being fulfilled). Some exceptions may be required, but as long as they are few and simple to delineate, the case for transfer is strong: it is the best option to lift individuals' abilities to control their own data for the purposes of participating more effectively in a data-based economy, and applying data for their own benefit. And via that, generating higher levels of trust and confidence in data analytics.

The term over which transfer should apply is an interesting question. In the United Kingdom, the data transfer scheme is limited to one year of data — apparently to limit data storage costs. This may not be a sound benchmark. Most businesses and governments retain data on their customers and clients well beyond a single year — including as required by other legislation. Moreover, when it comes to financial and insurance system participation, records go back decades and the whole of a consumer's history may be relevant to establishing risk. As such, a year should only be seen as a minimum required retention period; but where a business or government retains the data beyond that period, there appears to be no good reason not to require the whole of the digital record to be subject to the new right. We request feedback on this point.

Setting standards

To enable access and transfer of consumer data, industry-relevant standards determining the form of data transfer would be required. Standards are a crucial part of implementing data transfer or portability (chapter 6). At a minimum, data would need to be provided in an:

- electronic, portable and secure format (ACCC 2014), referred to in Harper et al (2015)
- accessible, machine readable, standardised, timely, interoperable and privacy protected format (CHOICE 2014; Harper et al. 2015; OIRA (US) 2011).

As discussed in chapter 6, options for the application of standards could include:

- *Leave implementation up to industry* — the Harper Review was concerned that too prescriptive an approach to standards could create significant costs. A larger concern is that, based on the UK midata experience, voluntary action is unlikely to achieve much.
- *Adopt a 'carrot and stick' approach* — for instance, as per the UK action to improve midata under the *Enterprise and Regulatory Reform Act 2013* (UK), retain a residual power to mandate portability but use it to encourage industry-consumer action, or otherwise motivate interoperability action at a high level.
- *Provide top-down guidance on implementation* — the EU GDPR explicitly states that the right to transfer data does not create an obligation for controllers to adopt or maintain systems that are technically compatible, and the right to transfer data from one provider to another only applies where this is 'technically feasible'; and limiting data portability to that which is provided in a 'commonly used' format could encourage holdings to be made in uncommon formats — in practice this is likely to limit the scope of data transfer in the European Union (Bapat 2013).

An additional option is to mandate that data be provided in an Application Programming Interface (API) format. This is being implemented in the United Kingdom's banking sector (appendix E). Provision of data in an API format appears to most closely achieve the objective of allowing consumers to derive value from their data, and most closely satisfies the criteria of improving the availability and usefulness of consumer data (criteria 2 and 3) as it allows consumers to, among other things, access their data in real time.

Nevertheless, we seek advice from Inquiry participants on whether this approach could prove to be infeasible if broadly applied (for instance, if it could impose excessive costs on small business relative to the benefits to consumers) or otherwise be insufficiently flexible.

Choice over how and when your data is collected and used

People often do not know fully what information is being collected about them. Even when 'consent' has been given it may well not be genuine consent, due to lack of understanding of the terms and conditions, or the 'take it or leave it' nature of them (chapter 4; Law Institute of Victoria, sub. 184). Some people may be less concerned about their privacy than others — but even they could benefit more through greater choice and the ability to exercise more use of and control over the data held about them (criteria 4 and 5) — for instance, when a business is on-sold and the use of the data changes.

Under Australia's existing privacy legislation, collection of personal information (Australian Privacy Principle (APP) 3) is allowed as long as the information is reasonably

necessary for, or directly related to, one or more of the entity's functions or activities, and it is collected by lawful and fair means. Consent to data collection is only required if the information is 'sensitive information'.[49] An individual must be notified of this collection if it is reasonable to do so (APP 5).

Existing privacy legislation provides that personal information cannot be used or disclosed (APP 6) for a purpose other than for which it was collected unless:

- it is a related purpose and the individual would reasonably expect it to occur
- the individual has consented
- it is required or authorised by law
- a specified public interest exception applies, for instance:
  - use or disclosure is necessary to prevent a serious threat to life, health or safety
  - the user is conducting health research or compiling health statistics.

Specific provisions apply to direct marketing uses — people have to specifically opt in to direct marketing, and be given the ability to opt out in each communication (APP 7).

Public interest uses

While no one questioned that privacy and individual choice were important values, some participants to this Inquiry (for instance, Telethon Kids Institute, sub. 5) argued that these provisions needed to be examined to ensure they enabled public interest uses of personal information that provide net benefits to the community (criterion 1). For instance, we recognise a compelling public interest in information about child sexual abuse being shared with appropriate bodies, or in births, deaths and marriages being recorded. Similarly:

> … it can be difficult for community organisations to obtain informed consent from consumers to share data with other agencies. This can be because people are reluctant to give consent, lack of trust in the system, or because they do not have capacity to consent as a result of age, disability, mental illness, drug and alcohol addiction or other reason. … If a person refuses consent for their information to be shared, organisations are forced to either refuse service to a vulnerable person or search for a 'work-around.' Governments need to work with the community sector to identify more realistic alternatives when consent to share information is refused. (Joint Council of Social Service Network, sub. 170, p. 20)
>
> Unpublished research undertaken by the Monash Centre for Occupational & Environmental Health and the Michael Kirby Centre indicates that some groups, such as parents of disabled children, assume that data concerning their children is being linked for the purposes of health research and are disappointed to learn that it is not. Participants from a number of different

[49] Defined in the Privacy Act as information or an opinion about an individual's: racial or ethnic origin, political opinions, membership of a political association; religious beliefs or affiliations; philosophical beliefs; membership of a professional or trade association; membership of a trade union; sexual preferences or practices; criminal record.

> population groups are happy for their data to be made available for research as long as the research is not-for-profit and the researchers are employed by reputable research institutions. (Monash University, sub. 133, p. 3)

The recent Caldicott (2016) report on consent to data sharing in the UK National Health Service further outlines some of the ways benefits can be limited when a model of individual choice is strictly applied to public interest situations, such as information sharing between health care providers (box 8.2). The reforms proposed were essentially designed to 'nudge' people into making decisions that would benefit them.

Box 8.2 The pitfalls of relying too much on consumer consent

The United Kingdom places significant emphasis on getting consumer consent for information sharing, which means that, very often, information is not shared between health service providers in a way that benefits the consumer.

The National Data Guardian recommended the development of a new consent/opt-out data sharing model for health and social care, to 'nudge' consumers into giving consent where it benefits them. Her view was that information was essential to improve the safety of care, including through research, to protect public health, and support innovation. She also considered there were benefits to joining health data with other types of information to provide better services to people.

Under the proposed consent/opt-out data sharing model, instead of having to opt-in, people would be able to opt out of their data being used beyond their own direct care:

- opt out of data used for purposes connected with providing local services and running the National Health Service
- as a separate decision, opt out of their personal data being used for purposes beyond their direct care (that is, to support research and improve treatment and care)
- give explicit consent, for instance, to be involved in research, and
- the consumer's wishes would apply unless there is a mandatory legal requirement or an overriding public interest.

Source: Caldicott Review (2016).

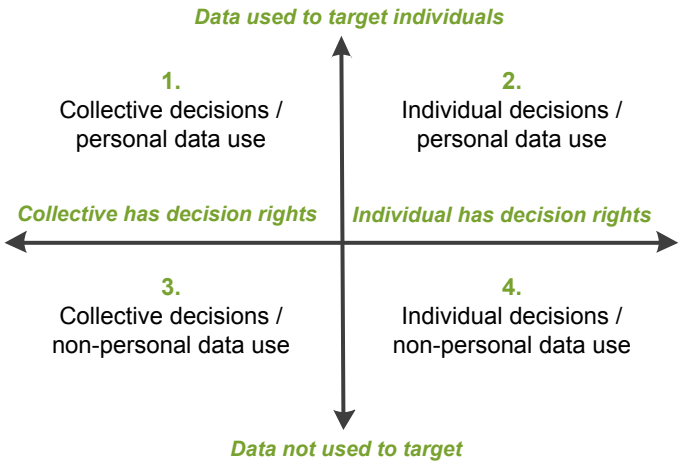
Balancing individual choice and the public interest

The right to control one's individual data can be context specific. Where there is a clear public interest, the decision *rights* about personal data may need to sit with the government, as the administrator of public services and the entity best positioned to determine and balance public interests (figure 8.1). This implies that where data may be used to understand and influence collective decisions and to deliver public benefits — for instance, for the delivery of social services, public health or safety — it may be socially detrimental (in contrast to criterion 1) to have the dataset decimated or the use prevented by individuals opting out. This idea is reflected in Australia's existing legislative

frameworks, which recognises that there are circumstances in which individuals should not be able to choose how their information is used.

As discussed in chapter 5, the Commission has supported extending the exceptions that allow identifiable information to be used for public interest health research without consent to cover public interest research in other fields — as long as the *published* data is confidentialised. Provision may need to be made for other compelling public or national interest uses where this does not already occur (chapter 9).

Figure 8.1 **Different data use scenarios for protecting privacy**



Source: New Zealand Data Futures Forum (2014, p. 18).

Options for greater consumer choice — collection

One area where confidence in data use could be improved is to have greater certainty that the law supports an individual's ability to withdraw from data collection, despite earlier agreeing to do so (criterion 4). Such a right to withdraw would be the obverse of the act of consent to participate in collection. This could require amending Australian Privacy Principle 6 to provide that an individual may withdraw their consent at any time to the collection or disclosure of their personal data, unless an exception applies (such as when collection is administratively necessary). But withdrawing consent would thus only apply to participation that was voluntary in nature — if information were collected without consent, there would be no change in current arrangements.

A different option could be expanding the range of circumstances in which a person can request cessation of collection of personal information to encompass situations where they have not given consent. However, a sensible list of exceptions would need to be designed to ensure that collection can still occur where it is required by law, is necessary for the performance of a contract, or some other public interest consideration applies (much like the exceptions overseas that apply to processing more broadly, which are discussed below).

Options for greater consumer choice — use and disclosure

If there were specific uses or disclosures that were of public policy concern, such as the practice of businesses on-selling identifiable personal information, one option could be to build on direct marketing provisions requiring businesses to give consumers the ability to specifically opt in or opt out of their personal data being put to that use. These powers could be extended beyond direct marketing to require that consumers must specifically choose (opt in) to have a business disclose their personal data to data mining companies, or to other companies in an identifiable form.

In the United Kingdom and the European Union, in addition to direct marketing controls, individuals have the right to object to processing of their personal information:[50]

- In the United Kingdom, this only applies if processing *would cause unwarranted and substantial damage or distress*. It does not apply if the individual has consented to the processing, or the processing is necessary for the performance of a contract, to fulfil a legal obligation, or to otherwise protect the individual's 'vital interests'.
- In the European Union, individuals have the right to object to any processing of their personal data, and the data processor has the onus to demonstrate that they have a legitimate interest in continuing that processing (which can include processing for a public interest purpose).

Both these provisions are drafted so as to balance the rights of the individual against other public interest considerations.

Additional rights in Australia to prevent processing based on distress, or the ability to appeal automated decisions might help engender a certain degree of community confidence (criterion 4), but are likely to be costly for businesses to implement and for the community to enforce (potentially contrary to criterion 1). It would also be important, particularly if an EU-style right was adopted, to ensure that exceptions were drafted broadly enough to permit legitimate public interest processing to continue — for instance, health care provision.

Finally, with respect to automated decisions — decisions taken using personal data processed *solely* by automated means — the United Kingdom and the European Union require individuals to be notified when an automated decision has been taken about them, and give the right to ask for the decision to not to be made or to be reconsidered. The European Union has included additional restrictions on taking decisions that may be discriminatory (for example, on the basis of race). Exceptions apply where decisions are authorised or required by legislation, or relate to a contract with the individual.

---

50 'Processing' under the *Data Protection Act 1998* (UK) is defined as obtaining, recording, or holding the information or data or carrying out any operations or set of operations on the information or data, including: organisation, adaptation or alteration. retrieval, consultation or use; disclosure; alignment, combination, blocking, erasure or destruction.

Giving individuals a right to due process may help provide confidence in big data analytics, particularly if these are used to ensure that significant decisions are not made on the basis of inaccurate information about individuals. But rights of appeal require resources to use them. And it would be important not to stymie the development of big data analytics.

The right to delete

The idea of destroying information seems intellectually unhealthy, as well as economically undesirable — historically, societies that have chosen it have generally not improved either personal or national welfare. The current interest in data destruction is of a more prosaic kind, driven by the unwise creation of electronic records rather than the fear of knowledge. Still, there are echoes, particularly in some calls for extension of deletion rights to official records.

Under existing privacy legislation, personal information must be deleted or de-identified when it is no longer required, but there is no obvious enforcement mechanism.

The right to request deletion of personal information could take a number of forms:

- An enforcement mechanism for existing APP 11 obligations to *destroy or de-identify personal information when it is no longer required*.
- An individual could be given the *right to request removal or deletion* of information that the individual has submitted to a service provider, directly from the service provider. This is consistent with better business practice as identified by Facebook (sub. 172).
- Finally, regulation could establish a *right to be forgotten*, or *right to erasure*. The European Union's approach includes — with great practical difficulty — deletion of information held by third parties, subject to various public interest and freedom of expression considerations.

Leaving aside the question of definitions, discussed earlier, an enforcement mechanism added to the existing APP 11 may add value in encouraging consumer trust, although the Inquiry has received few complaints about current practice. While the ALRC (2014) has previously raised concerns about this being difficult to enforce, the Office of the Australian Information Commissioner (OAIC) (2014) has since produced guidance on securing personal information and the implementation of existing deletion obligations, but no mechanism for individuals to personally enforce this has been developed.

However, a right to delete comes at a cost if applied to datasets where wide population participation is essential to maintaining quality of the resource (such as Census data) or conversely where there is only a small population of relevant participants and thus data to show rising or falling trends could be rendered inaccurate by a few withdrawals (for example, some public health trends such as suicide rates in small population groups).

Practical issues also arise. The data may not be stored in one place and complete deletion cannot always be guaranteed (Facebook, sub. 172) — although it might be sufficient for the information to be put 'beyond use' as per the guidance issued by the UK Information Commissioner.[51] In some cases businesses need to retain information for administrative purposes — for instance, to resolve unpaid, closed accounts, or to action 'change of mind' reactivations.

Further issues arise when personal information has become part of a database and subject to value-adding by the holding business. It may also not be desirable to mandate deletion in this circumstance, as at that point it is questionable if it is in substance personal data. Further, when the data relates to multiple individuals (for instance, a group photograph) resolving issues of deletion becomes complex (Facebook, sub. 172).

Bernal (2014) suggested a number of exceptions that could be applied to the right to delete — public interest reasons (medical records), administrative necessity (electoral rolls), archival reasons (newspapers), free expression, and security purposes (criminal investigations). The wider the list of exceptions, of course, the more reasonable becomes the proposition that the right is poorly conceived. It is difficult to conceive of an exhaustive list of exceptions, given how difficult it is to predict what value data may have in the future, in contributing to societies' knowledge. These are known unknowns — the best way to address this is to err on the side of retaining data.

Extending the right to deletion to apply to third parties is also particularly difficult, given the speed and breadth of dissemination the Internet allows — and data provenance recording is only a very partial solution to this. The nature of today's smart phones and data tracking technology is both so embedded and so pervasive that it is hard to conceive of the effective redaction of data covering both a service provider *and* third parties who were once authorised to use or download that data. Individuals must be aware: once distributed, data is most likely irretrievable.

We understand that other work is being done in this space to address specific concerns — for example, criminalising the non-consensual distribution of intimate images in New South Wales and the Commonwealth. This may create better incentives (discouraging initial dissemination) and be more enforceable than creating a right to deletion (which necessarily applies only after dissemination has occurred).

Overall, the Inquiry is not convinced of the public benefit nor practicality of a right to delete.

---

[51] For instance, information will be 'beyond use' if it has been deleted with no intention on the part of the data controller to use or access this again, but which may still exist in the electronic ether. Information will also be 'beyond use' if it should have been deleted but is in fact still held on a live system because, for technical reasons, it is not possible to delete this information without also deleting other information held in the same batch (ICO (UK) 2014).

Transparency

Along with creating legal rights for consumers, policy makers must also consider ways to facilitate the use of these rights. If individuals are not aware of their rights in respect to data management, or if access to data remains a convoluted process, reforms will be ineffective. Giving consumers greater choice is useless unless that choice can be exercised in a meaningful way. Likewise, transparency is important in public and private sector information handling practices — it promotes accountability and helps build trust (NetApp Inc, sub. 166, The Law Society of NSW, sub. 160) (criteria 4, 5, and 7).

Meaningful choice requires that consumers have genuine understanding of how their information is collected and used, and genuine control over how it is being used, so they can make an informed decision in a way that reflects their preferences. This includes being informed of the terms and conditions of a service, and being informed when information is being collected about them and what it will be used for — including when information is collected via monitoring or tracking, through website searches, or promotions. But gaps remain in community knowledge regarding the right to access personal information (OAIC, sub. 200). Evidence also suggests that in many cases consumer consent is not meaningfully given — giving consumers the legal right to exercise choice will have no effect if they do not have the information and understanding to be able to meaningfully use it.

Better disclosure allows consumers to make an informed choice about which businesses are going to be beneficiaries of individual data. This can create incentives for businesses to alter their information management practices to deter consumers switching away from businesses with poor information handling practices (criteria 6 and 8).

Ideally, disclosures should be simple and easy to understand but businesses vary in their practices (chapter 4). Default options remain important.

More information and a requirement for plain English would enable better decision making, but probably only to a small extent. Consumers are already prone to not read the terms and conditions before they sign up (chapter 4).

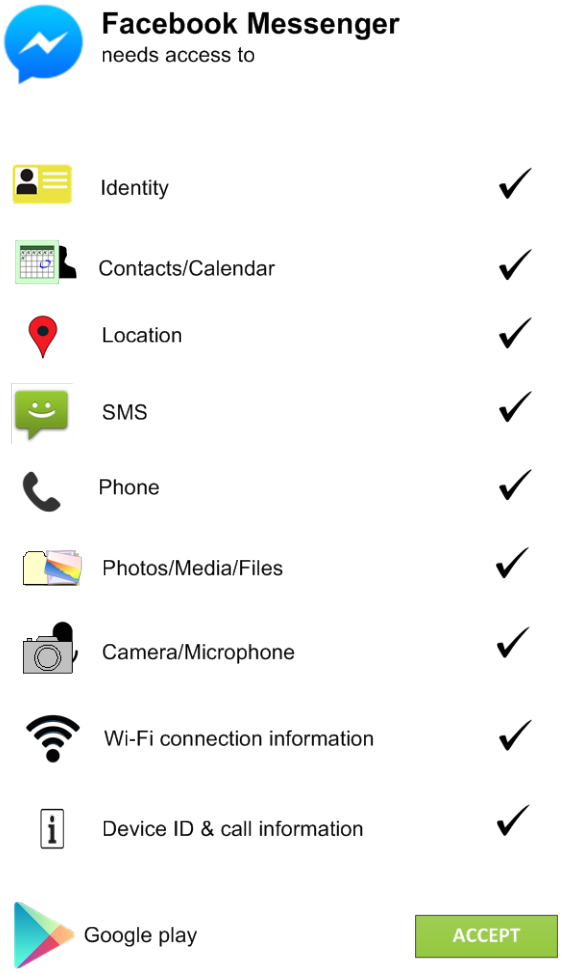
Another option would be amendments to the existing APP 5 to require businesses to clearly state in their collection notices if:

- personal data will be disclosed to another entity in identifiable form
- whether it will be linked with your personal data from other sources and, if so, in identifiable form or not
- how such disclosures will benefit the consumer and business respectively, and
- the measures taken to limit re-identification of your personal data.

But the sheer density of most terms and conditions suggest that adding to their volume may not add to their utility. Thus the Inquiry does not favour that approach. Separate, opt-in consent for transferring identifiable information to another entity could be another addition, but the same uncertainty of impact applies.

A better choice would be to mandate a simpler form of disclosure. One option could be to build on existing good industry practices and require a standard form of disclosure that is simple and easy to understand — for instance, like that which is shown in Android apps (figure 8.2), the disclosure indicates that Facebook Messenger needs all of the following permissions. Standardising disclosure in this way (but not necessarily with 'all or nothing' agreement to data use) would help in enabling machine-readable privacy policies.

Figure 8.2 **An example of more effective disclosure**



Source: Leadbetter (2014).

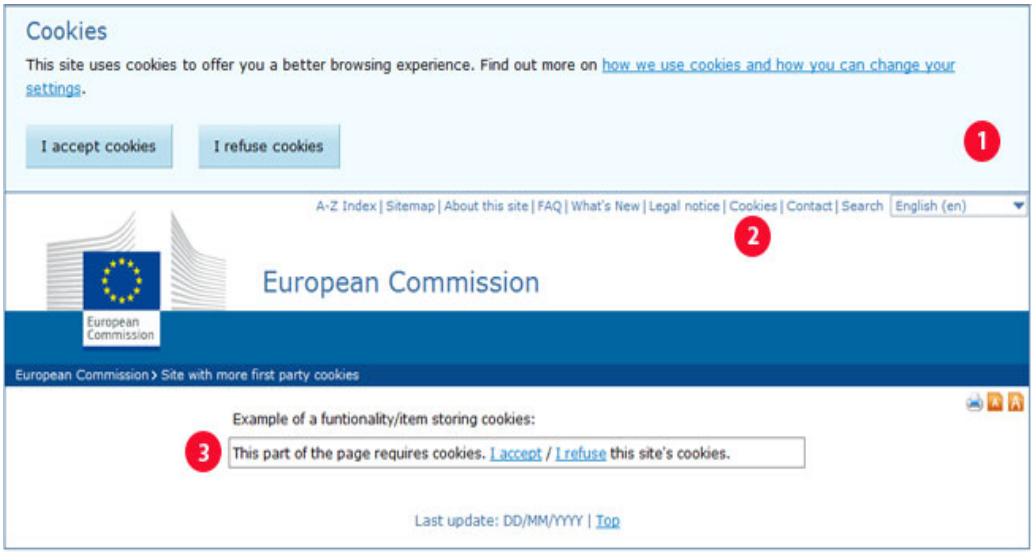
Another model could be to draw from consumer law and implement a check-the-box obligation, before signing up. Specific boxes to be checked could require consumers to acknowledge:

- cookies will be used to track their web use
- the business can aggregate and on sell their data

- the business will remove their identity via a process agreed with the OAIC before they do that.

However, a similar reform in the EU requiring consumers to give active consent before their information is collected by cookies has been less than effective (figure 8.3).

Figure 8.3 **Active consent is not always more meaningful consent**



Source: EC (2016).

We are seeking more information on what methods of disclosure are likely to improve outcomes for consumers (criteria 4, 5, and 8) and the most appropriate method of implementation (for instance, voluntary versus mandatory standardisation).

INFORMATION REQUEST

The Commission seeks views on what methods of disclosure would be most likely to result in consumers making a meaningful choice about how their personal information is being used, and how these disclosure requirements might best be implemented.

Safeguards

We recognised the importance of managing privacy risks in criterion 8. Privacy and information security is a responsibility that should be shared between governments, service providers and individuals (ACMA 2013). Promoting trust and transparency (criteria 4 and 7) in dealing with personal information can therefore encompass a range of measures, by the public or private sector:

- Individuals can and should take more responsibility for the security and disclosure of their data.
- Service providers can adopt higher standards in maintaining their customers' privacy.
- Governments can adjust their regulation, or implement new policies to enhance individuals' trust.

Options for improving the default contract

Consumers are typically offered privacy policies by companies on a take it or leave it basis — in other words, with little or no control over the terms of use of a particular service. Some participants to this Inquiry have argued there is scope to improve the default contract terms (Xamax, sub. 3).

Consumers who are concerned about the terms of service can perhaps switch to a different provider — for instance, DuckDuckGo is an alternative search engine that does not link search history to an individual (it does not personalise search results). Yet Google allows individuals to delete their search history at will, simply because they face similar market pressures. Thus, consumer concern can drive changes in business practices. And collecting personal data in the course of providing a product or service is not by itself unreasonable — indeed, it can be an inherent part of using a service such as Facebook or Fitbit.

If there were specific terms and conditions that were clearly unfair, there could be merit in reviewing the Australian Consumer Law framework to consider what additional consumer contract rights may be appropriate in a digital age. But no submissions to this Inquiry have suggested that this is a problem. Should submissions in response to this draft Report suggest further issues, we will consider the matter further.

Finally, some participants have noted that data ethics codes can play a role in driving better practices across industry (Australia Post, sub. 174; Datanomics, sub. 129). While these are positive developments, it is not clear to the Commission that the government has a role to play here.

Engaging all businesses

Trust and transparency can be strengthened by ensuring all businesses are covered by relevant regulations. Currently, there is no legal requirement for most businesses with annual turnover of under $3 million to protect their clients' privacy, despite technological advances that improve the way privacy is managed (for instance, confidential computing technologies and privacy by design). As pointed out by the OAIC:

> In today's increasingly data-driven economy, there are a number of economic sectors in which some small businesses have significant holdings of personal information and use this personal information, which in some circumstances is often sensitive information, in their business activities, such as online dating, or mobile app developers. It may be timely to re-examine the application of the small business exemption in this context. (sub. 200, pp. 12–13)

Many new start-up businesses, in particular, are focused on innovative uses of data and information, yet could well have turnover in the early years that puts them outside the scope of the Privacy Act.

Dealing with data breaches

The Australian Government is already working towards expanding the responsibilities placed on organisations holding personal information through the expected introduction of mandatory data breach notification legislation, and there are already laws in place that prohibit hacking (chapter 6). As noted by the Australian Computer Society (sub. 134), this can have positive effects on business' incentives (criterion 6):

> [Mandatory data breach notifications] will improve transparency and the broader public's understanding of benefits and risks associated with personal data and privacy. Data breach laws enhance the public good with industry likely to react faster, patch holes, and prevent more breaches.

The Australian Government has also recently moved to create a new offence to criminalise intentional re-identification of de-identified personal information that has been made publicly available (chapter 5).

Proactive privacy — innovation by service providers

Responding to community expectations that personal information is handled securely, service providers — public and private sector — can adopt new standards for handling data.

One approach that is gaining increased prominence is 'privacy by design', whereby organisations embed considerations of privacy into the design of a project and into their organisational processes and practices, including using privacy preserving technologies where appropriate (appendix B). This 'culture of privacy' is preferable to belatedly realising privacy is an issue and having to reverse-engineer privacy requirements.

Agencies and organisations can benefit from privacy by design (chapter 5), because management of privacy risks becomes part of their day-to-day operations, while individuals gain the reassurance that best-practice approaches are being used to handle their personal information (criterion 4).

Privacy by design is recognised in Canada and Australia as leading practice, and it has been endorsed as good practice by the Victorian Government (CPDC (Vic) 2014), the Office of the Information Commissioner — Queensland (sub. 42), the Office of the Privacy Commissioner — NSW (sub. 173), the OAIC (sub. 200), and the Law Institute of Victoria (sub. 184), among others. Other organisations could consider this approach in future, as well as other privacy preserving technologies (see appendix B for further details on technological developments).

Confidence in de-identification processes

Design for any new policy approach to support sharing and release of data would accordingly then have the retention of privacy as a central obligation. The more confidence consumers have that they are not able to be identified, the more confidence they will have in data sharing and release. As discussed in chapter 5, a significant step in this regard could include certification of de-identification measures taken by government agencies and private entities wishing to share and release data.

8.4 Policy options for sharing and release of public sector data

Elements of a public data framework

Previous chapters have shown that fundamental reform of Australia's public sector data infrastructure is needed (to achieve criteria 1, 2, 3, and 7). In developing options for reform, we have examined data management policies across governments in Australia and a number of countries overseas, as well as suggestions by stakeholders, in order to find good practices that Australia could adopt. Although each jurisdiction has taken a different approach (table B.1) there are some common themes:

- All jurisdictions have *open data policies* that make non-sensitive data open by default to encourage government departments to make public the data that they can (by law).
- In many cases, these jurisdictions had a *policy 'leader'* that built a coalition of support around the proposed reforms — for example, in the United Kingdom, the then Prime Minister wrote to all the departments encouraging them to release data on an 'open by default' basis (chapter 3).
- Some Australian states (NSW and South Australia) have *umbrella legislation* designed to simplify the myriad of legislative provisions that govern the sharing and release of public sector data.
- Some jurisdictions have adopted a *more centralised approach* to data sharing and release (United Kingdom, New Zealand, New South Wales), although they vary in the extent of centralisation — notably, the United Kingdom's Administrative Data Research Network has a release centre in every *country* in the United Kingdom, and the Office of National Statistics in the United Kingdom is responsible for driving overall public sector data standards.

Thus, experience in other jurisdictions indicates that key elements of public sector data infrastructure reform are clear leadership, reformed policies and legislation, and institutional change, to signal that permission is granted for an active policy of data sharing and release in Australia's public sector.

DRAFT FINDING 8.2

There is no shared vision amongst public sector data holders in Australia on how to consistently deliver widespread data sharing and release. The community — current and future — is entitled to expect such a vision. Comprehensive reform of Australia's data infrastructure is needed to signal that permission is granted for active data sharing and release and that data infrastructure and assets are a priority. Reforms should be underpinned by:

- clear and consistent leadership
- transparency and accountability for release and risk management
- reformed policies and legislation
- institutional change.

Risk management not risk aversion

An additional theme clearly apparent to us in the course of this Inquiry is the risk aversion across the Australian public sector. Public sector reluctance to share or release data is typically premised on the belief that release poses significant risks, including:

- re-identification of persons or businesses within datasets
- misuse of the data resulting from misunderstanding of its quality or meaning
- inappropriate exposure of commercially sensitive information
- reputational damage due to release of information about agency or government operations.

This risk aversion overlooks that different types of data have widely different risks, and release of some data will pose very low risks. Further, dealing with risk by not sharing or releasing the data at all means that the risk is not actually being properly managed. In fact, it appears that most data issues arise from poor data storage and handling procedures and/or negligent or deliberate misuse of data. Adopting an explicitly risk-based approach would likely reduce the number of breaches compared with the status quo. Adopting a risk-based approach would allow Australia to 'have its cake and eat it too' — it would allow Australia to maximise the benefits of its data (criteria 1, 2, and 3) and minimise potential harm to Australians (criterion 8) through robust risk assessment and management processes.

By applying an explicit risk-based approach to data access, government agencies would be required to clarify and manage the nature of data risks. Risk would be assessed based on both the likelihood, and probable consequence of, data breaches:

- Where consequences are non-trivial but likelihood is remote, agencies can still share or release, with mitigation strategies adopted as required.

- Where likelihood of breach and its consequence are considered high, access to the data needs to be carefully managed (National Statistical Service 2013).

A key benefit of this approach is that it allows for the lowest risk approaches to be identified (consistent with data security policies) and to therefore occur routinely (chapter 5). It will be important to allow for the exercise of judgment in policy and processes to account for different data types and different data users (thus satisfying criterion 9). This is the essence of a risk-based approach.

DRAFT FINDING 8.3

By applying a risk-based approach to data access, government agencies can establish a sound basis for where further risk mitigation effort is necessary and for moving early to the sharing or release of low risk data, while building and retaining the trust and confidence of users and the wider community.

Legislative instruments could be used to explicitly endorse the use of a risk-based approach to data release. South Australia employs one such model (appendix C). Such legislation can acknowledge that sharing information between government departments in a secure environment presents a different level of risk to releasing information to the private sector. It could also suggest various mitigation strategies, and designate specific agencies to oversee departments' implementation processes.

Cultural change through a legislative approach

The Australian Government, and most state and territory governments have adopted a policy of making non-sensitive public sector data open by default (appendix B), but the existing legislative framework for data management does not support this intent.

There are a wide range of legislative instruments that persist in restricting data sharing and release (chapter 5). This complex legislative framework undoubtedly contributes to the focus on lore, as opposed to law, as a means to prevent expanded data release and use (Australian Institute of Health and Welfare, sub. 162; Queensland Government, sub. 207; Research Australia, sub. 117). The Commission has experienced these first hand in its own efforts to research particular topics, notwithstanding its ability to draw on its own Act for access to data. There is no doubt that some move beyond a current interpretation of existing legislation will be needed if serious improvements in data sharing and release are to be delivered and the Australian (and most state and territory governments') policies on openness is to be delivered (criteria 2 and 7).

Two Australian jurisdictions — New South Wales and South Australia — appear to be leading practice in this regard. New South Wales has adopted, and South Australia is in the process of adopting, overarching legislation designed to authorise release of public sector

information and clarify the legislative obligations around data sharing and release. This legislation has the following features:

- overarching legislation that cuts through a myriad of secrecy provisions by having authorisations and prohibitions all in one place
- provides clear permission to take a positive approach to release of data, subject to particular safeguards as per New South Wales
- explicitly endorses a risk-based approach, as per the South Australian bill currently before parliament (appendix C).

The New South Wales approach in adopting legislation to ensure data collectors and guardians knew that change had been authorised appears to have been highly effective, albeit placing emphasis on data sharing within government. Separate legislation supports proactive release of government information in New South Wales (*Government Information (Public Access) Act 2009* (NSW)).

Advantages and disadvantages of a legislative approach

Overarching legislation would strengthen the authority of institutions currently weighing up requests for access to identifiable information by clarifying the legislative framework applicable to the request and addressing legislative complexity. This would be in line with the previous recommendations of the Australian Law Reform Commission (2010) to simplify the 506 secrecy provisions in 176 pieces of Australian Government legislation.

The use of a dedicated legislative instrument dealing with data access at the Commonwealth level could have several other *advantages* including:

- greater clarity and permanence of approach across the whole of government — clarifying 'lore' versus 'law' and less exposure to policy fashions
- a more coherent and systematic approach to data sharing, linkage and release
- a stronger signal of commitment by government to increased and improved data use and exchange, and a consequent reorienting of views towards data as an asset, to be catalogued and maintained, rather than simply an overhead cost
- the ability to cut through the plethora of regulatory impediments to data sharing and release, as described in chapters 3 and 5.

A legislative approach presents several potential *disadvantages* that should be minimised in the design of the legislation, including:

- a need for careful design to avoid an approach that is unable to adapt to future shifts in technology or data practice (in contrast to criterion 9)
- unanticipated effects at the interface between existing legislation and a new, dedicated data sharing legislation

- the creation of an opportunity to demonise data use and exploit community wariness for political gain.

These are mostly problems of poor implementation; the concept is not damaged by the need to be aware of these cautions.

To be most effective, any new legislation should consider closely new state and territory legislation and allow for consistency of operation where possible among the Australian, and state territory and local governments without jeopardising overall reform. It should also take into account Commonwealth constitutional powers (particularly limitations for state instrumentalities), while providing as broad a coverage as possible, and an opportunity to share data equally, without the contention that can come from current processes and their direct and indirect impact on Commonwealth-State financial relationships. An Implementation Plan with clear strategies to address this would be essential to effective movement in this area. Chapter 9 discusses these matters further.

Cultural change through institutional reform

Incremental institutional reform is unlikely to ever be effective

If Australia is to realise the full value of its data holdings it will be important to find ways to achieve meaningful and long-lasting cultural change in the way Australia's data is shared and used. By itself, legislative reform is unlikely to be effective in achieving this. Even with permission to release granted by the legislative reform discussed above, this would not guarantee *actual* release (criterion 2).

Although leadership from a central agency could encourage incremental cultural reform, data-holding agencies would remain the masters of their own level of commitment to centrally authored change. While this approach would be the path of least resistance and would not require data custodians to relinquish control of their datasets, the Commission does not think this model of incremental cultural change could be effective. Complex approval processes, fragmented data releases, distrust within and between jurisdictions, and a general culture of risk aversion would likely remain — the precise problems that triggered this Inquiry in the first place.

These issues require a comprehensive solution — namely, institutional reform that delivers a widespread *determination to release*, and coordination and cooperation between the Australian, and the state and territory governments.

In evaluating models for institutional reform, we have taken into account the principle of *subsidiarity*, which is that decisions should be made at the lowest, least centralised level possible. Another important consideration is whether the establishment of any new entity delivers significant additional benefits (*additionality*) to justify the outlays required and the flow on effects.

Addressing fragmented data releases

Allowing data custodians to themselves decide what data should be shared and release can lead to highly fragmented datasets that are of reduced value, particularly when compared, for instance, to the value that can be derived from an integrated, longitudinal dataset (criterion 3). Models for coordination in release are evaluated below.

Coordination at the jurisdictional level

Given the differences between jurisdictions in legislative and institutional structures for data sharing and release (appendix B), the simplest approach to coordinating the sharing or release of data holdings is to do so within a particular jurisdiction. This is the approach currently adopted by New South Wales. The NSW Data Analytics Centre acts as the repository of thinking and agent of change, facilitating data sharing between agencies, and acting as the central data analytics centre for New South Wales by conducting analytics on agencies' data and providing them with the results. It is an attractive model, although focused on interagency sharing for the most part. As it keeps control of state data within the state, there is likely to be lower resistance to implementing this approach compared with other approaches discussed below.

But although this progress within jurisdictions is desirable in its own right, jurisdictions are unlikely to design and implement agencies similar to the NSW Data Analytics Centre in a coordinated fashion. And the benefits of data holdings are enhanced when, for instance, linkage can occur across jurisdictional boundaries (Data Linkage Branch (Department of Health WA), sub. 13; SA NT DataLink, sub. 123).

Coordination at the sectoral level

Many important datasets (such as health and education) cut across jurisdictional boundaries. Compared with focusing on coordinating data sharing and release within a particular jurisdiction, there are likely to be more gains from designing a release authority around sectoral datasets, even if they cut across jurisdictional boundaries (i.e. to achieve datasets that achieve Commonwealth-State integration). In practice, issues that the community is interested in, such as health service provision, will require data from multiple jurisdictions to assess them effectively.

The Australian Institute of Health and Welfare (AIHW) is a good example of a release authority that achieves sectoral coordination. It has a national cross-jurisdictional remit. One of its more recent key roles has been to build performance indicators and targets for national agreements on, for instance, Indigenous health and welfare reform, between the Australian and the state and territory governments.

If the AIHW model were extended, other release authorities with relevant sectoral expertise could be appointed to develop sectorally consistent national datasets. To avoid

the bottlenecks that have plagued data integration thus far, it would be important to have a sufficient number of release authorities able to undertake this role. In order to build coalitions of trust between jurisdictions, data collectors would need to be confident that the release authority would act in the interests of all data providers — for instance, the AIHW has representatives from different sectors and jurisdictions on its board. Ideally, the release authority would, in principle at least, not be the funding responsibility of any one collecting agency, and clear objectives for data curation and release would need to be set by participating agencies at the start.

In summary, this model for release authorities is attractive (particularly compared with a more centralised approach, discussed below) because it inherently facilitates trust within a sector (via sectoral expertise) and between jurisdictions (via cooperative governance).

Coordination by a single national entity

A centralised approach would involve setting up one agency to be the central sharing and release point of government datasets across multiple sectors and jurisdictions. Since setting up a new entity when the Australian Bureau of Statistics (ABS) has indicated it is open to undertaking the role (sub. 94) would seem obtuse, the analysis of this option focuses on the ABS as a role model.

The argument that has been presented to us is that this model would create a clear structure of responsibility for data sharing and release. It has also been argued that this would enable better data mining and analytics, once all data is linked and readily accessible from a single source. According to the ABS (sub. 94, p. 16), a 'centralised model facilitates the management of risk, security and costs'.

It is plausible that Commonwealth-level sharing of data would be improved by the ABS's well-deserved reputation for technical competence, which may in consequence ameliorate some *intra*-governmental distrust. And individual state and territory bodies may also trust the ABS more than a Australian Government agency counterpart.

But although there are benefits from adopting a more centralised approach that cuts across jurisdictions, it does not follow that a completely centralised approach should be adopted. A single national body has no in-built mechanism to build trust within a sector and between jurisdictions (criterion 4). Indeed, its broad remit may well risk it being perceived with suspicion as 'external' to a particular sector, rather than a trusted broker of data with the necessary sectoral expertise.

There are further problems that come from a centralised body — problems with culture and incentives against action. For the ABS, the motivation to minimise risks associated with data access in order to protect its trusted position as the primary national statistical collector would be strong. This is not irrational and unreasonable behaviour, it is common sense to prefer not to jeopardise core activities when presented with new activities. And while at one level the lowest risk sounds good, the purpose of institutional reform is to

shift organisational cultures away from risk aversion and towards the well-judged and managed taking of risk, otherwise criteria 2 and 7 will be frustrated. Finally, depending on the other priorities of an agency with the span of the ABS, there is a risk that it too may become — simply by virtue of limited resourcing — an additional bottleneck in data release.

In summary …

Basing sharing and release authorities within particular jurisdictions is likely to be the most easily implementable option, but cannot be expected to occur comprehensively (that is, across the nation) within any useful timeframe or in a coordinated fashion. Moreover, it does not contain a mechanism to overcome sectoral or jurisdictional mistrust, or legislative restrictions on sharing data between jurisdictions — which is particularly relevant where there are split-service delivery arrangements like in health and education.

Having sector-specific release authorities that span jurisdictions allows both impediments to be directly addressed as they would have the accountability and incentives to deliver these outcomes. The principle of subsidiarity would indicate that only decisions about datasets that have some national significance should be made at a national level.

Further, not all release authorities would be involved in integrating Commonwealth-State data but where this occurs, cooperative efforts would be required to develop the model. A central Australian Government control model is unlikely to induce the confidence of states and territory governments in the way that a jointly governed entity could.

Trust between data custodian and user (or between custodians, in the case of sharing) is essential in all circumstances (criterion 4). Changing the law can assist in altering the culture of agencies, but is insufficient to deliver trust in sharing.

Many options, reflecting the nature of working relationships, are likely to be used to build and retain trust as opportunities to access data grow. And there are already models in place to build upon, such as creating a cross-jurisdictional working group, or better use of COAG — although in practice it seems improbable that Ministers will want to determine which entity is a release authority. Another option would be to establish a cooperative governance structure for release authorities (such as exists already in the AIHW) to involve all levels of government in the decision making process. This is an attractive model, particularly when combined with an institution that has sufficient sectoral expertise to grant it legitimacy. Chapter 9 outlines a way forward.

Different models for interoperability with release authorities

There are clear benefits from taking a more coordinated approach to developing datasets. But greater coordination does not mean that data collectors need to relinquish control over their data altogether. This section evaluates the pros and cons of different institutional

models to achieve interoperability (criterion 3) (box B.9) — but this analysis is preliminary in nature and responses to this draft Report may alter the Commission's final view.

Broadly speaking, there are several ways a release authority could be designed:[52]

- *Centralised:* a single, dominant provider of data dictates the presentation and access methods of the data — for instance, the ABS Census.
- *Aggregated:* An aggregator obtains the data in whatever way they can and transforms it to ensure consistency, providing a uniform interface for users of the data and bearing the cost of doing so — for instance, the Atlas of Living Australia.
- *Brokered:* Data providers make their data available via an API in whatever form is most convenient for them. The broker dynamically accesses the data and transforms it to a uniform interface for users. The cost is primarily shared between the data provider and the broker.
- *Federated:* Data providers make their data available via an API using an agreed set of community schema and terms. Users can access this API directly. The cost here is primarily born by the data provider and is the lowest cost overall (across users and providers) as long as there are sufficient users (Box et al. 2015).

Submissions to this Inquiry (box 8.3) mainly focused on the benefits of an aggregated model versus a federated model.

The benefit of an *aggregated model* is that it makes analysis, processing, and standardisation of the data easier. However, having all the data stored in one location poses significant data security risks (contrary to criterion 8) — it has the potential to act as a 'honeypot' for hackers. Further, an aggregated model requires the aggregating body to understand and deal with all the issues and inconsistencies of the data on an enduring basis (as updates are provided). Establishing an aggregated model would also appear to involve a substantial investment in infrastructure and development given the scale of data integration and storage required. Aside from cost — which would need to be benchmarked against other alternatives, rather than against an expectation of no spending at all — as noted earlier, an aggregated approach contains no inherent mechanism to overcome jurisdictional or sectoral mistrust, a primary goal of enduring reform and cultural change; and a preferable aspect of an enduring solution.

By contrast, while a federated model could result in a less standardised approach to data curation (and would require regular conformance checks), it is much more likely to offset jurisdictional distrust compared with a more centralised approach. It also offsets many data security issues, given that there is no central 'honeypot' for the data.

[52] Another interoperability model is point-to-point, where a user gets the data in whatever way it works for them. In this scenario the cost of translation occurs for each user and total cost is high if there are many users. Since this model implies that a core function of a release authority, adopting best practice in a sector, does not exist, it is not discussed further in the text (Box et al. 2015).

Box 8.3 Participants' views on models of interoperability

- The Australian Bureau of Statistics (sub. 94) envisaged a whole-of-government model, where the central data agency will have access to government agencies' datasets, and will link them to a centrally stored 'spine', such as an address or a business register. This linked dataset would be the 'single source of truth' (p. 16) for enduring linked datasets, contributing to program evaluation, research and official statistics. Data produced by the central entity would be open to the public, or accessible by trusted users in the case of identifiable information, according to the type of data being released, the risk of release and how that risk is managed (for instance through trusted user models).
- The Department of Social Services (sub. 10, p. 7) suggested that a whole-of-government 'data analytics hub' would operate as 'an infrastructure-light, minimally resourced and nimble coordinating entity [that] would focus on managing appropriate data linkage arrangements and maintaining agreements between agencies, while custodians remain responsible for key data assurance and maintenance functions, leveraging current investments in business intelligence and physical infrastructure'. The Department believes that such a model is scalable, and more datasets can be added to the registry as required, to include Commonwealth as well as state and territory datasets.
- The Australian Institute of Health and Welfare (AIHW) (sub. 162, pp. 14-15) saw a role for supporting data curation across government, ensuring the datasets use common definitions and formats so that is easily shared and linked, but argued this could be done through a central agency, a small number of agencies, or done collaboratively through shared resources and recording portals and with open supporting infrastructure and software. It argued the final approval for sharing and linkage should remain with the individual data custodian for each dataset (p. 15):
  - A light touch approach is possible here where each dataset is used differently, in a fit for purpose and resource appropriate way, however for this data system to be flexible enough for all purposes it should also have the capability to handle the most sensitive and highly curated datasets. Interoperability issues can be addressed with … data transmission checks and … data standards.
- The work of the Data Linkage Branch in Western Australia supported a somewhat similar federated data system approach. The branch acts as a central hub for data linkages using data from up to 13 separate sources across the Western Australian Government. It assesses requests for data integration, and works with custodians to obtain their approval to release data (Data Linkage Branch, sub. 13).
- The Australian Taxation Office (sub. 204, pp. 5-6) commented that standardised data storage and consistent metadata would increase system interoperability, and would reduce the need for individual agencies to maintain the facilities' hardware and software. It also supported a cooperative approach across multiple agencies to improve data connectivity.

However, while allowing the data to be curated by the original data custodian is likely to increase the regularity of the updates to the data (*federated model*), giving the release authority the ability to curate the data (*aggregated model*) could provide it with a secondary revenue source to help support and retain its capability.

We are seeking views on this point.

Ensuring safe sharing and release

Regardless of the institutional change that takes place, mechanisms will be needed to ensure the system functions well and that data is shared and released in line with best practices.

Trusted access models

Public sector data is shared and released for a variety of purposes. Sometimes it needs to be shared with other agencies or private or not for profit entities for service delivery (think health or education). In such cases, the data shared may be identifiable. In most other cases — for policy development or research purposes — data can be shared within the public sector and with the research sector in a de-identified manner. This sometimes, but not always, occurs through trusted access models at present (chapter 3). And where data is publicly released (for instance, on data.gov.au) an additional level of confidentialisation needs to be applied to manage risks such as re-identification.

Trusted access models provide certain approved individuals greater access to sensitive data, either directly or via a remote connection. Two important principles guide the process of gaining trusted access. First, individual users must be securely and appropriately identified (as opposed to issuing organisational licences to access data, which limit data custodians' knowledge of individual users). Second, the safeguards around data access are determined based on a holistic examination of users' needs and the capacity of users and systems to keep the data safe. Trusted access models involve an assessment of data users' capabilities, secure computing arrangements and legal undertakings to ensure information is handled securely in a risk-based way — they manage privacy risks (criterion 8) in a flexible way (criterion 9).

While trusted access models are a promising development and appear to be gaining in popularity (chapter 3), they do not appear to be as widely used as they could be. But any expansion of trusted access models would need to address the following issues:

- *Who should be able to use trusted access models?* Most examples of trusted access model use in Australia presented to us appear to be used by the public and research sectors (chapter 3). But with the increasing delivery of government services and outsourcing of expertise to the private and not-for-profit sector, there may be cases where access should be extended to these sectors as well in a carefully controlled manner for specified purposes. Denmark includes in its trusted user model individuals working in the private sector; and includes business data in the datasets that can be accessed by trusted businesses from other sectors.
- *At what level should approval be granted?* At present, approval for use tends to be granted to individual users who are generally required to sign a range of legal undertakings and undergo output checks. Some participants have suggested granting approval to trusted organisations rather than individual. While this makes managing staff turnover on a project easier, it makes accountability for the individual data use

more difficult if it is not clear who is actually using the data. Approval also tends to be granted for a fixed period of time. A requirement for periodic renewal can make longitudinal work difficult (for example, comparing five-yearly releases of data).

- *What type of data should trusted access models be used for?* In Australia, the majority of datasets provided via trusted access models are de-identified data about individuals (mostly demographic and health data). While in most cases providing access to de-identified information will be sufficient, in some cases trusted access models could be used for access to identifiable information (for instance, for policy delivery purposes). The advantage of such an approach is the existence of greater controls on who uses the data and how it is used which is likely to build trust. However, establishing trusted access models can be costly, so any expansion would need to take into account how they should be funded, and whether users should be charged on a cost-recovery basis as this can be quite expensive.

Chapter 9 proposes reforms designed to achieve greater use of trusted access models in line with the wider implementation of a risk-based approach to data sharing and release.

Capability

Concerns about the ability of a release authority to build and maintain capability to, for instance, administer a trusted access model, might reduce trust in the overall institutional framework (criterion 4). One option to resolve this is to use an accreditation approach, with accreditation for data users awarded by a separate entity.

An accreditation mechanism could be used to certify that data sharing or release entities have the capability to perform their intended function, or that their internal processes (such as de-identification processes) are best practice. Accreditation would remove the uncertainty around technical and legal requirements that leads to what the Department of Social Services (sub. 10) termed 'agency inaction' and limited data sharing and release. In administering this accreditation process, it would be important to strike a balance between being restrictive enough to engender trust, but not so restrictive as to create bottlenecks in the sharing and release process (criterion 1). The accreditation body would wish to maintain contact with latest developments in such de-identification techniques as part of its normal remit, and relationships with release authorities would need to be formed early for this purpose. This accreditation path could be adopted for Commonwealth-only release authorities (such as the Australian Taxation Office and the Department of Social Services), but further consideration once the model is well-defined would guide the path of implementation between the Commonwealth and the states and territories.

Implementation through leadership

Establishing an institutional framework for data sharing and release will not be enough. It will be important to establish a body responsible for implementation of reforms to Australia's data infrastructure, and for leadership in building coalitions of trust between

jurisdictions and sectors. Submissions to this Inquiry and international experience have suggested several possible models that could achieve this.

A designated agency with responsibility for policy oversight will be important to maintain accountability for progress and outcomes and champion the necessary cultural change. There are already centres, units or departmental areas across government, at the Commonwealth and state and territory levels, that oversight data policies and agendas. At the Commonwealth level, this includes the Public Data Branch within the Department of Prime Minister and Cabinet, the Public Sector Data Branch at the ABS, and the recently established Data61 in CSIRO. And states and territories have also been active through the establishment of data analytics centres (in New South Wales and, prospectively, Victoria) or departmental oversight of whole-of-government agendas (such as in Victoria through the Department of Treasury and Finance) (appendix B).

Given the breadth of government data holdings and the multitude of legal and technical considerations in sharing and release of diverse datasets, the agency will need to be sufficiently authoritative to genuinely lead. It must be able to judge both sides of the resource question: an agency simply focused on reducing spending and with limited experience in developing legislative reforms and explaining them in public before, for example, a Parliamentary Committee, is unlikely to be effective. Finally, a transparent accountability mechanism (criterion 7) will assist with implementation — for instance, having agencies publicly report on the implementation of the reforms. A way forward is proposed in chapter 9.

8.5 Greater openness by the research sector

As much as data custodians and legislators should be adapting to the ubiquity and essentiality of data analytics opportunities through increased openness, so should researchers. Researchers are an important group of data generators and users in their own right. Within research fields, increasing access to, and sharing of, data has the potential to generate significant gains and is clearly currently less than optimal.

Much research is publicly funded. But in the vast majority of cases, neither the public nor the bulk of any research community has access to datasets generated by each project, despite there being a clear public interest in this occurring (criteria 1 and 2).

Without more open access to research data, reproducibility is a major issue. Further, there can be significant duplication in data collection which is a waste of resources — linking data is a costly and time consuming process. Increasing the re-use of research data — which includes *not* destroying unique datasets at the end of research projects unless demonstrably preferable (chapter 5) — can improve the reliability of scientific findings through replication studies and increase the rate of scientific discovery through more thorough interrogation of results. We have found in chapter 3 that all research data should be made widely available to trusted researchers and identified some barriers that can

prevent this occurring. Given the strong public interest in making research data widely available, any countervailing considerations do not negate the principle, but are considerations that should be taken into account in reform design. This section discusses options for reform to make research data more widely available (criterion 2).

Reforms to open up re-use of research data

Strengthening government policies

There is currently limited onus on researchers to make their data available following the completion of research projects — chapter 3 outlined some of these existing arrangements.

For instance, the *Australian Code for the Responsible Conduct of Research* (NHMRC, ARC and UA 2007) requires only that that researchers *should* make data available for use by other researchers unless it is prevented by ethical, privacy or confidentiality considerations. It could instead mandate that researchers *must* make their data available in the absence of a compelling reason. In practice, this approach would be blunt and heavy handed given the complexity and variety of research activities. Such an approach would require a much broader range of exceptions than is currently in place.

Instead, a set of principles could be developed that provide guidance on how to resolve these competing considerations. Some work to achieve this is currently on foot. The National Collaborative Research Infrastructure Strategy (NCRIS) is currently running a consultation on Australia's 2016 National Research Infrastructure Capability Roadmap, and an Open Access Working Group (involving the Department of Education and Training; the Department of Industry, Innovation and Science; the Department of Health; the Australian Research Council; and the National Health and Medical Research Council) has also been established to look at this. It is too early to comment on this process, but the lead central agency should not lose sight of it. Australia does not yet appear to have a clear champion of this policy of greater openness.

Conditions of funding

As a primary sponsor of research, there is serious potential for the Australian Government to influence data sharing practices by making data sharing a condition of funding. This does not seem to occur as much as it could. General statements of positive intent can be found but results cannot.

A stronger stance would be to *require by default* researchers to make their data accessible in the absence of legislated barriers or timing considerations, and funding to research institutions could be prioritised based on their record of making data open. Funders using Commonwealth money should monitor and annually publish tables showing the availability of data following the completion of research projects, and ethics committees

should not be used to impose barriers to wider use of data within a trusted research community.

Whilst this approach may have an tangible impact, it should be noted that its scope would be limited by the proportion of research that is government funded relative to what is funded by other parties. Nevertheless, this is still a significant amount — about $8.6 billion each year, of which about $2.9 billion goes to universities, and another $1.8 billion is provided to publicly funded research agencies such as CSIRO (NCOA 2014).

Building on existing journal publication requirements

In recent years, journals have been increasingly requiring researchers to provide access to data. *Nature*, for example, requires authors to 'make materials, data, code, and associated protocols promptly available to readers without undue qualifications'. In a similar vein, *Biostatistics* has implemented a multi-criteria reproducibility rating, with separate scores for the availability of data, code, as well as whether an editor was able to reproduce the result. In cases where publishing authors are not the primary data custodians, requiring them to register their use of the data on a portal could build on this and may improve recognition for researchers that generate and provide data. However, while these initiatives are encouraging, given the plethora of journals based overseas, a more comprehensive and coordinated approach may be difficult to implement from Australia's perspective.

Institutional issues

Research institutions are currently required by the Australian Code for the Responsible Conduct of Research (NHMRC, ARC and UA 2007) to have policy on the retention of research data. They are also required to provide secure research data storage and record-keeping facilities. There is less published guidance on responsibility for maintaining data and ensuring it is suitable for subsequent reuse.

Making data available for reuse can be a resource intensive process that requires specific skills and experience. However, the amount of resources required can also be exaggerated. Many of the unique datasets discussed here are snapshots in time, thus curation may not be as relevant — but access is, which is far less expensive. Where it is seen to be desirable to curate and update the research data (criterion 3), dedicated responsibility for curation would require funding.

Available options for allocating responsibility for data curation and sharing vary according to their degree of centralisation. Expanding the role of the Australia National Data Service (chapter 3) to provide data storage and sharing services could provide a centralised approach. This may offer efficiencies over current practices. It may also enable streamlined procedures for access. However, given the varied and specialised nature of research datasets, a centralised approach may still require significant input from data generators for metadata consistent with particular datasets. At the other end of the spectrum, a federated

approach would place responsibility for data curation and sharing with researchers or the institutions in which they perform their research. Such an approach would offer fewer benefits from scale, but would allow a more flexible approach to the curation and sharing of data and perhaps provide greater incentives for cultural change within the research institution.

This is roughly what is meant to happen now, but does not — ways to provide additional incentives to do this (such as prioritising funding based on openness) would be needed. Potentially, additional funding within research grants could be required, given the resourcing curation and sharing require. Finally, re-use of sensitive researcher data could be facilitated by access to trusted user models for this purpose — availability of trusted access models for this purpose is currently limited.

There is room for significant improvement in the research sector. Researchers often complain about the lack of openness of public sector data, but their sector remains vastly behind the public sector in terms of openness and availability of data — the pot calling the kettle black, as it were.

8.6 Greater openness in private sector data

As discussed in chapter 4, the private sector collects, stores and uses a vast amount of data, and is almost certainly now the dominant controller of data in the economy. A number of participants to this Inquiry have suggested situations where there may be a case for additional government intervention in support of data sharing or release (Netapp Inc., sub. 166; SA NT DataLink, sub. 123). These situations can be broadly grouped as follows:

- Insufficient business-to-business sharing:
  - concerns about monopoly holdings of data and misuse of market power
  - insufficient private sector exchanges occurring.
- Insufficient information released for the community benefit:
  - insufficient information disclosed for markets to function properly (Business Council of Australia, sub. 191)
  - situations where private sector incentives to release data do not reflect the broader community benefits from its release (Medibank Private, sub. 98).

The key focus in our consideration of these concerns is the delivery of net benefits for the community (criterion 1) while preserving commercial incentives to collect and add value to data (criterion 6).

Business-to-business sharing

Market power concerns

Having access to a large amount of data can give a business — particularly a large, vertically integrated business — a degree of market power (chapter 4). Some participants have argued that such market power could be used to deter new entrants to a market. This section discusses existing and potential arrangements that could address market power, and then looks at the extent to which private sector data exchanges occur already.

Application of part IIIA

Part IIIA of the *Competition and Consumer Act 2010* (Cth) (CCA) allows third party access applications to be granted to infrastructure or its equivalent necessary to deliver a service. The access arrangements only apply to services, as defined by section 44B of the CCA.[53] State and territory governments also have generic, industry-specific and facility-based access regimes. Under both the federal and state regimes, if an access application is granted, a price for access must be determined.

The law is unclear on whether these access regimes could be applied to data holdings — the intellectual property exclusion will not cover all data (appendix C).

Leaving aside definitional debates, there appears to be no case for applying the national access regime to data. The Commission has previously indicated that the national access regime should only be applied where there is a lack of competition induced by the existence of a natural monopoly (PC 2013b). Data holdings do not appear to display the characteristics of natural monopoly. Actual barriers to the collection of data are few; rather, what is displayed in markets where very few parties or only one party has access to most of the data is that technology — usually combined with high quality of service — has created a period of dominance but that the ability of a new competitor to enter and start collecting its own data is not prevented by any natural factor. Rather it is the best idea that dominates — until it does not.

As investigated at length in other parts of this Report, consumers may need regulatory support to enable them to move quickly (with their data) to the next good idea, but the use of Part IIIA is about firms accessing other firms' monopoly assets. And as noted above the relevant policy criteria for intervention do not appear to be met.

Extending the National Access Regime to explicitly cover data would have as a consequence several significant downsides. First, since application of the access regime is

[53] 'Service' is defined in the CCA as a service provided by means of a facility and includes: (a) the use of an infrastructure facility such as a road or railway line; (b) handling or transporting things such as goods or people; (c) a communications service or similar service; but does not include: (d) the supply of goods; or (e) the use of intellectual property; or (f) the use of a production process. except to the extent that it is an integral but subsidiary part of the service.

itself a judgment that trades off the scale efficiency of a single supplier under conditions of natural monopoly for the dynamic efficiency potential from increased competition, there will almost certainly be efficiency losses if conditions in forcing access to data do not meet those characteristics. Moreover, if experience with application for access to essential infrastructure in the current Part IIIA regime is any guide, disputes over access can be long running and entail significant legal processes and costs. Thus, further net cost is likely (contrary to criterion 1).

Part IV of the Competition and Consumer Act

Part IV of the CCA already deals with general claims of misuse of market power, other than those involving natural monopoly. It contains several broad provisions prohibiting conduct deemed likely to lessen competition. These include:

- agreements that restrict dealings or affect competition (section 45)
- misuse of market power (sections 46 and 46A)
- acquisitions that would result in a substantial lessening of competition (section 50).

In addition, the CCA prohibits some specific conduct outright, including:

- agreements containing 'cartel provisions' (Division 1)
- price fixing between competitors (section 45C)
- boycotts (sections 45D, 45DA, and 45DB)
- exclusive dealing (section 47)
- resale price maintenance (section 48).

There is some jurisprudence indicating that section 46 can be used to create an access regime as per *NT Power Generation Pty Ltd v Power and Water Authority* (2004) 219 CLR 90. Several expert commentators — including Professor Corones (quoted in SELC 2006), Nielsen and Nicol (2008) and Lawson (Lawson 2008) — have noted that the remedy provisions of the CCA are broad enough to enable this. For example, under section 80 of the CCA, the court can grant an injunction on the terms it considers appropriate, including 'requiring a person to do an act or thing'. Section 87 gives the court broad powers to make any order it thinks appropriate. Section 46 has been used several times to extract access to copyrighted information, although this is not common (PC 2013a).

Private sector data marketplaces

There are a number of existing private sector mechanisms that already allow business-to-business data sharing (chapter 4).

While bilateral commercial agreements and data marketplaces are increasing data access and sharing, it may still be the case that broader access than that enabled by these

platforms could deliver public benefit. Any action to increase access to privately held data would, however, need to be premised on a clear articulation of net benefits to the community (criterion 1) and a demonstration that access to the relevant data is not able to be secured through other means including through existing private sector data marketplaces and platforms.

Westpac Banking Corporation (sub. 197) argued the government should do more to support the voluntary sharing of de-identified aggregated information and insights in a controlled manner through the use of private marketplaces and bilateral arrangements. Rather than mandating the use of specific mechanisms of data-sharing such as APIs, it suggested the government should, as a first step, encourage organisations to independently develop mechanisms for facilitating data-sharing for both customers and third parties, such as requiring machine-readable rates and fees. It considered the government and relevant regulators would then continue to monitor data-sharing developments in the private market.

Westpac Banking Corporation (sub. 197) also argued that the use of bilateral commercial agreements and emerging data marketplaces are the most effective mechanism to foster competition, innovation and the economic incentives required for continued investment in aggregated datasets and insights within robust security and privacy controls.

While this could be disputed on various grounds unrelated to data management, when it comes to data (the subject of this Inquiry) the comments are consistent with how assets are managed in markets. And the principle that underpins this Inquiry is that data is now a valuable asset — not an overhead, or a privacy issue, but an asset. It is desirable to see government interventions in markets occur (where justified by public interest considerations) in a fashion that as far as possible enables market participants to act as agents in markets would. In this context, it is essential that organisations retain the ability to actively vet and, where applicable, control third party access to their aggregated datasets and insights and, at their discretion, price access under commercial arrangements that reflect the commercial value of data in a data-sharing transaction.

The case for assisting new entrants *in any market* with access to competitors' assets is poor to non-existent in the absence of clear evidence of misuse of market power or adverse outcomes from the actions of a natural monopolist.

And even then, intervention would need to be a better way of solving the problem than the one put forward by this Inquiry — empowering consumers to direct their data to third parties (chapter 9). In the nascent fintech sector, consumers could under reforms outlined in this Report themselves choose to rapidly build up the data holdings of new entrants.

Looking across the economy more generally than fintech, access to the data holdings of incumbent firms would undoubtedly help a new firm, but this is not a socially or economically desirable role for governments. As long as business-to-business sharing already goes on, and consumers or markets are not suffering obvious harm from misuse of power, governments should not intervene at the firm level (criteria 1 and 6).

Releases that could benefit the community

The above has discussed business-to-business arguments. But there can be cases where greater private sector information sharing or release can improve market outcomes for *consumers* (criterion 1).

Alleviating information asymmetry

Information asymmetry often exists between buyers and sellers in a market when one party has more information than the other — suboptimal outcomes can occur when information is not available that is necessary for a market to function better. Existing regulatory disclosure obligations (for instance, to the ASX) and the proliferation of technology (for instance, the establishment of Whitecoat to publish fees and consumer reviews of some healthcare providers) can go some way towards addressing this. However, many participants to this Inquiry commented that in some sectors these disclosure obligations are insufficient to enable effective consumer choice, and private sector incentives are to maintain this information asymmetry (chapter 4).

Enabling wider public benefits

Another situation raised was where some of the data collected by the private sector may have the potential to deliver *significant public benefits* if shared more widely than dictated by the incentives of the private owners.

In some cases, business may want to share this information because they wish to be a good corporate citizen, but need regulatory or other permission. Google (2016) sharing Android mobile locations with emergency services when an emergency number is dialled in the United Kingdom is one such example. But cases where this may not occur include where:

- research into the causes and treatment of cancer is assisted by the collection and collation of data on all cancer diagnoses from the variety of private and public sources of such data, including hospitals, pathology laboratories and radiotherapy centres
- the management and coordination of the health system is assisted by the data that private hospitals provide to Medicare on medical services and pharmaceuticals (chapter 3).

Specific cases like these will need specific solutions. Where regulation simply could not envisage at the time it was drafted the opportunities available today, that regulation should be reformed.

Exceptional circumstances do exist

In sum, there are limited specific situations where the private sector lacks incentives to disclose information that is necessary for markets to function properly, or where there are

significant positive public benefits from this disclosure (criterion 1). In such cases there may be a role for government action to ensure broader access to data in the public interest. The benefits and costs of any such action should be assessed beforehand to ensure the least negative effect on incentives to continue to collect, curate and deliver consumer benefits and shareholder gains from data holdings (criterion 6).

A broad and effective approach to this at the level of datasets relevant to the national interest is addressed in chapter 9. Other less-than-national level data reforms will presumably be addressed by the Australian, state and territory, and local governments under their professed aims of improving data openness.

Implementing reforms involving private data

There are a number of existing arrangements designed to secure access to data that could benefit the community. Governments at both the Commonwealth and state and territory level presently buy data from the private sector. They have cooperative arrangements in some areas, such as prices monitoring and transport, to access and use private sector data as a supplement to the data they collect. Governments also fund, in part or full, a number of projects which the private sector then delivers — many of which generate significant data holdings. For example, infrastructure projects may generate significant data, such as geospatial data, usage data, and financial data — for instance, the UK Government now requires building information modelling to be transferred under its construction contracts (a data source enabling substantial project efficiencies and engineering exchanges) (BIMTG 2016). And regulatory disclosure requirements are likely to cover many of the remaining cases.

Options to build on these existing arrangements are evaluated below.

Much private sector information of public interest could be released

The Commission recommended in chapter 4 that where governments enter into contracts with the private sector that involve the creation of datasets, there should be consideration of the strategic significance of data prior to contracting — particularly where access to data might be necessary to determine whether contractual obligations have been fulfilled.

Under this approach, contract conditions would differ from situation to situation, with data access and value add potentially becoming a point of comparative advantage in tenders. Further, rights to access and disseminate data can be tied to funding conditions for private sector organisations undertaking government projects. The benefit of this approach is that the terms of release can be negotiated as part of the contract.

Expanding regulatory disclosure obligations

Information necessary for the efficient functioning of markets is already required to be disclosed under existing regulatory arrangements — either to a regulator (such as the Australian Prudential Regulation Authority or the Australian Securities and Investments Commission) or to the market (for example, under disclosure obligations to the ASX). The Comprehensive Credit Reporting scheme has been put in place to address the information asymmetry between lenders and borrowers in credit markets, but is not operating effectively yet (chapter 4). There is a wide range of regulators interested in the matter of keeping firms well informed, in the interests of efficient markets.

Red Energy and Lumo Energy submitted:

> The cost to industry implementing changes to its existing systems to facilitate greater access to data needs to be carefully considered. System changes to facilitate such increases in data availability will likely result in increased costs to consumers. (sub. 63, p. 2)

Similarly, ANZ commented:

> Increasing the availability of data would impose costs on current data custodians, particularly if the format of the data availability were prescribed. The cost of increased availability would vary with the quantum of data made available and the nature of its availability (sub. 64, p. 21).

While more could be done, the trade-off is between the additional cost of information provision and the ability of parties to take advantage of it.

Intermediaries can help consumers take advantage of information provided. For instance, comparison websites attempt to undertake the role of trusted intermediary with varying levels of success — conflicts of interest can arise, as can impediments (deliberate or coincidental) cause by the quality of data held by consumers.

These schemes can still work well — in the United Kingdom, such sites have been very effective in car insurance markets, for example — but are not a complete solution. At the consumer level, the Australian Government also makes some effort at private health insurance (privatehealth.gov.au) and energy provider comparison information (energymadeeasy.gov.au) and some state and territory governments do so as well. And in the private sector, Energy Tailors (chapter 2) helps consumers compare energy plans to get a better deal given their consumption patterns.

Introducing a right to transfer data, as discussed earlier in this chapter, allows consumers to overcome information asymmetry by allowing third parties to do the work of assessing what is relevant information. While it is desirable always for consumers to be fully informed themselves, it is often not a practical option.

Government-brokered release

There are several existing schemes where the government itself has a role in acting as the 'trusted broker' for the sharing and release of private sector data for the public benefit. For

instance, Geoscience Australia (sub. 211) curates private sector petroleum exploration data, which is released after a five-year moratorium period to allow companies time to profit from their exploration.

Trusted government bodies have also demonstrated they are able to share and release data relating to the private sector in a way that preserves confidentiality — the Business Longitudinal Analytical Database Environment is being developed by the ABS and the Department of Industry, Innovation and Science. It integrates administrative data from the Australian Taxation Office with collected survey data from the Business Characteristics Survey, Economic Activity Survey, Business Research and Development Survey and intellectual property data (Department of Industry, Innovation and Science, sub. 69).

These trusted broker schemes can help overcome private sector coordination problems and facilitate data sharing and linkages. But there would have to be a compelling public interest for government to involve itself here.

Private sector consultation and involvement will be crucial

Where cost-benefit analysis indicates that enhancing access to data will result in net gains for the community (criterion 1), action is justified to ensure that those benefits are realised. Government should consult with business to ensure the design and implementation of any schemes are as cost effective as possible. Private sector cooperation and involvement will be crucial in the implementation of any reforms, particularly to ensure the reforms minimise the effect on private sector incentives to collect and use data (criterion 6), and minimise any additional costs (such as small businesses earing a proportionately higher cost of software development to enable data release).

One option to achieve this might be the Australian Government appointing a data advisory council which would consist of members from different professional backgrounds who would facilitate increased data sharing and provide guidance on how best to achieve data access and sharing. A data advisory council could provide a useful mechanism to ensure cross-sectoral representation in data issues and have the ability to respond flexibly to data issues as they arise and the capacity to swiftly recommend solutions to emerging problems (criterion 9).

New Zealand's Data Futures Partnership is an example. Established in 2015 by the New Zealand Government, it is described as an independent 'cross sector group of influential individuals', dedicated to driving the overall work program and core deliverables of the Partnership. Among the Partnership's activities are brokering activities, identification of key problems and issues relating to data, provision of advice, and the formulation of solutions to system-wide problems. The Partnership's work is directed by a Working Group appointed by Cabinet and supported by a secretariat housed at Statistics NZ (NZ Data Futures Partnership 2015).

At this point, the Commission is not proposing such an entity, but seeks further advice.

9 A framework for Australia's data future

Key points

- This chapter presents the Commission's preferred data reform approach. At the centre of the proposed reforms is a new *Data Sharing and Release Act*, and an associated structure for granting access to datasets.
  - Together, these will provide a much needed framework for developing the fundamental asset that is digital data access and use, as an equal and complementary element to existing legislative instruments concerned primarily with privacy and security.
  - The approach takes account of the significant differences in data types and associated risks, while preserving private sector incentives for innovative use of data and strengthening individuals' trust and confidence in use of their data.
- A key element is to increase the control of individuals over their data.
  - Individuals' existing power to view and suggest edits to data held on them by business or government should be structured legislatively as a right to data access.
  - It should be extended to offer a right to have a data holder transfer, safely, an electronic copy of data it holds on the individual to another party.
  - Individuals should also be granted the right to opt out of data collection unless collection is necessary to satisfy legal obligations, forms part of a publicly funded dataset or is otherwise part of a broader dataset used for public benefit purposes.
  - Where data has been previously collected, it is considered not technically feasible or practical for individuals to stop the legitimate use of that data or have it deleted.
- Ensuring that more high quality data is available for use by governments, researchers, businesses and individuals is critical.
  - Under the proposed reforms, high value datasets that are likely to generate spillover benefits for the community would be designated as National Interest Datasets. These datasets would only contain non-confidential or de-identified data.
  - Access would be granted on an ongoing basis to trusted users initially, with the ultimate objective of public release in a form that is safe.
  - Existing arrangements for the sharing of identifiable data with trusted users would also be streamlined.
  - Data that is not about individual people or businesses would be made more readily available for broad use, by governments, consumers, businesses and the research community.
- While recent progress in policy and practices around data management are noted, the recommended reforms are intended to create a complete framework that is capable of enduring beyond current policies, personnel and institutional structures.

This chapter outlines the Commission's recommended approach to improving the availability and use of public and private sector data in Australia. The recommendations draw on the principles and options for reform outlined in Chapter 8. In the Commission's view, while increasing access to and use of data, the recommended approach will also ensure that individuals and businesses retain trust in the way data is collected, stored, managed and used, and remain motivated to collect quality data into the future, as discussed in chapter 10.

9.1 What is directing the recommended approach?

New technologies and analytical techniques are generating more data and increasing the scope to make use of data in innovative ways. Increased access to data can facilitate the development of ground-breaking new products and services, as well as incremental improvements in existing products and services. The potential value of data is tremendous; but so too is the scope for Australia to forgo much of this value by not enabling its use under the misconception that this would minimise risks to privacy.

The structures under which public and private sector data are generated, stored and used in Australia are ad hoc and not contemporary. The need to deal, head on, with contemporary data issues and enable opportunities will not diminish.

Incremental change to current data management frameworks will not suffice; fundamental and systematic changes are needed to Australia's data management frameworks. This conclusion is based on a number of findings:

- The nature of data sources and data analytical techniques is evolving rapidly, and will continue to do so.
- As data standards and metadata improve, digital data will be transferred across the economy, between sectors and across national boundaries with increasing ease. To ensure coverage is comprehensive and understandable, data management frameworks need to be consistent across the economy.
- The range and volume of datasets across the economy, held in the public or private sector, that *could* potentially be made more widely available, is monumental. While there have been noticeable increases in the sharing and release of data in recent years, these releases remain 'a pimple on the pumpkin' of data release possibilities. Incremental changes in the data management framework to date have failed to deliver a culture of making data available for widespread use. Other countries are making better use of their public and private data than Australia.
- As new sources, types and uses of data emerge, so too do issues that go to the fundamental rights of individuals to data held about them, and the need to maintain trust and confidence in data collection, handling and use.

The remainder of this chapter details a recommended approach that enables a fundamental change to the way governments, business and individuals handle and use data. This approach builds on recent progress in policy and practices around data management but is also intended to be capable of enduring beyond current policies, personnel and institutional structures. It takes account of the significant differences in data types and associated risks and uses, and recognises that incentives and trust have to be maintained.

The staging of reforms will be important — some will take time to implement and further consultation is required. Nevertheless, there is much that can be achieved in the near term that will deliver immediate benefits, and also lay the foundation for further progress.

Different data, different uses and risks

How data is collected (voluntarily or otherwise), the characteristics of data (personal, de-identified or non-confidential), potential uses, and the extent to which benefits derived from the use of data accrue to an individual user or the broader community, all influence the potential risks that attach to data sharing and release. So, for example, personal health data that is either identifiable — or if de-identified, could be easily re-identified — would fall at the high end of a risk spectrum. At the other end of the spectrum, routinely collected data of a program or process would, in many instances, be considered low risk.

Data management frameworks must reflect and address the attendant risks associated with different datasets. We are of the view that different data uses can require different data use solutions and different rules. This view parallels the findings of the New Zealand Data Futures Forum in its review of New Zealand's data framework (2014, p. 18). Such a tailored approach has significant merit, subject to the proviso that adequate safeguards are in place, and that it operates with clear guidelines.

In particular, rigorous assessment of *genuine* risk is needed to inform the development of effective risk strategies and controls. There are undoubtedly areas of ambiguity where it is debatable what the 'right' approach is to data access. Further complicating this is that risks are not just related to data characteristics but also vary considerably with who is using the data and for what purpose. To be clear, in situations where risks remain unacceptably high, even after the adoption of known safeguards, it is envisaged that data would *not* be released.

While recognising and managing risks is essential, it is also important that the value of data access be afforded due weight in considerations about the level of risk to be accepted. In the course of this Inquiry, we have become aware of many instances where risks are given too much weight and potential benefits of data access are given insufficient consideration, resulting in opportunities foregone (box 9.1).

The Commission's preferred approach is aimed at dealing more effectively with risks, but also at improving arrangements where ambiguities *do not*, in fact, exist in reality. There remains significant room for improvement regarding these less contentious cases — and

there is broad agreement that current frameworks are not working well. To put it starkly, we should err on the side of releasing datasets of high potential value to the community — with appropriate controls — and those that are low risk.

Box 9.1 Participant's views on trade-offs

Telethon Kids Australia:

> Unjustified fear of breaches of privacy and confidential data being used inappropriately, in spite of evidence of this being extremely low. (sub. 5, p. 5)

John D Matthews:

> Australian developments have been inhibited by concerns about whether it is possible to guarantee privacy protection. However, the experiences in Australia and overseas show that data aggregation projects have not led to breaches of privacy, and that any theoretical risks are more than justified by the countervailing public benefits. (sub. 36, p. 15)

The University of Sydney:

> We recognise the need to maintain privacy, security and confidentiality of public--derived data. However, for non--sensitive data existing limitations to access seem unnecessary … Removing unnecessary barriers to non--sensitive data must be a listed priority for this Inquiry. (sub. 35, pp. 1–2)

Sax Institute:

> It has been our experience, that the assessment of potential benefit of the data access is not given the same weight as the assessment of risk. This may be due to the historical lack of interaction between researchers and policymakers that results in poor alignment between government policy and research aims but there is great opportunity for this gap to be closed in the future with greater data availability and research that would then feed back into government policy and services. (sub. 56, p. 4)

Incentives and trust have to be maintained

The recommended framework is intended to foster the beneficial uses of data while managing the potential risks and costs. Maintaining a credible trust framework (such as that proposed) will be necessary to get widespread agreement for reform. Likewise, ensuring the recommended approach does not undermine incentives for businesses, academics and others to continue to collect data, and engage in value adding processes, is critical.

There will always be trade-offs between enabling access to data and managing potential risks associated with that access. The potential effects of a data release that is unintended or does not afford the desired confidentiality to individuals or businesses can be significant — but most often they are able to be, and should be, managed. The significant opportunity costs of limiting access in the name of privacy also need to be directly addressed. The solution lies in developing clear and well-understood policy objectives underpinned by robust risk management processes. A challenge in developing a more coherent approach is striking a balance between policies that provide a framework of safeguards and broad direction, while not obstructing much needed innovation and investigation.

Skeleton of the recommended approach

There are four key elements to the Commission's recommended approach:

1. Giving individuals more control over data held on them.
2. Enabling broad access to datasets that are of national interest.
3. Sharing publicly funded identifiable data with trusted users.
4. Releasing non-personal and non-confidential data for widespread use.

The Commission's multifaceted approach (figure 9.1) recognises the spectrum of risk associated with different types and uses of data, and the corresponding need for different risk controls and approaches to apply. It is acknowledged that the confidentiality of some data held by businesses is critical to their commercial operations, and so to avoid undermining commercial incentives, the availability of some data will remain a commercial decision. Where the risks associated with release cannot be effectively mitigated, the Commission's approach would not involve release, for example, in the case of genuinely commercial in confidence data, or data that is integral to the operations of governments and the security of the country. For the remaining bulk of data, the recommended approach to improving sharing or release is detailed below, and reflects a sliding scale of release strategies and controls commensurate with the potential risks and benefits of potential release.

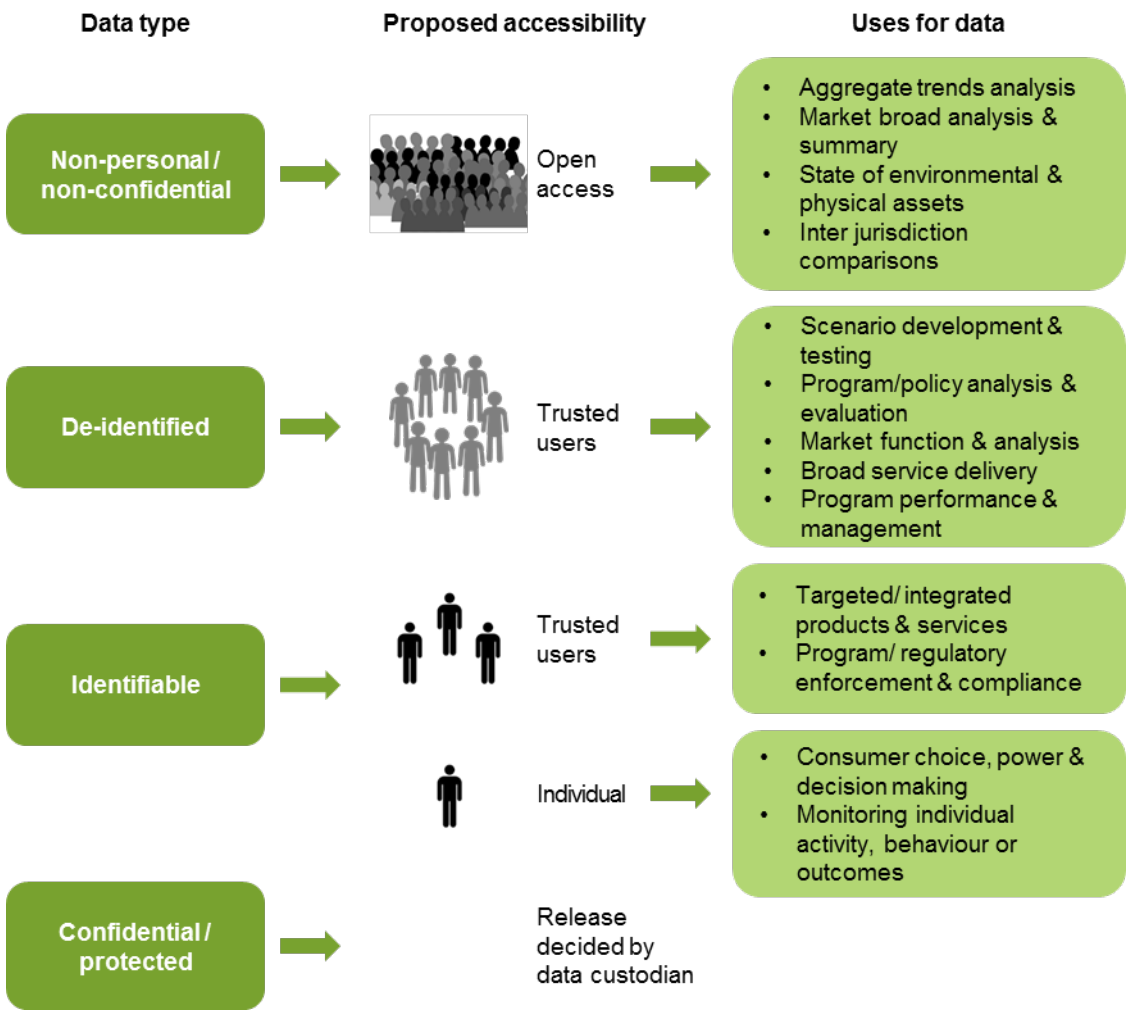
9.2 Element 1 — Giving individuals power in data held on them

The scope to provide consumers with more access to digital data held about them is considerable. And, in the face of the ubiquity of collection, the Commission considers that consumers should also have greater say — within limits — on what is collected about them.

Increased access and transferability gives individuals more control over the information held about them, affords individuals more choice about the products and services they consume, and is an avenue to improve market competitiveness. While consumers arguably already have access power, it can be unclear how to exercise this.

The Murray financial services Inquiry (2014) and the Harper competition policy review (2015) found that giving consumers more access to data about themselves would give them the power to make better purchasing choices. This in turn would benefit the economy as a whole. The reforms set out below include significant changes to access arrangements for individuals. These access arrangements are intended to apply to both public sector and private sector data. As noted in chapter 1, we see no legitimate basis for distinguishing between these in affording consumers more power to access and control data on themselves.

Figure 9.1 **Framework of the recommended approach**



The previous chapter detailed several options in the area of individuals' data access. These included giving consumers ownership over their personal data, and the right to be forgotten. The Commission's view is that these options would deliver questionable, if any, overall benefits, and, most fundamentally, would be unlikely to cope with the dynamic and rapidly changing environment in which data collection on individuals is occurring. The following sections discuss the Commission's preferred approach to designing a Comprehensive Right over personal data.

An updated definition of consumer data

Improved control over collection and access to personal information requires clarity on what is meant by it. The current situation is too complex. The new Comprehensive Right should apply to:

- personal information as defined in the *Privacy Act 1988* (Cth)
- all files posted online by the consumer
- all data derived from consumer's online transactions or Internet-connected activity
- other data associated with the transactions or activity that is relevant to transferring data to a nominated third party.

To be effective, in turn, a revised definition of consumer data covering these facets needs to be practical and implementable.

Given issues associated with different definitions of personal information for different purposes, defining 'consumer data' in the *Acts Interpretation Act 1901* (Cth) would guarantee wide application. Consequential amendments to other Commonwealth legislation would ensure harmonisation across federal laws. This definition would need to take account of other relevant definitions, such as applies to protected information under the *Tax Administration Act 1953* (Cth).

While this updated definition will better facilitate consumer rights over information about themselves, it is also likely to entail some associated regulatory impacts and costs. Businesses would need guidance around what steps would be considered reasonable to take to ascertain an individual's identity. If complex and time consuming steps are involved to 'reasonably ascertain' an individual's identity, for example, then businesses that collect data on their customers could be forced to invest more money into improving the way they store, retrieve and supply this data to their customers (Clayton Utz 2015).

While acknowledging these costs and the need to limit them, evidence from other countries that have more expansive, but credible, approaches to the definition and treatment of personal data, in particular the United Kingdom, suggest that they are manageable. As discussed in earlier chapters of this Report, the countervailing benefits of such a reform, particularly in the facilitation of greater consumer choice and the provision of a richer underlying information set upon which to base such choice, will be significant. Further, such benefits are likely to accrete over time.

The definition of consumer data for the purposes of the proposed Comprehensive Right, and associated oversight arrangements, will need to be framed in such a way as to take account of the extent of data transformation. In many instances, data is transformed by the holders of data to a point where it is no longer classifiable as consumer data as such.

For example, if the data point was received from an individual and subsequently remains substantially unaltered, such that it is able to be linked within the systems of the entity holding the data back to that individual, then it is desirable for the purposes of a Comprehensive Right that it be defined as 'consumer data'. Conversely, in instances where data is transformed to a point where it is no longer able to be re-identified as being about an individual within the entity's systems, it should no longer be classifiable for the purposes of a Comprehensive Right as 'consumer data'.

In implementing this approach to consumer data, a possible risk is the creation of incentives for entities to keep only de-identified data to avoid requirements to provide consumer access. Introducing an appeals process, and associated regulation, connected to the implementation of the Commission's proposed approach may merit further consideration, but would bring with it costs. Further views are sought on this and alternative remedies.

DRAFT RECOMMENDATION 9.1

The Australian Government should introduce a definition of consumer data that includes:

- personal information, as defined in the *Privacy Act 1988* (Cth)
- all files posted online by the consumer
- all data derived from consumers' online transactions or Internet-connected activity
- other data associated with transactions or activity that is relevant to the transfer of data to a nominated third party.

Data that is transformed to a significant extent, such that it is demonstrably not able to be re-identified as being related to an individual, should not, for the purposes of defining and implementing any Comprehensive Right, be defined as consumer data.

The definition of 'consumer data' should be provided as part of a new Act regarding data sharing and release (Draft Recommendation 9.11). Given the need for this definition to have broad applicability, it should also be included within the *Acts Interpretation Act 1901* (Cth). Consequential amendments to other Commonwealth legislation would ensure harmonisation across federal laws.

INFORMATION REQUEST

Further views are sought on the effects of providing access to consumer data, as defined. In particular, views are sought on the potential creation of incentives for deliberate de-identification of data holdings to avoid providing access, and whether effective and low cost remedies to such behaviour could be introduced.

Access and edit powers

Under the Commission's proposed plan, consumers would retain the power to view what information a business or government agency holds on them and request edits or corrections for the sake of accuracy. As per existing provisions, the capacity to have data edited would be a right (box 9.2) to request specific edits, not a right to have the edits actually agreed to unless they are correcting something wrong or misleading in the data.

Consumers would also have a right to be informed of disclosure of data by a data holder to third parties; and a right to be informed, and request reconsideration, of automated

decisions to ensure decisions are made on the basis of correct information. These rights could not, by law, be invalidated or expunged by Terms and Conditions, similar to the way other consumer protection laws cannot be invalidated by purported contracts or conditions of sale.

As detailed in appendix C, the Privacy Act already gives consumers powers to access data about themselves and request corrections, but consumers are generally not aware of who is holding what information on them and for what purposes and the format of access is not specified. As a consequence, few consumers are aware of the powers they have or how to exercise them (chapter 8). The Commission's recommendation neither reduces nor increases existing access and correction powers, but rather, formally defines them as a 'right', along with increased consumer protections elsewhere, to create a Comprehensive Right. This new Comprehensive Right would be enforceable in the same way that existing powers are — via complaint to the OAIC.

Box 9.2 **What do we mean by 'rights'?**

The Commission is proposing new data rights for consumers around data access, correction and transfer. These would be implemented in a manner consistent with the existing Australian legal framework for other rights. Australia has no Bill of Rights at the federal level, as exists in some other countries. Instead human rights are protected in the Constitution, common law and legislation (Acts passed by the Commonwealth Parliament or State or Territory Parliaments).

The Government would legislate (as the UK did in the limited sphere of midata), to impose obligations on businesses to provide in industry-relevant electronic form a customer's data back to that customer or on (safely) to that customer's authorised third party.

Source: Australian Human Rights Commission (2006).

There are many aspects of a consumer right that will need to be clarified in the final Report of this Inquiry. One that has been indicated already is in situations of personal incapacity, disability or death, or regarding data held by a business that has ceased to trade.

We welcome input on what access rights should look like in such circumstances.

Right to stop collection

Under the Commission's recommended approach, consumers would be able to request that a data holder stop collecting information on them (that is, they can 'opt out' of a collection process). This capacity to opt out would be subject to a number of exceptions — specifically, individuals would not be able to have collection cease if the collected information is required for a public interest purpose or the performance of a contract. These situations may include information:

- necessary for continued delivery of a product or service to the individual
- necessary to satisfy legal obligations or legal claims

- necessary for public benefit purposes (such as the maintenance of public health and safety, administrative purposes such as tax collection, security reasons, or collection for official statistical purposes)
- part of a national interest dataset (discussed further below)
- required for archival reasons — that is, keeping an accurate and useful historical record (such as newspaper archives and journalism).

The opt out right should be actionable by an individual at any time, subject to the above exceptions. It is envisaged that given the above exceptions, individuals would not be able to opt out of collection by a public sector agency, or by a private sector entity collecting data on behalf of a public sector agency.

The right to stop data collection would not include the option to have historical data deleted or use of it cease. This provision recognises that once data is integrated into a dataset or analysis, it is often costly or infeasible to extract it. It is also intended to ensure that individuals' opt out decisions do not decimate the investment that data holders have made in datasets and, in some cases, ensure that information on past activity is available to inform future activity (information on past medical procedures of an individual would be necessary for future medical treatment, for example).

We have recommended elsewhere that all public interest research be enabled by exceptions in the privacy legislation (chapter 5). Provision would also need to be made for the collection, use and disclosure of national interest datasets (discussed below). We welcome feedback on any other public interest uses of data that may be hindered by existing arrangements.

Data transfer

The right of individuals to request that data about themselves be copied from one data holder to another is a key additional power that the Commission is recommending be afforded to individuals.

The capacity for individuals, as consumers, to copy their data between service providers is an integral part of facilitating competition in markets and reducing barriers to market entry. It will also, like the power to opt out, improve consumers' ability to control their data and so should serve to lift confidence over time that they too — along with governments and businesses — can choose how and when to use their own data.

In some circumstances, the consumer may see benefits in having a copy of data about them provided to an entity that is not a competitor (for example, provision of medical records to a life insurance company or provision of utility payment information to a credit provider), or authorise a third party to transfer it on their behalf. This too is envisaged in this reform.

A right to transfer data is inherent in EU data protection regulations and is enabled in the UK with regard to regulated businesses, such as electricity providers, financial services and phone providers (chapter 4). In Australia, this concept is at best half-heartedly present in some industries (such as banking) but absent in others (such as insurance). And in all cases, the recipient of the data is the consumer. This has within it the seeds of its own ineffectiveness: consumers generally will not know what in their data is of greatest relevance to service providers. Moreover, to use their data today in Australia, consumers are expected not only to extract it from a provider but also to upload it to another provider or comparison site, and to bear the responsibility (or frustration) when the form in which the data was downloaded does not suit the new provider or site. This is self-evidently a barrier to effective competition.

Underlying this right, and maintaining incentives for data holders, is the idea that the right to data is a *joint right*, shared between the individual and the businesses or agencies that hold the individual's data.

The individual's decision to switch service providers does not alter the right of the initial service provider to the data that they collected while providing a service to the individual.

A new right to data transfer would encompass the following aspects:

- copying of data would be initiated by consumers making the request to their existing service provider to release a copy of their data to another identified service provider
- the data would be copied in machine-readable form (text files should be transferred as text files; numerical data should be transferred as numerical data) to a standard agreed by industry parties as effective and relevant to that industry
- the initial service provider would be permitted to retain a copy of the customer's data and continue to use it
- the transfer right would apply to data provided to a service provider by an individual, data relating to transactions between the consumer and the service provider, and other data necessary to transfer to a nominated third party.

All businesses and government agencies should be subject to these new data transfer requirements. In human services delivery an effective transfer right may be essential to future improvements to service quality — and if so will need to apply to both public and private data holdings.

It is vital that standards be developed, potentially on a sectoral basis, to implement data transfer. We consider that participants in each sector, rather than governments, are best placed to develop these standards, but establishing a Comprehensive Right is expected to motivate industry action. We are requesting feedback on whether an API format should be required (chapter 6).

Legislative changes would be required — these are discussed further below (section 9.6).

DRAFT RECOMMENDATION 9.2

Individuals should have a Comprehensive Right to access digitally held data about themselves. This access right would give the individual a right to:

- continuing shared access with the data holder
- access the data provided directly by the individual, collected in the course of other actions (and including administrative datasets), or created by others, for example through re-identification
- request edits or corrections for reasons of accuracy
- be informed about the intention to disclose or sell data about them to third parties
- appeal automated decisions
- direct data holders to copy data in machine-readable form, either to the individual or to a nominated third party.

Individuals should also have the right, at any time, to opt out of a data collection process, subject to a number of exceptions. Exceptions would include data collected or used as:

- a condition of continued delivery of a product or service to the individual
- necessary to satisfy legal obligations or legal claims
- necessary for a specific public interest purpose (including archival)
- part of a National Interest Dataset (as defined in Draft Recommendation 9.4).

The right to cease collection would not give individuals the capacity to prevent use of data collected on the individual up to the point of such cessation.

Costs

Data transfer may entail additional costs to the data collector/holder, including the process of developing standards, and any additional storage requirements. We envisage that data holders may levy a fee for providing access and/or limit the number of free access opportunities per year. Any charges for editing data should take into account that the data holder also benefits from edits that contribute to an accurate and complete set of information on the individual customer. The charging for such access, correction and transfer rights would need to take into account the additional costs incurred by data holders, but also the scope for businesses to use such charges to limit data transfer in practice.

Any charges levied by data holders for access, editing and transfer of data should be monitored by the ACCC, with the methodology used by businesses transparent and reviewable on request by the ACCC.

Oversight will be needed

The recommended individual data access model will require governance and oversight arrangements. We consider that, in the first instance, this could best be achieved by building on existing oversight, complaints handling and dispute resolution mechanisms for personal information. At present there is a range of industry-specific regulators that exercise their powers jointly with the OAIC where those industries involve collection, disclosure or use of personal information (for instance, ACMA as regulator for telecommunications, ASIC as corporate regulator, APRA as prudential regulator and Commonwealth Ombudsman as Private Health Insurance Ombudsman). These other regulators have various enforcement powers relating to their respective regulatory responsibilities.

We are proposing the following key regulatory responsibilities:

- The ACCC would play a key regulatory role with regard to individual data access, given that competition and consumer policy lies at the heart of the proposed changes. In performing this role, it should consult closely with other regulators, and particularly with the OAIC.
- The OAIC would continue to have overall responsibility for overseeing and driving privacy compliance and enforcing privacy complaints in the new system. It would also act as a backup where no ombudsman or dispute resolution scheme already exists.
- At an operational level, complaints handling and dispute resolution would be by existing external dispute resolution schemes in various sectors (for example: the Financial Ombudsman Scheme and the Credit and Investment Ombudsman scheme in the banking sector; ACMA and the Telecommunication Industry Ombudsman in the telecommunication sector). These bodies ensure consumers have access to a convenient, speedy and independent avenue of redress for complaints or other issues that might arise between an individual and organisations. In the case of banking and utility sector external dispute resolution schemes, they also have responsibilities for credit reporting (including recognition by the OAIC).

The effectiveness of these arrangements will be critical to the operation of the Commission's proposed framework. While some supplementary resourcing will be necessary, it is likely to be hard to define where it will be needed until these arrangements are settled in detail. The OAIC is likely to have a clear case for extra support. For other agencies, there may well be time to consider what if any additional support is needed after early experience with implementing reforms. Higher level oversight of the proposed regulatory functions will also be required and, given the primacy of economic issues within the framework, may best reside at the Commonwealth level with a central economic agency such as the Treasury.

To a large extent the individual access model proposed is simply building on existing consumer powers, although it is apparent from Inquiry participants' submissions that knowledge and use of existing powers is not widespread. The ACCC and state and territory

offices of fair trading are well positioned to advise and educate consumers regarding these powers, and to monitor any charging regimes used by data holders.

While the changes proposed aim to enable consumers to exercise more control over the collection and use of data on them, the onus remains on individuals to make responsible choices on who they provide personal information to and for what purposes. This must include the decision regarding the exercise of the right to require transfer of data. We recognise that ensuring consumers are able to make meaningful, informed choices is important — chapter 8 requests more information on how to achieve this.

DRAFT RECOMMENDATION 9.3

The Australian Government should provide for broad oversight and complaints handling functions within a reformed framework for individual data access. Key roles should be accorded to the Australian Competition and Consumer Commission (ACCC), the Office of the Australian Information Commissioner (OAIC), and to existing industry ombudsmen.

Any charging regimes, policies or practices introduced to address costs associated with data access, editing or transferability should be transparent and reasonable. The ACCC should be responsible for monitoring and assessing the reasonableness of charges applied. The ACCC, supported by state and territory Fair Trading Offices, should also educate and advise consumers on their new rights in regard to data access and collection.

For specified datasets (such as in banking) the relevant ombudsman scheme would need to be expanded to deal with disputes.

9.3 Element 2 — Access to datasets of national interest

Wider access to high value datasets across and between sectors — public, private, not-for-profit and academia — and jurisdictions has the capacity to deliver considerable benefits (chapter 2). These benefits often extend well beyond the initial data collector or holder, because digital data can be combined, re-used, transformed and added to in ways that create new and additional value without diminishing the value to the initial collector and holder of the data. In the absence of proactive efforts to create frameworks to facilitate wider access, existing barriers and impediments will ensure that such benefits are not realised.

The Commission has given considerable thought to establishing a framework whereby wider access would be enabled to high value, National Interest Datasets. The intention is to promote the development of a valuable suite of datasets — some of which are released publicly; others that will, at least initially, be shared (rather than released) across all Australian governments and with a small group of other trusted users (as defined below).

The Commission recommends that a process be established whereby public and private datasets are able to be nominated for designation as a National Interest Dataset.

The Commission intends that the designation of a dataset as a National Interest Dataset would take precedence over existing and future restrictions to access, which apply to all data contained in the dataset. This would include legislative and other program specific requirements that data be used only for the purposes for which it was collected; that it not be retained for ongoing use or re-use; and or that individual consent is required for use. In other words, this process is intended to 'cleanse' valuable data of existing or future encumbrances on its broader use, while also maintaining appropriate protections around privacy and confidentiality.

Datasets contributed to the suite of National Interest Datasets would emerge cleansed of anachronistic restrictions, and be shared or released through sectoral Accredited Release Authorities, subject only to contemporary safeguards and restrictions.

What is a National Interest Dataset (NID)?

National Interest Datasets (NIDs) are a subset of high value datasets. Their use is likely to generate benefits to the community beyond those derived by just the data holders and data contributors. These community-wide spillover benefits may have been identified in prior research or program evaluation within the relevant sector, through use of datasets with comparable features or circumstances in other sectors or overseas, or may be inferred from the interest in or demand for access to particular datasets.

In some cases data that would form part of a NID is likely to be held across different sectors. And, with many services now split in delivery between public and private sectors (such as health and education), and complete outsourcing by governments of the operation of others (such as public transport and electricity generation), inclusion of private sector data in the suite of national interest datasets is essential. Acknowledging this, the Grattan Institute (sub. 12, p. 8) for example, stated:

> A significant proportion of private hospital activity is subsidised by the taxpayer through the private health insurance rebate. The public also has an interest in comparing attributes (eg. quality or efficiency) of the two – public and private- hospital sectors.

While there are some important and obvious initial examples of *possible* NIDs, such as land use data, business register data, property and transport data, data on service provider performance, financial systems data and data on public infrastructure projects, others may be less immediately obvious but become clear candidates over time. What is important is that a process and legislative structure be established to encourage these datasets to emerge for broader use in a simple and adaptable manner.

To enable potential community benefits to be realised, the suite of NIDs must extend beyond the low hanging fruit of spatial data and aggregated activity data to include access

to de-identified datasets that are integral to social service delivery and decision making, as well as key privately held datasets relevant to these functions.

State and territory, as well as Commonwealth, datasets are relevant to the creation of NIDs, and the process must specifically allow for states and territories to volunteer their data, and for it to emerge under the same conditions of future use as Commonwealth data with which it is integrated.

The framework established must promote the inclusion of such data, in a manner that is consistent with community confidence in its use.

INFORMATION REQUEST

The Commission seeks further views on datasets that are of national interest and that could feasibly be designated as such under the process proposed.

What would access look like?

The approach proposed by the Commission represents a marked expansion in data access in Australia that would provide significant opportunities for research and innovative market development and improve delivery of public services.

In contrast to existing arrangements for access to significant datasets, the approach recommended envisages *open release* to the general public of a wide range of NIDs. Some of these datasets may already be publicly available, but the process of designating them as NIDs would not only improve their discoverability but ensure they are adequately maintained into the future, as a national asset.

A number of datasets designated as NIDs would not be publicly released in the first instance, but would be made available to accredited trusted users. The approach aims also to expand the range of data users that would be considered trusted (box 9.3), expand the range of datasets that they might access, and the types of uses to which the data can be put. *Sharing* of these NIDs with approved trusted users is a first step — a trial of the approach — with *public release* the ultimate objective that increases the use, and therefore the value of the dataset. In other words, the intention is to build confidence that the proposed approach to increasing access works well, beginning with trusted users, and then to gradually widen access, all the way demonstrating that this can be done without compromising individual identification.

NIDs that contained identifiable data would only be made available in de-identified form, reducing the risks to individual persons and businesses, and avoiding the need for ethics committee approvals. Risks associated with data transfer and storage of de-identified data would be managed through use of approved secure computing environments.

Box 9.3 Trusted users

Trusted users are individuals or organisations that are accredited as capable of responsibly accessing and using National Interest Datasets that are not yet made available through open access. In particular they have:

- coverage under the Privacy Act
- governance structures and processes that minimise and address the risk of inappropriate use or release of information
- access to appropriate secure technology/technological infrastructure and facilities.

Under the Commission's proposed model, trusted users would be vetted and accredited by the National Data Custodian (discussed below).

Organisations from which trusted users might come would include all Commonwealth and State and Territory agencies; all universities; corporations and not for profit organisations, and a range of other government funded research bodies.

Trusted user status would cease when a project is completed (in the case of identifiable data access); and could be suspended pending review if a breach occurs by a user in the organisation and/or working on the same project.

This approach is risk based, and recognises that researchers who are well known in their fields and are employed by trusted entities are likely to be less risky. It creates extra incentives, beyond just reputation, for data user organisations to want to avoid data breaches.

More specifically, where data (public or private) forms part of a NID that is shared in the first instance, rather than publicly released, the Commission recommends that:

- Access be granted on an *ongoing* basis to approved government personnel (in a Commonwealth, state or territory government agency) and to approved trusted users in a secure computing environment. Access that is ongoing will better enable innovative investigation with data, as well as longitudinal research with data that is updated infrequently.
- There be *few limitations* on the nature of the project for which the data could be used (with limits only being applied in circumstance where risk of re-identification could not be effectively eliminated).
- The output from the dataset *may* be reviewed prior to release, on a risk-assessed basis.
- Responsibility for appropriate use would rest with the trusted user, with clear and significant consequences for any breach of this trust.

The designation of National Interest Datasets

The Commission proposes that a national statutory office holder position, the National Data Custodian (NDC), be established.

The purpose of the NDC is to accredit release authorities, encourage datasets of value to be nominated for designation as National Interest Datasets (box 9.4), and implement and oversee the process whereby datasets deemed to be of sufficiently high value (that is, capable of delivering high net community benefits) are deemed to warrant designation for broader access and use.

Box 9.4 Potential sources of National Interest Datasets

There is a range of sources for datasets that might potentially be designated as NIDs. They *could* include, for example, land use data, business register data, property and transport data, data on service provider performance, financial systems data and data on public infrastructure projects. Initially, the focus would be on known Commonwealth datasets that could readily be designated for broader access, and their State counterparts if available. Some collations of data collected by local governments may also be found suitable for designation.

Over time, it is expected that on the back of greater transparency regarding data holdings and engagement with stakeholders, the NDC would more proactively prioritise public datasets for designation.

In terms of private sector datasets, there are several avenues through which these could be 'acquired' for inclusion in designated NIDs:

- Data collected in the course of meeting regulator requirements could automatically be incorporated by requirement. Similarly, organisations might incorporate such data where they believe it may enable them to use data more broadly without breaching regulatory requirements related to the data collection
- Data collected through the course of the provision of services on behalf of the government should automatically be included
- New contracts and funding agreements would incorporate appropriate terms and condition consistent with these expectations for data collection, access and sharing.

Other private datasets of national interest could be purchased.

In regard to datasets held by States and Territories, it is possible that future funding agreements with the Commonwealth would incorporate transfer, storage and curation costs related to data.

Specifically the NDC would:

- engage with data holders and users to select potential datasets for designation (that is, to consider what is available, what are the sensitivities and how the dataset could best be 'acquired')
- assess the value of nominated State and Territory or private sector datasets, negotiate with relevant parties, and recommend those sets of sufficient potential value for designation
- develop and oversee processes that remove or override existing and future restrictions on access and use
  - at the Commonwealth level this could be achieved through the use of disallowable instruments

- determine which Accredited Release Authority (ARA) is best placed to manage data curation, value add, storage and access
- determine the funding to be allocated to the ARA for the management, curation, storage and access of the relevant dataset.

Critically, the NDC has to be capable of instigating disallowable instrument processes, plus managing issues of funding and its allocation, and broader stakeholder engagement. Having the right culture, skills and expertise will be critical to the success of this office.

Funding the ongoing maintenance of National Interest Datasets

Datasets that are of national interest should be recognised as strategic assets, with their curation funded accordingly to ensure their quality is maintained. As these datasets would, by definition, be capable of generating significant public benefits, their maintenance should be funded by taxpayers upon designation. Given the costs involved, further consideration of user charging for external parties (that is, external to the contributors) is valid, and would best reside for decision with the NDC. Pricing of access to researchers should follow the guidelines proposed in Chapter 7. The independent review recommended in that chapter should be commenced as an early part of the implementation plans for these reforms, so that by the time the initial NIDs are established, pricing should be a settled issue. The NDC is also likely to play a key role in encouraging and facilitating the development of standards for dataset curation and storage.

There are several options for funding the improved curation of datasets (chapter 7). We consider the most accountable path is for the NDC to assign an ARA (box 9.6) with responsibility for collation and curation of a nationally significant dataset, and channel government funding to the authority to undertake this function. That is, funding for the creation and maintenance of quality NIDs would be attached to the prospective dataset, with ARA then either allocated particular datasets or awarded them under tender.

Other main operational features and requirements

Extensive community and stakeholder consultation is also expected to be an important aspect of the dataset designation process. Having a system for selecting and funding the ongoing maintenance of NID assets would help build consensus and cooperation between sectors and jurisdictions. This would build on existing work at COAG to select a spine of essential public sector assets.

To enhance community consultation in the process and ensure ongoing input, it is possible that a deliberative forum such as a parliamentary committee could be established to review declarations made and make proposals for future nomination and designation. A mechanism of this kind would ensure that detailed consideration of the existing pool of datasets from which designated sets can be drawn continuously occurs. It would also open to public scrutiny arguments against declaration.

Custodians of the data that contribute to a NID would, through their role, have a substantial impact on the ease, speed and cost of securing the dataset for wider access. For example, in regard to public sector data:

- If the data are held entirely within one Commonwealth Government agency, then the primary steps for designation would involve dataset curation, selection of an Accredited Release Authority, any de-identification and linkage necessary, and the process of determining who would be trusted users.
- If the data is held across multiple government agencies or multiple governments, then in addition to the above steps, designation will also involve determining which agency has responsibility for ongoing dataset updates and curation.

As flagged in box 9.4, governments will learn over time to build into any private contractual and funding arrangements mechanisms by which they can get shared access to data generated, for future inclusion in NIDs. This may include the need for the de-identification of commercially sensitive data, and avoiding compromising incentives for private sector entities to collect and value-add to data. On this matter, we also we welcome further stakeholder views.

DRAFT RECOMMENDATION 9.4

The Australian Government, in consultation with state and territory governments, should establish a process whereby public and private datasets are able to be nominated and designated as National Interest Datasets (NIDs).

Datasets (across the public and private sector) designated as NIDs would satisfy an underlying public interest test and their release would be likely to generate significant community-wide net benefits. Designation would occur via a disallowable instrument on the recommendation of the National Data Custodian.

NIDs that contain non-sensitive data should be immediately released. Those NIDs that include data on individuals would be available initially only to trusted users and in a manner that retains the privacy of individuals and/or the confidentiality of individual businesses. The in-principle aim should be for these de-identified datasets to be publicly released in time.

The process to designate datasets as being of national interest should be open to the states and territories in order to cover linked datasets, with negotiations undertaken to achieve this.

For community confidence, consideration should be given to use of a deliberative forum, such as a parliamentary committee, to take community input on and review nominations made, and to make proposals for future designations.

DRAFT RECOMMENDATION 9.5

The Australian Government should establish an Office of the National Data Custodian, as a new function within the Government to have overall responsibility for the implementation of data management policy.

Specifically, the National Data Custodian (NDC) would have responsibility for broad oversight and monitoring of Australia's data system, recommending the designation of National Interest Datasets, and accrediting Release Authorities and trusted users within the reformed data system.

DRAFT RECOMMENDATION 9.6

Selected Australian and state/territory government agencies should be accredited as Release Authorities by the National Data Custodian. In considering applications for accreditation, the National Data Custodian should consult a wide range of parties and ensure Accredited Release Authorities (ARAs) have sectoral expertise. The current model used by the National Statistical Service for appointing data linkage authorities should be considered in developing a model upon which to base this process.

ARAs will be responsible for:

- deciding (in consultation with initial data custodians) whether a dataset is available for public release or limited sharing with trusted users
- collating, curating and ensuring the timely updating of National Interest Datasets.

ARAs will also perform an important advisory role in regard to technical matters, both to government, and to the broader community of data custodians and data users.

DRAFT RECOMMENDATION 9.7

Trusted users should be accredited by the National Data Custodian for access to those National Interest Datasets (NIDs) that are not publicly released. Trusted users should be drawn from a wide range of potential entities, including: all Australian Government and state and territory government agencies; all Australian universities; and other entities (be they corporations, not-for-profit organisations or research bodies) that are covered by privacy legislation.

The default position should be that someone from one of these organisations would be approved for access unless the National Data Custodian transparently specifies a reason, on consideration, of why this should not occur.

For trusted users of NIDs, trusted user status should provide an ongoing access arrangement, with few restrictions on what could be done with the data. Trusted user status for NIDs should cease when the user leaves the approved organisation or be suspended if a breach occurs by any other trusted user in that same organisation and/or working on the same project.

INFORMATION REQUEST

The Commission seeks further views on the establishment of a Parliamentary Committee to take community input on possible national interest datasets, to review nominations made, and make proposals for future designations. Views are also sought on practical alternatives.

9.4 Element 3 — Sharing identifiable data with trusted users

A further important element of the Commission's recommended approach is the sharing of publicly funded *identifiable* data with trusted users, in secure computing environments. While this has many of the same governance features as the approach discussed in the previous section for designated NIDs, it is primarily about data that is not suitable for designated national interest status, given its persisting identifiable nature.

This third element is therefore about providing greater access for entities to data that does not have a public face, but is nevertheless of critical value to a smaller group of trusted users.

Data that identifies individual persons or businesses is already shared in a limited way with trusted users. There is scope to increase this sharing, within careful parameters. This data is typically used for targeted program and product/service delivery, regulatory compliance, and for research where there are necessarily very small samples involved.

The Commission's recommended approach would streamline access to identifiable data within and between Australian governments, and for a limited range of other trusted users. But additional sanctions and high security environment would be added features of this form of access.

What would access look like?

The Commission's model for sharing identifiable data with trusted users is focused on both streamlining existing processes for access approval and increasing carefully the range of users and uses to which these datasets may be put. Key features of the model include that:

- access be granted on a *project-specific* basis to approved personnel (box 9.3) in either Commonwealth or State/Territory government agencies and to approved researchers
- existing prohibitions on the linking of particular datasets be lifted or reviewed
- access would occur in a secure computing environment (this could build on a model such as the SURE system, overseen presently by the Sax Institute)

- the NDC, in consultation with data custodians, would provide a list of approved uses for the dataset to the Accredited Release Authority
- the project for which the data could be used would be subject to the list of approved uses provided by the NDC and require approval of the accredited release authority and, if relevant, an ethics committee
- if a project does not satisfy the data custodian's list of approved uses, the applicant would be able to apply to the data custodian for special access
- all the output from the dataset may be reviewed prior to project completion to ensure confidentiality requirements have been satisfied
- responsibility for ensuring use of the dataset is consistent the project's approach would rest with the trusted user, with the highest level of sanction for any breach being loss of access for the institution.

To the extent that data custodians are willing to pre-approve data uses and hand final approval over to accredited release authorities and ethics committees, the recommended reforms would substantially streamline access to data. Risks associated with data transfer and storage would be managed through the use of approved secure computing environments (box 9.5).

Box 9.5 Secure computing environments

The main features of secure computing networks currently used in Australia are discussed in appendix B. The SURE system is the most well-known and widely used of these networks.

The Commission sees potential to build on the experience with these networks to develop a more expanded secure computing environment. This should facilitate access in an environment disconnected from external servers, and subject to rigorous vetting and usage monitoring. Such a system would, if feasible given further considerations of practicality and cost, ideally permit users to access data remotely from their own computer in their own workplace.

The requirement that all main uses of identifiable Commonwealth data be referred to the NDC for approval is likely to mean these uses are considerably less than uses of NIDs. However, the range of data applications that necessitates identification of individuals is also much less, so we do not consider the need for such approval to significantly limit data applications by trusted users.

In a similarly transparent fashion to that envisaged for NIDs, a listing of all datasets that are potentially available to share with trusted users; the relevant data custodian and ARA for that dataset; and a contact point would be included on a website, such as on data.gov.au. This would enable potential users of these to know of a dataset's existence and how to be approved for access to it.

DRAFT RECOMMENDATION 9.8

Arrangements for access by trusted users to identifiable data held in the public sector and by publicly funded research bodies should be streamlined and expanded by the Australian Government. The National Data Custodian should be given responsibility to:

- develop, in consultation with data custodians, a list of pre-approved uses for a dataset, and make decisions on access to data for projects not consistent with the pre-approved uses list
- grant, on an approved project-specific basis, trusted user access to personnel from a range of potential entities, including: all Australian Government and state and territory government agencies; all Australian universities; and other entities (be they corporations, not-for-profit organisations or research bodies) that:
  - are covered by privacy legislation
  - have the necessary governance structures and processes in place to address the risks of inappropriate data use associated with particular datasets, including access to secure computing infrastructure.

Access would be granted for the life of the specific approved project. Trusted user status for use of identifiable data would cease when the user leaves the approved organisation; a project is completed; or if a breach occurs in that same organisation and/or project.

Associated reforms

In addition to the proposed broad overarching reform of access arrangements for identifiable data, a number of additional measures would assist in providing for a more effective and streamlined approach, in particular within the research sector.

Prioritising research funding based on openness of research data

We have recognised a compelling public interest case for greater re-use of researchers' data, but progress on this has been limited to date. To create incentives for re-use, research funding to institutions should be prioritised on the basis of their record in making data available to trusted researchers on conclusion of the research project.

DRAFT RECOMMENDATION 9.9

Public research funding should be prioritised on the basis of progress made by research institutions in making their researchers' data widely available to other trusted researchers on conclusion of research projects.

We have recommended in chapter 3 that the ARC be responsible for maintaining a public register of research data that is available for re-use. This register will provide transparency for implementation of this reform.

Reforming ethics committee processes

Ethics committees perform a valid oversight role, and bring much needed sectoral and subject expertise to bear, but numerous Inquiry participants attested to the delays in the approval process (acknowledging that some of this delay is due to data custodians rather than ethics committees).

There have been efforts to implement aspects of mutual recognition, (detailed in chapter 5). However, as discussed recently by the Senate Select Committee on Health (2016, pp. 47–53), progress remains slow.

While we do not, on the weight of evidence provided to the Inquiry to date, see a compelling case for the wholesale abolition of the ethics committee system, substantive reform of the type envisaged by the recent Senate Committee report is required, and would provide a much needed mechanism for swifter approvals. Accordingly, the Commission recommended in chapter 5 (Draft Recommendation 5.4) that reform of ethics committees arrangements continue, in particular around the mutual recognition of approvals. These reforms should reduce delays associated with current sequential nature of approval through multiple ethics committees, as well as reduce the extent to which researchers are subjected to creeping scope of an ethics committee's role.

Exceptions to consent requirements when data is used for research

The Commission considers that the existing exceptions in privacy legislation (sections 95 and 95A of the Privacy Act) allowing sharing of personal information for health and medical research purposes without obtaining the consent of individuals, should be extended to cover public interest research more generally. This would substantially streamline approval processes for access by trusted users to identifiable data, without increasing risks to individuals.

This approach would be consistent with the Office of the Australian Information Commissioner submission to this Inquiry (sub. 200), and the Australian Law Reform Commission (ALRC) recommendations (2008 Recommendation 65-2). It is also a consideration of the current Productivity Commission's National Education Evidence Base Inquiry (PC 2016).

As part of implementing these broader exceptions for research, and as per Draft Recommendation 5.2, we consider that the separate sets of rules under sections 95 and 95A of the Privacy Act should be combined. The elements of the National Statement on Ethical Conduct in Human Research dealing with privacy should also be aligned with the Privacy

Act and the Research Rules to minimise confusion for institutions, researchers and Human Research Ethics Committees.

Retention of linked datasets

One comparatively simple way to substantially increase the value of linked datasets is to retain (rather than delete after initial use) linked datasets and linkage keys as an ongoing resource for future research work. This would require changes that would allow for the retention of datasets for further research work, and the corresponding linkage keys used to create datasets. Such an approach would be consistent with practice in other countries, including the UK and EU.

Re-use of already linked datasets can be further enhanced by also retaining value added by researchers. Specifically, the Commission recommended (Draft Recommendation 5.3) that workable options be considered for enabling researcher efforts at dataset clean-up to be incorporated into the linked dataset, for the benefit of future users of the dataset. We welcome stakeholder feedback on how this might best be achieved in practice.

INFORMATION REQUEST

The Commission seeks further views on the most practical ways to ensure improvements to linked datasets are available for subsequent dataset uses.

9.5 Element 4 — Release of other data for widespread use

Release of data that is not related to individual people or businesses should be routinely available for use by governments, consumers, businesses and the research community. This includes information that, while it may identify individuals, is already in the public domain in some form. Data that identifies private property ownership, for example, may be of significant value to local governments and private sector entities, if presented in a collated and accessible manner. A realistic assessment of the risks associated with public release of identifiable information that is already publicly available in a less accessible form, should be undertaken.

Some of this data may well be held by the private sector entities and the same approach is desirable in those cases as well, acknowledging that it will be a matter for those entities to determine.

Where such data is held in the public sector or has been collected through public funding (such as may be the case with data generated in academic research) it is difficult to see why it should not also be released publicly in a timely manner. Some research funders

adopt such an approach in principle but action often does not match words. Governments have failed to press the issue with vigour.

We consider that there should be a shift in emphasis from only releasing non-confidential data on request for particular projects, toward actively pushing data out in a coordinated way. We recognise also that in some instances — such as where a data variable is derived from other raw data, usefulness of the derived data may be enhanced if its release is accompanied by the release of key code or models used to derive it. In principle, all non-confidential datasets in fields where there are burgeoning opportunities and capabilities would be opened up and released, as resources and sectoral demand allow.

This may mean that the average quality of such datasets is lower than it would otherwise have been had data custodians delayed release with a view to adding value. Chapter 7 noted that data should be released generally with the least value add unless a full list of conditions was met (Draft Recommendation 7.1). The Commission recommends that ARAs be tasked with assisting data custodians with these processes where required, and facilitating timely updates and ongoing dataset maintenance. Release authorities should, accordingly, be funded for this task where it applies.

Such an approach has the potential to make a marked difference to the range and volume of data available for decision making, innovative activity and improved service delivery in the community. As noted elsewhere in this Report, there is a disincentive to release data that may be critical of an agency's own performance. Nevertheless, it is imperative that this data be released to enable improvements in program management and service delivery.

DRAFT RECOMMENDATION 9.10

All non-sensitive public sector data should be released, consistent with release priorities and as resources allow, with curation, provision of metadata and adherence to agreed standards resourced as specified in Draft Recommendation 7.4. A realistic assessment of the risks associated with public release of identifiable information that is already public in a less accessible form, should be undertaken by all governments.

Data that could be used for program or agency performance management purposes should not be withheld from release.

9.6 Legislative and institutional changes required

To enable effective implementation, the recommended reforms described above will require an overhaul of the existing governance structure for Australia's data. The Commission has provided a preliminary view on the key entities and their roles within its proposed framework (box 9.7). As with most large shifts in the policy paradigm, there will be gaps in this model, and further views on alternatives are welcome.

Box 9.6 Key institutions and roles in the new Framework

The Commission's recommended reforms require government to take up important new functions to enable the opportunities from data to be realised. This will require the establishment of a new national position, and the authorisation of some additional functions by existing institutions.

The National Data Custodian (NDC)

The new position of the NDC creates a role that parallels for data access and use, the role of Australia's Information Commissioner for data protection. The NDC would have responsibility for broad oversight of the operation of the national data system, and be involved in designation of datasets of national interest, potentially with designation by disallowable instrument. The NDC would also accredit release authorities and trusted users within the reformed data system.

Accredited Release Authorities (ARA)

ARAs will largely be existing public sector agencies that already release data but would now be funded to take on additional responsibilities as an ARA. ARAs would play an important role in deciding whether a dataset is available for public release or limited sharing with trusted users, curating, ensuring the timely update and maintenance of datasets, and supporting the linkage of national interest and other datasets for release.

ARAs could conceivably be either Australian Government or state/territory government entities with sectoral expertise. They would have a long track record of trusted data management in their particular sectoral areas of focus, and be given a national remit for that data. As discussed in chapter 8, the Australian Institute of Health and Welfare is a possible model for a sectoral ARA with inter-jurisdictional involvement.

It is envisaged that ARAs would also perform an important advisory role on technical matters, both to government, and to the broader community of data custodians.

Other existing institutions with additional roles

Existing institutions with important new roles in the reformed data management system include:

- Australian Competition and Consumer Commission — oversee the Framework around consumer access to data
- Office of the Australian Information Commissioner — continue its fundamental role regarding privacy and handling of privacy related data complaints
- Administrative Appeals Tribunal — possible role to assess disputes and appeals regarding data sharing and release
- Australian Bureau of Statistics, Australian Institute of Health and Welfare and CSIRO Data61 — key advisory and support roles, in addition to their existing functions within the data infrastructure.

However, the functions that will need to be undertaken, as described above, are required if a systematic increase in data access and sharing is to take place (as it should). Some of the key legislative changes likely to be needed are also clear from the above discussion (table 9.1). Several other aspects of implementation are discussed in greater detail in chapter 10.

Table 9.1 **Proposed changes to key legislative instruments for different types of data**

Proposed new Data Sharing and Release Act	*Privacy Act 1988*	*Acts Interpretation Act 1901*
General		
• Government agencies to share and release data with other government agencies. • Require sharing between government agencies and other sectors. • Limited exceptional circumstances specified. • Possible role for AAT to assess disputes and appeals.	• Existing rules apply, to the extent not inconsistent with new Act. • Office of the Australian Information Commissioner (OAIC) continues to regulate existing rules.	
1. Giving individuals more control over data held on them		
• Give individuals new Comprehensive Right • Governance arrangements for new Comprehensive Right. • New powers for ACCC to oversee Framework around consumer data. • (If required) expand powers for ombudsmen and dispute resolution schemes, to deal with disputes.	• Existing rules apply, to the extent not inconsistent with new Act. • OAIC continues to regulate existing rules. • New power for OAIC to assist ACCC where no ombudsman or dispute resolution scheme.	• Create new definition of 'consumer data' within the Data Sharing and Release Act and include in the Acts Interpretation Act. • Consequential amendments to other Commonwealth Acts to ensure harmonisation.
2. Enabling broad access to datasets of national interest		
• Functions and powers of Office of the National Data Custodian (NDC). • System to appoint entities as Accredited Release Authorities (ARAs). Related functions and powers. • Process for nomination and designation of National Interest Datasets.	• (If required) specify that OAIC to develop and publish guidance on best practice de-identification processes. • (If required) new power for OAIC to certify that entities are using best practice de-identification processes.	
3. Sharing publicly funded identifiable data with trusted users		
• (If required) process for approval of trusted users. • (If required) legislation to support national system of ethics approvals accreditation. • Authorise retention of datasets and linkage keys. • Allow identifiable information to be used for all research determined to be in public interest (incorporating and expanding sections 95, 95A, 95AA of the Privacy Act).	• Existing rules apply, to the extent not inconsistent with new Act. • OAIC continues to regulate existing rules. • (If required) specify that OAIC to develop and publish guidance on the inputs required to establish a public interest case for expanded research exception.	
4. Releasing non-confidential data for widespread use		
• Legislation to support release of all non-sensitive public sector data (if required).	• N/A	

Legislative changes needed to implement the recommended reforms will primarily involve changes to existing Commonwealth privacy legislation as well as the creation of new legislation to facilitate data sharing and release. For clarity at this stage of the Inquiry, the new legislation is referred to as the 'Data Sharing and Release Act'.

This Act would be a Commonwealth piece of legislation applying across Australia to all digital data, by using the Commonwealth's communications power under the Constitution (s 51(v)) covering postal, telegraphic, telephonic and other like services. Other heads of power may also apply, such as the Commonwealth's power over external affairs (which forms the basis of the Privacy Act), trade and commerce, banking and insurance, and corporations. All of these Commonwealth heads of power, even when used together, are subject to restrictions on covering state governments and instrumentalities. It would therefore be 'umbrella' legislation, and would make redundant clauses in other dataset specific and program specific legislation and replace them with contemporary protections that are consistent across datasets.

This approach would aim to provide consistency across jurisdictions for the new schemes — but state and territory involvement will clearly be vital to achieve this. In particular, where states and territories opt in to have datasets designated as having national interest status, separate state and territory legislation may be required to cover state governments and instrumentalities. A concerted and genuinely national effort will therefore be needed if the proposed reforms are to be provided with the legislative support required at all levels of government.

An Act of this kind would facilitate the consideration of issues around data access via an alternative lens — one that has the ability to focus on *data as an asset*, not a threat or risk — rather than that provided by existing legislation, notably the Privacy Act.

It would also play an important role in giving *permission* to government agencies to share and release data and, in turn, drive cultural change in the attitude of the public sector to data and its release (as discussed further in chapter 10). So, for example, the Act should contain revised (rather than current draconian) penalties on the extent to which data custodians carry responsibility for the sharing of data in circumstances of authorised use under the new Act. Where custodians take reasonable steps to prevent misuse, they should not be liable for misuse by others, and the new Act would stipulate this.

The *Privacy Act 1988* (Cth) would remain a principal piece of legislation governing privacy, retaining core privacy safeguards in areas such as the Australian Privacy Principles and credit reporting provisions (Part IIIA).

This approach also allows scope for the Australian Government to address recommendations made by the Belcher Review and Hawke Report for a simpler and more coherent legislative framework for managing and accessing government information (Belcher 2015, p. 123, Hawke 2013, p. 22).

As outlined above, the Commission also proposes the introduction of a new definition ('consumer data') into the *Acts Interpretation Act 1901* (Cth).

RECOMMENDATION 9.11

The Australian Government should introduce a *Data Sharing and Release Act* which includes the following:

- Provisions requiring government agencies to share and release data with other government agencies and requiring sharing between government agencies and other sectors.
  - These provisions would operate regardless of all restrictions on data sharing or release contained in other legislation, policies or guidelines.
  - The provisions may be waived in limited exceptional circumstances, and the Act should specify what these circumstances are.
- Strengthened provisions on access to data by individuals, including rights to access and edit data about them, a right to have data copied and transferred, and a right to request that collection cease.
- Provisions establishing the Framework for the governance of Comprehensive Rights of consumers, access to National Interest Datasets, approval of trusted users, and accreditation processes for Release Authorities.

10 Trust: the foundation of the new data framework

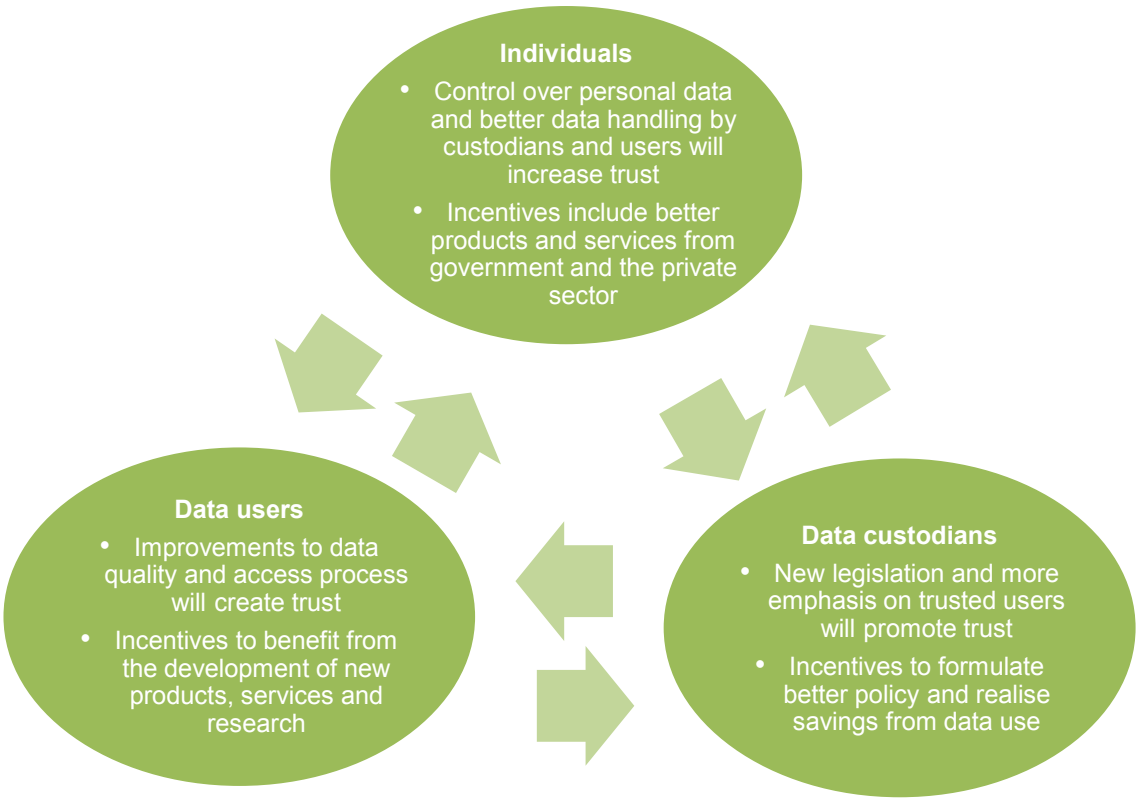
Key points

- In seeking to enable better use and sharing of data, governments must be mindful of community expectations and preferences, and deliver a framework for data sharing and release that manages risks. To do otherwise would erode the confidence of the community, data custodians and analysts, and jeopardise further progress in the use of data.
- The Commission's proposed reforms are aimed directly at retaining trust while enabling access to data. We propose moving from a data system based on risk aversion, to one based on increasing trust through greater participation and realisation of the benefits that better use of data can offer.
- Legislative change is important to achieve this, but Australia's experience to date has shown that it is not enough on its own — community expectations go beyond legislation. The Commission's proposed reforms create institutions and processes to ensure risk management practices are in place to maintain confidence in how data is handled, and to enable the community to realise the benefits of data use.
  - For individuals, trust and control over their data are interlinked. The Commission has recommended giving individuals stronger rights to determine how their data is used. New and existing institutions would ensure that individuals are able to exercise their rights.
  - For data custodians, trust in the legal framework within which they operate is crucial. The recommended new Data Sharing and Release Act would create a clearer operating environment and give permission for the safe release of data. Agencies charged with release responsibility would support custodians in sharing and releasing data.
  - For data users, the Commission has recommended improvements to access and pricing policies, as well as ways to improve the quality of datasets that are made available. These reforms will reduce the overly bureaucratic and inconsistent processes to access public sector data and help re-instate users' trust in the data system.
- Incentives for data holders to share and release data must be apparent and consistent. Under the Commission's proposed framework:
  - Transparency and exercisable rights would give individuals more knowledge of who holds what information about them, and which other entities value this information. This provides them with both the means and incentive to share and use the data that entities hold about them.
  - Dedicated funding for datasets that are of national interest is intended to ensure data custodians are motivated to undertake ongoing curation, maintenance, and sharing or release of these datasets. Improved access to both public and private sector data will enable some entities to realise additional commercial benefits through data analysis and use.
  - Data users would have clear guidelines and incentives to operate within the bounds of the new trust-based system — breaching these guidelines would come at a cost to them and their organisation, and erode the considerable benefits of more open data.

In previous chapters, the Commission proposed reforms to improve the availability and use of Australia's data for public benefit, while maintaining community trust and managing the potential risks involved. Strengthening trust in the system and increasing use of data are not trade-offs — the two go hand in hand. People are more willing to share information when they trust how it is being used and can see benefit in increased access to their data — and as more data is made available, the benefits are likely to grow, which builds public trust in the system and supports continued use of data.

Achieving this, however, requires more than the legal and institutional reforms presented in previous chapters. Data access and use is a unique area; one that requires strong social licence for policy to be successful. To maintain their social licence, data custodians and users need to operate within these updated legal boundaries, while at the same time complying with community expectations and ensuring community trust is not compromised (section 10.4). Therefore, meaningful change will only be achieved if *all* participants in Australia's data system realise the opportunities that the Commission's new framework offers them, and the responsibilities it creates, in ways that allow them to build trust in each other and the framework as a whole (figure 10.1).

Figure 10.1 **Trust and incentives in Australia's new data system**



Trust is a crucial enabler of economic and social interactions, particularly when these occur online or involve private data. Researchers have identified two key aspects of the data system that contribute to trust:

- *institutions* that certify the trustworthiness of the individuals using the data and put in place guarantees that data will be used appropriately
- *processes* that establish positive experiences with data use, ensure the usefulness and ease of use of data, and most importantly, manage the risks involved in data access (Warkentin et al. 2002).

Managing risk is a concern that affects individuals, data custodians and data users. In the current environment, data custodians take an overly cautious approach to risk management in order to avoid risk. As a result, data sharing and release are kept to a minimum and substantial opportunities to improve community wellbeing through data use are wasted. The Commission's proposed new framework for data access establishes a system that is based on genuine risk assessments, which seeks to address community expectations and concerns regarding the potential risks associated with data release, but also recognises the value and public interest benefit that increased data access can bring. This risk-based approach can play a pivotal role in creating trust, as individuals can trust that their data remains secure while data users are able to access and use it for positive purposes.

In describing the reforms that created New Zealand's extensive data infrastructure, the NZ Data Futures Forum (2014, p. 60) concluded that '[t]rust is the oil that can make the data-use machinery really work'. This applies equally to Australia, and across all sectors of the economy. This chapter discusses the ways in which the Commission's proposed reforms create institutions and processes that would strengthen trust and incentives, to get the most value out of Australia's data assets.

10.1 Increasing access to data for all Australians — what's in it for us?

Strengthening people's trust in data and institutions

Earlier chapters of this Report have provided a large number of examples of the payoffs possible from making better use of data than is currently the case. The potential benefits are diverse, extending from diagnosing health problems earlier, designing social policies that are much more effective, to providing consumers with more accurate information on which to base their choices.

But the key qualifier, when it comes to data that contains details on individuals, is control. Control over one's personal data is a cornerstone of trust in the systems that collect and handle that data (New Zealand Data Futures Forum 2014). This is true irrespective of whether the data is collected and held by government or by private companies:

> We realise that we won't be able to achieve [our goals] without the *trust* of the people we serve if they don't have confidence that they can *control* the information they share on our platform. (Facebook, sub. 172, p. 2, emphasis added)

In many instances individuals do not (and cannot) have full control over their personal data once it is directly or indirectly collected. To ensure the data system operates more effectively, governments need to create an environment that gives people *as much* control as possible (Warkentin et al. 2002).

Currently, public and private data custodians offer individuals limited scope to control the personal data collected about them, due partly to lack of clarity around access provisions in the *Privacy Act 1988* (Cth). The Australian Government (DPMC 2016) has undertaken a series of communication initiatives to build and maintain community trust but these measures have had only limited success in creating a culture of trust around the use of personal information (chapter 5).

The Commission's proposed framework goes much further in ensuring that individuals have control over their data, by:

- creating a Comprehensive Right, bundling the right to access and request edits to information collected about them with the ability to direct data holders to transfer a copy of this information to a third party (Draft Recommendation 9.2),
- creating a strong inalienable right to opt out from data collections in a broad range of circumstances (Draft Recommendation 9.2), and
- introducing a new, broader legal definition of consumer data, to allow a greater range of situations where consumers can exercise their rights (Draft Recommendation 9.1).

The Commission's proposed framework also embeds institutions and processes aimed at building and retaining trust, and enabling more people to benefit from access to data — directly and indirectly.

- New institutions would be created (such as the Office of the National Data Custodian) and existing institutions (such as the Australian Competition and Consumer Commission) would be given additional powers to ensure that individuals are able to exercise their data-related rights (Draft Recommendations 9.3 and 9.5).
- Implementation of standards for data sharing in the private sector — either instigated by entities or facilitated by governments — would enable consumers to exercise their right for data access and transfer (Draft Recommendation 6.2).
- The creation by Australian Government agencies of comprehensive, easy to use data registers, with published plans to improve the quality of datasets would make government and research data easier to find and use (Draft Recommendations 3.1, 3.2, and 6.1). Data users would be able to access accurate, useful and usable data, without needing a degree in statistics.

Giving individuals greater control over who has their data and how it is used will increase competition and at the same time will help improve business practices for the handling of consumer data. Consumers who are concerned about the way their data is being used would have avenues to shift their personal data to a different provider; or require entities to stop collecting data about them — a leverage point in its own right.

Participation in managing data will, over time, alert people (and businesses) to the value that is increasingly being generated from data, and benefits that they forgo, if they do not play an active part in the data ecosystem. Encouraging participation via stronger rights is a necessary step in raising awareness and consequently, in lifting confidence in data sharing.

These measures will not eliminate the potential risks associated with increased access and use of personal information, risks that even now are created every time a mouse is clicked or a mobile phone operated. However, collectively, the changes proposed by the Commission will provide more effective mechanisms for managing the additional aspects of such risks that come from data sharing. For example, agencies would be able to leverage the substantial technical knowledge of CSIRO, the Australian Bureau of Statistics, and the Office of the Australian Information Commissioner, to ensure that they handle individuals' data in accordance to best practice. By doing so, these mechanisms are more likely to strengthen individuals' trust in the organisations handling their data.

Incentives for individuals: creating tangible benefit from increased access to data

When people become confident that they are able to benefit from sharing their data, they are far more likely to share data directly, or to consent to data being used for secondary purposes.

Individuals are already supplying their personal information to private companies in exchange for benefits — such as access to free mobile phone applications or personalised discounts offered through brand loyalty programs, or the chance to win free holidays, concert tickets or other prizes. Further, the vast majority of Australians have already indicated that they would consent to their data being used in medical research, because they see a benefit in improved healthcare (chapters 4 and 5). On the public sector side, the time for Australian governments to tell people how data is being used — or could be used — to their benefit, is long overdue. For example, the provision of welfare benefits would not be possible without the recipients providing their personal information; but this information could also be used to develop social policies that are more effective in identifying and tackling long-term disadvantage.

The Commission's new proposed framework would increase access to, and use of, information across all domains. The potential benefits for individuals are substantial: for example, removing the need to supply the same personal information time and time again when dealing with government agencies that currently do not share data; switching bank

accounts much more easily; and benefiting from more accurate credit reporting (Draft Recommendation 4.1).

But it is not only identifiable data that can offer benefits. The Commission has recommended that all non-sensitive public sector data should be released (Draft Recommendation 9.10). Publishing more government data in an accessible format, and in some cases, in real time (for those areas such as public transport, where real time decisions may need to be made), can be very useful both to individuals and businesses. Enabling consumers to benefit from their data would help build community support for Australia's reformed data infrastructure.

10.2 Data custodians — replacing a culture of risk aversion with trust and incentives to share

New legislation would allow data custodians to operate in an environment of trust

Lack of trust is a significant factor contributing to the current state of paralysis in Australia's data system. Data custodians, particularly in the public sector, have limited trust in the processes and the legislation that purport to allow them to release data. They also have limited trust — because there is little experience on which to base such trust — that the users requesting access would use the data appropriately.

The Commission's proposed framework establishes a trust-based environment for data custodians, underpinned by new legislation as well as new, fit for purpose institutions and processes.

- The Data Sharing and Release Act would create an environment of greater legal certainty for data custodians, clarifying what they can and cannot do (Draft Recommendation 9.11). Further, the Office of the Australian Information Commissioner (OAIC) would have the authority to certify custodians' de-identification processes, allowing them to trust that they are releasing data lawfully (Draft Recommendation 5.1).
- The National Data Custodian (NDC) would be given powers to strengthen and expand trusted user models. For data custodians, accreditation of trusted users would simplify the process of determining who can be trusted with data and make approval processes more transparent and predictable for those who want to access and use data (Draft Recommendations 9.5 and 9.7).
- Sector-based Accredited Release Authorities (ARAs) would help build trust by assuring data custodians within each sector that those sharing and releasing data have sufficient sector-specific expertise. ARAs will provide data custodians with the relevant technical expertise to enable them to make the best use of their data, and implement a

risk-based framework to prioritise datasets for curation, sharing and release (Draft Recommendation 9.6).

- Existing secure access models — such as the SURE computing environment — would need to be expanded, to allow more trusted users to access government data (Draft Recommendation 9.8).
- The processes to designate datasets as National Interest Datasets would provide transparency for data custodians and the community, and would support increased access to, and use of, Australia's most important data assets (Draft Recommendation 9.4).

The Commission recognises that there is a distinct difference between *sharing* of data within government, or with other trusted users, and *releasing* it into the public domain. These two activities have different risk profiles, and a different level of trust in the potential data user is required for the custodian to enable access to the data.

The recommended adoption of a risk-based approach to data *sharing* should see greater sharing of identifiable data, via trusted user models, to approved entities. Greater use of trusted user models would shift the focus away from managing risk by confidentialising, perturbing, or simply not releasing the data; instead, the focus would be on recognising that the risk of data sharing varies substantially, based on the type of data and user. This is expected to allow better risk management and more sharing of usable data.

Simplified legal obligations (provided by the new umbrella legislation) combined with the support of ARAs, would make it easier for data custodians to know with certainty whether or not data *release* is an option, and ensure they have the capability to get datasets into a format suitable for release. Rationalising the complex web of legislative provisions is likely to have efficiency gains simply by reducing the time (and legal fees) spent interpreting legislation.

Private sector data custodians

Data custodians in the private sector are often required to provide data to governments, either for compliance purposes or statistical collections. Trust in the agencies that collect and use this data is vital. As the Australian Bureau of Statistics (sub. 94, p. 2) points out, one of its roles is 'sustaining the high level of community trust to protect the privacy and confidentiality of personal and business information required under legislation (which in turn is fundamental to the ABS's ability to collect high quality information from businesses and people)'.

The Commission's proposed framework would support government agencies in maintaining private sector trust through clearer legislative requirements, improved processes and best-practice approaches to data security. Changes to the legislative environment would also facilitate further development of private sector data sharing models, such as those of Data Republic and Quantium (subs. 176, 187).

Creating incentives for data custodians

In Australia's current data system, public sector data custodians face substantial disincentives when considering increased data sharing and release. In the most extreme situation, they can face imprisonment if data is released illegally; and even where data can be legally released, they may incur substantial costs that are not covered by operational budgets (chapter 5). In addition, it is not unknown for data release to be opposed for political motivations.

A number of aspects of the Commission's new model address these disincentives, including funding issues.

- The proposed new Data Sharing and Release Act would give agencies permission to share or release data safely, rather than hold onto it out of fear of legislative consequences. This is intended to create an environment for data custodians in which data sharing and release are the norm, rather than the exception, with benefits from greater data use made apparent to the broader community.
- The Commission recommends that government agencies should refrain from adding value to data before release in most circumstances (Draft Recommendation 7.1). This would prevent agencies from spending time and money unnecessarily, and would speed up data release.
- Separate funding would be made available for necessary (justifiable) curation of non-sensitive public sector data before release, as well as for the ongoing maintenance and updating of National Interest Datasets (Draft Recommendations 7.4, 9.6 and 9.10). This would provide an incentive for government agencies to release more data, as additional resources would be made available for data curation activities beyond an agency's operational requirements.

Increasing the use of data would generate insights that can improve the policies formulated by the agencies responsible for managing data, and by the government as a whole. As more agencies realise budgetary savings and implement more effective policies as a result of improved data use, others will have an incentive to similarly make greater use of their data — and the Commission's proposed approach creates the legal and institutional framework to make this possible.

Private sector data custodians

Past experience, from both Australia and overseas, has shown that the private sector can develop new products and services far beyond what can be currently conceived, if given more access to data (chapters 2 and 4). Active data sharing between private sector entities is widely practiced in new (Uber, Airbnb) and old (finance, retailing, airlines, and more) service industries, driven by commercial incentives.

The government role in this activity is to establish not just that privacy is well-managed, but also that consumers can exercise a degree of choice and control over how data about them is managed.

Other than that, the Commission has recommended that future contracts between the public and private sector — contracting out public functions or projects — should take commercial incentives into consideration, to ensure governments can access data created by private sector contractors (Draft Recommendation 4.2).

Private sector entities have large data holdings that may form part of National Interest Datasets. Some of this data may be commercially valuable, and the Commission recognises that appropriate incentives will be needed if private sector data is to be contributed to National Interest Datasets. Measures to preserve private sector incentives where some of their data holdings may be deemed to be of national interest include:

- requiring provision of data by the private sector as a condition of receiving taxpayer funding and/or providing goods or services under contract to government, or
- purchasing data from the private sector as required.

Academic institutions

Universities and other academic institutions hold large datasets generated through their research work. There is a significant public interest in open access to data generated by researchers, and progress has already been made in this area (Universities Australia, sub. 90). To build upon this work, and create further incentive for academic data custodians to share their data, the Commission has proposed:

- that public funding for research be prioritised to those institutions that improve access to their data holdings (Draft Recommendation 9.9),
- that publicly funded entities involved in the generation of data, such as the Australian Research Council, should publish up-to-date registers of datasets that they commission or hold (Draft Recommendation 3.2), and
- that funding priority be given to academic institutions that implement mutual recognition of approvals issued by accredited human research ethics committees (Draft Recommendation 5.4), which would encourage academic institutions to adopt approval processes that simplify access to identifiable information.

Increasing access to research data will enhance the efficiency of Australia's research efforts. Universities Australia (sub. 90, p. 4) has submitted that 'Australia's ongoing research success will depend on its capacity to optimise the use and reuse of research data'. From the community's perspective, this represents a more efficient use of public funds invested in research.

10.3 Data users need strong incentives to maintain system integrity

Better quality data will give data users confidence

The benefits of increased access to data would only be realised if the data released is of high quality. Data must be accurate and presented in a usable format in order for users to generate any value from it (chapter 6). Conversely, 'poor quality data yields poor quality decisions' (CoreLogic Asia Pacific, sub. 102, p. 3). If the data released is known by collectors to be of poor quality and users are misled by its weaknesses, they could lose trust both in the data and the institutions releasing it. The Commission has recommended ways to improve data quality through standardisation and curation, and has also designated funding for these activities (Draft Recommendations 6.1 and 9.10).

The Commission has proposed a number of reforms to public sector institutions that would build users' trust in Australia's reformed data system. The Commission has also called for substantial improvements to the processes used to access data, which should eliminate much of the frustration expressed by data users in submissions to this Inquiry(chapters 3, 5 and 6). Key improvements proposed include:

- Changes to the Privacy Act would expand the legal exceptions made for access to identifiable information, to cover all public interest research. Further, the requirements to destroy linked datasets and linkage keys at the completion of projects that use Commonwealth data would be abolished (Draft Recommendations 5.2 and 5.3).
- Government data would be easier to find and use, by improving public data registers and ensuring data and related metadata are machine readable (Draft Recommendation 3.1).

Strengthening data users' incentives through better access processes

Difficulty finding and using much of Australia's data, combined with the fact that requests for dataset access can take many years to elicit all the approvals necessary from data custodians and ethics committees, means that some data users currently have little incentive to participate in Australia's data system. Local researchers have reported, for example, having to rely on overseas data to answer questions that are of relevance to Australia (SSCH 2016).

In the past, the Australian Government has made a number of declarations of its commitment to open data. However, these have failed to materially alter the overly bureaucratic data access system that potential data users must contend with. Such an environment erodes trust. Implementing the Commission's proposed framework would increase data users' trust in the system, prompting increased use of data.

As outlined above, the Commission recommends substantial improvements to data access processes, which would create an incentive for further data use.

- Requests for public release of government data would be centrally managed (Draft Recommendation 2.1), simplifying access for data users and increasing data custodians' accountability. Approval processes put in place by data custodians would also be streamlined, to ensure that requests for data access are dealt with in a timely and efficient manner. To maintain accountability, data custodians would be required to report annually on their handling of requests for access (Draft Recommendation 5.4).
- In this new data framework, the accreditation as a trusted user confers a status and opportunities that the user and their institution would be loath to lose. As such, researchers would face increased *disincentives* to use data in ways that may prove harmful to individuals or the community. At the same time, the Commission's proposed framework sets up a pathway and standards via which users would qualify for such status, creating an incentive to display behaviour that warrants community confidence.
- The capacity to integrate data would increase, as state-based data linkage units would be able to apply for accreditation to allow them to link Commonwealth data (Draft Recommendation 5.5).
- Reforms to pricing policies would strengthen consistency across government agencies, and tend to lower the cost of data for researchers. Minimally processed public sector datasets would be made freely available or priced at marginal cost of release, and agencies would work towards lowering prices on datasets, where there is strong demand from data users (Draft Recommendations 7.2 and 7.3). This would remove a substantial barrier to research.

In the Commission's proposed framework that is based on trust, the incentives to operate within the legal and institutional boundaries are considerably stronger than under the current system of largely ad hoc arrangements.

10.4 Maintaining the social licence to collect and use data

Data analytics is a unique field of policy — where custodians and users need a 'social licence to operate', and society's expectations go beyond the requirements of regulators:

> [I]ndustries with significant environmental impact may find that operating within the law but outside the boundaries of social approval can result in corporate damage — for example, by having a negative impact on a company's brand or provoking new and restrictive regulation …
>
> Some analogies have been drawn between a social licence for the mining of minerals and for the process of data mining. Regarding these and similar activities, what the social licence emphasises is the possible need for those (whether they are public or private bodies, or specific occupational groups) undertaking activities likely to provoke public disquiet to go 'beyond compliance' with legal requirements (Carter, Laurie and Dixon-Woods 2015, p. 2).

Recent overseas examples highlight the importance of maintaining a social licence when increasing access to and use of data, particularly for personal information. In the United Kingdom, the recent failure of an initiative to use health data collected by GPs was attributed to a lack of social licence. While the care.data project had all the required legal authorisations, it ultimately failed to get off the ground because it did not gain the community's trust — the options to opt out of the data collection were poorly communicated, and the public viewed the initiative as one that compromised confidentiality and primarily benefited commercial interests (Carter, Laurie and Dixon-Woods 2015).

In contrast, Statistics NZ attributes the success of its Integrated Data Infrastructure (IDI) project, which uses many datasets that contain personal information, to its social licence:

> Because the IDI is linking and using data to an extent previously unseen, the question of social licence is a crucial one. Recent research on social licence and perceptions about data collection and use showed the New Zealand public has an expectation that their data is used, but used wisely and for public good. It also showed that Statistics NZ has strong reputation as 'data expert' and trusted custodian of public information. However, this should not be taken for granted – social licence, once earned, can be easily lost.
>
> Statistics NZ is therefore continuing to engage the public in an informed debate about social licence, intended to move them from naïve to informed trust. (Statistics NZ, sub. 62, p. 2)

The need to maintain a social licence for data collection and use is a central consideration for stakeholders in this Inquiry (ABS, sub. 94; Australian Institute of Marine Science, sub. 171; ATO, sub 204; Department of Industry, Innovation and Science, sub. 69; Department of Social Services, sub. 10; Office of the Australian Information Commissioner, sub. 200; University of Melbourne, sub. 148).

The Commission's proposed framework includes a number of reforms that would reinforce the social licence that the public, private and academic sectors must have in order to keep expanding their use of data. Central to this is strengthening individuals' control over consumer data. Telstra (sub. 88, p. 11) has argued that:

> While Australia's privacy principles play an important role in creating confidence and trust by giving consumers an element of control over how their data is used, this confidence and trust can be fragile and weaken quickly.

By introducing the right for control over consumer data, the Commission's proposed reforms can contribute substantially to community trust in the data system. Stronger institutional oversight and data security measures included in the reforms would also support an increased sense of trust in the community.

In addition, clear communication between *all* participants in the data system, including articulation of the purposes for which data is used and the ways it benefits the community, is a necessary condition to maintaining a social licence (Carter, Laurie and Dixon-Woods 2015). The care.data example is a sobering case study of what can happen when the benefits of data sharing are not appropriately communicated to the community. The UK

Government spent more than two years and £7.5 million on the project before it was scrapped (Knapton 2016). Beyond the waste of funds and time, this project was also a wasted opportunity to discover ways to improve people's health. Learning from this experience, the Australian Government must ensure that it engages the community in implementing the Commission's proposed reforms.

10.5 Finally, a word on implementation

The Commission's proposed framework represents a fundamental change to the way Australia manages its data. The Commission has made a number of recommendations that can be implemented in the short term to improve data availability and use — release of public sector data should be a priority for governments. But achieving the large-scale change needed to position Australia to make the most of its data now and in the years ahead will require time and detailed planning.

The implementation plan for the new data framework should consider the sequence of actions required to achieve the reform, as well as steps necessary in communications and engagement with state and territory governments, business and the community. As a matter of priority, the Australian Government should reform data access in the public sector, and make substantial changes to its open data agenda to align with developments in competing economies. Cabinet should take account of any data aspects of policy issues it considers. In addition to improving policy decisions, such a move would demonstrate the Australian Government's commitment to data use and assist in putting sharing and release of data at the centre of government deliberations, at least at the Commonwealth level.

The Government's implementation plan should put in place clear ownership and accountability structures, including an emphasis on transparency against milestones, which are a vital part of successful reform. A requirement at the Australian Government level for data custodians to provide details of their progress would assist in the ongoing oversight provided by the OAIC and other regulators within the Commission's proposed data framework (Draft Recommendations 3.1 and 5.4).

Some ad hoc planning for increased data access and use has already occurred in different parts of governments, and various demonstration projects are aiming to establish frameworks for future progress. However, there is limited coordination between these initiatives (chapter 5). Large-scale change that will deliver the potential benefits of increased access to data will only be possible with a clear implementation plan, managed by a central agency that is held accountable for coordinating progress across governments.

While many discussions of open and big data focus on the role of technology (such as applications, software capabilities and cloud computing), the skills of the people accessing the data are paramount in ensuring it is used effectively and appropriately. Developing the capabilities of data custodians and users — from individuals accessing their own data to

analysts managing complex mathematical models — is critical to getting the most value out of increased access to data:

> While some data analysis is ceded to algorithms, especially the grunt work of processing and calculating, direction and interpretation is still largely the preserve of a human analyst. Drawing on their skills, experience, and knowledge, researchers and analysts make decisions concerning where to focus attention, how to frame and undertake analysis, and make sense of the findings and act upon them. People then remain key actors in building, maintaining and running data-driven projects. (Kitchin 2014, p. 160).

There are many organisations across the economy, such as the Australian Bureau of Statistics, the Australian Institute of Health and Welfare, and Data61, that already have substantial levels of expertise upon which to draw when developing both the technological and human capabilities required to achieve better utilisation of data.

The Commission is of the view that, as with many other required elements of its proposed framework, such capability development will be greatly assisted by a reformed system that signals a new, more dynamic stance on data sharing and release.

We have (deliberately so) not been specific in this draft Report on some of the alternative technology approaches that could be taken to implement aspects of the proposed framework. Depending on Inquiry participant views on the proposed framework, we may address this part of the implementation in the final Inquiry report.

As it implements the new data framework, the Australian Government should commence a national conversation on data, considering its collection and use in a way that benefits the entire community. There have been various consultations on these topics in the past, but they lacked the impetus and government commitment that is required to truly engage the community in a discussion on the immense potential — and possible risks — of increased access to and use of data.

Public consultation will be essential to secure citizens' trust in the new data framework. And while it will take time to achieve meaningful change, the Government should commence this process as soon as possible. It can leverage its own commitment to public sector data reform to secure action and engagement for consumer data reform — and together, reform in these areas will deliver substantial benefits across the community.

A Inquiry conduct and participants

This appendix describes the stakeholder consultation process undertaken for the Inquiry and lists the organisations and individuals that have participated.

Following receipt of the terms of reference for the Inquiry on 21 March 2016, an initial circular advertising the Inquiry was distributed to industry organisations and individuals and the Inquiry was advertised in national newspapers.

The Commission received 211 public submissions (table A.1) prior to the release of this draft report, including: 57 from industry or stakeholder organisations; 43 from governments or government agencies; 38 from academics or research groups; 36 from businesses; 27 from individuals; and 10 from not-for-profit or other non-business groups. All public submissions are available on the Inquiry website.

In addition, the Commission held separate discussions with around 100 businesses, business groups, academics, government agencies and individuals in Australia and overseas (table A.2), as well as a roundtable discussion with academics (table A.3).

The following public documents have been prepared by the Commission so far in this Inquiry:

- Issues paper — released 18 April 2016
- Draft report — released 3 November 2016

The Inquiry final report will be provided to Government by the 21 March 2017 and is to be released publicly within 25 parliamentary sitting days from that date.

The Commission welcomes further contributions to the Inquiry from interested individuals or groups. Public hearings will be held in Melbourne on 21 November 2016 and in Sydney on 28 November 2016. Submissions and comments on this draft report close on Monday 12 December 2016. Further details on registering for hearings and making submissions can be found on the Inquiry website.

Table A.1 **Public submissions received**

Participant	*Submission no.*
Sam Toady	1
Phoensight	2
Xamax	3
Ryan, Bruce	4
Telethon Kids Institute	5
Rannila, Jukka	6
Tyro Payments Limited	7
Australian Dental Association	8
Centre for International Finance and Regulation	9
Department of Social Services	10
Centre for Policy Development	11
Grattan Institute	12
Data Linkage Branch (Department of Health WA)	13
Federation of Victorian Traditional Owner Corporation	14
Integrated Marine Observing System, University of Melbourne	15
Federation of Ethnic Communities' Councils of Australia	16
Cleary, Michael	17
Department of Employment	18
IEEE SSIT Australia	19
Department of Prime Minister and Cabinet	20
Centre for Big Data Research in Health, UNSW	21
Australian Research Council	22
NSW Bureau of Crime Statistics and Research	23
Dr Hazel Moir	24
Australian Statistics Advisory Council	25
Lorica Health	26
InFact Decisions	27
Bryan Kavanagh	28
Chartered Accountants ANZ	29
Archerfish Consulting	30
Timothy Finney	31
Australian Government Environmental Information Advisory Group	32
Semantic Identity	33
Australian Payments Council	34
University of Sydney	35
John D Mathews	36
Department of Agriculture and Water Resources	37
Ecosystem Science Council of Australia	38
Health Services Research Association of Australia & New Zealand	39
VANZI	40
Curtin University	41
Office of the Information Commissioner - Queensland	42

(continued next page)

Table A.1 (continued)

Participant	*Submission no.*
Cooperative Research Centre for Spatial Information	43
Australian Payments Clearing Association	44
Federal Chamber of Automotive Industries	45
Australian Government Linked Data Working Group	46
Leading Age Services Australia	47
Australian Public Service Commission	48
White Label Personal Clouds	49
University of New South Wales	50
The BetterStart Child Health And Development Research Group	51
Australian Institute of Tropical Health and Medicine	52
Australian Communications Consumer Action Network	54
Lisa Schutz	55
Sax Institute	56
Australian Financial Markets Assoc.	57
Commonwealth Grants Commission	58
Pia Waugh	59
Australian Indigenous Governance Institute	60
Tyro Fintech Hub	61
Statistics New Zealand	62
Red Energy and Lumo Energy	63
ANZ	64
KimMic International	65
Insurance Council of Australia	66
SPUR powered by Landgate	67
The George Institute for Global Health	68
Department of Industry, Innovation and Science	69
Open Data Institute Queensland	70
Heart Foundation	71
Atlas of Living Australia	72
Financial Institutions & Management Advisory	73
NPS Medicine Wise	74
Alzheimer's Australia	75
Capital Markets Cooperative Research Centre	76
QIMR Berghofer Medical Research Institute	77
Johnson-Johnson	78
University of Melbourne	79
NSW Government	80
Consumer Action	81
A future beyond the Wall – ARC Linkage Project	82
Community Insight Australia	83
Australasian Open Access Strategy Group	84
CREBP – Bond University	85

(continued next page)

Table A.1 (continued)

Participant	*Submission no.*
Australian Chronic Disease Prevention Alliance	86
ARCA	87
Telstra	88
Swipezy Pty Ltd	89
Universities Australia	90
Victorian Alcohol and Drug Association	91
Health Geography Study Group – Institute of Australian Geographers	92
Australian Bankers' Association	93
Australian Bureau of Statistics	94
Australian Unity	95
Australian Institute of Superannuation Trustees	96
Council of Australian University Librarians	97
Medibank Private	98
Department of Health	99
Australian Library and information Association	100
Surveying and Spatial Sciences Institute	101
CoreLogic Asia Pacific	102
Australian Centre for Financial Studies	103
Cancer Australian	104
The National Liveability Study and Place Health and Liveability	105
Judy Allen and Carolyn Adams	106
Financial Rights Legal Centre	107
John Mills	108
Australian Longitudinal Study on Women's Health	109
Population Health Research Network	110
Impact Investing Australia	111
Australian Restructuring Insolvency and Turnaround Association	112
Australian National Data Service	113
National Archives of Australia	114
Health Research Institute, University of Canberra	115
AURIN – The University of Melbourne	116
Research Australia	117
Clean Energy Regulator	118
Western Sydney University	119
The Department of the Environment and Energy	120
Australian Injury Prevention Network	121
Australian Projections Pty Ltd	122
SA NT DataLink	123
Equality Rights Alliance	124
Chacko Thomas	125
National Health and Medical Research Council	126
Australian Energy Council	127

(continued next page)

Table A.1 (continued)

Participant	*Submission no.*
IAG	128
Datanomics	129
Open Source Industry Australia	130
Federation University Australia	131
Customer Owned Banking Association	132
Monash University	133
Australian Computer Society	134
Dun & Bradstreet	135
AGW Lawyers & Consultants	136
Uniting Church in Australia – Synod of Victoria and Tasmania	137
Center for Data Innovation	138
Australian Data Archive	139
AGL Energy	140
Cancer Council Australia	141
Australian Privacy Foundation	142
Origin Energy	143
Bruce Sweeting	144
Chris Doulton	145
Australian Business Roundtable for Disaster Resilience and Safer Communities	146
Property Council of Australia	147
University of Melbourne	148
Name withheld	149
Name withheld	150
Name withheld	151
Name withheld	152
Name withheld	153
Name withheld	154
Mark Rennick	155
Adrian Bennett	156
Australian Automobile Association	157
Australian Energy Market Commission	158
Australian Hotels Association	159
The Law Society of NSW	160
CSIRO	161
Australian Institute of Health and Welfare	162
VEDA	163
ANZLIC – the Spatial Information Council	164
Law Council of Australia	165
NetApp Inc	166
CHOICE	167
Department of Immigration and Border Protection	168
Australian Property Institute	169

(continued next page)

Table A.1 (continued)

Participant	*Submission no.*
Joint Councils of Social Service Network	170
Australian Institute of Marine Science	171
Facebook	172
Office of the Privacy Commissioner - NSW	173
Australia Post	174
Commonwealth Bank of Australia	175
Data Republic Pty Ltd	176
ASX	177
Association for data-driven marketing & Advertising	178
Data Governance Australia	179
National Native Title Tribunal	180
Aboriginal Health Council of WA	181
FinTech Australia	182
Australian Private Hospitals Association	183
Law Institute of Victoria	184
AUSTRAC	185
Brotherhood of St Laurence	186
The Quantium Group Pty Ltd	187
IoT Alliance Australia	188
National Computational Infrastructure	189
Name withheld	190
Business Council of Australia	191
National Aboriginal Community Controlled Health Organisation	192
NHMRC Centre of Research Excellence in Offender Health	193
Lisa Doyle	194
Australian Securities and Investment Commission	195
University of Tasmania	196
Westpac	197
Bureau of Meteorology	198
Health Informatics Society of Australia	199
Office of the Australian Information Commissioner	200
Department of Infrastructure and Regional Development	201
Department of Foreign Affairs and Trade	202
People with Disability	203
Australian Taxation Office	204
Tasmanian Government	205
Actuaries Institute	206
Queensland Government	207
Jan Whitaker	208
Attorney–General's Department	209
Australian Criminal Intelligence Commission	210
Geoscience Australia	211

Table A.2 **Consultations**

.id
Academy of the Social Sciences
Australian Bankers' Association
Australian Bureau of Statistics
Australian Centre for Financial Studies
Australian Communications and Media Authority
Australian Government Attorney–General's Department
Australian Government Australian Bureau of Agricultural and Resource Economics and Sciences
Australian Government Australian Law Reform Commission
Australian Government Department of Defence
Australian Government Department of Education and Training – National Research Infrastructure Council (NRIC)
Australian Government Department of Finance
Australian Government Department of Health
Australian Government Department of Human Services
Australian Government Department of Immigration and Border Protection
Australian Government Department of Industry, Innovation and Science
Australian Government Department of Social Services
Australian Government Department of the Prime Minister and Cabinet
Australian Government Open Access and Licensing Framework
Australian Government The Treasury
Australian Institute of Health and Welfare
Australian National Data Service
Australian Payments Clearing Association
Australian Privacy Foundation
Australian Prudential Regulation Authority
Australian Research Council
Australian Retail Credit Association
Australian Taxation Office
Australian Unity
Australian Urban Research Infrastructure Network
Bureau of Meteorology
Bureau of Transport, Infrastructure and Regional Economics
Business Council of Australia
CRC Data to Decisions, Deakin University
CSIRO Data61
Customer Owned Banking Association
Data Republic
Deloitte
Dominello, Victor – NSW Minister for Innovation and Better Regulation
Facebook
Federal Chamber of Automotive Industries
Financial Rights Legal Centre
Geoscience Australia

(Continued next page)

Table A.2 (continued)

Google
Haikerwal, Mukesh
Health Informatics Society of Australia
Health&
Melbourne Institute
Microsoft
National Archives of Australia
National Australia Bank
NSW Centre for Big Data Research in Health
NSW Data Analytics Centre
NSW Department of Premier and Cabinet
NSW Transport
Office of the Australian Information Commissioner
Open Access Working Group
Public Health Research Network
Quantium
RateSetter
Research Australia
Sax Institute
Shuetrim, Geoff
Semantic Consulting
SocietyOne
SSIS Data Services
Stanley, Fiona
Telstra
Telstra Health
Torque Solutions / Velocity Frequent Flyer
Tyro Payments
University of New South Wales – Heather Gidding
University of Sydney – Fabio Ramos
University of Technology Sydney – Data Arena
University of Wollongong – School of Engineering and Information Science
University of New South Wales School of Public Health and Community Medicine
Veda
Victoria University – John Houghton
Victorian Office of the Commissioner for Privacy and Data Protection
Victorian Department of Premier and Cabinet
Victorian Department of Treasury and Finance
Xamax
Western Australia Department of Health – Data Linkage Branch
World Wide Web Consortium (W3C) (Australia) / Australian Government Linked Data Working Group

(Continued next page)

Table A.2 (continued)

New Zealand

Land Information New Zealand
Ministry of Education
Ministry of Health
Ministry of Justice
Ministry of Transport
Motu
New Zealand Productivity Commission
New Zealand Treasury
Office of the Privacy Commissioner
Statistics New Zealand
Victoria University of Wellington, School of Government — Miriam Lips

United Kingdom

Farr Institute of Health Informatics Research – Ruth Gilbert
UK Statistics Authority – Ed Humpherson

Table A.3 **Roundtable details and participants**

20 June 2016 — University of Melbourne

Children's Bioethics Centre and Centre for Health Equity	Professor Lyn Gillam
Department of Microbiology and Immunology	Professor Lorena Brown
Office for Research Ethics and Integrity	Dr Daniel Barr
Department of Surgery, Austin Campus and Faculty of Medicine, Dentistry and Health Sciences	Professor Arthur Shulkes
Deputy Vice Chancellor (Research)	Professor James McCluskey
Victorian Comprehensive Cancer Centre; Herman Chair of Cancer Medicine, Medicine, Dentistry and Health Sciences	Professor Jim Bishop
Head of Melbourne School of Population Health	Professor Terry Nolan
Architecture, Building and Planning, Lecturer in Urban Analytics	Dr Gideon Aschwanden
Department of Computing and Information Systems	Professor Lars Kulik
Department of Computing and Information Systems	Professor Chris Leckie
Research Computation Strategy and Infrastructure Services	Dr Steven Manos
Melbourne School of Population and Global Health	Professor John Mathews
Melbourne School of Population and Global Health	Professor Janet McCalman
Melbourne Networked Society Institute	Professor Thas Nirmalathas
Melbourne Law School	Professor Megan Richardson
Melbourne Institute of Applied Economic and Social Research	Professor Anthony Scott
Pro Vice-Chancellor Research Collaboration and Infrastructure	Professor Liz Sonenberg
Australian Urban Research Infrastructure Network	Ms Emma William

B Australia's public sector data infrastructure

This appendix gives an overview of the key components of Australia's public sector data infrastructure. This is a large topic and the appendix is not intended to be comprehensive. Rather, it aims to give a reader unfamiliar with the operation of Australia's public sector data infrastructure a sketch of the existing *institutional* and *technical* frameworks. It works in conjunction with chapter 3, which provides broader detail on public sector data collection and access, and appendix C, which aims to sketch the *legislative* frameworks governing Australia's data holdings. Table B.1 summarises the discussion that occurs in appendixes B and C.

Each section in this appendix provides a mix of technical detail, where required; detail regarding current arrangements in Australia; and information on recent developments and innovations of note.

B.1 Public sector open data

Institutions responsible for public sector open data

Australian Government

Responsibilities relating to data policy and practices are distributed throughout Australian Government agencies, and many agencies are responsible for multiple roles in this field. Key bodies include:

- The Department of Prime Minister and Cabinet (DPMC) has a whole-of-government responsibility for driving the implementation of the Australian Government's open data policy. In partnership with the Commonwealth Scientific and Industrial Research Organisation's (CSIRO) Data61, it is responsible for overseeing and implementing the data.gov.au site.
- The National Statistical Service (NSS) is a community of government agencies, led by the Australian Bureau of Statistics (ABS). It publishes a range of guidance, including on deidentification, and publishing information development plans. The ABS also coordinates statistical activities relating to collection, compilation, analysis and distribution of statistics.

Table B.1 Components of public sector data infrastructures

	Jurisdictions that have implemented this	*Jurisdictions that have not implemented this*
	Policy framework	
Open data policy?	All jurisdictions except NT	NT (some data is made open, however)
Ownership of policy?	UK – Letter from Prime Minister to departments and agencies encouraging data to be made open NZ – Bill English, Minister of Finance AUS – Dept of Prime Minister & Cabinet (includes Digital Transformation Office) ACT – Chief Minister, Treasury & Economic Development Directorate (Chief Digital Officer) NSW – Dept of Finance, Services & Innovation QLD – Dept of Premier & Cabinet SA – Dept of Premier & Cabinet (Office for Digital Government) TAS – Dept of Premier & Cabinet (Office of eGovernment) VIC – Dept of Treasury & Finance; Dept of Premier & Cabinet WA – Dept of Premier & Cabinet	
	Legislative Change	
Umbrella legislation?	NSW, SA	All others
Other legislative reforms?	NZ – Privacy Act (research, information-sharing agreements) UK – bill before Parliament but has not been passed QLD and NSW – FOI legislation follows a 'push' model: disclose by default, so FOI application should be last resort	VIC (although Information Technology Strategy 2016–2020 moots reform of FOI legislation), ACT, AUS, NT, TAS, WA
	Institutional change	
Central institution holding and releasing data?	NZ – Integrated Data Infrastructure (IDI) NSW – Data Analytics Centre WA – Data Linkage Branch, similar to IDI VIC – not created yet but ICT Strategy 2016–2020 includes establishment of a state data agency	UK (more devolved), ACT, AUS, NT, QLD, SA (current reform may see one established), TAS
Risk-based approach?	SA (has explicitly adopted a 'five safes' approach for within-government sharing of data)	Many piecemeal moves towards greater use of trusted access models
Non-restrictive licensing?	All jurisdictions moving towards Creative Commons licenses for public sector data. Progress and policy very mixed: NZ policy is strongly pro-CC (agencies are directed to take NZGOAL into account) while in WA the policy only weakly recommends CC.	

Sources: appendix B; appendix C; chapter 3; chapter 5; chapter 6; GOV.UK (2010); Government of South Australia (2016b); McColl (2010); New Zealand Government (2016); OGCIO (WA) (2015); UK Parliament (2016); Victorian Government (2016); *Government Information (Public Access) Act 2009* (NSW); *Right to Information Act 2009* (Qld).

- The Digital Transformation Office (DTO) has responsibility for, among other things, making it easier for people to prove who they are to government online, and creating cloud.gov.au, to make delivering and operating government services easier.
- AUSGOAL is responsible for developing licences and driving the adoption of Creative Commons licences.

- The Australian Government Information Management Office (AGIMO) (now within the Department of Finance) previously had a role in developing standards. More broadly, the Department of Finance has produced guidance on charging for data services, a range of ICT policies and standards including security, business continuity, authentication and identity management, and a data centre strategy, and is responsible for implementation of the cloud storage policy under the Protective Security Policy Framework (PSPF).
- The Australian Signals Directorate (ASD) and the Attorney General's Department (AGD) are responsible for the PSPF, which covers data security matters and guidance on when cloud services can be used.
- The National Archives of Australia (NAA) has a responsibility for standardising metadata (insofar as it relates to their archival functions) and driving the implementation of the Digital Continuity Policy (appendix C)
- ANZLIC is the peak organisation for spatial data in Australia and New Zealand, and provides access to the wealth of spatial data and services provided by a wide range of organisations in the public and private sectors.

There are also a number of data-related groups within the public service. The Secretaries Data Group and the Deputy Secretaries Data Group promote public data initiatives across Australian Government entities, and the Data Champions Group promote the use, sharing and re-use of data. The Intergovernmental Committee on Surveying and Mapping provides leadership, coordination and standards for surveying, mapping and charting, and national datasets. Other groups responsible for data policy include the Open Access Working Group, and the Science Agencies Data Stewardship Working Group (chapter 3). Additionally, the Council of Australian Governments (COAG) has a role to play in agreeing to the collection of National Minimum Data Sets in areas such as health, education, and disability services.

State and territory governments

Each state and territory government also has governance structures in place regarding data collection, sharing and release.

In New South Wales:

- The NSW Data Analytics Centre (DAC) was established within the NSW Department of Finance, Services and Innovation to facilitate data sharing between agencies to inform more efficient, strategic, whole-of-government evidence-based decision making by leveraging internal and external partnerships so that the right capabilities, tools and technologies are applied.

In Victoria:

- The Department of Premier and Cabinet administers DataVic.

- The Department of Treasury and Finance is responsible for data policy, and driving open data in Victoria.

In Queensland:

- The Queensland Spatial Information Council provides a strategic forum for the spatial information industry. It is comprised of representatives from the professional, academic, industry and government sectors and the community.
- The Department of Premier and Cabinet is responsible for Queensland's open data strategy.

In South Australia:

- The Office for Digital Government within the Department of Premier and Cabinet is responsible for South Australia's digital transformation, and implementation of South Australia's open data policy.

In Western Australia:

- The Department of Premier and Cabinet (DPMC) has developed the WA whole-of-government Open Data Policy.
- The Office of the Government Chief Information Commissioner is an independent agency in the WA Government established to address whole-of-government ICT issues and future directions.
- Landgate is responsible for implementation of Western Australia's open data policy.

In Tasmania:

- the Office of eGovernment is responsible for leading the development of an ICT strategy for the Tasmanian Government and building governments statistical assets and capability through Stats Matter, which is Tasmania's open data initiative
- the Tasmanian Government Statistical Policy Committee is responsible for monitoring progress of the implementation of Tasmania's open data policy.

In the Northern Territory:

- the Northern Territory has not announced an open data policy
- the Northern Territory Land Information System and the Department of Mines and Energy are responsible for spatial data.

In the Australian Capital Territory:

- the Office of the Chief Digital Officer is responsible for the ACT's open data policy.

What open data does Australia provide?

Accessing public sector data occurs in a number of ways presently.

A large amount of public sector data is available via government operated websites (table B.2). For the Australian Government, for example, data.gov.au provides both a registry and repository[54] for open data. In addition to open datasets, the data.gov.au catalogue includes unpublished data and data available for purchase. Data.gov.au contains a Toolkit, which is designed to assist agencies with practical information on how they can make their data open, how they can build agency capabilities, and the benefits of open data. It provides hosting for tabular, spatial and relational data with hosted APIs and the option for agencies to link data and services hosted by other government sources. Similarly, all states and territories (except for the Northern Territory) have their own open data websites.

Public sector data is also made available on a wide range of other websites, not just the designated open data ones. As is discussed in greater detail in chapter 3, this proliferation can mean that the data obtained across disparate sites is frequently not interoperable — fragmentation of data releases is common, which can prove problematic (chapter 3).

Real time data

Timeliness is an important factor in data quality, given that for any forward-looking use of data — all other things being equal — the most up-to-date data will have the greatest accuracy. For data that is openly released to the public, timeliness affects the viability of: commercial applications of the data (using the data to improve or target existing products and services; *creating* products and services such as apps from the data itself, such as apps) and the value individuals can derive from personal uses of the data to improve their lives (whether they use the data at its original source or via a data-driven app).

Currently, a handful of public sector data types are published in real-time: that is, as a new data point is inputted into a government agency's dataset, it is automatically and immediately — or with a small delay — published on an open data platform. The effect is that of a 'live stream' of data, which does not rely on human labour to update the published dataset. If this is done using a real-time API, the data can be 'pushed' directly into any interoperable app, rather than needing to be manually inputted from a document format such as HTML, XLS or CSV.

Much non-confidential data has the potential to drastically improve public safety by being released in real-time. For example, real-time streams of fire danger ratings, maps of currently burning fires, or wind speed and direction measurements (for both fires and storms) could help people and towns plan their hazard management strategies earlier and more appropriately.

[54] Data is stored in a repository (storage, mechanism to read and write) and findable via a registry (a set of things defined by a context). Registries come in different forms —sometimes there can be a registry of registries, or meta-registry that contain a register of registries. Regardless, the important point is that storage of data and searching for data do not have to be done in the same place.

Table B.2 **Selected Australian public sector open data**

Jurisdiction	*Websites that data is available from*
Commonwealth	• data.gov.au – open data catalogue • find.ga.gov.au – spatial data catalogue • National Map • Soil and Landscape Grid 9CSIRO0 allows download of soil and landscape attributes to a GIS client • abs.gov.au – Australian Bureau of Statistics • RecordSearch – National Archives of Australia • various individual public sector bodies, including the Bureau of Meteorology, GeoScience Australia, Reserve Bank of Australia, ABARES, BITRE, Tourism Research Australia, and the Great Barrier Reef Marine Park Authority all publish data on their websites
New South Wales	• data.nsw.gov.au - NSW open data catalogue • the NSW Spatial Data Catalogue is the central repository for the publication of metadata describing NSW Local and State Government Spatial Data • OpenGov NSW lists other information published by NSW Government agencies, including annual reports and open access information • NSW Spatial Information Exchange (SIX) provides a high resolution map of NSW with flood imagery and lot boundaries https://maps.six.nsw.gov.au/ • State Records NSW allows state archival records to be searched • NSW Government Information Asset Register provides searchable metadata and contact details for a list of core-value information assets, including datasets, held by NSW Government agencies (accessible only to NSW Government employees) • other bodies, such as Land and Property Information NSW and Tourism NSW and the NSW SES service provide data and/or spatial data on their websites.
Victoria	• data.vic.gov.au is the Victorian Government data directory • data.melbourne.vic.gov.au/data is the City of Melbourne's open data platform • vicroadsopendata.vicroadsmaps.opendata.arcgis.com is the VicRoads open data website • Vic Spatial Datamart is the main spatial data site in Victoria. VicMap is available as an API • Geovic (energyandresources.vic.gov.au) is a free web mapping application that allows users to search geospatial databases and display the results as maps or tables • Forest Explorer (Department of Environment, Land, Water and Planning Victoria) maps forests and forest recreation tracks • VICNAMES allows search of all registered and recorded place names in Victoria • Public Record Office of Victoria provides a searchable database of Victorian Government archival data • GDA2020 is being developed as a new modern datum for Australia • other bodies, such as Tourism Victoria, also provide data on their websites
Queensland	• data.qld.gov.au is the Queensland Government open data website • QSpatial is the Queensland spatial data catalogue • MinesOnline maps provides spatial information relevant to the mining and resources industry • maps of environmentally sensitive areas are on the Department of Environment and Heritage Protection website • SmartMaps allows access to current information about Queensland property boundaries, valuation and sales data

(continued next page)

Table B.2 (continued)

Jurisdiction	*Websites that data is available from*
Queensland	• QTopo allows online access to topographic maps • Queensland Globe is an interactive online tool that can be opened inside Google Earth to view and explore Queensland spatial data and imagery • a range of geological maps are available through Geological Survey Queensland • QDEX allows submission and search of company statutory reports and access to maps and publications from the Geological Survey of Queensland and to other government publications including departmental annual reports and the Queensland Mining Journal • the Queensland State Archives allows online search of their archival holdings • other bodies, such as Tourism Queensland, also publish data on their websites
South Australia	• DataSA is the South Australian Government's open data portal. They collaborate with data.gov.au to share metadata about government datasets, which allows the two sites to be interoperably searchable • NatureMaps and WaterConnect are online mapping applications provided by the Department of Environment, Water and Natural Resources • Location SA Map Viewer allows search and display of a huge range of government data on a road or satellite base map • population projects and demographic information are available from the Department of Planning, Transport and Infrastructure website • South Australian Resources Information Geoserver contains a number of online map applications, spatial data, drilling, and geochemistry data • other bodies, such as Tourism SA, also publish data on their websites.
Western Australia	• Open Data WA is the WA Government's open data portal • Landgate SLIP Enabler is an open data platform for location-based information. Datasets can be accessed through SLIP on a wide range of maps. Most of the data is publicly available, although some is available only by subscription and some are restricted for use only between agencies. Most data published through SLIP is now also searchable through data.wa.gov.au • the State Records Office of WA provides a searchable catalogue for the vast majority of its holdings. • other bodies, such as Tourism WA, also publish data on their websites.
Tasmania	• some Tasmanian Government open data is published on data.gov.au • Land Tasmania manages Tasmania's foundation spatial datasets, including The List (which shares Tasmanian location-based information), TASMAP (maps of Tasmania, including national parks and bushwalks) and the Tasmanian Imagery Program (a strategy for the ongoing acquisition and delivery of imagery (remotely sensed) data for Tasmania • the Tasmanian Environmental Protection Authority shares some data such as real time air quality monitoring • the Tasmanian Fire Service publishes a map of controlled burns and bushfire alerts • Mineral Resources Tasmania publishes a wide range of geological and geospatial data • Forestry Tasmania publishes a map of some of the geographic information system data that they use to manage the Permanent Timber Production Zone land • Glenorchy City Council has a new mapping and spatial data sharing website that documents spatial data that is licensed under Creative Commons

(continued next page)

Table B.2 (continued)

Jurisdiction	*Websites that data is available from*
Tasmania	• the Department of Primary Industries, Parks, Water and Environment publishes TASVEG, which is a comprehensive digital map of Tasmania's vegetation, including sub-Antarctic Macquarie Island. • Tasmanian Archives Online allows online search of their archival holdings. • other bodies, such as Tourism Tasmania, also publish data on their websites
Northern Territory	• some open data from the Northern Territory is published on data.gov.au • the NT Spatial Data Broker is a subscription service offered by certain NT Government agencies allowing private users to download vector (for example, shapefile/tab file) spatial data • the Department of Mines and Energy holds a range fo data on mining titles and mining maps • the Northern Territory Geological Survey (Department of Mines and Energy) releases geological and geospatial data • The Northern Territory Archives Navigator allows online search of their archival holdings • Tourism NT provides data on the number of visitors to the NT • Natural Resources Maps NT is a web mapping tool to discover, research and map natural and cultural research data
Australian Capital Territory	• data.act.gov.au contains ACT open government data and spatial data • Archives ACT allows online search of archival records • ACT map allows people to access ACT Government location information • other bodies, such as VisitCanberra, also publish data on their websites.

This would be a significant outcome because Australia's national Emergency Alert system currently relies on telephone coverage, and therefore a range of factors can lead to emergency text messages not being received by people in an affected area (Emergency Alert nd).

Some data relating to potential natural hazards is already published in real-time. The Queensland Department of Natural Resources and Mines, for example, operates a Water Monitoring Information Portal (WMIP) that provides real-time stream height and stream flow values from water monitoring sensors at stations throughout Queensland, as well as groundwater levels from monitoring bores equipped with data loggers (Queensland Government 2016). In South Australia, the Country Fire Service operates an RSS feed of current and daily fire incidents at data.sa.gov.au (and linked to data.gov.au). Similarly, the Tasmania Fire Service has an RSS feed of current bushfires and alerts at fire.tas.gov.au, also linked to data.gov.au.[55] And the South Australian Fire and Emergency Services Commission offers a smartphone app called 'Alert SA' that incorporates real-time data from more than a dozen government agencies to provide users with warnings and

[55] RSS (Really Simple Syndication) is a system for embedding semi-structured news snippets into a notification 'stream' for subscribers, using a variety of XML formats. Websites publish these snippets to inform subscribers of new content at the website. Subscribers receive the notifications in real-time with an RSS reader, and can receive multiple streams at once if they use an aggregator that continuously 'polls' the subscriber's chosen RSS websites for notifications.

information about hazards in their proximity and selected zones (Government of South Australia 2016a). One agency publishing water quality data is the New South Wales Department of Primary Industries' Office of Water, which offers a smartphone and computer app that, along with providing similar stream height, stream flow and groundwater level data to Queensland, also shows water salinity, temperature and turbidity measurements (NSW Government 2016).

Some other types of non-confidential data do not carry such significant public safety or health outcomes, but could support increased productivity and spur innovation. Two related examples are real-time public transport data — that is, live train, tram and bus location status and corresponding arrival times factoring in delays and cancellations — and traffic congestion data. Publishing this data openly allows commuters to better utilise their time by spending less time waiting in traffic or at public transport stops and more time working productively. In Australia, several jurisdictions already publish one or both of these types of data online or incorporate them into smartphone and computer apps, and some (not all) also provide a live data feed to third-party developers through open APIs.

Other public sector data that could support increased productivity if published in real-time might include the waiting times at various government agencies — such as Centrelink, Australia Post offices, or state and territory vehicle licensing centres. Public hospital emergency department waiting time data — which is published in near real-time by the Western Australia Department of Health (Department of Health (WA) 2016) — could also be of significant benefit. Meanwhile, real-time feeds of some types of data could also encourage innovation and greater engagement with parts of the public sector and, if monetised, could eventually result in tax revenue for governments. Examples of data types that have been used to create monetised, recreational apps overseas include surf/swell data and food service health code violation data.

While the potential benefits to openly publishing some types of public sector data in real time are numerous and include supporting health and safety, productivity, and innovation, there are also potential drawbacks to some real time data. Examples include the risk that incorrect, inaccurate or low-quality data (perhaps resulting from faulty sensors, incorrect calibration, or damage to equipment) could be published and could mislead or unnecessarily alarm the public, or the risk that data could be misinterpreted:

> …[M]embers of the public may identify pollution problems based on data that isn't credible because it wasn't collected using established Environmental Protection Agency (EPA) or state monitoring protocols. … [Also] data may be collected from a monitoring device of which the EPA or the state is unaware, or real-time data may be incorrectly interpreted for standards that are based on longer-term averages, such as the daily average for a particular pollutant. (Saiyid 2016)

The normal approach to minimising this risk would be to mark data as provisional and automatically generate quality flags based on machine processing of raw data. In this case, the quality controlled data would be released at a later stage. Thus there can sometimes be a trade-off between the quality and timeliness of the data. One option for agencies to

minimise the delay brought about by these processes (and therefore reduce the time between data collection and publishing) may be to set and achieve broad standards in the data collection stage, since less work would then be required to make the data usable. However, this comes with its own issues of technology and training costs.

Along with the issues discussed for non-confidential data — potential delays to publication or sharing brought about by the need to spend time on data quality assurance, standardisation, and ensuring data is fit-for-purpose — the issue of privacy protection is also one that could cause delays. Some identifiable data, such as that in the Australian Criminal Intelligence Commission's National Criminal Intelligence System (NCIS, the successor to CrimTrac), does not require de-identification before being shared with agencies because the point of the dataset is to identify individuals (and the NCIS is legislated for outside of the *Privacy Act 1988* (Cth) (Privacy Act). Other identifiable data is subject to the APPs and as such must often be de-identified before being disclosed to another entity.

Depending on whether de-identification is done manually or automatically, and the methods used to de-identify the data in either case, this step could present a significant delay in the process of making data shareable between agencies. If this data were to be shared in *close* to real-time, or even if agencies wanted to establish an automatic sharing system with small delays built-in, the data would need to be de-identified automatically (via data management software) instead of manually. Agencies may be apprehensive about automated 'on the fly' de-identification techniques failing or missing steps, though the ABS's TableBuilder applies these techniques with success.

The feasibility of any real-time data publication scheme must therefore be considered on a case-by-case basis.

Application Programming Interfaces (API)

The Application Programming Interface (API) is a concept from computer programming. APIs describe the names, protocols and input, outputs and their representations, of systems that provide specified functions. They allow computer software to communicate with other computer software in controlled ways for an outcome without knowing how the software chooses to implement that outcome.

APIs allow machines to communicate with each other quickly, efficiently and reliably. They allow content that is created in one place to be dynamically posted and updated in multiple locations on the web, mobile, TV, etc. Thus, one benefit of APIs is that they can support real time data (discussed earlier).

APIs are most useful when a resource needs to be regularly or frequently accessed or a resource needs to be combined with other resources. Most data providers would hope that both of these are true for their resources, so machine accessibility must be required in most cases. To be useful for machines, the resource must be provided in a way that:

- allows a machine to read it in a machine accessible format (an API)
- uses terminology which is of broader use or at least provides transformations to a more usable set of terminology
- has a context which is either well known or well described in a machine accessible way.
- engenders trust in the resource via a validation suite, provenance trace or a certification process.
- is interpretable for humans — descriptions and analytics that explain the data to humans are required.

When considering APIs, there are four types of uses:

- Public: an agency makes information and services available to almost anyone to use for building their own applications. These APIs are built on top of public information and services. Applications can be used commercially. Developers can, for example, create a mashup that uses government data — like Census block data — as a supporting part of an application.
- Private: Organisations use APIs across offices and divisions to share data to improve access and efficiency. These APIs are built on internal information and services. Amazon famously required all data and functionality to be available *only* through APIs. This created modular services that could be re-used easily and prepared Amazon for innovation, such as easily deploying apps to smartphones.
- Hybrid: Some APIs are available both externally and internally. The organisation can limit access to some information to the public and make more available for internal use or to use with specific partners. It is important to understand these uses and to apply appropriate security, legal and technical rules, depending on the use.
- Authorised: APIs that have restricted access. It may be restricted by the network such as an internal API. It may also be restricted to authorised connections from correctly authenticated users. For example a bank may allow use of an API from a known list of other financial institutions (Fielding 2000; GSA nd).

APIs are mostly either read-only or read-write. A general way to differentiate the two is whether the underlying material is information that is meant to be broadcast (read-only) or a service that is meant to allow a consumer to interact with the government and supply information, such as submitting an online form (read-write). Security considerations will differ just as they do with web pages (static vs. a system interface). The organisation's security team will be an important part of API design and deployment.

A common architecture in web based APIs is Representational State Transfer (REST). REST is preferred by many because it is based on the familiar HTTP web protocol. REST also typically supports content negotiations — that is, a data representation can be returned to a program while a browser-supported representation (such as HTML) can be returned to browsers or web search engines. Both can be generated from the same source. This allows

web search engines to access the web pages and search them like any other so people can find resources while machines can interact with REST APIs (box B.1). One of the key characteristics of the REST API is that it is stateless, which means the server does not keep track of the context (or state) of a set of operations. Everything required to understand the context is transferred to the caller (or client). This make the server implementation much simpler (Fielding 2000; GSA nd).

Web based API formats are commonly either Extensible Markup Language (XML) or JavaScript Object Notation (JSON). JSON is increasingly popular with developers due to its speed, ease of use, and wide acceptance.

The API becomes standard (by an official process or not) across a community, when a community reaches an agreement on its form. A standard becomes effective when there is community adoption. Agreements covering the usage patterns, content and API of information enable easy flow of information around the community. Application development effort can then be focused on utilising that information.

Box B.1 APIs and linked data

Linked data is an ecosystem that uses REST to link data objects on the web. Objects are referenced using a Uniform Resource Identifier (URI). Providers of linked data ensure the objects returned have links to other data objects (possibly by other organisations on other web sites) so the chain of objects can be followed. Linked data allows the knowledge about an object (its properties) to be distributed. Different organisations can describe properties of the same objects and the combination of all those properties is considered the full property set. Linked data follows the open world assumption which essentially says that a class of objects can be described that has strict relationships between the properties and the class, however, any instance of the object may not necessarily have the required properties. Any missing properties are assumed to exist elsewhere.

For example, an object representing Manchester Railway Station can have properties relating to its physical layout and properties relating to its train traffic. These may be held by multiple organisations and if each use the same URI for Manchester the properties can be combined. If there are conflicts in the properties then this will be clear and must be resolved at some stage. Even if the two organisations used different URIs for Manchester Station a third party could make an assertion that they are the same. A view of information including that assertion could attempt to combine properties.

The ability to make assertions is freeing and dangerous. It is freeing in that third parties can act as integrators of existing data on the web. It is dangerous in that erroneous assertions can be made. Tracing the provenance and trust in the asserter is critical in this case.

Source: Heath and Bizer (2011).

APIs are essential for timely and repeatable functions but they come at a cost. They require a computer programmer to tell the computer how to follow the API for the desired outcome. Most users will not wish to deal at this level so the code using the API must be built into more usable applications specific to the task at hand (such as web browsing). The cost of embedding these programs must be spread across a big enough user base so that the

cost to a user is reasonable. APIs are therefore most effective at scale where there are many users in a community and many programs use the API.

Examples of public sector data being provided via APIs

APIs can take many forms — the most functional enables a third party to query a dataset at a granular level. The publishing of these types of APIs allows developers in agencies and from outside government to build apps, widgets, websites, and other tools based on government information and services. Allowing consumers to get the information they need from many places, not just the government website. For example, in the United States, the National Weather Service publishes an API that makes weather data available to developers within and outside of the organization. The API offers real-time access to data so that an app can automatically access the latest information instead of requiring a developer to return to the agency's website and manually copy each update. This supports an enormous and innovative range of products that present up-to-date weather information to the public (GSA nd).

Some public sector data is made available via an API. For instance, in the United States:

- the Federal Aviation Administration provides travel websites and mobile apps with live airport status and delay information through its Airport Service API
- the Pillbox API from the National Library of Medicine powers third party mashups that serve consumers who need to quickly identify an unknown pill
- the Sunlight Foundation's Scout project contacts the Federal Register API to provide alerts and notifications for formal government action.
- the United States requires public sector agencies to establish central online resources for outside developers and publish their open data in structured machine-readable formats by default, with data.gov pulling much of its data from agency websites continuously (Office of the Press Secretary (US) 2012).

However, implementation varies across jurisdictions and agencies, and can be problematic. For instance:

- Data.gov.au uses Comprehensive Knowledge Archive Network (CKAN) to register its metadata and store some of the registered data. CKAN does expose an API to its repository but the documentation for that is hard to find on data.gov.au. Where spatial data is registered, then data.gov.au sets up OGC WFS or WMS services to provide those APIs over the data (DPMC 2015).
- The NAA has an API for World War I service dossiers. It uses REST and is only available to authorised persons. The NAA has also registered two datasets with data.gov.au which are available for download. The NAA is running the Government's continuity 2020 program where in principle three it expects public records to be available via APIs so that software like National Map can access them (NAA 2015; National Archives of Australia 2014).

- The Queensland Government has data available on data.qld.gov.au which also uses CKAN. The tide data for example is available for download, visualisation and available through a REST API which is documented on their data page. The format of the tide information is CSV and while that is machine readable it lacks the use of a formal standard for observational data such as ISO 19156 Observations and Measurements. It also combines the measurement date and value in the one column which renders their visualisation tool unusable. The metadata download for the tide gauge (content metadata) is available in a custom text format with no obvious API (Queensland Government 2015).
- The NSW government provides API access to its land and property spatial data. The data is provided through REST, SOAP (Simple Object Access Protocol) and WMS APIs which it documents. From 1 July 2016 NSW Land and Property Information is now offering titling data through third parties on a fee for service basis (DFSI (NSW) nd; Land and Property Information 2016).

B.2 Research data

Not all government data is indexed on a registry or made publicly available. For example, the National Centre for Longitudinal Data, which is operated by the Australian Government Department of Social Services, does not make its data public. Restricted access to government data is discussed further in chapter 3. This section discusses specific arrangements for research data.

Ethics committees

Researchers wishing to access the data that custodians hold must undergo a stringent application process requiring approval from each data custodian and also from a Human Research Ethics Committee (HREC) that certifies that the study is valid and in the public interest.

Institutions that undertake research 'with or about people, their data or tissue' are responsible for ensuring that research they conduct, or for which they are responsible, is ethically reviewed in accordance with the National Statement on Ethical Conduct in Human Research (NHMRC, ARC and AVCC 2007) (appendix C). Institutions may establish their own processes for ethical review or use those of another institution.

The National Statement provides that ethical review can be undertaken at various levels depending on the degree of risk involved in the research. Research involving more than a low level of risk must be reviewed by a HREC. The National Statement expressly provides that research proposing to use personal information in medical research without consent and research using health information without consent must be reviewed by an HREC. Research involving 'no more than a low level of risk' may be reviewed by a non-HREC ethical review body, such as a departmental committee, or a subcommittee of an HREC.

Research involving 'negligible risk' and the use of existing collections of data or records that contain only non-identifiable information may be exempt from review.

HRECs must be composed and function in accordance with the National Statement. The minimum membership of an HREC is eight: a chairperson; at least two lay people (one man and one woman) who have no affiliation with the institution; at least one person with knowledge of, and experience in, the professional care, counselling or treatment of people; at least one person who performs a pastoral care role in the community; at least one lawyer; and at least two people with current research experience. The primary responsibility of HREC members is to decide whether a proposal meets the requirements of the National Statement and is ethically acceptable.

Both the Section 95 and 95A Guidelines under the Privacy Act (appendix C) provide a detailed framework within which HRECs must consider the privacy implications of research proposals involving the use of individuals' personal or health information. In particular, HRECs must consider, and may approve, research proposals seeking to use personal or health information without consent, on the basis that the public interest in the research substantially outweighs the public interest in maintaining the level of privacy protection.

The Guidelines require that, before making a decision, an HREC must assess whether it has sufficient information, expertise and understanding of privacy issues, either among the members of the HREC or otherwise available to it, to make a decision that takes proper account of privacy. The Section 95A Guidelines note that it may be necessary to appoint additional members with specific expertise in some circumstances. It is important to note that, although an HREC may give approval for a research proposal to proceed, the final decision to release personal information to researchers is not made by an HREC, but by the relevant data custodian (ALRC 2008).

HRECs are usually established by organisations that conduct research involving humans, such as universities and hospitals. Not all organisations that conduct research have their own HREC — some use the services of HRECs that are based within other organisations. They may also use HRECs that are established by organisations that do not conduct research, but have established an HREC to provide the service of ethical review to researchers who do not have an HREC at their own organisation or who are not associated with an organisation. There are more than 200 HRECs in institutions and organisations across Australia (NHMRC 2016).

The National Statement on Ethical Conduct in Human Research (NHMRC, ARC and AVCC 2007) sets out the requirements for the composition of a HREC and the relevant ethical principles and values by which research should be designed and conducted and to which HRECs should refer when reviewing research proposals. It also identifies requirements and responsibilities for: institutions/organisations in establishing HRECs; researchers in submitting research proposals to HRECs; and HRECs in considering and reaching decisions regarding these proposals and in monitoring the conduct of approved research. In some circumstances, HRECs charge fees for considering research applications

(NHMRC 2016). However, the National Health and Medical Research Council (NHMRC) cautions:

> Professional judgement is needed in the interpretation of this guidance document as no single document adequately captures the full range of legislation, standards and guidelines that apply to human research. Good practice in research governance depends on those with research governance responsibility being appropriately skilled and experience and working in an environment that enables them to use their professional judgement effectively. (NHMRC 2011, p. iii, 2014).

HRECs are encouraged to register with the NHMRC. A list of registered HRECs is available on the NHMRC website.

Moves towards harmonisation

The concept of research governance has grown from being considered an ancillary responsibility of the HREC to one that is the responsibility of the institution where the research is being conducted. This is because research governance encompasses both ensuring adequate ethical review and institutional considerations about undertaking research, in the context of the institution's policies, strategic priorities, expertise, resources, contractual arrangements, financial issues and approach to risk management. Thus, although some research governance activities supporting collaborative research may be coordinated across centres (for instance, the application for a single ethical review), each institution remains responsible for authorising the commencement of research, and for the appropriate governance of research activity at each institutional location where the research is carried out.

In the national approach to single ethical review, site assessment and project authorisation are the responsibility of each institution participating in a multi-centre human research project while ethical review is provided by only one HREC using certified ethical review processes (NHMRC 2011).

Data integration

Data integration involves combining data from different sources and providing users with a unified view of that data — for instance, to produce new datasets for statistical and research purposes. *Data linkage* is that part of the process that involves creating links between records from different sources based on common features in those sources (NSS nd). *Data curation* is a broad term used to describe the active and ongoing management of data throughout its lifecycle (CLIR 2014).

A major advantage of data integration is that it allows further insights into data that is already available, so it is a cost effective and timely way of gathering more information in order to help improve social, economic and environmental wellbeing. It also reduces the

duplication of information collection from people and businesses, as integration projects make use of existing information which was collected from them for other purposes.

It is worth noting that there can be privacy-preserving benefits from this process — for instance, in the way data linkage is carried out (box B.2). Prior to the development of data linkage systems, most research using linked data involved the use of identified data. In other words, researchers received both the personal information (for example, name, address, date of birth) as well as the related content information (for example, diagnosis and treatment). The data linkage process and the use of coded linkage keys to better protect a person's privacy has meant a significant change in the way health and health related research is carried out. Identified data is now only rarely released to health researchers as it is not needed for many projects. Data linkage has also supported an expansion in population-based health research in an environment where personal details can be protected. Finally, the access and use of these research datasets is strictly controlled and managed — usually through a trusted access model such as the Secure Unified Research Environment (SURE), discussed earlier.

More recently, CSIRO has developed technology that allows data to be linked and subsequently analysed in an encrypted form. As a result, data holders can allow external users to analyse their data and add value by linking it with other datasets without revealing underlying raw values. While promising, the use of this technology is in its early stages (CSIRO 2016b).

Institutional arrangements

Several institutions undertake data integration in Australia. These institutions operate at both state and federal levels.

The Population Health Research Network

The Population Health Research Network (PHRN) was established to build a nationwide data linkage infrastructure capable of securely and safely managing health information from around Australia. With the establishment of the PHRN, data linkage units and managing nodes now operate across every State and Territory in Australia, allowing population health research to be carried out more thoroughly and more effectively. The PHRN is a national network comprising a Program Office located in Perth, Western Australia, a Centre for Data Linkage located at Curtin University in Western Australia, a secure remote access laboratory located at the Sax Institute in New South Wales (box B.3) and a network of Project Participants and Data Linkage Units located in each State and Territory.

Box B.2 The process of data linkage

Broadly speaking, there are two techniques for linking entities: deterministic and probabilistic. Deterministic linkage follows prescribed rules on how identifying attributes of entities match. Probabilistic linkage gives a degree of confidence that two entities match based on both deterministic rules and similarities in properties. Probabilistic linking is slower but generates a higher rate of matches. Regardless of the linkage technique used, it is good practice to separate personal identifying information (such as name and address) from a dataset and store it separately, generally in a different system and encrypt it. Separated records are given a token to allow record correspondence to be kept. Identity token systems reduce the risk of inadvertent identification, but do not on their own completely prevent re-identification.

Once a project is approved, the data custodians and the staff at the Data Linkage Unit (DLU) work together to determine which records are required for the study to ensure the minimum amount of information is provided to the researcher. The data linkers then use the Linkage IDs to create Project Linkage IDs that are specific for the approved study. They then send the Project Linkage IDs, along with the Record IDs of the required records, to the data custodians.

Using the Record IDs, the data custodians extract the required records from their collections and replace the personal information of each record with its matched Project Linkage ID. The researcher is then provided with the content data of each record and its corresponding Project Linkage ID by each data custodian.

Data custodians provide regular updates of the personal information and Record IDs to the data linkers. The data linkers then check the new data against the existing personal information to see if they already have Linkage IDs for these records using a statistical probability method.

Using the Project Linkage ID, the researcher can determine which records from different datasets belong to the same person without having access to the personal information in order to create a merged dataset for their analysis. The other portion of the record containing the health, education or other data (known as content information) remains with the data custodians, meaning that the data linkers never have access to this data. Linkage IDs are stored on secure computer servers and can only be accessed by authorised DLU staff.

In the AIHW new tokens are minted on a project basis and mapped to the original token. This means the token identity of the record is not preserved across projects.

There are positives and negatives to this approach. A project's disclosure risk can be assessed on the data it will use and not the data it might use in future projects. The down side is that if a researcher adds value to the datasets then that value is lost for future projects. For example in medical records the disease is often mentioned in a free text field written by the doctor. If a researcher analyses these fields and uses a controlled term for the disease, the relationship to the record only survives for the life of the project.

Finally, care must be taken not to confuse similar properties or combine them in inconsistent ways. For example if a person's annual payment summaries were linked to a person from multiple employers it may seem desirable to simply add up the gross payments. It is better to keep the records independent and linked, and from there, create a new property which combines the values.

Source: PHRN (2011a).

Box B.3 The Secure Unified Research Environment (SURE)

The Secure Unified Research Environment (SURE) is a high-powered computing environment developed to help make best use of our national knowledge base. It is helping to bring researchers together from across Australia and the world to collaborate on large-scale projects tackling major health and social issues such as population ageing, diabetes and mental health.

It has been purpose-built as Australia's only remote-access data research laboratory for analysing routinely collected data, allowing researchers to log in remotely and securely analyse data from sources such as hospitals, general practice and cancer registries.

SURE was established with funding from the Australian Government National Collaborative Research Infrastructure Strategy (NCRIS) as part of the Population Health Research Network (PHRN).The PHRN is a collaboration that was set up in 2009 to further develop Australia's data linkage capabilities.

SURE provides a uniquely simple and secure solution to the problem of transferring and storing sensitive data contained in registries, routinely collected datasets and cohorts for the purposes of research and analysis. This data is currently underused in research projects due to privacy concerns. SURE users can access this type of data through virtual workspaces and while they access the data from their own computers, the data never leave SURE.

The data cannot be copied, downloaded or transmitted by email or other means. Researchers can take their analyses from SURE but not the original data. All inputs and outputs are vetted through a unique 'curated gateway' for compliance and the SURE system records and archives all transactions for future reference.

SURE security features

SURE is the only facility that operates as a central, secure, online destination for analysing sensitive human research data, and allows data custodians tight control over what they make available and to whom. SURE access is strongly authenticated requiring three different factors of authentication. Regular on-site and off-site backups of data are made. All off-site backups and archival data are encrypted prior to being transferred to secure off-site storage

All users are required to undertake training on issues of privacy, ethics, information security and statistical disclosure control prior to gaining access to SURE, and sign a deed outlining the terms and conditions of using SURE

Source: Sax Institute (nd).

The PHRN is developing and testing leading-edge technology to ensure the safe and secure linking of data collections while working to protect peoples' identity and privacy. The PHRN is also developing mechanisms for the secure exchange of linkable data between those who hold the data collections and the researchers who receive approval to analyse the linked data. The PHRN data linkage units do not conduct their own health or health related research (PHRN 2016). They do not have a remit for capturing value added to data by projects.

The PHRN commenced operations in 2009 with a $20 million allocation of funds from the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) program. This initial funding covered a four year period from 2008-09 to 2011-12. Subsequently, the PHRN received further funding from the Australian Government's

national research infrastructure programs, with cash contributions totalling over $42 million in in the periods 2008-09 to 2015-16. In addition to the Australian Government cash contribution, government, research institutes and universities have provided significant cash and in kind contributions. In December 2015, the Australian Government announced that it will allocate $1.5 billion over 10 years from 2017-18 for the NCRIS program. The 2016 NCRIS roadmap issues paper raises the question of NCRIS research facilities operating on a user pays basis (PHRN nd).

SA-NT Data Link was established in 2009 as part of the PHRN. It is a collaboration between the Northern Territory and South Australia partners and supports population based data linkage research to inform many areas of policy and service development within South Australia and the Northern Territory, and nationwide. Providing access to accurate and unbiased information held by Government agencies and other organisations that can better inform research, policies and practices is a central role of SA-NT DataLink. Research using de-identified data linked for large or entire populations are much more inclusive, representative and unbiased, and also more cost effective and efficient than conventional studies based on sampling. SA-NT DataLink:

- enables the linkage of administrative and clinical datasets providing an evidence base for researchers and policy makers to better understand and monitor the population health and wellbeing impacts of policy and investment decisions for South Australians, Northern Territorians and other jurisdictions.
- provides the infrastructure to make available access to information for researchers who wish to collaborate within and across jurisdictions in population based research.
- can provide access to sources of information for analysis to assist the achievement of SA and NT government strategic plan targets and COAG goals.

As part of building the population research infrastructure in South Australia and the Northern Territory, SA NT DataLink has been established with very high levels of security and with practices that protect the privacy of individuals (SA-NT DataLink 2016).

The WA Data Linkage Branch has operated within the Department of Health (WA) since 1995 and is also part of the PHRN. The WA Data Linkage System (WADLS) is capable of securely linking data collections from a wide range of sources to enable approved activities including policy, planning and research. The Data Linkage Branch (DLB) manages the WADLS. The WADLS is one of only two well-established and enduring data linkage systems in Australia, the other being the Centre for Health Record Linkage (CHeReL) in New South Wales.

The rapid progress of data linkage centres elsewhere in Australia has, and will continue to benefit from, the lessons learnt and shared by DLB and CHeReL. From its modest beginnings in 1995, the WADLS is now one of the most comprehensive, high quality and enduring linkage systems worldwide. It currently comprises a linkage infrastructure spanning over 400 data collections (both infrastructure and ad-hoc) (Data Linkage WA 2016), representing 88 million links (for about 4.1 million people) dating back to 1945 (Data Linkage Branch (Department of Health WA), sub. 13).

Commonwealth integrating authorities

Under previous approaches, researchers were often responsible for data merging and data access. Under the new Commonwealth arrangements for data integration, integrating authorities carry out these tasks as part of their responsibility for the end-to-end management of statistical integration projects involving Commonwealth data.

In 2010, Australian Government Portfolio Secretaries endorsed seven high level principles for data integration as well as a supporting set of governance and institutional arrangements (box B.4). The principles support the flow in value added data from project to integration authorities and even to data custodians in certain circumstances.

These arrangements only apply to high risk integration of Commonwealth data (as discussed in box B.5) — that is, when data from multiple data custodians are involved *and* the intended user is someone other than the data custodian(s) *and* there is a benefit from applying the Commonwealth data integration arrangements. That is, not all data integration projects involving Commonwealth data have to be done by accredited Commonwealth integrating authorities.

The four steps to become an accredited integrating authority are:

- Self-assessment: an agency applies for accreditation by completing a self-assessment against eight criteria (box B.6).
- An audit by an independent third party to substantiate the claims made against the eight accreditation criteria in the self-assessment are factual (largely through documentary evidence).
- A decision by the Oversight Board on whether to grant the agency accreditation to undertake high risk data integration projects, based on their self-assessment and the audit report.
- Inclusion on a published list of accredited Integrating Authorities, together with a summarised version of the integrating authority's application (with commercial in confidence information removed by the successful applicant) and a summary of the audit report.

Box B.4 High level principles for data integration

- *Strategic resource:* Responsible agencies should treat data as a strategic resource and design and manage administrative data to support their wider statistical and research use. This principle aims to maximise statistical and research use of existing and new Commonwealth datasets.
- *Custodian's accountability:* Agencies responsible for source data used in statistical data integration remain individually accountable for their security and confidentiality. This principle ensures that data custodians recognise their continued accountability for their data within integrated datasets and establish adequate controls over the use of personal or other sensitive data in data integration projects.
- *Integrator's accountability:* A responsible integrating authority will be nominated for each statistical data integration proposal. This principle sets out the responsibilities of integrating authorities to manage the data integration project from start to finish in line with the agreements made with data custodians and requirements as part of approval processes.
- *Public benefit:* Statistical integration should only occur where it provides significant overall benefit to the public. This principle ensure there is a demonstrated ability to produce significant outputs from the integrated dataset and an independent assessment is made that the public good outweighs the privacy imposition and risks to confidentiality.
- *Statistical and research purposes:* Statistical data integration must be used for statistical and research purposes only. This principle requires that where data integration is approved and implemented for statistical and research purposes, it is not then used for regulatory purposes, compliance monitoring, or service delivery. This helps to ensure that the risk of breaches of personal information and the potential impact of any advertent breach remain low.
  - There must be no feedback of information relating to individuals or individual businesses, from the statistical data integration project back to the originating administrative sources, unless that feedback was derived from a single source and is returning the same data to that source.
- *Preserving privacy and confidentiality:* Policies and procedures used in data integration must minimise any potential impact on privacy and confidentiality. This principle ensures that privacy and confidentiality are preserved to the maximum extent possible.
  - Operational, administrative and personal identifiers should be removed from datasets as soon as they are no longer required to meet the approved purposes of the statistical data integration. Access to potentially identifiable data for statistical and research purposes, outside secure and trusted institutional environments should only occur where: legislation allows; it is necessary to achieve the approved purposes; and meets agreements with source data agencies.
  - Once the approved purpose of the project is met, the related datasets should be destroyed, or if retained, the reasons for and necessity of retention documented, and a review process set up. If such retention was not part of the initial approval process, re-approval of the decision to retain is required. Archiving of statistically integrated datasets should be restricted to confidentialised datasets.
- *Transparency:* Statistical data integration will be conducted in an open and accountable way. This principle ensures the public is aware of how Commonwealth government data is being used for statistical and research purposes.

Source: NSS (2010).

Box B.5 Scope of the Commonwealth Government integration arrangements

A project is in scope if it meets **all** of the following criteria:

1) is statistical and research in nature;

That is, integration for non-statistical purposes (such as delivery of services to particular individuals, compliance monitoring, incident investigation or regulatory purposes) is out of scope as these activities have different processes and legislative requirements governing them.

AND

2) has cross portfolio status;

That is, involves two or more data custodians, where at least one is Commonwealth.

AND

3) involves users beyond the Commonwealth data custodian(s) participating in the project;

For example, the intended use by other Commonwealth agencies, state and territory governments, academic researchers or the public.

AND

4) derives a benefit from the application of the Commonwealth data integration arrangements.

That is, utilising a structured framework to maximise the use of public data assets, safeguard privacy and maintain trust in Government around managing data appropriately for statistical and research purposes.

Risk framework

Where a data integration project is assessed as high risk post mitigation, the integrating authority must be accredited.

The Risk Assessment Guidelines provide a platform to assess the risk of harm to a data provider and the risk of a reduction in public trust in the Australian Government and its institutions as a result of a breach. They are based on assessing the likelihood of a breach, and the severity of any potential consequences of a breach.

Data custodians can decide that the assessment guidelines on risk dimensions are not valid for their particular context. However, deviations from the assessment guidelines must be explained in the risk assessment.

Source: NSS (nd).

Box B.6 Criteria for accrediting an integrating authority

The eight criteria integrating authorities must meet to gain accreditation are:

1. ability to ensure secure data management
2. demonstrated ability to ensure that information that is likely to enable identification of individuals or organisations is not disclosed to external users
3. availability of appropriate skills
4. appropriate technical capability
5. lack of conflict of interest
6. culture and values that ensure protection of confidential information and support the use of data as a strategic resource
7. transparency of operation
8. appropriate governance and administrative framework.

Source: NSS (nd).

It is a requirement of the High Level Principles that a single integrating authority be identified for each statistical data integration project to manage the project from start to finish. Any consortium/partnership or outsourcing arrangements needs to be accredited by the Oversight Board (discussed below), subject to certain exceptions (NSS nd).

So far, three Commonwealth integrating authorities have been accredited — the Australian Bureau of Statistics, the Australian Institute for Health and Welfare, and the Australian Institute for Family Studies.

- The ABS is Australia's national statistical agency and has undertaken linkage projects using data on education enrolment, registers of births deaths and marriages and income tax, among others (ABS 2016; NSS 2016).
- The Australian Institute for Health and Welfare (AIHW) is a Commonwealth government agency responsible for releasing information on health and welfare. It has undertaken integration projects involving data on immunisation, mortality, pharmaceutical benefits and diabetes, among others (AIHW 2016; NSS 2016).
- The Australian Institute for Family Studies (AIFS) is the Australian Government's key research body in the area of family wellbeing. To date, the Australian Institute for Family Studies has no integration projects published by the National Statistical Service (AIFS 2016; NSS 2016).

Some participants to this inquiry have complained that the accreditation process for Commonwealth integrating authorities and the low number of bodies that have been accredited to date poses a 'bottleneck' for projects. This is further discussed in chapter 5.

The Commonwealth integrating authorities are overseen by the Cross Portfolio Data Integration Oversight Board. Among other things, this Board is responsible for providing

advice to help manage the risks of particular data integration projects. In practice, the Board:

- Has ten working days following registration of the project and receipt of the risk assessment to raise any concerns about the project with the data custodians or integrating authority. These concerns relate to the management of systemic risks of data integration
- Has no authority to approve or delay integration projects. Approval is given by data custodians
- Ensures that the risk mitigation strategies proposed when a project is registered are implemented. To ensure this, the Oversight Board may request a review or one or more of a data custodian's integration projects
- May work with data custodians and integrating authorities to improve their risk assessment processes
- Can delegate its review functions (National Statistical Service 2013).

Re-use of research data

The Australian Government is the dominant provider of funds for public research infrastructure and facilities, particularly in the national and landmark categories. Funding is provided through a range of programs administered by departments, the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC), and through direct budget allocations. Each of the programs has different characteristics depending on the scale, intended impact and allocation mechanism. For instance, NCRIS was established (in the 2004-05 Budget) to develop and fund national research infrastructure projects. It differed from previous research infrastructure funding initiatives in that it had: an emphasis on collaboration from the outset; the strategic identification of capabilities through a consultative road-mapping process; the strategic rather than competitive process for funding allocation; use of a facilitation process to develop capability plans; and the provision of funding for skilled staff and operating costs. State and territory governments also provide significant funds for research and research infrastructure, as do philanthropic organisations (National Research Infrastructure Council 2010).

Those who fund research have a role in driving the conditions under which it will be undertaken, including the use and re-use of research data. NCRIS in particular have played an important role, particularly in driving open access to research — for instance, through implementing the Strategic Framework for Research Infrastructure Investment Principles and the recommendations of the Australian Research Data Infrastructure Strategy on access to and use of research data. NCRIS are also currently running a consultation process to develop the 2016 National Research Infrastructure Roadmap. NCRIS is also responsible for funding the Australian National Data Service (ANDS), which is responsible for facilitating the re-use of research data (box B.7). Finally, the Australian Government

Department of Education and Training is convening the Open Access Working Group, which will run consultations on developing principles for open access to researchers' information at a later date (chapter 3).

Box B.7 **ANDS**

- The Australian National Data Service (ANDS) is a partnership led by Monash University, working in collaboration with the Australian National University (ANU) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO).
- It is funded by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).
- Since being formally established in 2008, ANDS has supported numerous research data projects around Australia. It has also been playing an important role in the international research data community.
- ANDS' flagship service is the Research Data Australia discovery portal where you can find, access and re-use data for research from Australian research organisations, government agencies and cultural institutions.
- A core purpose of ANDS is to make Australia's research data assets more valuable for researchers, research institutions and the nation through:
  - **Trusted partnerships:** working with partners and communities on research data projects and collaborations
  - **Reliable services:** delivering national services to support discovery, connection, publishing, sharing, use and re-use
  - **Enhanced capability:** building the data skills and capacity of Australia's research system.

Source: ANDS (nd).

B.3 Preserving privacy

Methods of access are designed to manage risk, particularly the risk of re-identification. There are a number of main methods currently being used to manage such risks.

Reducing the risk of re-identification before providing the data

CSIRO has identified four broad techniques to preserve privacy. All come with a trade-off in usability but the trade-off is different for each:

- *Restricting access*: Allow only authorised people to access data.
- *Restricting data*: Ensure the data is safe before allowing access and allow only select datasets to be analysed together.
- *Restricting queries*: Restrict the questions that can be asked of the data and do not reveal the data at all.

- *Restricting output*: Allow only safe results to be release from a system based on less restrictive queries (O'Keefe and Rubin 2015).

Traditionally, restricting data has been the main way that risk of re-identification is managed and privacy is preserved. In this regard, three broad techniques are generally used:

- *De-identification:* where the personal identifiers are removed from the data and typically replaced with a token. One example could be removing the name and address for personal data. A person can still be identified if the dataset is combined with other data where the person is identified (such as data from the Australian Tax Office (ATO)). Deidentified data is how the ABS stores the census internally.
- *Confidentialisation:* In addition to de-identification, the data may be changed so that an entity is no longer likely to be identifiable when combined with other data, but the data still adds up the right way. The ABS makes this data available via CD to authorised persons (CURFS).
- *Aggregated confidentialised data.* The data can be further aggregated (typically by spatial area) until at least a small number of people (the ABS uses three) are reported in aggregate in an area. This data is publically available from the ABS website.

Recent developments of note

There have been significant recent innovations in this area. For example, synthetic datasets make the data safe so that no private information is released. It does this using a technology known as Differential Privacy, developed at Microsoft (Dwork 2006). Differential privacy adds random noise to the data in a way that many analyses can achieve similar results as they would have achieved with the original data but individual records might not reflect the original data. When noise is added there will be some analyses where the results are unusable. The CSIRO (sub. 161) is working with the federal government on releasing a significant government dataset with this technique (Anonalytix nd). Another example is the ABS Table Builder which allows de-identified data to be accessed and the resultant tables are confidentialised on the fly and aggregated as needed. This is a form of restricting output to keep it safe while allowing arbitrary combinations of data.

There has also been a shift away from solely focusing on managing the risk of re-identification that has occurred with the greater adoption of the five safes risk assessment framework from the United Kingdom, particularly by the ABS. The framework focuses on five orthogonal risk axes:

- Safe people: Can the researchers be trusted?
- Safe projects: Is the purpose of use appropriate? What analysis is being done?
- Safe settings: Does the access environment prevent unauthorised use?
- Safe data: Can the data disclose identity?
- Safe outputs: Are the statistical results non-disclosive (Desai, Ritchie and Welpton 2016)?

More recent approaches using this framework have been based on the realisation that, if one or two of the axes introduce higher risk, the overall risk of disclosure may still be low, because there are multiple ways risk can be managed. In practice this has resulted in the ABS and others (such as the work being undertaken by the Department of Social Services, sub. 10) exploring virtual laboratories or trusted access models which provides more risky data in a safer environment. For example, a remote access environment is under development at the ABS. It shares many cores features with SURE: researchers login remotely to a virtual computer hosted by the ABS and data cannot be digitally exported, with any outputs requiring manual inspection for risk of statistical disclosure. The ABS has commenced trials of a virtual DataLab, including providing several agencies with access to data created by the Multi-Agency Data Integration Project (ABS, sub. 94).

Other privacy-preserving technologies, such as confidential computing techniques, are also being developed as technology advances (box B.8). CSIRO has developed a confidential computing platform that uses a partial homomorphic computation system to enable secure multi-party computation (CSIRO 2016b). It demonstrates that important knowledge can be extracted from data linkage without sharing raw data values between users.

Box B.8 Confidential computing techniques

In the 1990s the Canadian government collected data from across the country called the Longitudinal Labour Force File. Its goal was to improve government services. After a public outcry around privacy it was dismantled. If a virtual data collection can be established where only certain operations are allowed to operate over the collection then it is much for likely to be palatable to the public.

Confidential computing leaves data where it is securely governed and provides cryptographic means to ask specific analyses. It is a form of restricting queries and output while using unmodified data underneath. To enable this some cryptographic techniques are commonly used:

- Zero knowledge proof: Being able to prove another party knows something without knowing what they know. For example Alice knows the secret password to a door. Zero knowledge proof techniques can enable Bob to prove Alice knows the secret password without Alice revealing the password. However anyone else observing will not know if Alice truly knows the password or whether she is colluding with Bob. This could be usefully applied to test the honesty of the service without revealing its data.
- Oblivious computing: Being able to access encrypted data without revealing the pattern of access to the data.
- Homomorphic computation: Being able to calculate over encrypted content. For example addition and multiplication.
- Secure multi-party computation: Being able to derive a result from data across multiple secure parties while only revealing the answer and the validity of the answer to the parties.

Sources: CSIRO (sub. 161); Hack Canada (nd); Lindell and Pinkas (2008); Naone (2011); Simons Institute for the Theory of Computing (2013); Quisquater, Guillou and Berson (1990).

However, for this technique to be effective the data must be clean and comparable, placing responsibility with data custodians to effectively curate their holdings. Moreover, this technology is still in its early stages (full homomorphic computation was invented in 2011), and the scope of operations that it can perform are limited. However, as it is refined and optimised, the breadth of operation that can be securely performed should increase.

B.4 Data storage and security

Good metadata is essential to be able to find the data

Metadata allows data to be found, much like a card catalogue at a library. It typically contains a summary of the contents of the data holdings — 'data about data'. Ideally, it describes the content, format, quality, currency and availability of data in a consistent and meaningful way. While metadata is useful for cataloguing single documents, it is most important for managing a large body of data, particularly when it is generated by multiple authors.

Metadata can be critical to correct interpretation, and a good example is the spatial coordinate system. Many public sector agencies use Latitude and Longitude (lat/long) as part of their data collections, but these are not fixed concepts and this matters when something is surveyed again and the data compared. The Australian plate is moving at over 7 centimetres a year and therefore its recorded place in terms of latitude/longitude changes. In response to these changes, from 2017 Australia will use a new Geodetic reflecting its new location, but for those datasets that do not explicitly state the coordinate system they use, there will be confusion — the latitude/longitude created under the new Geodetic will be over a metre distant relative to the Australian coast to the same latitude/longtitude under the old measurement standard (GDA94).

Broadly speaking, there are two types of metadata managed by a registry.

- *structural metadata*: this details the set of terms used in a record, such as person, name, age, income and the values associated with the fields.
  - Structural metadata is critical to machine discovery and integration. It describes what the fields mean and what values can be expected. The relationships between fields can be hierarchical: for example, 'net income' being a more specific form of income. These properties can be independently related to objects or things. So income can be a property of a person or a business.
  - It is often expressed as a data dictionary, such as in ISO 11179, or an ontology. This describes the information model of the data and metadata, which is critical to maintaining a consistent dataset that can be made interoperable.
- *content metadata*: that describes the context of the data, who last created it, when, how, what from (provenance) and the environment in which the data was captured. Much of this metadata is domain specific. For example a scientific field instrument will have

different metadata to a statistical survey. However some aspects, particularly around authorship and provenance, have been standardised (Brodaric and Gahegan 2006).

Standardising the metadata — that is, establishing a common vocabulary — makes it possible for users to find and remix data in a clear and structured way. Ideally, there would be a single ontology, with all data custodians consistently expressing their metadata using it (box B.9).

Box B.9 Levels of interoperability

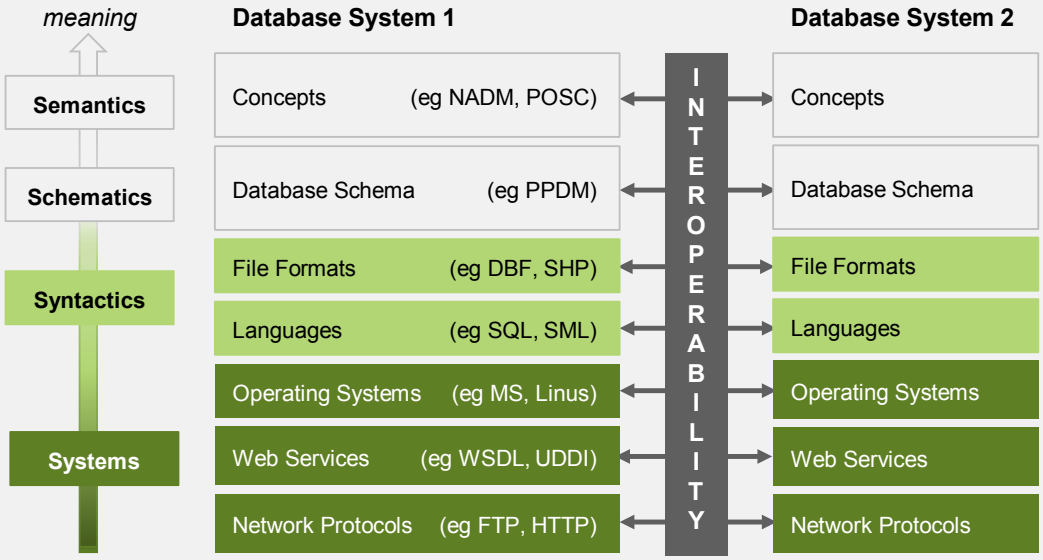
Interoperability allows resources to flow freely between entities in a way that allows it to be easily accessed, read and compared. Humans can be part of the direct functioning of an interoperability infrastructure. They allow an infrastructure to deal with complex representations that need interpretation but come at a significant time cost. For efficient operation humans are best as the final recipients and designers of interoperability infrastructure.

Interoperability is possible when each machine has the same understanding of the resource transferred on the four levels below. Some software systems are flexible. A single resource governed by a custodian could be expressed as linked data, a WFS service and a database endpoint. It could be shipped as XML, CSV and RDF and express the schematic and semantic content using slight variations based on the requirements of the representation. This will allow a resource to be more flexible in its interoperability requirements — but regardless two machines must find a match on the following four aspects.

For machine to machine interaction interoperability has four layers:

1. Systems: Mechanisms for interactions between machines
2. Syntactics: The shape of the container of the resource
3. Schematics: Terms used to describe the types of object or features and their properties.
4. Semantics: The values use for their properties

It is not sufficient to merely document all four levels to guarantee interoperability. A social agreement is necessary on what things actually mean.



Sources: Brodaric and Gahegan (2006); Box et al. (2015).

In practice this is not possible across different portals, platforms and jurisdictions, particularly given the philosophical disagreement that is often apparent as to the meaning and structure of terms (box B.10).

Box B.10 An example of inconsistency across registries

Three registries holding some of the same metadata records are:

- Research Data Australia: a registry of data in Australia.
- Data.gov.au: a registry of government data.
- Bioregional Assessment (BA) Programme: a registry of assessments of biological impact.

A random dataset was chosen to explore these sites: The Bioregional Assessment area v03 of the Galilee subregion as produced on the 11 July 2016 (Bioregional Assessments 2016)

From this register the Bioregional Assessment area v03 can be chosen. Here the bioregional assessment (BA) site links to the metadata record on data.gov.au. (Data.gov.au 2016).

The web site used to keep its own metadata data record and this can be found using google. Preserving URLs even when they are no longer required is good practice. Instead of showing their own record the BA site could have immediately redirected the page to data.gov.au. Instead the BA's own record is displayed.

The web page has: a link to the full metadata entry (but that requires a login); the data licence and the licences of the datasets used to create the data; and a mechanism to download the metadata and data (a zip file of shape files). It is machine readable but perhaps not state of the art. An OGC WFS service may have been a better choice for example; an API — but it is a little hard to find. There is also structural metadata both as human readable and a formal language (OWL). These are also hard to find.

When looking at the equivalent entry on data.gov.au much of the metadata that was hidden behind a login on the BA site is now visible. There are: links to the dataset history; a list of dataset ancestors; a citation of the dataset on the BA site through clicking on the link though clicking on the link redirects back to data.gov.au; the license history however the formatting and link to the BA description of the license is now poor. The bioregional assessment site defers to data.gov.au to host its metadata however Google can find the web sites own entry. There is a mechanism to download the metadata and the data. The zip file has the same content and the metadata is available in machine-readable forms.

The metadata download from data.gov.au shows that the list of dataset changes, citation and ancestor datasets are embedded as a wiki-style markup and not as metadata properties. The dataset ancestors do not appear at all. The structural metadata from the BA site shows that the ancestry and changes were rich structured data that are lost in the data.gov.au registry.

Going over to the ANDS site, the metadata is harvested from data.gov.au. All the content is there but ANDS does not understand the wiki like markup in the data.gov.au machine-readable metadata record and renders the content without formatting (Research Data Australia nd).

BA, data.gov.au and ANDS are lead government agencies in handling data and metadata. They are doing a good job on the whole. There is however greater need for coordination and harmonisation so the richness of the metadata can be made available. If leading agencies are struggling then others will be further behind.

Where such disagreement occurs, and is difficult to resolve, the best that can be hoped for is that the significant communities of interest are internally consistent, and it may then be possible to create connections or crosswalks to other communities (box B.11).

Varying degrees of access to important metadata are also observed. Some agencies (such as the NAA and the BoM) have provided access to their data via Linked Data. This provides a web orientated access to data and typically comes with explicit references to Structural Metadata through links to ontologies. The New South Wales Government has also established a metadata registry that is accessible by NSW public servants only (NSW Government nd).

Box B.11 How BoM resolved interoperability challenges

The Bureau of Meteorology faced the challenge of bringing together data from 28 different agencies on water storage level, volume and % full to deliver to the public water storage volume and % full.

The data was provided by agencies as:

- Hydstra exports (Hydstra is a leading water data management system)
- WDTF. An XML format developed for transferring water data.
- A variety of CSV and other formats.

Storage data was perceived as a relatively simple problem, but it was not.

There were inaccuracies in the time and spatial location of the data. Some of these were able to be screened with a rigorous format such as WDTF but not all agencies had the capacity and knowledge to implement a WDTF based solution and so they sent what they had. In some cases the data was incomplete and then agencies made their best estimate, resulting in one case of a water storage appearing 100km from where it actually was.

It was in the metadata though that things were really complex. WDTF supported metadata about the storage of deadStorageLevel (the level at which the water cannot be accessed) and full storage level (the level at which the storage is full) along with a datum to tie the levels to the Australian Height Datum (AHD). In addition the data provider would send separately a table relating storage height to volume. This would allow the Bureau to verify the data being sent. Once the table was sent it was realised that the Bureau view of a storage was simplistic at best. And extended consultation with agencies led the Bureau to develop an effective storage diagram. It then allowed the Bureau to work with agencies to ensure the correct provision of data. It is not sufficient to simply state what is required the data user and the data provider need to have the same interpretation of the content.

Source: Anderson et al. (2010).

The role of machine learning

Humans can sift through material written to other humans to determine the suitability of resources, clean them up and merge them together for specific purposes. However, when dealing at scale, humans can also introduce errors and the logistics of dealing with enough

people to handle the data is problematic. Also, when dealing with confidential or private data, trusted programs can remove the possibility of humans accidently discovering private information. Machines are very efficient at repetitive tasks where the context is well defined. Machine learning techniques are allowing machines to guess the schematic and semantic relationships between datasets where sufficient data and metadata exists (Data61 2016). This is a best guess which might then allow a human to make more positive connections. Unless there is a way to record these connections then this work cannot be re-used. Linked data provides a way of recording this.

Data provenance

Smart systems will also describe the chain of processing that resulted in in the data's metadata, and this traceability of resources is called provenance (W3C 2013). If provenance is captured and published at a fine enough scale then this information can be used to assist machines in knowing how to automatically combine resources for certain outcomes. It might also be used to provide guidance by resource providers on how their resources are best used. Another application of data provenance is when observations are made through samples (which is a common technique in statistics and other communities). Care must be taken when converting a set of samples to a dataset reflecting complete coverage of the domain.[56] as described above). Recording provenance metadata is critical to developing trust in derived products.

Agencies such as CSIRO and Geoscience Australia are currently employing provenance tools and storage in geophysical modelling and bioregional assessments (see, for example, CSIRO nd).

Data storage

Methods of storage

Data storage is a general term for archiving data in electromagnetic or other forms for use by a computer or device. Different types of data storage play different roles in a computing environment. In addition to forms of hard data storage, there are now new options for remote data storage, such as cloud computing, that can fundamentally change the way users access data.

Cloud computing involves computing over the internet. A public cloud is on the public internet and a private cloud is on a private network (or virtual private network). Cloud

[56] A coverage is defined as having a domain, defining its extent, and a series of ranges for properties which are of interest. Coverages are commonly used in satellite imagery where the domain is the spatial extent of the image and the ranges are the spectra captured by the satellite. Census data is also a coverage with the population being the diffuse entity domain can be spatial and the ranges being the responses to the survey questions.

computing has general benefits in flexible use of resources, economies of scale, duplication and being able to allocate costs to activity. It can also be beneficial in dealing with confidential data which would not normally be shared between parties. By choosing a neutral cloud resource the combined activity can be performed and then the cloud evaporated. Care must be taken to ensure the cloud host is indeed neutral.

Broadly speaking, there are four deployment models (Mell and Grance 2011):

- *Private cloud*. The cloud infrastructure is provisioned for exclusive use by a single organisation comprising multiple consumers (for example, business units). It may be owned, managed, and operated by the organisation, a third party, or some combination of them, and it may exist on or off premises.
- *Community cloud*. The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (for example, mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organisations in the community, a third party, or some combination of them, and it may exist on or off premises.
- *Public cloud*. The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.
- *Hybrid cloud*. The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (for example, cloud bursting for load balancing between clouds).

Capability provided to the consumer can also vary:

- Software as a Service (SaaS). The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (for example, web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.
- Platform as a Service (PaaS). The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

- Infrastructure as a Service (IaaS). The capability provided to the consumer is to provide processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (for example, host firewalls) (Mell and Grance 2011).

The Australian Government released its Information Security Guidelines including Cloud in 2014 (Department of Finance 2014). These endorse using cloud storage as the default option, but in most cases this will be limited to onshore clouds. According to the guidelines, use of an offshore cloud resource should be risk assessed on criteria including legal powers to restrict data access; complications from multiple simultaneous legal jurisdictions; inability to monitor operations; and differences in business and legal cultures.

Onshore clouds have lower risks but the cloud service may still be in the hands of a third party.

Data security

There are four main security measures, which are discussed in detail below. These are:

- Authentication: identifying who is accessing the resource.
- Authorisation: providing a defined set of roles or operations the entity can perform on the resource.
- Encryption: ensuring only those with the necessary keys can see what is happening.
- Duplication: ensuring identical copies of the resource are available elsewhere.

Authentication

Establishing identity is fundamental to questions of trust and access. People have identity in their interactions with Government and the private sector. Things, sensors and computers also have identity, both their own and those they carry out tasks for. Authenticating an entity's identity, or choosing to ignore it, is the first step in security. Thus it is not just people that need to be authenticated. Authentication is important as to whether to trust the data reported by the devices in the internet of things.

The sort of information provided can determine the level of trust in the entity's identity (the person is who they a say they are). Rather than just a black and white yes or no for identity (as is common with computer or phone logins), there is the possibility of transmitting the context of level of trust within the authentication, and then the entity can determine if that level of trust is sufficient for the tasks authorised.

One potential technique for encoding that context is Vectors of Trust (which is a proposal with the Internet Engineering Task Force (IETF)), which have the following components:

- Identify Proofing: Indicates how thorough a check of credentials was made. This describes the likelihood that the entity provided their own credentials.
- Credential Usage: The type of credentials used to provide proof on identity (more than one type can be provided).
- Credential Management: The credibility of the provider of the credentials (for example, the passport office). How well that provider managers the credentials.
- Assertion Presentation: How well the identity can be transmitted from the place of authentication to the place of use without being corrupted (Johansson 2016).

An efficient, effective identity verification and authentication system helps ensure that data is only accessible to the correct individual. Digital identity verification is already a common feature of many services, such as online banking or government services. As a result, Australians have multiple digital identities, using different login and password combinations (ACMA 2013).

The Australian Government has introduced a number of policies that seek to improve the security of digital identities and the efficiency of verification systems, mostly directed at compliance rather than consumer interest.

- The Document Verification System, managed by the Attorney-General's Department, allows public and private sector entities to verify the accuracy of government-issued identity documents (AGD nd).
- MyGov allows the creation of an identity, meant to represent the individual. This identity is then linked to the individual's relationship in government by first establishing an equivalent identity with each agency in government and linking them up.

The Australian Government's DTO is redeveloping MyGov and the way identity is managed in government, bringing in a Trusted Digital Identity Framework (TDIF). It will draw on lessons from the UK's Verify Service, which verifies service works by having authorised companies digitally check personal information held by the individual against data held by verification services. For example, if details from a passport are provided they are passed to a UK government passport checking service, and the service returns a 'yes' or 'no' as to whether they are consistent. The use of this sort of information is fairly common. For example a credit card company may ask for date of birth and the value of recent transactions to verify a person's identity. There is a legal requirement that verifying agencies do not retain the information used for verification (2016).

The Financial System Inquiry (Murray et al. 2014, p. 153) concluded however, that 'Australia's current approach to identity management results in significant process duplication, as individuals apply to, and government and businesses undertake to, verify and re-verify identities at multiple points'.

The Digital Transformation Office (DTO 2015) is working towards developing new identity verification frameworks, and several private companies have developed platforms to manage online identities (New Zealand Data Futures Forum 2014).

Authorisation

Once the entity's identity is known they can interact with the service in accordance with authorisation given. This is both the authorisation by the service, what the entity can do on the service, and the authorisation by the entity, what the service can glean from the entities' context, such as content on their mobile phone. The authentication of services — that the service used is the one expected — is typically done through some form of public key infrastructure such as that provided by the Department of Human Services for the health sector (DHS 2016).

Personal Identity Management Systems can be used to provide prescribed authorisation to personal content (either on devices, personal entry or the web). The Internet of Things will see an explosion in devices able to understand and communicate a person's context. A framework will be needed to ensure the devices are authorised to share the data with the right entities. A system where an individual authorises their Personal Identity Management Systems to interact with only trusted services (like only calling skype contacts) could reduce the incidence of phishing and waterhole attacks.[57]

Encryption

At the core of encryption technology are public and private keys. An entity's public key is designed to be shared broadly. Anyone can encrypt content with a public key, but only the person that holds the private key can decrypt it. This allows resources to be encrypted by an entity for the sole viewing of another party. Public key encryption is relatively slow so what typically happens in an exchange is a public key is used to encrypt a temporary encryption key. That key is sent to the other entity and used to encrypt the ongoing conversation.

Duplication

Duplication helps engender trust in the validity of content. By ensuring there are many copies of the content on the web corruption of a single item can be checked against the others.

Data security threats

Data security is important for storage and for securely transferring files.

[57] Phishing is defined as the attempt to obtain sensitive information such as usernames, passwords, and credit card details (and sometimes, indirectly, money), often for malicious reasons, by masquerading as a trustworthy entity in an electronic communication (Confluence nd). Waterhole attacks a computer attack strategy, in which the victim is a particular group (organization, industry, or region) (DDS IT Security nd). In this attack, the attacker guesses or observes which websites the group often uses and infects one or more of them with malware. Eventually, some member of the targeted group gets infected.

For access and storage, modern security systems rely heavily on public and private key encryption technologies. If an entity's private key is stolen then their identity is effectively stolen. Any content encrypted with their public or private key can be decrypted, and anything stored in immutable blockchains (box B.12) cannot be encrypted again with a new key. It will be immutably compromised. Unfortunately, systems such as bitcoin store private keys on home computers and phones, where the likelihood of someone having their private key stolen is high.

There are a multitude of suggestions being put forward for utilisation of blockchain technology. The technology shows promise for where proof of transactions is important. Private trials are underway in the banking and agriculture sectors. The Australian Government is also working with CSIRO to understand how blockchain might benefit Australia and undertake publicly funded trials (CSIRO 2016a).

The SANS institute promote a layered security model, where the defences need to be broken down one by one until the system is compromised. This can make it harder to compromise individuals (SANS Institute 2016).

The Australian Crime Online Reporting Network (ACORN) lists three main areas of cyber attacks:

- Unauthorised access or hacking: When an entity's computer or device is used without permission. Access may be from phishing or water-hole attacks where messages are sent to access illegitimate web sites often impersonating real ones. Some attacks are personalised, sometimes faking emails from known friends.
- Malware: Unauthorised access can lead to malicious software being installed or users may invite malware to opening malicious attachments or links. Malware can damage devices, steal private keys and recruit machines as bases for other attacks.
- Denial of service attacks. If enough compromised machines can be recruited then a denial of service attack can be launched to overload the beneficial service and bring it down.

ACORN have also raised identity theft as a significant issue. If enough information is known about someone then another person may attempt to steal their identity. That is they can provide credentials to verifying bodies that they are the person whose identity they wish to steal. If enough credentials are provided the verifying authority may accept the fake person as the real one (ACORN 2015).

The incidence and severity of data breaches means that ensuring effective data security arrangements are in place remains a real and pressing concern (chapter 6).

Different organisations have different arrangements in place for securely transferring files. For instance, the Australian Government has the physical infrastructure in place for secure government-to-government sharing via Fedlink, which was designed to allow government agencies to securely share information up to and including the non-national security classification of PROTECTED and the national security classification of RESTRICTED. It creates authenticated, encrypted links between participating agencies to create a secure Virtual Private Network (VPN) across the Internet. An accredited certificate authority

(CA) provides digital certificates and the public key infrastructure (PKI) necessary for a scalable and manageable VPN (Department of Finance 2016).

Box B.12 **Blockchain technology**

Blockchain is the technology that records transactions for the bitcoin cryptocurrency, and is thus the technology underpinning it. All participants in the blockchain have public and private cryptographic keys. A sender requests an amount of bitcoin to be sent to the owner of a public key, by using an address based on the public key. Parts of the transaction are encrypted with the public key so that only the custodian of the corresponding private key can utilise the funds in the transaction for further transactions.

In due course (it can take hours) the transaction is associated with a block which forms part of a block chain. A block contains a hash (an encoding) of the previous block so the blocks form an immutable chain. Blocks are created independently by different entities in a network. The entities synchronise the block chain which may result in transactions initially being associated with a block and then later removed to maintain consistency. Eventually a transaction becomes part of a block and becomes locked in the immutable chain.

This does not mean the transaction was correct, for example the wrong amount may have been paid. However the transaction is immutable and shared so that parties in the transaction have immutable evidence as to what happened. Immutability can be breached though if more than 50% of participants conspire to change it.

Bitcoin has one blockchain which was over 70GB and growing at 2.5GB/month. Bitcoin users on smart phones either rely on a third parties to participate in the blockchain on their behalf or keep only the parts of the blockchain that are relevant to them. The latter is more efficient but means the user does not participate in the immutability of the blockchain.

Contracts can be embedded into transactions to trigger the exchange of fund or cancellation of the transaction base on certain conditions. In trade for example it has opportunity to work with the Internet of Things (IoT). Where IoT sensors are attached to high value items the sensors can track the state and location of the item during its transit. This data could be transmitted to the IoTs and stored in a blockchain, making the data captured immutable. When the item reaches its destination the embedded contract could make the payment, possibly modified by the state of the item in transit. None of this guarantees the data captured is true but an immutable record of everything is kept so recourse can be later taken.

However, a report into blockchain by the Open Data Institute (ODI) highlights that blockchain has reached a reasonable level of maturity for bitcoin but is highly immature for other applications. In terms of the Gartner curve of technology development, ODI places blockchain ascending the hype curve. There is likely more hype to come, followed by some failed or less than successful trials until the best use of the technology is found.

It is unlikely all of these proposed uses will be satisfied by a single blockchain, such a blockchain would be unwieldy. The ODI suggests some blockchains will belong to private networks while others will be in separate public networks. Some of these networks will need to be linked. There are also potential issues with the level of privacy offered on transaction content. Some content must be visible to ensure the transaction is allowed. Privacy preserving computation may help this. Together the ODI shows there are significant technical hurdles still to be overcome before blockchain becomes embedded in the transaction landscape.

Sources: Bitcoin (2016); Eyers (2016); Hames Smith et al. (2016).

Another example is SUFEX (box B.13), which is used by the PHRN to securely transfer files.

Box B.13 **SUFEX**

SUFEX is a secure file transfer service used by the Population Health Research Network (PHRN) and its stakeholders. SUFEX provides users with a secure file exchange service and is not a file storage solution. The service is provided through the PHRN (established by the National Collaborative Research Infrastructure Strategy (NCRIS)) and has been designed, implemented and hosted by the Centre for Data Linkage (CDL), a national node of the network.

Security

SUFEX uses the Accellion Managed File Transfer application which provides various security features:

- Files are encrypted using AES 128-bit encryption
- Files are encrypted in transit and at rest
- Secure links generated by a double 128-bit MD5 token
- Links have a limited lifespan and access to the file is blocked after its expiration
- Users must identify themselves before they can download a file
- Recipients download files via an HTTPS/SSL connection
- Prevents forwarding links by verifying if a user is on the original recipient list of the email for downloading a file
- Logging of Recipient email address, time of access and IP address makes it easy to audit activity.

SUFEX features include:

- Access to the service through a simple and intuitive user interface
- Support for large file transfer
- Automatic file encryption/decryption
- Email notifications
- Download receipts
- Individual file reports

Secure file transfer is a simple four step process:

- Sender uploads files
- Email sent to recipient with link to files
- Recipient downloads files
- Sender receives notification of download

Cost

Under current funding arrangements, the service is offered free to PHRN partners and their stakeholders.

Source: PHRN (2011b).

C Australia's legislative and policy frameworks

This appendix gives an overview of the main legislation and policy frameworks governing data availability and use in Australia. It is not intended to be comprehensive but, rather, a general indication of the types of requirements that may apply to data sharing and release in the public, private and research sectors.

C.1 Data sharing and release instruments

Data sharing and release — general

Legislation providing a general authorisation

New South Wales is the only jurisdiction that has passed legislation to encourage proactive sharing between government agencies. The *Data Sharing (Government Sector) Act 2015* (NSW) (box C.1) provides that a government body is authorised to share government sector data with the Data Analytics Centre or another government agency to enable policy-related data analytics work to be carried out. However, this authorisation does not override existing privacy and other safeguards.

The South Australian Parliament has tabled legislation to enable the 'Sharing of Public Sector Data' including enabling de-identification, sharing of data at unit record level, an authorising environment that covers personally identifying information and the release of State Government data to non-government agencies (such as the university sector) (box C.2).

Box C.1 The Data Sharing (Government Sector) Act 2015

The *NSW Data Sharing Act* received assent in 2015. The Act established a Data Analytics Centre (DAC) and contains a number of key mechanisms through which the DAC and other government entities facilitate data sharing within the government sector:

- A government sector agency (other than the DAC) is, subject to certain conditions, authorised to share government sector data that it controls with the DAC or other agencies, to enable data analytics work to be carried out to identify issues and solutions regarding policy making, program management and service planning and delivery
- The Minister may direct an agency to provide specified government sector data that it controls to the DAC if the Premier advises the Minister that it is required for advancing Government policy
- The Minister may also direct a government sector agency to provide the DAC with information about the number and kinds of datasets that the agency controls and the kind of information collected in those datasets
- The DAC is authorised, subject to certain conditions, to share with agencies the results of the analytics work that it undertakes
- Subject to approval of the Premier and relevant Ministers, direction may also be given to State-owned corporations to provide data.

The Act also provides for a number of data-sharing safeguards, particularly regarding health and personal information, and confidential or commercially sensitive information.

Amendments to the legislation allow the Government to make data-sharing agreements with the Commonwealth, other states and territories, local government and non-government organisations (SA Premier 2016). These amendments are in line with data-sharing recommendations of the South Australia Child Protection Systems Royal Commission, led by former Supreme Court Judge Margaret Nyland. It is worth noting that, unlike most other Australian jurisdictions, South Australia does not have legally binding privacy principles — this is discussed later.

In terms of legislation covering proactive release of government information, New South Wales, Queensland and Tasmania have adopted freedom of information regimes that encourage proactive release of government information and aim to make bringing formal applications a last resort. These are discussed later.

Box C.2 **Public Sector (Data Sharing) Bill 2016 SA**

The *Public Sector (Data Sharing) Bill 2016 SA:*

- authorises agencies to provide public sector data, other than exempt public sector data, that they control to other public sector agencies for any of the following purposes:
  - to enable data analytics work to be carried out on the data to identify issues and solutions regarding government policy making, program management and service planning and delivery by the agencies
  - to enable public sector agencies to facilitate, develop, improve and undertake government policy making, program management and service planning and delivery by the agencies
  - as per Ministerial direction or such other purposes prescribed by regulation.
- specifies that the trusted access principles must be applied in respect of the sharing and use of public sector data to ensure the sharing and use of the data is appropriate in all the following circumstances:
  - *safe projects*: the purpose for which data is proposed to be shared and used must be assessed as appropriate having regard to the public interest in the proposed use and any risk of loss, harm or detriment to the community if the sharing does not occur
  - *safe people*: a proposed data recipient must be assessed as appropriate having regard to whether the proposed recipient: is in possession of the relevant skills and experience to effectively use the data; will restrict access to the data to specified persons with appropriate security clearance; whether the data provider will be able to engage with the data recipient to support the use of the data for the purpose; and whether other persons or bodies in addition to the data recipient are invested in the outputs of the project and the motivations of those persons or bodies to be so invested
  - *safe data*: data to be shared and used for a purpose should be assessed as appropriate for that purpose having regard to: whether the data is of the necessary quality for the proposed use (such as being accurate, relevant and timely); whether the data relates to people; if the data contains personal information, whether the personal information is necessary for the purpose for which the data is proposed to shared and used or whether the data should be de-identified (and if so, the risk of re-identification)
  - *safe settings*: the environments in which the data will be stored, accessed and used must be assessed as appropriate having regard to storage and access arrangements
  - *safe outputs*: the publication or other disclosure of the outputs must be assessed as appropriate having regard to the publication or disclosure's: nature, likely audience, likelihood and extent to which it may contribute to the identification of a person to whom the data relates; and whether the results will be audited and whether that process involves the data provider.

The Bill also proposes to establish an Office for Data Analytics to undertake data analytics work on public sector data received from across the whole of Government and to make the results of that work available to public sector agencies, to the private sector and to the general public as appropriate.

At the time of writing the Bill had passed the House of Assembly and had been read a first time by the Legislative Council.

Open data policies

The Australian Government, and all state and territory governments except for the Northern Territory, have open data policies.

The Prime Minister issued a Public Data Policy Statement (2015) (box C.3) and the Australian Government Department of Prime Minister & Cabinet issued *Guidance on Data Sharing for Australian Government Entities* (2016) which requires Australian Government entities to:

- when an entity requires arrangements to be formalised in writing, establish data-sharing arrangements through a letter of exchange between entities (rather than memorandums of understanding or deeds of arrangement)
- share data by default with other Australian Government entities, unless there are ongoing insurmountable legislative barriers or risks to privacy, security or confidentiality
- consult responsible expert groups and the Public Data Branch at the Department of the Prime Minister and Cabinet when determining the extent of legislative barriers and other risks
- foster a culture of trust and collaboration between entities
- where possible, provide data in a format that is machine-readable, high quality and complies with agreed open standards, with as few restrictions on use as possible.

Other relevant policies are as diverse as the Australian Government Smart Cities Plan (DPMC 2016), and the Digital Transformation Office's (2015) *Open Data* design guide.

All the state and territory governments, with the exception of the Northern Territory, have also announced open data policies:

- New South Wales (2016) released an updated Open Data Policy in 2016 as part of its Information Management Framework, replacing the 2013 NSW Government Open Data Policy (2016). The policy outlines NSW's vision to release better data in accessible, consumable formats with metadata and quality statements, release data faster using automated processed, standard data categories and trusted user models, and to release more data and make it discoverable through central portals.
  - Supporting this, NSW has established the Data Analytics Centre to facilitate data sharing between agencies and manage whole of government analytics projects (appendix B).

Box C.3 Australian Government public data policy statement

Australian Government entities will:

- make non-sensitive data **open by default** to contribute to greater innovation and productivity improvements across all sectors of the Australian economy
- where possible, make data available with free, easy to use, high quality and reliable Application Programming Interfaces (**APIs**)
- make **high-value** data available for use by the public, industry and academia, in a manner that is enduring and frequently updated using high quality standards
- where possible, ensure non-sensitive **publicly funded research** data is made open for use and reuse
- only charge for **specialised data services** and, where possible, publish the resulting data open by default
- **build partnerships** with the public, private and research sectors to build collective expertise and to find new ways to leverage public data for social and economic benefit
- **securely share data** between Australian Government entities to improve efficiencies, and inform policy development and decision-making
- **engage openly with the states and territories** to share and integrate data to inform matters of importance to each jurisdiction and at the national level
- **uphold the highest standards of security and privacy** for the individual, national security and commercial confidentiality
- ensure all **new systems** support discoverability, interoperability, data and information accessibility and cost-effective access to facilitate access to data.

At a minimum, Australian Government entities will publish appropriately anonymised government data by default:

- on or linked through data.gov.au for discoverability and availability
- in a machine-readable, spatially-enabled format
- with high quality, easy to use and freely available API access
- with descriptive metadata
- using agreed open standards
- kept up to date in an automated way
- under a Creative Commons By Attribution licence unless a clear case is made to the Department of the Prime Minister and Cabinet for another open licence.

Requests for access to public data can be made via data.gov.au or directly with the government entity that holds the data. If access to data is denied by an entity, users may appeal the decision using the public request functionality available through data.gov.au.

Source: Australian Government (2015).

- Victoria: As a result of a 2009 Parliamentary Inquiry into Improving Access to Victorian Public Sector Information and Data (EDIC (Vic) 2009), the Victorian Government released the DataVic Access Policy. This policy supports the sharing of Government data at no, or minimal, cost to users (2012), with guidelines released in August 2015 (2015). Victorian Government data is available at data.vic.gov.au in re-usable, accessible, understandable and shareable form. The website also enables users to suggest datasets that are not yet available at the website.
  - Supporting this, Victoria has announced (Andrews 2016) its intention to establish a data agency to facilitate information sharing between agencies — including to help address the information management process gaps identified in the Victorian Royal Commission into Family Violence. The Information Technology Strategy 2016–2020 aims to open up government information and data to businesses, universities and the community. Victoria's strategy may also address gaps identified in the Victorian Auditor-General's Office review of public sector information (2015).
- Queensland: In October 2012, the then Premier of Queensland (Bligh 2012) announced an 'open data revolution' for the Queensland Government with the aim of releasing as much government data as possible to encourage the private sector to develop innovative new services and solutions using the data. The strategy set out some broad principles for open data release and required statutory bodies to publish an Open Data Strategy including a roadmap to release datasets.
- South Australia (2013) released a Declaration of Open Data in September 2013, committing the government to proactively release data, and stating that all public sector agencies will be expected to develop open data strategies that include specific actions and report on their progress. The Office for Digital Government is leading the open data agenda and is working with agencies to proactively publish open and accessible datasets on the website Data.SA.
- Western Australia has an Open Data Policy (April 2015) (WA Land Information Authority 2015) and is developing a framework for accessing location information (nd) as part of the WA Location Information Strategy (nd). Landgate is the lead agency for implementation of these policies, responsible for both open and spatial data. Its leadership is intended to allow WA to implement it open data initiative by building on its progress in spatial data .
- Tasmania: The Tasmanian Government's Open Data Policy (2016) is designed to help facilitate the release of 'appropriate, high-value' datasets by Tasmanian Government agencies to the public — these will be prioritised to align with demand from the public and industry as determined by stakeholder consultation, followed by high-value datasets determined by agencies, to contribute to better service delivery in Tasmania and, finally, datasets identified as holding potential to address emerging opportunities and challenges. The Tasmanian Government Statistical Policy Committee is responsible for monitoring implementation of this policy. An open data website is yet to be developed (appendix B). This work is supported by the Stats Matter strategy (2015), which is a long-term strategy to build Tasmanian Government statistical assets and capability led by the Office of eGovernment.

- ACT: In June 2011, the Chief Minister made a Ministerial Statement on Open Government, which was intended to take a broad approach to enhance the openness of the way the ACT is governed. An open data portal commenced in 2012. Part of this Open Government initiative is the Proactive Release of Data (Open Data) Policy (2015), released in December 2015 by the Office of the Chief Digital Officer. It mandates provision of as much Prioritised Data to the community as is practicable (given resource limitations).

Specific sharing

Specific legislative sharing provisions are discussed under secrecy provisions, later in this appendix.

Memoranda of understanding

Memoranda of understanding (MOUs) also play a role in facilitating information sharing between agencies — but within existing legislative constraints, because MOUs have no legal force. MOUs vary in style and form, but some examples include:

- those documented by the Australian Law Reform Commission (ALRC) (2010a) that enable the sharing of domestic violence information
- NSW Government blank MOU template (currently available on the Department of Finance, Services and Innovation website)
- the MOU between the Australian Federal Police and the APRA
- the MOU between WA government agencies for sharing of domestic violence information
- the MOU between the ACT Environment and Sustainable Development Directorate and the Victorian Essential Services Commission MOU.

In March 2016, the Department of the Prime Minister and Cabinet released the *Guidance on Data Sharing for Australian Government Entities* (2016) which indicated a move away from MOUs to letters of exchange between entities. This was because MOUs were considered to be unnecessarily complicated and time consuming, take several years and multiple agreements to establish and, despite all that, not be legally binding — although it is not desirable that they should be.

C.2 Privacy and personal information handling principles

Privacy legislation

The Australian Constitution does not clearly state whether the regulation of personal information is the responsibility of the Australian Government or state and territory governments. This means that all jurisdictions are able to enact privacy laws. The Commonwealth Privacy Act (section 3) states that the Australian Parliament does not intend to 'cover the field' in relation to the protection of personal information.

The *Privacy Act 1988* (Cth) is the principal legislation in Australia governing privacy. The Privacy Act preamble states that the constitutional basis for the acct was the Australian Government's power to make laws in relation to 'external affairs' (under section 51(xxix) of the Australian Constitution).

There is no 'right' to privacy in Australia in the same way there is in some other countries — Australians cannot sue to remedy an invasion of privacy. Rather, Australia's privacy legislation sets out a number of fair information handling principles that must be followed when handling personal information.

Commonwealth Privacy Act

The Privacy Act applies to Australian Government agencies, and all businesses and not-for-profit organisations with an annual turnover of more than $3 million.[58] Entities covered by the Privacy Act are termed 'APP entities'. The Privacy Act generally does not cover state and territory government agencies (including state and territory public schools and public health care facilities), individuals, universities (other than private, incorporated universities and the Australian National University), small businesses, media organisations, and registered political parties.

Personal, sensitive and credit information

The Privacy Act applies to personal information (box C.4), defined as information or an opinion about an identified individual or an individual who is reasonably identifiable — for instance, someone's name, address, or medical records. The Privacy Act does not apply to de-identified information — information that is no longer about an identifiable individual or an individual who is reasonably identifiable (section 6).

[58] Subject to some exceptions, including private sector health providers (which includes private schools and childcare centres, because they are considered to be collecting health information), businesses that sell or purchase personal information, credit reporting bodies, contracted service providers for a Commonwealth contract, and employee associations.

Special protections apply to personal information that is:

- *Sensitive information* — includes information or an opinion about an individual's racial or ethnic origin, political opinions, membership of a political association, religious beliefs or affiliations, philosophical beliefs, membership of a professional association or trade union, health information, genetic or biometric information (section 6).
- *Credit information* — includes the consumer credit liability information about the individual, repayment history, credit applications, and default and insolvency information (covered by the Part IIIA Credit reporting and credit reporting code, appendix E).

Box C.4 Definitions of personal information

- Commonwealth — Information or an opinion about an identified individual, or an individual who is reasonably identifiable, whether the information or opinion is true or not, and whether the information or opinion is recorded in a material form or not.
- New South Wales — Information or an opinion (including information or an opinion forming part of a database and whether or not recorded in a material form) about an individual whose identity is apparent or can reasonably be ascertained from the information or opinion. It does not include information about a person who has been dead for more than 30 years.
- Victoria — Information or an opinion that is recorded in any form and whether true or not, about an individual whose identity is apparent, or can reasonably be ascertained.
- Queensland — Information or an opinion, whether true or not, and whether recorded in a material form or not, about an individual whose identity is apparent, or can reasonably be ascertained from the information or opinion.
- South Australia — Information or an opinion, whether true or not, relating to a natural person or the affairs of a natural person whose identity is apparent, or can reasonably be ascertained, from the information or opinion.
- Western Australia — Information or an opinion, whether true or not, and whether recorded in a material form or not, about an individual, whether living or dead, whose identity is apparent or can reasonably be ascertained from the information or opinion; or who can be identified by reference to an identification number or other identifying particular such as a fingerprint, retina print or body sample.
- Tasmania — Any information or an opinion in recorded format about an individual whose identity is apparent or reasonably ascertainable from the information or opinion and who is alive or has not been dead for more than 25 years.
- Northern Territory — Government information that discloses a person's identity or from which a person's identity is reasonably ascertainable. Personal information ceases to be covered five years after an individual's death.
- Australian Capital Territory — Information or an opinion about an identified individual, or an individual who is reasonably identifiable, whether the information or opinion is true or not, whether the information or opinion is recorded in a material form or not.

Disclosure

Personal information cannot be used or disclosed for a purpose other than that for which it was collected, unless the secondary purpose is related to the primary purpose (for an Australian Government agency directly related), and the individual would reasonably expect that disclosure (Australian Privacy Principle (APP) 6). (For sensitive information, the secondary purpose must be directly related to the primary purpose). The reasonable expectations test is quite broad, and permits, for example, web scraping or data broking business models, where collecting personal information is consistent with the purpose of their business (as long as the collection is done fairly and lawfully). The test is subject to a number of exceptions, such as where the information is required to be disclosed by law, where it is required for enforcement-related activities, where it is required to prevent a serious threat to health and safety, where it is required to provide a health service or for public health or public safety research purposes, and when an emergency declaration is in force and the information is necessary to provide a person with assistance.

Access and correction

If an Australian government agency or organisation holds personal information about an individual, the agency or organisation must, on request by the individual, give the individual access (APP 12) to that information. There are a number of exceptions to this requirement. For instance, organisations are not required to give individuals access to their personal information if giving access would be unlawful, have an unreasonable impact on the privacy of other individuals, relates to a commercially sensitive decision-making process or could reasonably pose a serious threat to the life, health or safety of any individual or to public health or public safety.

An agency or organisation must also take reasonable steps to correct (APP 13) personal information it holds, to ensure it is accurate, up-to-date, complete, relevant and not misleading, having regard to the purpose for which it is held. The entity must correct personal information it holds where it is satisfied, independently of any request, that the information is incorrect. An entity must also correct personal information on request from an individual. In some cases, it may not be lawful to correct the information, such as Commonwealth records over 15 years old (*Archives Act 1983* (Cth), section 26). In these cases, or other circumstances where the agency or organisation refuses to correct the information, the individual may request that the entity put a statement with the information to flag to users that it is incorrect.

An agency must respond to a request for access or correction within 30 days after the request is made, and an organisation must respond 'within a reasonable period' after the request is made. For access requests, both agencies and organisations should give access to the information in the manner requested by the individual if it is reasonable and practicable to do so. If this is refused, the entity should take reasonable steps to give access in a way that meets the needs of both the organisation and the individual.

Agencies are not permitted to charge for access. Organisations are allowed to charge for access request but the charge must not be excessive and must not apply to the making of the request. Neither agencies nor organisations are permitted to charge for correcting personal information or for associating the statement with the personal information.

Deletion

Under Australian law, individuals do not have the right to request deletion of their personal information. However, under APP 11, if an entity holds personal information about an individual and no longer needs it, it must take such steps as are reasonable in the circumstances to destroy the information or to ensure that the information is de-identified (box C.5).

Box C.5 **Australian Privacy Principles**

The Privacy Act governs the collection, use and disclosure of an individual's personal information. An individual has the right to have their personal information collected, stored, used and disclosed in a way that complies with the Australian Privacy Principles (table C.1), and may complain to the Office of the Australian Information Commissioner if that does not occur.

The Privacy Act sets out the Australian Privacy Principles, which that govern the collection, use and disclosure of personal information. Among other things, they stipulate:

- *Collection* — Personal information should be collected fairly and lawfully (APP 3, APP 4).
- *Notification* — An individual should be notified when personal information is collected about them (APP 5).
- *Disclosure* — Personal information cannot be used or disclosed for a purpose other than that for which it was collected, unless the secondary purpose is related to the primary purpose (for an Australian Government agency directly related), and the individual would reasonably expect that disclosure (APP 6). (For sensitive information, the secondary purpose but be directly related to the primary purpose).
- *Direct marketing* — Personal information may not be used for direct marketing unless the individual has consented, or it is impractical to obtain their consent (for personal information that is not sensitive). The *Spam Act 2003* (Cth) and the *Do Not Call Register Act 2006* (Cth) also regulate certain direct marketing communications — these are administered by the Australian Communications and Media Authority.
- Outsourcing — APP entities within Australia are responsible for maintaining privacy of data they transfer/outsource overseas, except where the overseas country has similar laws about privacy protection and enforcement/compliance.
- *Data quality and security* — an entity must take reasonable steps to ensure the personal information it holds is accurate, complete and up to date. It must also protect the information from misuse, interference, loss, unauthorised access, modification, or disclosure (APP 10).
- *Access and correction* — an entity must, on request by an individual, give access to information or correct it if inaccurate, out-of-date, incomplete, irrelevant or misleading.

Source: Australian Privacy Principle Guidelines (2015).

Enforcement

The Australian Information Commissioner (AIC) has a range of compliance and complaint handling powers including being able to make binding determinations, including a declaration that an individual is entitled to compensation (OAIC 2016a). The AIC can also accept enforceable undertakings from APP entities (set out in Parts IV, V, VI and VIB of the Privacy Act).

The AIC can bring court proceedings to enforce these determinations in the Federal Court or the Federal Circuit Court. The Office of the Australian Information Commissioner (OAIC) may also apply to the Federal Court or Federal Circuit Court for a civil penalty order. OAIC determinations and decisions not to investigate (or to further investigate) a privacy complaint may be appealed by regulated entities and individuals to the Federal Court or Federal Circuit Court on questions of law.

States and territories

Privacy legislation in New South Wales, Victoria, Queensland, Tasmania, the Northern Territory and the Australian Capital Territory regulates the collection, use and disclosure of personal information by state and territory public sector agencies (table C.1). South Australia does not have statutory privacy protection. Instead, the handling of personal information by public sector agencies is regulated by the *Information Privacy Principles Instruction 2013* (SA), contained in the South Australian Department of Premier and Cabinet circular no. 12 and the *Information Sharing Guidelines for Promoting Safety and Wellbeing*, also issued by the South Australian Government. The state public sector in Western Australia does not currently have a legislative privacy regime, although the *Freedom of Information Act 1992* (WA) provides for access to documents and the amendment of 'personal information' in a document held by an agency that is inaccurate, incomplete, out-of-date or misleading. The *State Records Act 2000* (WA) also affords some limited protection of privacy. For example, no access is permitted to medical information about a person unless the person consents, or the information is in a form that neither discloses nor would allow the identity of the person to be ascertained (section 49).

State and territory privacy legislation is also often expressed to apply to private sector organisations contracted to provide services to the state or territory — for instance, the *Information Privacy Act 2014* (ACT), section 9. This means that these organisations may be subject to both Commonwealth and state or territory privacy laws.

Privacy legislation in states and territories contains privacy principles that are similar but not identical. The Commonwealth Australian Privacy Principles have only been adopted in the ACT. This inconsistency between states and territories leads to different rules:

- *Definitions of personal information*: personal information is defined as being information or an opinion that could identify or reasonably identify an individual, but these vary between jurisdictions (box C.4).

- *Sensitive personal information:* The ACT and federal legislation contains more stringent rules relating to the collection of sensitive personal information.
- *Child protection*: All states and territories (except Western Australia) allow the disclosure of personal information without consent where it is necessary to lessen or prevent a serious threat to the life, health or safety of an individual. However, in New South Wales and Victoria, the disclosure must be necessary to prevent or lessen a serious *and* imminent threat to the life or health of an individual. Both the Wood (2008) Report and the ALRC (2010a) Family Violence report identified that this exception does not allow for circumstances in which there is progressive abuse and neglect of a child, or where the risk of harm is in the medium to long term, not imminent.
- *Cross-jurisdictional sharing:* New South Wales, Victoria, Tasmania and the Northern Territory impose restrictions on agencies transferring information to an individual or organisation outside the state or territory (New South Wales privacy legislation section 19(2), Northern Territory Information Privacy Principle 9, Tasmanian Personal Information Protection Principle 9, Victorian Information Privacy Principle 9.1).

There is separate health privacy legislation in New South Wales, Victoria and the ACT. The ALRC (2008) recommended that the federal Privacy Act be amended to apply to private sector organisations to the exclusion of state and territory health privacy legislation. The Australian Government (2009) accepted this recommendation in principle while undertaking to work with the states and territories to progress the matter.

New South Wales

The *Privacy and Personal Information Protection Act 1998* (NSW) (PPIPA) sets out Information Protection Principles (IPPs) that outline how NSW public sector bodies should manage personal information (table C.1). The PPIPA applies to NSW public sector agencies, statutory authorities, NSW universities, local councils, and other bodies whose accounts are subject to the Auditor-General. State-owned corporations such as RailCorp and Sydney Water are not included.

The NSW Privacy Commissioner has the power to investigate and mediate complaints made against agencies. The Commissioner also has responsibilities to promote the adoption of, and monitor compliance with, the Information Privacy Principles (IPPs), to prepare and publish guidelines, and to provide advice, conduct research, and educate the public on privacy related matters.

Health information is regulated separately under the *Health Records and Information Privacy Act 2002* (NSW).

Victoria

In Victoria, the *Charter of Human Rights and Responsibilities Act 2006* (Vic) sets out the basic rights, freedoms and responsibilities of all people in Victoria. Section 13 provides that a person has the right not to have their privacy, family, home or correspondence unlawfully or arbitrarily interfered with. Individuals are not able to directly enforce this right. However, all new legislation in Victoria must be accompanied by a statement of compatibility with this Act. Courts are required to, as far as possible, interpret laws in a way that is compatible with human rights. Public authorities must not act in a way that is incompatible with a human right or, in making a decision, fail to give proper consideration to a relevant human right.

The *Privacy and Data Protection Act 2014* (Vic) contains Information Privacy Principles (IPPs) (table C.1) which apply to all information held by the Victorian public sector (including the police and a contracted service provider). An agency may enter into an information usage arrangement with other federal or state agencies or the private sector that modify the application of an IPP to allow information to be used for a public purpose. These information usage agreements must be approved by the Privacy and Data Protection Commissioner and relevant Minister/s. These arrangements were introduced to allow information sharing to occur between agencies where it is in the public interest to do so, such as child protection programs. The Commissioner also has the power to specify that an act or practice of an agency is consistent with the privacy legislation, and an agency that acts in good faith in accordance with the certification will not convene the legislation.

An organisation may also depart from the IPPs where a determination has been made that there is a substantial public interest in doing so (Divisions 5 and 6 of the Privacy and Data Protection Act).

Health information is covered by the *Health Records Act 2001* (Vic), which sets out a number of Health Privacy Principles that are similar to the IPPs. The Act is administered by the Health Services Commissioner.

Queensland

The *Information Privacy Act 2009* (Qld) contains Information Privacy Principles (table C.1) that govern how Queensland Government agencies (other than health agencies) collect, store, use and disclose personal information. The Act applies to government departments, local governments, statutory authorities, government owned corporations and universities. The Act is supported by the *Information Privacy Regulation 2009* (Qld) and the *Right to Information Regulation 2009* (Qld).

The Office of the Information Commissioner handles privacy complaints.

Health information is governed by the National Privacy Principles (set out in schedule 4), that broadly correspond to the federal APPs in the Privacy Act.

Table C.1 **Comparing privacy principles across Australian jurisdictions**

	Cth	*NSW*	*Vic*	*Qld*	*SA*[a]	*WA*[b]	*NT*	*Tas*	*ACT*
Open and transparent management of personal information	✓	✓	✓	✓	×	×	✓	✓	✓
Sensitive information	✓	✓	✓	✓	✓	×	✓	✓	✓
Right to anonymity/pseudonymity	✓	×	✓	✓	✓	×	✓	✓	✓
Notification of collection	✓	✓	✓	✓	✓	×	✓	✓	✓
Purpose test for use / disclosure	✓	✓	✓	✓	✓	×	✓	✓	✓
Direct marketing restrictions	✓	×	×	×	×	×	×	×	✓
Cross border disclosure	✓	×	✓	×	×	×	✓	✓	✓
Government-related or unique identifiers	✓	×	✓	×	×	×	✓	✓	✓
Data quality	✓	✓	✓	✓	✓	×	✓	✓	✓
Data security	✓	✓	✓	✓	✓	×	✓	✓	✓
Access and correction	✓	✓	✓	✓	✓	✓	✓	✓	✓

a Circular only — not legislative. **b** WA does not have privacy legislation.

South Australia

South Australia does not have statutory privacy protection. The South Australian Department of Premier and Cabinet circular no. 12 (September 2013) contains the Information Privacy Principles Instruction (table C.1).

The Instruction applies to 'public sector agencies' as defined in section 3(1) of the *Public Sector Act 2009*, other than agencies specifically excluded by the circular (State records of South Australia nd).

The Privacy Committee of South Australia (Privacy Committee) handles privacy complaints and is responsible for overseeing the implementation of the Information Privacy Principles Instruction by South Australian public sector agencies. A copy of the Proclamation of the Privacy Committee can be found at the end of the Information Privacy Principles Instruction. Privacy complaints can also be dealt with by the Ombudsman SA or, in the case of a privacy complaint relating to police matters, the Office of the Police Ombudsman. Complaints that relate to health and community services can also be made to the Health and Community Services Complaints Commissioner.

Western Australia

The state public sector in Western Australia does not currently have a legislative privacy regime, although some privacy principles are provided for in the *Freedom of Information Act 1992* (WA). This Act provides for access to documents and the amendment of

'personal information' in a document held by an agency that is inaccurate, incomplete, out-of-date or misleading. (table C.1) The definition of 'personal information' is similar to the definition under the federal Privacy Act except that it also includes information about an individual who can be identified by reference to an identification number or other identifying particular such as a fingerprint, retina print or body sample. The WA Ombudsman (2013) has also issued Guidelines for the Management of Personal Information.

The *State Records Act 2000* (WA) affords some limited protection of privacy. For example, no access is permitted to medical information about a person unless the person consents, or the information is in a form that neither discloses nor would allow the identity of the person to be ascertained (section 49). Neither the *State Records Act* nor the *Freedom of Information Act 1992* (WA), however, deal comprehensively with privacy issues associated with the collection, storage and use of personal information by agencies (ALRC 2008). The Information Commissioner deals with complaints about decisions made by agencies in respect of applications for amendment of personal information.

Tasmania

The *Personal Information Protection Act 2004* (Tas) contains Personal Information Protection Principles (table C.1) which regulate personal information held by the Tasmanian public sector, including the University of Tasmania.

The Personal Information Protection Act is subordinate to other legislation where its provisions are inconsistent with those of another act.

The Tasmanian Ombudsman may receive and investigate complaints.

NT

The *Information Act* (NT) sets out Information Privacy Principles (table C.1) for managing personal information.

The NT Information Commissioner is responsible for regulating privacy compliance and handling complaints.

ACT

Section 12 of the *Human Rights Act 2004* (ACT) provides that all individuals have the right not to have unlawful or arbitrary interferences with their privacy, family, home or correspondence or have their reputation unlawfully attacked. The Act also imposes a duty of consistent interpretation in respect of other legislation. Under the Act, when a court is interpreting an ACT law it must adopt an interpretation 'consistent with human rights' as far as possible.

The *Information Privacy Act 2014* (ACT) includes Territory Privacy Principles which regulate how personal information is handled by ACT public sector agencies (table C.1).

The Office of the Australian Information Commissioner handles some of the functions of the ACT Information Commissioner under an arrangement between the ACT Government and the Australian Government. These functions include handling privacy complaints against, and receiving data breach notifications from, ACT public sector agencies, and conducting assessments of ACT public sector agencies' compliance with the Information Privacy Act.

Health information is regulated under the *Health Records (Privacy and Access) Act 1997* (ACT).

Other information handling principles and codes

Codes under the Privacy Act

Credit reporting code

The Privacy (Credit Reporting) Code 2014 (CR Code) is a mandatory code that binds credit providers and credit reporting bodies (box C.6). The CR Code supplements the provisions contained in Part IIIA of the Privacy Act and the Privacy Regulation 2013. Importantly, a breach of the CR Code is a breach of the Privacy Act.

Market research code

The Privacy (Market and Social Research) Code 2014 is a code developed by the Association of Market & Social Research Organisations (AMSRO). The code sets out how all AMSRO member organisations adhere to the Australian Privacy Principles in the Privacy Act. The code sets out how the APPs are to be applied and complied with by AMSRO members in relation to the collection, retention, use, disclosure and destruction of personal information in market and social research.

Box C.6 Organisations with consumer credit reporting obligations

Under the *Privacy Act 1988* (Cth), credit reporting bodies and credit providers have privacy obligations in relation to consumer credit reporting. Businesses that handle individuals' credit reports may also have obligations under the Act.

The laws that regulate the handling of personal information relevant to consumer credit reporting are contained in the Privacy Act (primarily part IIIA), the Privacy (Credit Reporting) Code and the Privacy Regulation 2013. Under the Privacy Act, credit providers include:

- banks
- organisations or small business operators if a substantial part of their business is the provision of credit (for example, building societies and credit unions)
- retailers that issue cards in connection with the sale of goods and services
- organisations or small business operators that supply goods and services where payment is deferred for at least seven days (for example, telecommunications companies and energy and water utilities)
- certain organisations and small business operators that provide credit in connection to the hiring, leasing, or renting of goods.

Real estate agents, general insurers, and employers are not regarded as credit providers.

Credit reporting bodies are organisations whose business involves handling personal information so that they may provide other entities with information regarding the credit worthiness of an individual. The three main credit reporting bureaus in Australia are Dun & Bradstreet, Experian and Veda.

Source: OAIC (2016b).

Codes approved by the ACCC

Credit reporting – Principles of Reciprocity and Data Exchange

The Principles of Reciprocity and Data Exchange (PRDE) are a set of industry-developed data exchange rules to support the move of Australia's credit reporting towards a comprehensive system (appendix E). The PRDE was developed through extensive consultation with members of the Australian Retail Credit Association and other key stakeholders.

The intention of the PRDE is to create a clear standard for the management, treatment and acceptance of credit related information amongst signatories. The PRDE only apply to consumer credit information and credit reporting information. The PRDE facilitates sharing of credit reporting information among signatories by setting up a reciprocal data exchange. It has also been developed to create an open and standardised system for the management, treatment and exchange of positive data. This is achieved through the Reciprocity, Consistency and Enforceability provisions.

The Australian Retail Credit Association formalised this arrangement through the Principles of Reciprocity and Data Exchange, which were approved by the Australian Competition and Consumer Commission in December 2015 (authorisation for five years).

Codes administered by the ACMA

The Australian Communications and Media Authority (ACMA) has developed codes of practice for radio (Commercial radio codes of practice and guidelines), television (Commercial television industry codes of practice), community broadcasting (Community broadcasting codes of practice) and online organisations (Online Codes). The ACMA has also developed privacy guidelines for broadcasters (Privacy Guidelines for Broadcasters 2011). More information can be found on the ACMA website.

Codes administered by the ASIC

The ePayments Code

Users of electronic payment facilities in Australia are protected by the ePayments Code. This code regulates consumer electronic payments, including ATM, EFTPOS and credit card transactions, online payments, internet and mobile banking, and BPAY.

The Australian Securities and Investments Commission is responsible for the administration of the ePayments Code. It monitors subscribers' compliance with the code and review the code regularly.

Almost all banks, credit unions and building societies in Australia are subscribers to the ePayments Code. Other providers of consumer electronic payment facilities such as PayPal have also subscribed to the code.

ASIC has power to approve financial services codes

ASIC has the power to approve codes in the financial services sector. The Regulatory Guide 183 Approval of financial services sector codes of conduct (RG 183) sets out that such codes need to meet a number of criteria including consultation with stakeholders, enforceability, compliance monitoring and dispute resolution. To date, no financial services sector codes of practice have been submitted to ASIC for approval.

C.3 Secrecy provisions

Common law secrecy provisions

Breach of confidence

The equitable action for breach of confidence may be used to restrict the disclosure of information in certain circumstances. Obligations of confidence recognised and protected by the common law include those that occur in doctor-patient and lawyer-client relationships. The principle is that the court will 'restrain the publication of confidential information improperly or surreptitiously obtained or of information imparted in confidence which ought not to be divulged' — *Commonwealth v Fairfax* (1980) 147 CLR 39, at 50, citing Swinfen Eady LJ in *Lord Ashburton v Pope* (1913) 2 Ch 469, 475.

An action for breach of confidence may be brought to restrain disclosure by a third party who has received confidential information. The information may have been communicated in breach of a duty of confidence, or may have come into the hands of the third party by human error. While legal actions for breach of confidence most commonly relate to commercial or technical information held by private individuals and companies, the principles of breach of confidence can be applied to protect government information in some circumstances. However, different principles apply to restraining the disclosure of government information (ALRC 2010b).

Duty of loyalty and fidelity

The common law imposes a duty of loyalty and fidelity upon all employees. This duty arises from the contract of employment, but may also arise from a fiduciary obligation where the employee is in a special position of trust and confidence. In the context of confidential information, the duty of fidelity requires that an employee must not use information obtained in the course of his or her employment to the detriment of the employer (ALRC 2010b).

Statutory confidentiality and secrecy provisions

Commonwealth

Secrecy provisions restrict the handling of information. A few apply across the whole of government — for instance, section 70 of the *Crimes Act 1914* (Cth) for Commonwealth officers. But most are contained in specific legislation, for instance, the *Taxation Administration Act 1953* (Cth).

The ALRC identified 506 secrecy provisions in 176 pieces of legislation, including 358 distinct criminal offences, in its report *Secrecy Laws and Open Government in Australia* (ALRC 2010b). The ALRC recommended that all Commonwealth secrecy offences should be reviewed to determine whether criminal sanctions are warranted for the

unauthorised disclosure of information (Rec 11-1). In the interests of consistency and simplification, the ALRC also recommended a set of principles to guide the creation of new offences and the review of the existing 'plethora' of secrecy provisions (2010b, pp. 22, 24).

Secrecy provisions can cover any information — confidential, personal, business (including but not limited to commercial in confidence) indigenous sacred, law enforcement, national security information, or sensitive information. Entities whose conduct is regulated can include Commonwealth employees, organisations or individuals providing services for or on behalf of the Commonwealth, Commonwealth agencies, other specific categories of persons and/or organisations.

Secrecy provisions can regulate a range of conduct including disclosure of Commonwealth information (such as tabling in parliament), making a record, using, soliciting or obtaining (mere receipt or without authorisation) information, and unauthorised handling. The ALRC (2010b) recommended that secrecy provisions should generally include an exception for disclosures in the course of an officer's functions or duties (Rec 10-2). Also, the *Criminal Code Act 1995* (Cth) provides a defence of 'lawful authority' so that 'a person is not criminally responsible for [a Commonwealth] offence if the conduct constituting the offence is justified or excused by or under a law'. The Australian Government has yet to respond to the ALRC's report.

A subset of existing Australian Government secrecy provisions is at table C.2 covering family assistance, census, immigration and border protection, social security, tax and telecommunications. More detail can be found in the ALRC (2010b) report on Commonwealth secrecy provisions.

What this myriad of secrecy provisions means is that the legislative framework applying to a particular initiative can be very complicated — for instance, the legislation governing MyHealth Record (box C.7).

Box C.7 Regulation of MyHealth Record

Health Records Act

The *Health Records Act 2012* (Cth) sets out a registration framework for individuals and healthcare providers and a privacy framework (aligned with the *Privacy Act 1988* (Cth)) specifying which entities can access and use information in the system, and penalties for improper use. The unauthorised collection, use or disclosure of information in the MyHealth Record system, of healthcare identifiers or of other information collected in relation to either the MyHealth Record system or Healthcare Identifiers Service is subject to civil and criminal penalties

Healthcare Identifiers Act 2010

A foundation of the MyHealth Record system is the Healthcare Identifiers Service, which is established under the *Healthcare Identifiers Act 2010* (Cth). The MyHealth Records Regulation 2012 specifies additional information as identifying information and privacy laws that continue to apply to the disclosure of sensitive information. The Healthcare Identifiers Regulations 2010 provide additional detail and requirements regarding the operation of the Healthcare Identifiers Service. The PCEHR (Information Commissioner Enforcement Powers) Guidelines 2013 set out the Information Commissioner's general approach to exercising its enforcement and investigative powers.

Opt out and other rules

The Commonwealth Minister for Health can make MyHealth Records Rules to support the operation of the MyHealth record system. The Rules currently in force are: — requirements for registered entities in the system; and MyHealth Records (Assisted Registration) Rule 2015, which specifies requirements for registered healthcare providers that assist individuals to register (through 'assisted registration').

With the agreement of Australian health ministers, the Minister has initiated trials through the MyHealth Records (Opt-out Trials) Rule 2016 which took effect on 9 February 2016. The Health Minister also has authority to implement opt-out nationally but only if the opt-out trials provide evidence demonstrating the value of an opt-out system and with the agreement of the Australian Government.

Health data breaches

The 2016 changes to the MyHealth Act also introduced new requirements to notify the System Operator (from the Australian Digital Health Agency) of potential and actual data breaches.

Third party information

The 2016 MyHealth reforms also expressly authorised healthcare provider organisations to upload information to a MyHealth Record if it includes relevant information about a third party.

Governance

The Independent Advisory Council and Jurisdictional Advisory Committee was abolished July 2016 and the Australian Digital Health Agency became the MyHealth Record System Operator from July 2016.

Source: ADHA (2016a).

Table C.2 **Selected Commonwealth confidentiality and secrecy provisions**

	A New Tax System (Family Assistance) (Administration) Act 1999	*Census and Statistics Act 1905 and Statistics Determination 1983*	*Social Security (Administration) Act 1999*	*Taxation Administration Act 1953*	*Telecommunications Act 1997 and Telecommunications (Interception and Access) Act 1979*	*Australian Border Force Act 2015*
Criminal offence or civil	Criminal (2 years)	Criminal (2 years or 120 penalty units)	Criminal (2 years)	Criminal (2 years)	Criminal (2 years)	Criminal (2 years)
Secrecy provision/s cover information:	A person (s3)	Any information (s19)	A person (Social Security Act 1991, s23)	An entity, can include personal information (s355-30)	Specified information, can include personal information	Any information
• about						
• conduct by	Any person	The Statistician or ABS officer past and present	Any person	Any person	Specified persons	An 'entrusted person',
Prohibits	✓ (s163)	×	✓ (s203)	×	×	×
• Unauthorised access to information						
• Making a record	✓ (s164)	×	✓ (s204)	✓ (s355-25)	×	✓ (s42)
• Using information (internal)	✓ (s164)	✓ (s19)	✓ (s204)		✓ (s276, 277, 278)	×
• Disclosure of information (external)	✓ (s164)	✓ (s19)	✓ (s204)	✓ (s355-25)	✓ (s276, 277, 278)	✓ (s42)
• Soliciting information	✓ (s165)	×	✓ (s205)	×	×	×
• Offering to supply	✓ (s166)	×	✓ (s206)	×	×	×

(Continued next page)

Table C.2 (continued)

	A New Tax System (Family Assistance) (Administration) Act 1999	*Census and Statistics Act 1905 and Statistics Determination 1983*	*Social Security (Administration) Act 1999*	*Taxation Administration Act 1953*	*Telecommunications Act 1997 and Telecommunications (Interception and Access) Act 1979*	*Australian Border Force Act 2015*
Exceptions						
• Consent (APP 6.1(a))	✓ (s162(2)(f))	✓ Determination, cl 5)[a]	✓ (s202(2B))	×	✓ (s289, 290)	✓ (requires disclosure in accordance with consent: s42(2)(a), 47)
• Disclosure in the public interest	×	×	×	×	×	×
• Census or statistics	✓ (Guidelines, s7, s14, 14AA[b]	✓ (19(2))	✓ (Guidelines, s7, s14, 14AA	✓ (s355-65, Table 7, item 1)	×	✓ (s42(2)(a) s44, 45, 46(e))
• Disclosure of de-identified information	×	✓ (Determination, cl 7) (requires undertaking by recipient) [c]	✓ (s202(2AA))	×	×	×
• Disclosure of publicly available information	×	✓ Determination, cl 3)	×	✓ (s355-45)	×	✓ (If lawfully made publicly available: s49)
• Disclosure to avert threats to life or death	✓ (Guidelines, s8)	✓ (Determination, cl 5) (not personal or domestic nature)	✓ (Guidelines, s8)	✓ (s355-65, Table 1, item 9)[d]	✓ (s287, 300)	✓ (s42(2)(a), s44, 45, 46(d), 48)[e]
• In the course of functions and duties	×	×	×	✓ (s355-50)	✓ (s279, 296) [f]	✓ (s42(2)(b))

(Continued next page)

Table C.2 (continued)

	A New Tax System (Family Assistance) (Administration) Act 1999	*Census and Statistics Act 1905 and Statistics Determination 1983*	*Social Security (Administration) Act 1999*	*Taxation Administration Act 1953*	*Telecommunications Act 1997 and Telecommunications (Interception and Access) Act 1979*	*Australian Border Force Act 2015*
• Coronial inquiries, investigations, inquests	×	× (Only with consent or if de-identified: Determination, cl 5, 7)	×	×	✓ (s295Y)[g]	✓ (s42(2)(a), 44, 45, 46(c))
• Law enforcement	✓ (s162(2)(e), s167, Guidelines, s9)	×	✓ (Guidelines, s9)	✓ (s355-70, Table items 1, 3, 4, 6)[h]	✓ (TIA Act, s178, 179, 180D)	✓ (s42(2)(a), s44(1), (2))[b]

[a] Except if disclosure would be likely to enable identification of a person or organisation that has not consented to its disclosure. [b] Requires Secretary to be satisfied that the information cannot reasonably be obtained from another source and the recipient has sufficient interest in the information. [c] Recipients must give the Statistician an undertaking agreeing to requirements relating to use, disclosure and security of the information. [d] Limits disclosure to an Australian government agency. [e] Limits disclosure to federal, state and territory government agencies or authorities, police, coroner and office holders, or other prescribed bodies or persons, and for the classes of information prescribed in Schedule 1, Parts 1-9 to the *Australian Border Force (Secrecy and Disclosure) Rule 2015*. [f] Note these provisions include both permitted primary and secondary disclosures. [g] Must relate to an emergency or likely emergency. [h] All disclosures for law enforcement must be authorised by an SES officer, who is not a direct supervisor of the taxation officer.

States and territories

A range of legislation at the state or territory level contains confidentiality or secrecy provisions restricting sharing of information. As with Commonwealth legislation, these secrecy provisions can be found across multiple sectors including: child protection, criminal law (official secrets), education, health, juvenile justice, and public service legislation. Similarly, the provisions cover a range of information including personal information, business information, government information and official secrets. State and Territory secrecy provisions also regulate a range of conduct including access, recording, disclosure or communication of information. For instance, in Victoria, the *Crimes Act 1958*, the *Corrections Act 1986*, and the *Surveillance Devices Act 1999* all contain confidentiality provisions, and section 194 of the *Independent Broad-based Anti-Corruption Commission Act 2011* exempts information dealt with by the Commission from the Freedom of Information Act.

Breaches of many of the state and territory secrecy provisions are, under the law, criminal offences. For example, a 2016 study for the Royal Commission into Institutional Responses to Child Sexual Abuse considered that the ALRC Secrecy report was relevant at the state and territory level. For instance, the study indicated it would be worth considering whether such provisions need to be criminal offences. The provisions are generally intended to ensure that information obtained in the course of employment and other activities regulated by the legislation is only used and disclosed in appropriate and lawful circumstances. However, the study commented that imposing criminal sanctions on those working in the child protection and welfare sectors for breach of secrecy or confidentiality provisions is likely to encourage a risk-averse organisational culture when it comes to sharing information (Adams and Lee-Jones 2016). The Royal Commission study also echoed the ALRC (2010b) secrecy report recommendation to ensure that state and territory secrecy provisions include exceptions for the sharing of information in the course of an officer's functions or duties, or where authorised or required by law.

A subset of existing state and territory secrecy provisions is at table C.3, covering information related to health, education and official secrets.

Table C.3 **Selected state and territory secrecy provisions**

State	*General government secrecy*	*Education*	*Health*
New South Wales	*Government Information (Public Access) Act 2009* s 11: Schedule 1 lists overriding secrecy provisions in 27 different Acts covering a range of sectors.	*Education Act 1990* s 18A (4): publishing a ranking of particular schools according to results, or identifying a school as being in a percentile of less than 90 per cent in relation to results (except with the permission of the principal of the schools concerned) is an offence.	*Health Administration Act 1982* s 22: Disclosure of information obtained in connection with the administration or execution of the Act, except in prescribed circumstances, is a summary offence against the Act.
Victoria	*Crimes Act 1958* s 464ZGK: A person who has access to Victorian information; and (intentionally or recklessly causes the disclosure of the Victorian information other than as provided by this section — is guilty of a summary offence and liable to level 8 imprisonment (1 year maximum) or a level 8 fine (120 penalty units maximum). Does not apply to information that cannot be used to discover the identity of any person.	*Education and Training Reform Act 2006* S 2.5.52: It is an offence to publish or broadcast information identifying a complainant or contravening a determination (of a medical, informal or formal hearing panel for complaints).	*Health Records Act 2001* s 90: Current or former employees or delegates of the Office of the Health Services Commissioner commit an offence if they record, disclose or communicate any information acquired in the exercise of powers under the Act (except in prescribed circumstances).
Queensland	*Criminal Code 1899* s 85: Disclosure of official secrets. A person who is or has been employed as a public officer who unlawfully publishes or communicates any information that comes or came to his or her knowledge, or any document that comes or came into his or her possession, by virtue of the person's office, and that it is or was his or her duty to keep secret, commits a misdemeanour. Maximum penalty — 2 years imprisonment.	*Education (General Provisions) Act 2006* S 373: It is an offence for anyone involved in the administration of Chapter 13, Part 3 (Financial Data) to disclose protected information (financial statements of private schools) to any person, except in prescribed circumstances.	*Public Health Act 2005* Many similar provisions, following this structure. s 219: confidential information means information that has become known to a relevant person (a person who is or was a health service or public service employee) in the course of performing the relevant person's functions under this part (perinatal statistics). s 220: A relevant person commits an offence if they disclose confidential information, except in prescribed circumstances.

(Continued next page)

Table C.3 (continued)

State	General government secrecy	Education	Health
South Australia	*Criminal Law Consolidation Act 1935* s 238 improper conduct by public officials: Public Sector Act 2009 states that 'public sector employees must observe the public sector code of conduct' (s 6), including requirements not to access (or attempt to access) or disclose official information other than is required by law or where appropriately authorised in the agency concerned.	*Education and Early Childhood Services (Registration and Standards Act) 2011* — Education and Care Services National Law An individual who is, or was, a person exercising functions under this Law commits an offence if they disclose protected information (which could lead to the identification of a person, and which came to the individual's knowledge in the course of exercising functions under this law), except in prescribed circumstances.	*Health and Community Services Complaints Act 2004* s 75: A person commits an offence if they record, disclose or use confidential information (which could lead to the identification of a person) gained by the person through involvement in the administration of this Act (includes Commission staff, conciliators, mentors) except in prescribed circumstances.
Western Australia	*Criminal Code Compilation Act 1913* s 81: Disclosing official secrets: (1) Unauthorised disclosure means — (a) the disclosure by a person who is a public servant or government contractor of official information in circumstances where the person is under a duty not to make the disclosure; or (b) the disclosure by a person who has been a public servant or government contractor of official information in circumstances where, were the person still a public servant or government contractor, the person would be under a duty not to make the disclosure. (2) A person who, without lawful authority, makes an unauthorised disclosure is guilty of a crime and is liable to imprisonment for 3 years. Summary conviction penalty: imprisonment for 12 months and a fine of $12,000.	*School Education Act 1999* s 242: A person commits an offence if they disclose or make use of official information, except in prescribed circumstances. Official information includes all registers and documents of, in the possession of, or under the control of: the Minister; the department; the department CEO; the principal of a State school or a panel appointed under the Act.	*Health Act 1911* s 314: every person employed in the administration of this Part who does not preserve secrecy with regard to all matters that may come to his knowledge in the course of such employment, and communicates any such matter to any other person except in the performance of his duties under this Act, commits an offence (except in prescribed circumstances). *Health Services Act 2016* s 219: A person commits an offence if they collect, use or disclose any information obtained because of the person's employment under or for the purposes of this Act or any disclosure made to the person under this Act, except in prescribed circumstances.

(Continued next page)

Table C.3 (continued)

State	*General government secrecy*	*Education*	*Health*
Tasmania	*Criminal Code Act 1924* s 110: Disclosure of official secrets: Any public officer who discloses (except to some person to whom he is authorized to publish or communicate the same) any fact which comes to his knowledge, or the contents of any document which comes to his possession, by virtue of his office and which it is his duty to keep secret, is guilty of a crime.	*Youth Participation in Education and Training (Guaranteeing Futures) Act 2005* s 45: A person who is or was employed by the Department or by the Office of Tasmanian Assessment, Standards and Certification commits an offence if they record, disclose or allow access to prescribed information (personal information obtained in the course of administering the Act), except in prescribed circumstances.	*Health Service Establishments Act 2006* s 55: A person who is or was employed in performing functions related to the administration of this Act commits an offence if they disclose confidential information acquired in the course of performing those functions, except in prescribed circumstances.
Northern Territory	*Criminal Code Act 1983* s 76 Disclosure of official secrets: (1) Any person who, being employed in the public service or engaged to do any work for or render any service to the government of the Territory or any department or statutory body thereof, unlawfully communicates confidential information coming to his knowledge because of such position is guilty of an offence and is liable to imprisonment for 3 years. (2) If he does so for purposes of gain he is liable to imprisonment for 5 years.	*Education Act 2015* s 158: A person commits an offence if they obtain information in the course of performing functions related to the administration of private schools and intentionally engage in conduct that results in the information's disclosure, except in prescribed circumstances.	*Health and Community Services Complaints Act 1998* s 97: A person commits an offence if they record, disclose or use confidential information (disclosed in complaints; of personal concern to an individual; or about a complainant, user, provider or investigator of a complaint) obtained through their involvement in the administration of this Act, except in prescribed circumstances.

(Continued next page)

Table C.3 (continued)

State	*General government secrecy*	*Education*	*Health*
Australian Capital Territory	Crimes Act 1900 — s 153: Disclosure of information by Territory officer (1) A person who, being an officer of the Territory, publishes or communicates, except to some person to whom he or she is authorised to publish or communicate it, any fact or document which comes to his or her knowledge, or into his or her possession, by virtue of him or her being an officer of the Territory and which it is his or her duty not to disclose, commits an offence. Maximum penalty: 50 penalty units, imprisonment for 2 years or both. (2) A person who, having been an officer of the Territory, publishes or communicates, without lawful authority, any fact or document which came to his or her knowledge, or into his or her possession, by virtue of the person having been an officer of the Territory and which, at the time when he or she ceased to be an officer of the Territory, it was his or her duty not to disclose, commits an offence. Maximum penalty: 50 penalty units, imprisonment for 2 years or both. (3) In this section: 'officer of the Territory' means — (a) a public employee; or (b) a person who performs services for the Territory or a territory authority.	Children and Young People Act 2008 S 846: An information holder (a person who is or was exercising a function under, or engaged in the administration of, this Act) commits an offence if they recklessly record or divulge (directly or indirectly) protected information (obtained because of the person's role as an information holder) about someone else, except in prescribed circumstances.	Health Act 1993 S 125: An information holder (a person who is or was exercising a function under, or engaged in the administration of, parts 4 or 5 of this Act) commits an offence if they recklessly record or divulge (directly or indirectly) protected information (obtained because of the person's role as an information holder) about someone else, except in prescribed circumstances.

Interactions with public interest information sharing provisions

There is also a vast swathe of legislative provisions requiring or permitting information to be shared in specific circumstances, usually where there is considered to be a public interest in that disclosure. For instance, Australia's privacy legislation allows identifying information to be shared without consent for a number of reasons, including: where it is required by law, where it is necessary for an enforcement activity, or for lessening or preventing a serious threat to the life, health or safety of any individual or to public health and safety (where it is unreasonable or impractical to obtain consent).

The government can also amend legislation to allow non-compliance with the privacy legislation for a particular purpose — for instance, in the child health and welfare context, chapter 16A of the *Children and Young Persons (Care and Protection) Act 1998* (NSW) expressly notes that the safety, welfare, and wellbeing of children and young persons takes precedence over the protection of confidentiality or privacy of the individual. In that special circumstance, information may be provided to or requested by prescribed bodies for investigative or service provision purposes relating to the safety and wellbeing of a child or young person.

Other, narrower, examples of disclosure include mandatory reporting of child abuse and neglect in all jurisdictions — see, for example, section 356 of the *Children and Young People Act 2008* (ACT), section 23 and 27 of the *Children and Young Persons (Care and Protection) Act 1998* (NSW), sections 15, 16 and 26 of the *Care and Protection of Children Act 2007* (NT), part 1AA of the *Child Protection Act 1999* (Qld), sections 364, 365, 366, 366A of the *Education (General Provisions) Act 2006* (Qld), sections 6, 10 and 11 of the *Children's Protection Act 1993* (SA), sections 3, 4, and 14 of the *Children, Young Persons and their Families Act 1997* (Tas), sections 182, 184 and 162 of the *Children, Youth and Families Act 2005* (Vic), section 327 of the Crimes Act 1958 (Vic), sections 124A and 124B of the *Children and Community Services Act 2004* (WA), sections 5 and 160 of the *Family Court Act 1997* (WA), and section 67ZA of the *Family Law Act 1975* (Cth). Additionally, there is a range of provisions either mandating or permitting child protection information to be shared with relevant bodies — including those within that jurisdiction, bodies that are interstate and with New Zealand. However, these vary widely between jurisdictions — see Adams (2016) for more detail on the provisions, and chapter 3 for a discussion on this point.

Reporting obligations also exist with respect to domestic and family violence. For instance, the *Family Law Rules 2004* (Cth) specify that a party must file a copy of any family violence protection order when a case starts or as soon as practicable after the order is made. However, this obligation to disclose acts as an exception to the general principles contained in family violence legislation in all the states and territories that prohibit publication of certain information involved in or associated with protection order

proceedings. The ALRC (2010a) report into *Family Violence: a National Legal Response* contains more details on this tension between confidentiality and information sharing.

These disclosure provisions are not standard across jurisdictions or within sectors, and significant variations are evident — see Australian Institute of Family Studies (2016) for one example in the child abuse information sharing space. One example of the variation that can occur is that some of these information sharing provisions mandate disclosure while others merely allow it to occur if the individual chooses. Thus, navigating these information sharing obligations can be complicated, particularly when countervailing obligations of secrecy or confidentiality apply.

Complex ethical issues often surround these disclosure provisions. For instance, there is a duty of confidentiality in the provision of healthcare that operates in addition to the general privacy legislation, but medical practitioners are *allowed* to disclose patient's genetic information, whether or not the patient gives consent, in circumstances where there is a reasonable belief that doing so is necessary to lessen or prevent a serious threat to the life, health or safety of their genetic relatives. While health practitioners have an ethical obligation to advise the patient to inform relatives of the diagnoses, they are under no legal obligation to disclose the information to genetic relatives themselves, whether consent is given or not — thus a medical practitioner will not be liable for non-disclosure.

However, where a medical practitioner is considering disclosure, the guidelines issued under section 95AA of the *Privacy Act 1988* (Cth) set out the legal requirements and ethical considerations that govern this decision. For instance:

- in determining whether there is a serious threat, the medical practitioner should take into account the nature of the condition, its associated risks and treatment or care options and the probability that a genetic relative may also have the condition or be a carrier of the relevant mutation
- in determining whether the threat can be lessened or prevented, the medical practitioner should take into account whether the condition is preventable or manifestations treatable (that is, whether the relatives can benefit from the information) and, if the disease is incurable, whether knowledge of the condition would allow optimal management.

Key points for good ethical practice suggested by the guidelines include holding discussions with the patient and trying to get consent where possible, considering arranging genetic counselling for the patient and/or their relatives, and considering ways of making the disclosure in the way that is least likely to identify the patient.

Another example of guidance includes *Providing support to vulnerable children and families* (2007) which is an information sharing guide for registered medical practitioners, nurses and people in charge of relevant health services in Victoria which outlines the reporting obligations that apply (box C.8).

Box C.8 Information sharing guide for health professionals: Victoria

As a registered medical practitioner or nurse you must:

- make a report to Child Protection if you form a reasonable belief that a child is in need of protection from physical injury or sexual abuse (a mandatory report).

As a registered medical practitioner or nurse, or as the person in charge of a relevant health service you must:

- provide information relevant to the protection or development of a child who is subject to a Children's Court protection order where properly directed to do so
- only share information as authorised by privacy legislation (such as the *Health Records Act 2001* and *Information Privacy Act 2000*) where you are not specifically authorised by the *Children, Youth and Families Act 2005* as described in this guide.

As a registered medical practitioner or nurse, or as the person in charge of a relevant health service you should:

- give priority to a child's best interests, including consideration of the need to protect a child from harm, protect their rights and promote their development
- seek consent, where this is possible, before sharing information and where this does not place the child or another person at risk
- exercise professional judgment — using your professional skills, knowledge and experience — when deciding what action to take in regard to a vulnerable child
- consult with your manager where you are unsure what to do and, if necessary, seek the advice of your professional association or union, medical defence insurer or legal counsel
- make a referral to a Child FIRST team where you have a significant concern for a child's wellbeing
- make a report to Child Protection where you form a reasonable belief that a child is in need of protection (registered medical practitioners and nurses *must* make a report to Child Protection where this involves physical injury or sexual abuse)
- share relevant information with Child FIRST or Child Protection workers to help them complete the assessment of a referral or report they have received
- share relevant information with Child Protection where a child is subject to Child Protection investigation, further Child Protection intervention or a Children's Court Protection Order.

As a registered medical practitioner or nurse, or the person in charge of a health service:

- you are protected when you share information in good faith with Child FIRST or Child Protection as authorised — you cannot be successfully sued or suffer formal adverse consequences in your work
- your identity will be protected, unless you consent to its disclosure or if disclosure is required by law.

If you work for, or at, an organisation you should generally consult with your manager before disclosing information about a child or their family without their consent — except in very urgent situations. Organisations have a legal responsibility to protect patient information from inappropriate disclosure. An organisation's policies on information sharing should therefore also be consulted.

Source: Department of Human Services (Vic) (2007).

C.4 Freedom of information

Commonwealth

Under the Freedom of Information Act *1982* (Cth) (FOI Act), individuals have the right to request:

- access to documents held by Australian Government ministers and most agencies, and
- that ministers or agencies amend or annotate any information held about the requesting individual.

The FOI Act also establishes an information publication scheme that requires agencies to publish details about their functions and structure online, and allows agencies and ministers to release documents that would be exempt under the FOI Act unless prevented by a secrecy requirement in another law.

The FOI Act gives the Australian community access to information held by government by requiring agencies to publish that information and by providing for a right of access to documents (table C.4). This presumption of openness and maximum disclosure has applied since the 2010 reforms to the FOI Act. It applies to most Australian Government agencies, and to some documents created or held by a contractor or subcontractor relating to the provision of services to the public or third parties on an agency's behalf.

The FOI Act only applies to information held in the form of a document. The definition of a 'document' in the FOI Act includes:

- any paper or other material on which there is writing or a mark, figure or symbol
- electronically-stored information
- maps, plans, drawings and photographs
- any article from which sounds, images or writing are capable of being produced.

The FOI Act does not cover documents that are otherwise accessible to the public. These include:

- documents available for access under the *Archives Act 1983* (Cth)
- documents open to public access subject to a fee or charge
- the library, historical and museum collections of the Australian War Memorial, National Library of Australia, National Museum of Australia, National Archives of Australia and the National Film and Sound Archive.

The FOI Act is not intended to restrict the circumstances in which government information can be released. An agency may disclose information without a request under the FOI Act, including information that would be exempt under the Act. An agency may also disclose

exempt information if a request is made under the Act, except where restrictions such as secrecy provisions prohibit disclosure.

A new statutory framework for proactive publication of information by government agencies was also established by the 2010 reforms through a new Information Publication Scheme and disclosure logs of documents released under the FOI Act (box C.9).

Box C.9 Information Publication Scheme

Since May 2011, the Information Publication Scheme requires all government agencies subject to the FOI Act to publish information that falls into the following categories:

- the agency's structure, functions (including its decision making powers and other powers affecting members of the public), its operational information (such as rules, guidelines and practices) and statutory appointments
- details of consultation arrangements for members of the public to comment on specific policy proposals
- the agency's annual reports
- details of officers who can be contacted about access to the agency's information or documents under the FOI Act
- information in documents to which the agency routinely gives access in response to requests under the FOI Act and information that the agency routinely provides to Parliament.

In addition, agencies *may* publish other information that they hold, such as statistical databases (OAIC 2011).

The Office of the Australian Information Commissioner (OAIC) is tasked with reviewing agencies' compliance with the IPS, but there are no direct enforcement measures in the Act (Stewart 2015).

There are two types of exceptions under the FOI Act — absolute exceptions and conditional exceptions. Absolute exceptions cover documents affecting national security, defence, informational relations, law enforcement and public safety — and Cabinet documents. Documents falling under a conditional exception must be released unless it would be contrary to the public interest to do so. Conditional exceptions cover a range of matters including personal privacy, business information and certain operations of agencies (such as audits, examinations and personnel management).

Personal information is protected from disclosure under the FOI Act where this would be unreasonable and contrary to the public interest. Where an agency decides to release personal information under the FOI Act (that is, where it is reasonable and in the public interest), the Privacy Act allows such release under APP 6.2(b).

The 2010 reforms to the FOI Act also made it easier for members of the public to make FOI requests: a request can be made by email; there is no application fee or charge for the first five hours of FOI processing; stronger regulation applies to agency observance of

processing time limits; and agencies are required to consult with FOI applicants before refusing access on the basis that the work involved in processing the request would substantially and unreasonably divert the agency from its other operations.

Government-business enterprises and the FOI Act

Government-business enterprises (GBEs) are exempt from the operation of the *Freedom of Information Act 1982* (Cth) in relation to documents about their commercial activities (section 7(2) and Part II of Schedule 2). For all other documents, GBEs do not automatically come under an exemption — they need to consider how the various exceptions apply to documents requested.

New South Wales

The *Government Information (Public Access) Act 2009* (NSW) (GIPA Act) was established to provide an open and transparent process for giving the public access to information from NSW public sector agencies and to encourage the proactive release of government information (table C.4).

The GIPA Act applies to government information. Government information is information in a record held by an agency, on behalf of an agency by a government contractor, or by the State Records Authority. A record can mean any document or source of information that is compiled, recorded or stored in printed or electronic form.

The GIPA Act:

- authorises and encourages the proactive release of information by NSW public sector agencies
- gives members of the public a legally enforceable right to access government information
- ensures that access to government information is restricted only when there is an overriding public interest against releasing that information.

The GIPA Act applies to all NSW government agencies, and also extends to Ministers and their staff, local councils, state-owned corporations, the non-judicial functions of courts, and to certain public authorities, such as universities.

The guiding principle of the GIPA Act is public interest. It is generally presumed that all government agencies will disclose or release information, unless there is an overriding public interest against doing so. Under the GIPA Act it is compulsory for agencies to provide information about their structure, functions and policies, and agencies are encouraged to proactively and informally release as much other information as possible.

Table C.4 **Comparing FOI laws across Australian jurisdictions**

	Cth	*NSW*	*Vic*	*Qld*	*SA*	*WA*	*NT*	*Tas*	*ACT*
Mandatory release for open access information	×	✓	×	×	×	×	×	×	×
Proactive release policy	✓	✓	×	✓	×	✓	×	✓	×
Information Publication Scheme	✓	✓	×	✓	✓	×	×	✓	✓
Public interest test[a]	✓	✓	×		×	✓	✓	✓	✓
Exceptions				✓					
• Cabinet	✓	×	✓	✓	✓	✓	✓	✓	✓
• Contrary to public interest	✓	×	✓	✓	1	×	✓	✓	
• National security and/or State relations	✓	×	✓	✓	✓	✓	✓	✓	✓
• Internal working documents	✓	×	✓	×	✓	✓	✓	✓	✓
• Law enforcement	✓	×	✓	✓	✓	✓	✓	✓	✓
• Legal professional privilege	✓	×	✓	✓	✓	✓	✓	✓	✓
• Contain personal information about others	✓	×	✓	×	✓	✓	✓	✓	✓
• Provided in confidence	✓	×	✓	×	✓	✓	✓	✓	✓
• Trade secrets	✓	×	✓	×	✓	✓	✓	✓	✓
• Secrecy provisions apply to the documents	✓	×	×	×	✓	×	×	×	✓

[a] South Australian Act provides many other exceptions that could encompass public interest.

Victoria

The Victorian *Freedom of Information Act 1982* gives individuals the right to request documents held by ministers, state government departments, local councils, most semi-government agencies and statutory authorities, public hospitals, and universities, TAFE colleges and schools. These documents might be created by the agency or supplied to the agency by an external organisation or individual. It is not only documents in paper form that are accessible. The word 'documents' covers a broad range of media including maps, films, microfiche, photographs, computer printouts, emails, computer discs, tape recordings and videotapes. Documents covered by the FOI Act are:

- documents about an individual's personal affairs, regardless of the age of the documents
- documents of a non-personal nature, not older than 5 July 1978
- documents held by a Council, not older than 1 January 1989.

The FOI Act also gives individuals the right to request that incorrect or misleading information held by an agency about them be amended or removed (table C.4).

The Freedom of Information Act allows an agency to refuse access to certain documents or information. These documents or information are often called 'exempt' documents, and can include Cabinet documents, some internal working documents, law enforcement documents, documents covered by legal professional privilege, such as legal advice., documents containing personal information about other people, documents containing information provided to an agency in confidence, documents containing information provided to an agency by a business, and documents which are covered by secrecy provisions in other legislation.

Documents that an individual might be able to obtain without an FOI application include those containing:

- an individual's personal information, such as personnel records
- information which is available publicly, such as on a public register
- information which is available for purchase (for instance, a criminal record check).

Queensland

Queensland's privacy and freedom of information legislation was reformed after the 2008 Solomon Review. The provisions were harmonised and a move towards more proactive disclosure of government information under the freedom of information legislation occurred.

The *Right to Information Act 2009* (Qld) (RTI Act) promotes the release of information and clarifies that formal access applications under the RTI Act should be used only as a last resort. Under the RTI Act, all public sector information is open to the community as a starting point (table C.4). Information can only be withheld by agencies if there is a good reason not to disclose it, such as protection of privacy. The starting point for all public sector documents is that they are open to the public. Agencies have an obligation to proactively release information, maximise disclosure, and otherwise provide administrative release. Administrative access is appropriate where any of the following apply:

- there is demand for access to the requested information
- there are no significant adverse effects as a result of disclosing the information, either generally or to particular applicants (this is discussed below)
- the information involved is of a kind that would be released if it was requested under the RTI Act, either generally or to particular applicants.

As a general rule, the sorts of documents that may be suitable for administrative release (that is, rather than release by FOI request) include those:

- provided to the agency by the person seeking access to them (for example, correspondence sent to the agency from the requester)
- provided by the agency to the person seeking access to them (for example, previous correspondence sent by the agency to the requester)
- that are publicly available
- that are routinely made available by the agency.

Agencies are not obliged to publish copies or details about information released under administrative access in their disclosure log. However, agencies are encouraged to consider publishing as much information as possible, where appropriate, in their disclosure log in the interests of openness and accountability. Please note, documents containing personal information about the requestor are not included in disclosure logs. Similarly, certain types of information are required to be deleted from information included in a disclosure log.

Access applications are considered a last resort. Possible criticism of the government, loss of confidence in the government or the mischievous use of information by applicants are factors that should be not be taken into account in deciding whether information is to be disclosed.

Formal applications for information can be made under the RTI Act for any document of an agency. RTI decision makers are required to have a pro-disclosure bias and documents must be released unless it is contrary to the public interest to do so.

When processing RTI and information privacy access applications, agencies are required to consider factors such as:

- whether there is any exempt information contained in the document
- public interest factors favouring disclosure and/or non-disclosure of the information
- whether consultation with third parties is required.

It is usually contrary to the public interest to disclose personal information about an individual to a third party other than that individual. The *Information Privacy Act 2009* (Qld) requires agencies to protect the personal information it holds and prevent it from being disclosed inappropriately. These two Acts work together to ensure that there is an appropriate balance between privacy protection and government openness. Personal information is protected unless there is a legal authority to disclose it. The RTI Act also requires agencies to provide details of:

- the information they will proactively make available
- how the information can be accessed
- the terms on which the information will be made available, including any charges
- the alternative formats in which information is available

- how to make a complaint when information included in the publication scheme is not available.

South Australia

The *Freedom of Information Act 1991* (SA) gives individuals a legal right to:

- request access to documents held by state government agencies, government ministers, local councils or state universities
- request the amendment of documents about them that are incomplete, incorrect, out-of-date or misleading
- seek a review of a decision made by a state government agency, government minister, local council or state university (table C.4).

While the aim of the South Australian Freedom of Information Act is to provide access to the maximum amount of information possible, some exceptions are necessary to ensure that people's privacy is not breached or that the proper administration of Government is not adversely affected.

Examples of documents that access may be refused to include:

- documents that would lead to an unreasonable disclosure of another person's affairs
- documents that contain trade secrets or information of commercial value
- documents affecting law enforcement and public safety
- documents subject to legal professional privilege or parliamentary privilege.

State Records of South Australia assists the Minister responsible for the Freedom of Information Act to administer the legislation by:

- giving general advice to members of the public
- giving advice to the Minister and government agencies
- drafting policy, guidelines and information sheets
- training staff from government agencies on how to manage freedom of information applications.

State Records does not process freedom of information applications. To access documents or amend personal documents held by State Government agencies, local councils or state universities, individuals can make a freedom of information application directly to the relevant organisation.

Western Australia

Part 4 of the *Freedom of Information Act 1992* (WA) establishes the Information Commissioner, whose main function is to deal with complaints about decisions made by agencies in respect of access applications and applications for amendment of personal information.

The Western Australian FOI Act gives the public a right to access government documents, subject to some limitations. The right applies to documents held by most state government agencies (such as departments, public hospitals, public universities and state government authorities), Ministers and local government. Together, these bodies are referred to as 'agencies' (table C.4).

Documents accessible under the FOI Act include paper records, plans and drawings, photographs, tape recordings, films, videotapes or information stored in a computerised form.

Some documents are protected from disclosure because their release would have an adverse effect on the private and business interests of individuals, or would hinder the proper functioning of government.

Under the FOI Act, agencies should only claim an exceptions when there are good reasons to do so and when the public interest requires nondisclosure, rather than merely because an exception is potentially available to be claimed. The onus is on the agency to show that its decision is justified.

Some of the exceptions in the FOI Act require an agency's decision-maker to decide whether disclosing certain information is, on balance, in the public interest. If the agency is required to consider the public interest, this usually means that information that would otherwise be exempt will not be exempt if its disclosure would, on balance, be in the public interest.

'Public interest' is not defined in the FOI Act. It can be a complex legal concept. Consideration of the public interest under the FOI Act is not primarily concerned with the personal interests of the particular access applicant or with public curiosity. The public interest is a matter in which the public at large has an interest as distinct from the interest of a particular individual or individuals. The question is whether, on balance, giving access to the information would be of some benefit to the public generally.

Deciding whether or not disclosing information would, on balance, be in the public interest test involves identifying and weighing the relevant competing public interests for and against disclosure of the information and deciding where the balance lies.

Agencies are required to assist applicants to obtain access to documents at the lowest reasonable cost.

The Information Commissioner is an independent officer reporting direct to the WA Parliament who deals with complaints about decisions made by government agencies under the FOI Act.

Tasmania

Section 7 of the *Right to Information Act 2009* (Tas) (RTI Act) gives any person a legally enforceable right to be provided with information in the possession of a public authority or a Minister, provided that it is not exempt information. The RTI Act promotes the proactive release of information by public authorities and Ministers, and refers to four types of disclosure:

- required disclosures, which are disclosures required by law such as annual reports.
- routine disclosures, which are those made by a public authority in relation to information it decides may be of public interest.
- active disclosures, which are disclosures in response to a request made other than under the RTI Act, such as an informal request for information by telephone.
- assessed disclosures, which are disclosures made in response to a formal request under the RTI Act for information in the possession of a public authority or Minister that is not otherwise available (table C.4).

The Ombudsman can also provide oral or written advice on the operation of the RTI Act to a public authority or Minister, either on the Ombudsman's own motion or on the request of a Minister or the principal officer of the authority. The Ombudsman is the review authority under the RTI Act.

Reviews relate to applications for assessed disclosure. Mostly, they occur at the request of the applicant for assessed disclosure, but review rights are also given by the Act to third parties who do not want information released. The Act gives the Ombudsman wide powers in relation to the conduct of reviews, including the power to give directions to the parties, and to promote settlement of a review application.

The Ombudsman is obliged to use these powers to resolve an application for review as soon as practicable after its receipt. Where the application cannot be resolved, the Ombudsman must ensure that a decision on the review is made as soon as practicable. The Ombudsman will normally only proceed to make a formal decision on an application for review when it is clear that there is no other way of resolving the issues between the parties.

Northern Territory

The *Information Act 2009* (NT) is the freedom of information and privacy legislation for the Northern Territory (table C.4). It applies to all public sector organisations including agencies, government-owned corporations, local governments, statutory corporations,

police, courts and tribunals (but not in relation to their judicial or decision-making functions), and contracted service providers (to the extent of the services they provide under their contract). The Information Commissioner is the independent officer appointed to oversee the freedom of information and privacy provisions, as well as to oversee public interest disclosures.

ACT

The *Freedom of Information Act 1989* (ACT) provides a general right of access to documents held by government agencies (table C.4). The Act requires decisions on access to be made promptly and at relatively low cost. It permits requests for documents to be refused for specific reasons related to the work of government or the interests of third parties, with all decisions subject to internal and external review. The Act provides a special right to complain to the Ombudsman about actions related to a request.

The Act provides three forms of review for those people who have sought access to documents under the Act, and are not satisfied with the response of an ACT Government department or authority to their request.

C.5 Copyright

Under the *Copyright Act 1968* (Cth), an author of a creative work has certain exclusive rights to control the use of their copyright material, including the right to copy, publish, communicate and publicly perform it. Holders of copyright also have certain moral rights — the right of integrity of authorship, the right of attribution of authorship and the right against false attribution of ownership.

Where there is copyright, exceptions allow certain uses of copyrighted material without the authorisation of rights holders. Australia's copyright system includes an exception for 'fair dealing' for research or study, criticism or review, parody or satire, reporting the news, judicial proceedings and professional advice. An exception also allows for temporary reproductions made in the course of communicating a work. Exceptions also allow Australians to record a television show on a video tape for their private viewing, or copy music to an mp3 player.

Unauthorised use of copyright material generally constitutes a civil infringement, requiring copyright holders to enforce their rights, usually in the Federal Court of Australia. Commercial-scale infringements of copyright are a criminal offence and prosecuted by the Commonwealth Director of Public Prosecutions. Copyright holders are able to seek an order requiring an Internet Service Provider to block access to an overseas website that facilitates online copyright infringement and the Australian Border Force has a role in detecting and seizing potentially infringing copyright-protected goods at the border.

There are no exceptions in the Copyright Act that cover data and text mining. Data or text mining processes involving the copying, digitisation or reformatting of copyright material without permission may give rise to copyright infringement. The reach of any fair dealing exceptions may not extend to text mining if the whole dataset needs to be copied and converted into a suitable format — such copying would be more than a 'reasonable portion' of the work concerned.

Databases

Copyright protects the form or way an idea or information is expressed, not the idea or information itself (*Breen v Williams* (1996) 186 CLR 71). Data compilations fall within the 'literary works' category of works protected under the Copyright Act. Literary works are defined as including a table or a compilation, expressed in words, figures or symbols (whether or not in a visible form). A factual compilation will be a literary work if it provides intelligible information, as opposed to a random collection of data (*Hollinrake v Truswell* (1894) 3 Ch D 420).

In the case of databases, this means copyright typically extends to cover original compilations of data, but not automated compilations of data, nor the underlying data. For example, telephone directories have previously been found to be subject to copyright: *Desktop Marketing Systems Pty Ltd v Telstra Corporation Ltd* (2002) 119 FCR 491. However, in two recent decisions courts have required that there be a human author involved in the reduction of the database to material form, and that there be some intellectual effort in the creation of that material form: *IceTV Pty Ltd v Nine Network Australia Pty Ltd* (2009) 239 CLR 458; and *Telstra Corporation Ltd v Phone Directories Co Pty Ltd* (2010) 194 FCR 142. These decisions narrowed the application of copyright to databases — automated databases will now generally be excluded from copyright protection.

Unstructured data poses particular challenges to traditional legal principles. Copyright has typically been the focus of protection of databases, but unstructured data in particular is not typically the province of the copyright lawyer, given the emphasis in copyright law on expression rather than ideas.

Creative Commons licenses

Where copyright exists, a use may be authorised through a licence granted by the copyright holder. The Australian Government's Public Data Policy Statement requires Australian Government entities to publish appropriately anonymised government data by default under a Creative Commons By Attribution licence unless a clear case is made to the Department of the Prime Minister and Cabinet for another open licence (box C.10). The state and territory governments have similar initiatives (see chapter 3). For instance, the Victorian Government's *Inquiry into Improving Access to Victorian Public Sector Information and*

Data (2009) recommended the Victorian Government make use of the Creative Commons licensing model for the release of public sector information (PSI). The Committee was told that Creative Commons licences can be appropriately used for up to 85 per cent of government information and data, providing a simple-to-understand and widely used system for the reuse of PSI. Remaining Victorian Government PSI should either not be released, or released under licences tailored specifically for restricted materials.

Box C.10 Terms of a Creative Commons licence

What is a CC licence?

A Creative Commons (CC) licence provides a simple standardised way for individual creators, companies and institutions to share their work with others on flexible terms without infringing copyright. It allows users to reuse, remix and share the content legally.

Offering one's work under a CC licence does not mean giving up copyright. It means permitting users to make use of the material in various ways and under certain conditions.

Licence terms: baseline permissions and core conditions

A CC licence sets out the uses that may lawfully be made of the copyright material and specifies the conditions that must be complied with when it is used.

There are six standardised CC licences. Each grants certain baseline permissions to users in advance, authorising them to use the material, provided they comply with core conditions and other general terms in the licence.

The baseline permissions granted by the CC licences permit the material to be copied, distributed, displayed and performed. Four of the CC licences additionally grant permission to users to use the CC-licensed material to create a Derivative Work (version 3.0 Australia licences) or Adapted Material (version 4.0 international licences) that may be copied, distributed, displayed and performed.

The core condition that applies to all six of the CC licences is the requirement that the author of the work is attributed – the Attribution condition.

The other core conditions are:

- Non-Commercial (NC)
- No Derivatives (ND)
- Share Alike (SA).

Source: Creative Commons Australia (2010).

Relevant Australian Government policies include the Australian Governments Open Access and Licensing Framework (AusGOAL) and the Australian Government Intellectual Property Rules.

C.6 Archives

Commonwealth Archives Act

Section 31 of the *Archives Act 1983* requires the National Archives of Australia (NAA) to make publicly available all Commonwealth records (box C.11) that are:

- in the open access period
- in the care of the Archives or in the custody of a Commonwealth institution
- not an exempt record (section 33).

Box C.11 National archives selection principles

The National Archives of Australia has adopted three principles to underpin its selection of Australian Government information for inclusion in the national archival collection.

1. Government authority, action and accountability

To keep information that provides evidence of the authority for the establishment and structure of the Australian Government and its agencies, and evidence of the deliberations, decisions and actions taken by the Australian Government and its agencies relating to key policies, functions and programs and significant issues faced in governing Australia.

2. Identity, interaction and rights and entitlements

To keep information that for individuals and communities: reflects identity and the condition and status of Australia and its people; provides evidence of ongoing rights and entitlements; or shows the impact of Australian Government activities on individuals and communities as well as their interaction with government.

3. Knowledge and community memory

To keep information that has substantial capacity to enrich knowledge and understanding of Australia's history, society, culture and people. We select information with the highest significance and value to communities and society.

Source: National Archives of Australia (2015).

Under the Archives Act, most Commonwealth records in the open access period are available for public access. Most records (98 per cent) are wholly released for public access, while 1.75 per cent are released with some exempt information deleted. All records will ultimately enter the open access period, although this period of time differs between types of records: after 20 years for most records; after 30 years for Cabinet notebooks; and after 99 years for census records (National Archives of Australia 2016).

Only 0.25 per cent of records are wholly withheld because they consist entirely of exempt information. If the NAA refuses access to a record, it is usually because it contains sensitive information or information that is not in the open access period. There is no time limit in the Act on how long a record may be exempt from release. Where a record is wholly withheld due to an exemption, a person may apply for access to that record. The NAA is

also able to reconsider records that have been wholly withheld and determine that the exemption no longer applies.

Under the Archives Act, it is an offence to destroy Commonwealth records without permission from the NAA unless destruction is specified in another piece of legislation or allowed under a normal administrative practice (box C.12). While section 31 applies to all Commonwealth records (and therefore any Commonwealth records an agency might hold that are in the open access period but which could have been destroyed, or may be destroyed after a longer period of time), the NAA only requires the permanent retention of records which are determined to be 'archival resources of the Commonwealth'.

Box C.12 **Normal administrative practice**

Normal administrative practice (NAP) allows agencies to destroy certain types of records in the normal course of business. Agencies do not need to contact the Archives for permission to dispose of records that fit within the scope of NAP. NAP allows agencies to manage the volumes of records they create and use every day in an efficient and accountable way. Records that can be considered for destruction using NAP fall into five broad categories:

- facilitative, transitory or short-term items including appointment diaries, calendars, 'with compliments' slips, personal emails, listserv messages and emails in personal or shared drives, emails that have been captured into a corporate records management system
- rough working papers and/or calculations
- drafts not intended for further use or reference – whether in paper or electronic form – including reports, correspondence, addresses, speeches and planning documents that have minor edits for grammar and spelling and do not contain significant or substantial changes or annotations
- copies of material retained for reference purposes only
- published material not included as part of an agency's records

The National Archives of Australia recommends a risk assessment to help agencies identify records that can be destroyed using NAP.

Source: National Archives of Australia (nd).

The NAA must consult with relevant entities about a request to access information that may be exempt. The Archives Act requires the NAA to make a decision and notify an applicant within 90 days of receipt of an access request, after which the decision is deemed to be a refusal. The applicant may seek internal reconsideration or review of a decision in the Administrative Appeals Tribunal.

The NAA is also responsible for administering the Digital Continuity Policy 2020 (box C.13), which is a whole-of-government approach to digital information governance. It aims to ensure that: information is managed as an asset; information is managed digitally; and agencies have interoperable information, systems and processes to improve information quality and enable information to be found, managed, shared and reused easily and efficiently.

Belcher Review recommendations

The recent Belcher Red Tape Review (2015) observed documents about peoples' personal information are available under both the FOI Act and the Privacy Act, leading to confusion about which access scheme should apply The Belcher review referred to the Hawke report recommendation for a comprehensive review of the FOI Act which should consider the interaction with the Archives Act and the Privacy Act (Hawke 2013, Recommendation 1).

Box C.13 Transitioning records to a digital format

Digital Transition Policy of 2011

The Digital Transition Policy of 2011 requires entities to move to digital information and records management and away from paper-based records management. Digital transition includes replacing paper-based processes with digital processes and limiting the creation of new paper records to reduce the costs of storing increasing quantities of paper records. While the NAA assists entities to observe elements of the policy, it has sought to limit the creation of paper records by not accepting paper based records that are created digitally after 1 January 2016.

As part of the digital transition, the NAA administers 'Check-up Digital', an annual survey to help entities gauge their digital information management maturity and set clear direction for improved digital practices. Better practice is highlighted through the NAA's Awards for Digital Excellence, which recognise and promote excellence and innovation in the management, use and reuse of digital information by entities.

Digital Continuity 2020 Policy

In May 2014, the NAA announced the development of the Digital Continuity 2020 Policy to build on the Digital Transition Policy. A Digital Continuity Plan provides practical advice to entities on managing digital information to ensure that it remains accessible and usable for as long as it is needed. The NAA has set non-binding digital continuity targets for 2020.

The review recommended that:

- the NAA publish its annual reports to government as part of the digital continuity policy
- the NAA work with entities to be more closely involved in policy development processes and decision-making forums on government information management, including digital transformation-related matters, to reduce the administrative burden arising from meeting their responsibilities under the Archives Act 1983; and ensure government information and data is usable for the future
- the Attorney-General's Department (AGD) work with the Archives to develop a proposal to amend the 90-day requirement for processing requests for access to information under the Archives Act 1983 to reduce the administrative burden by: changing the calendar day requirement to a business day requirement; and provide greater flexibility for the Archives to consult relevant entities on information
- AGD begin work with relevant entities to scope and develop a simpler and more coherent legislative framework for managing and accessing government information

during its life-cycle in a digital environment through staged reforms, commencing with legislation regulating archives (Belcher 2015, recommendation 18).

States and territories

All states and territories have archives arrangements:

- In New South Wales, State Records NSW operates under the *State Records Act 1998*.
- In Victoria, the Public Record Office Victoria operates under the *Public Records Act 1973*.
- In Queensland, the Queensland State Archives operates under the *Public Records Act 2002*.
- In South Australia, State Records of South Australia operates under the *State Records Act 1997*.
- In Western Australia, the State Records Office of Western Australia operates under the *State Records Act 2000*.
- In Tasmania, LINC Tasmania operates under the *Public Records Act 1943*.
- In the Northern Territory, the Northern Territory Archives Service operates under the *Information Act 2002*.
- In the ACT, the Territory Records Office operates under the *Territory Records Act 2002*.

C.7 Information security

Commonwealth

Australian Privacy Principle 11 requires Australian Government agencies or businesses with a turnover of more than $3 million to take reasonable steps to protect personal information from misuse, interference and loss as well as unauthorised access, modification or disclosure of personal information. Additionally, there are a number of other laws dealing with specific information security issues — for instance, the *Telecommunications (Interception and Access) Act 1979* (Cth) prohibits the interception of and access to telecommunications except where authorised in special circumstances.

The Australian Government has established the Protective Security Policy Framework (PSPF) (2012), which is managed by the AGD. Protective security encompasses:

- governance (Fraud Control Framework, accountability, risk management) — agencies must manage security risks to prevent harm to official resources and disruption to business objects

- personnel security (security clearances and the Australian Security and Intelligence Organisation contact reporting scheme to identify intelligence or hostile activity directed against Australia and its interests)
- information security — while AGD is responsible for the overall PSPF, the Australian Signals Directorate has specific responsibility for government information security implementation and monitoring, including implementation of controls in the Australian Government's Information Security Manual (see further below).

The PSPF provides appropriate controls for the Australian Government to protect its people, information and assets, at home and overseas. Governance arrangements and core policy documents in the PSPF describe the higher level mandatory requirements applicable to entities. Detailed protocol documents and guidelines support the personnel security, information security and physical security core policies. The protocol documents set out minimum procedural requirements. Some entities have specific security risks that will require them to apply more than the minimum requirements.

The PSPF applies to Non Corporate Entities. The principles of the PSPF are being extended to apply to corporate Commonwealth entities and wholly-owned Commonwealth companies that have received a government policy order under the *Public Governance Performance and Accountability Act 2013* (Cth) (PGPA).

There are 13 governance requirements in the PSPF (abbreviated to GOV-1 to GOV-13). Key governance requirements of the PSPF involve entities:

- applying risk-based principles and policies to manage the functions of an entity and the security threats its faces
- developing, implementing and maintaining protective security measures
- preparing, monitoring and reviewing security plans to ensure they address risks in the operating environment
- reporting annually to their portfolio minister on the level of entity compliance with the PSPF
- developing a culture of security through strong programs of security awareness and education to ensure employees fully understand their security responsibilities
- investigating security incidents promptly and with sensitivity.

The Belcher (2015) review found that the PSPF governance requirements appeared to be overly prescriptive and applied a compliance approach that may not assist entities to effectively engage with risk. The review considered the PSPF would benefit from being reviewed, particularly to adopt a more a principles-based approach, where possible, to be more consistent with the PGPA framework.

Information security manual

The Australian Signals Directorate (ASD) produces the Australian Government Information Security Manual (ISM). The manual is the standard that governs the security of government ICT systems. The ISM consists of three documents targeting different levels within each organisation, making the ISM accessible to more users and promoting information security awareness across government. Since April 2013, all Australian Government agencies have been required to comply with ASD's Top 4 Mitigation Strategies.

Cyber Security Strategy

The Australian Cyber Security Strategy has been developed over 18 months of consultation with more than 190 organisations and across business, government and academia, both in Australia and overseas. Government and private sector stakeholders set the strategic agenda and co-design initiatives within the strategy. This strategy has established five themes of action for Australia's cyber security over the next four years to 2020.

In November 2014, the Australian Government established the Australian Cyber Security Centre (ACSC). The purpose of the ACSC is to bring together existing cyber security capabilities and to provide a hub for greater collaboration and information sharing between the private sector, state and territory governments, and international partners. In doing so, the ACSC co-located several bodies, including the ASD's cyber security mission which provides advice and assistance to Australian government agencies, and the Computer Emergency Response Team Australia which is a point of contact in government for cyber security issues affecting major Australian businesses. The ASD leads the ACSC, sharing information and working closely with the Australian Security Intelligence Organisation, the Australian Federal Police, the Australian Signals Directorate, the Defence Intelligence Organisation and the Australian Crime Commission.

The Department of Finance and the Digital Transformation office have also issued a number of specific ICT and identity management policies and guidelines including:

- Identity Management for Australian Government Employees (IMAGE) Framework
- National e-Authentication Framework
- National Identity Proofing Guidelines — guidance for agencies for the identification of users
- Gatekeeper public key infrastructure framework
- Third Party Identity Services Assurance Framework.

More details can be found on the relevant website.

States and territories

State and territory information security frameworks are broadly similar to those established at the Commonwealth level, and similar telecommunications interception provisions apply — see for example the *Telecommunications (Interception and Access) Act 1987* (NSW) and the *Telecommunications Interception Act 2009* (Qld).

Key state and territory government provisions include:

- *New South Wales*: Information Protection Principle 5 requires NSW Government agencies to ensure that personal information is stored securely, and there are other specific acts. The NSW Government has adopted a Digital Information Security Policy as part of the NSW ICT Strategy.
- *Victoria:* Information Privacy Principle 4 requires personal information to be stored securely and the Victorian Government has adopted the Victorian Protective Data Security Framework — both of these are governed by the *Privacy and Data Protection Act 2014* (Vic).
- *Queensland*: the *Privacy Act 2009* (Qld) imposes requirements to keep information on the Queensland public sector. The Queensland Government has adopted the Queensland Government Information Security Policy framework.
- *South Australia*: the South Australian Information Privacy Principles (contained in a circular issued by the Department of Premier and Cabinet) require secure storage of personal information. The Information Security Management Framework addresses cybersecurity in the Government of South Australia, and consistent of 40 policies supported by 140 standards — it is aligned with the Australian Government Protective Security Policy Framework.
- *Western Australia:* the WA Ombudsman has issued Guidelines for the Management of Personal Information, which includes a requirement for secure storage, and the WA Government has adopted the Western Australia Whole of Government Digital Security Policy.
- *Tasmania:* Personal Information Protection Principle 4 requires secure storage of personal information, and the Tasmanian Government has adopted an Information Security Framework.
- *Northern Territory:* Information Privacy Principle 4 requires secure storage of personal information, and the Records Management Standards for public sector organisations in the NT require that records be stored securely.
- *ACT:* Territory Privacy Principle 11 requires the secure storage of personal information, and the ACT Government has adopted the Protective Security Policy and Guidelines.

C.8 Research governance

Broad guidelines for research

The National Statement on Ethical Conduct in Human Research (2007) and the Australian Code for the Responsible Conduct of Research (2007) are the two major guiding documents for institutions and researchers conducting high quality, ethical and sustainable research involving humans:

- The *National Statement on Ethical Conduct in Human Research* sets out ethical considerations and processes of research governance and ethical review.
- The *Australian Code for the Responsible Conduct of Research* describes best practice for institutions and researchers on, for instance, how to manage research data and materials, how to publish and disseminate research findings, how to conduct effective peer review and how to manage conflicts of interest. It also sets out a framework for handling breaches of the Code and research misconduct.

The *National Health and Medical Research Council (NHRMC) Research Governance Handbook* also provides guidance on the national approach to single ethical review (chapter 5).

Health and medical research without consent

In certain circumstances, the *Privacy Act 1988* (Cth) permits the handling of health information and personal information for health and medical research purposes, where it is impracticable for researchers to obtain individuals consent. This reflects that health information, as sensitive personal information, has extra protections placed around it by the Privacy Act, but health and medical research also has an important role in advancing public health.

The OAIC has approved two sets of *legally binding* guidelines issued by the NHMRC. Researchers must follow these guidelines when handling health information for research purposes without individuals' consent. The guidelines also assist Human Research Ethics Committees (HRECs) in deciding whether to approve research applications.

- Guidelines under section 95 of the Privacy Act set out procedures that HRECs and researchers must follow when personal information is disclosed from a Commonwealth agency for medical research purposes where the public interest in the research outweighs the public interest in the protection of privacy.
- Guidelines under section 95A of the Privacy Act provide a framework for HRECs to assess proposals to handle health information held by organisations for health research (without individuals' consent). They ensure that the public interest in the research activities substantially outweighs the public interest in the protection of privacy.

The section 95 and 95A guidelines apply when:

- the collection, use or disclosure of health information is necessary for research or the compilation or analysis of statistics relevant to public health or public safety
- it is impracticable to seek the person's consent before the use or disclosure (seeking a waiver of consent or implementing an opt out approach may be required for section 95)
- collection, use or disclosure is conducted in accordance with the relevant guidelines
- if, in disclosing the personal health information, the organisation reasonably believes that the recipient will not disclose it or personal information derived from it (for section 95A).

Guidelines under section 95AA allow the use and disclosure of a patient's genetic information to a genetic relative of that patient where the patient has not given consent but the health service provider reasonably believes there is a serious threat to the life, health or safety of a genetic relative of a patient and the use and disclosure is necessary to lessen or prevent that threat.

In 2008, the ALRC (2008) recommended that guidelines be able to be issued for all types of human research, not just medical research. This recommendation was accepted but not implemented.

Legislation governing ethics committee approval

Research involving humans: must be reviewed by a HREC or an institutional low risk review process in accordance with the National Statement on Ethical Conduct in Human Research. Further detail about ethics review processes is given in appendix B.

There is a wide range of other legislation and guidance that affects ethics committee approval and the conduct of research — requirements differ depending on what type of research is involved, and what jurisdiction the research is being conducted in. For example, within the health and medical research sector (NHMRC 2015), legislative requirements fall into the following categories:

- clinical trial notification and exemption schemes
- consent and impaired capacity to consent and research involving children
- embryo research
- gene technology and research using gene technology (that is, any technique for the modification of genes or other genetic material) and research involving gene and related therapies and stem cell-based cellular therapies
- ionising radiation — radiological procedures that are performed specifically for research
- removal of human tissue (excluding blood) from a living or deceased person

- coronial material
- research involving animals, which must be reviewed and approved by a properly constituted Animal Ethics Committee as being in accordance with the Australian Code for the Care and Use of Animals for Scientific Purposes 8th Edition 2013
- research involving genetically modified organisms, which must comply with all the requirements of the *Gene Technology Act 2000* (Cth) and Gene Technology Regulations 2001. Applicants should seek advice from their Institutional Biosafety Committee on the level of authorisation needed for any proposed GMO research
- use of carcinogenic or highly toxic chemicals, which must adhere to the National Occupational Health and Safety Commission guidelines, National Code of Practice for the Preparation of Material Safety
- use of cultured cell lines for research
- unapproved therapeutic goods, which must obtain an exemption under the *Therapeutic Goods Act 1989* (Cth)
- biotechnology researchers and other scientists seeking to gain access to genetic resources must comply with the Nagoya Protocol which establishes a legally binding framework for and to share any benefits from the use of genetic resources or traditional knowledge associated with those resources with the provider country
- controlled technology and the dissemination of intangible technology, which must comply with the *Defence Trade Controls Act 2012* (Cth)
- other jurisdiction-specific requirements — for instance, in NSW, the *Research - Ethical & Scientific Review of Human Research in NSW Public Health Organisations PD2010_055* provides that all research projects requiring access (including linkage) to statewide data collections owned or managed by NSW Health or the Cancer Institute NSW must be reviewed by the NSW Population and Health Services Research HREC. In Victoria, the *Charter of Human Rights and Responsibilities Act 2000* also applies to research conducted by a 'public authority'.

More general privacy and confidentiality obligations and data linkage requirements may also apply under *Privacy Act 1988* (Cth), the A*ustralian Institute of Health and Welfare Act 1987* (Cth), the *Health Records (Privacy and Access) Act 1997* (ACT), the *Public Health Act 1997* (ACT), the *Adoption Act 1993* (ACT), the *Health Records and Information Privacy Act 2002* (NSW) and the Statutory Guidelines on Research (2004) published under the NSW Privacy Act, the *Mental Health Act 2007* (NSW), the *Parliamentary Electorates and Elections Act 1912* (NSW), the Health Administration Regulation 2010 (NSW), the *Public Health Act 2010* (NSW), the *Information Act 2002* (NT), the *Cancer (Registration) Act 1988* (NT), the Public Health (Cervical Cytology Register) Regulations 1996 (NT). the *Public and Environmental Health Act 2011* (NT), the *Public Health Act 2005* (Qld), the *Privacy Act 2009* (Qld), the *Hospital and Health Boards Act 2011* (Qld), the *Mental Health Act 2009* (SA), the *Health Care Act 2008* (SA), the *Adoption Act 1998* (SA), the *Assisted Reproductive Treatment Act 1988* (SA), the *Children's Protection Act 1993* (SA), the

Controlled Substances Act 1984 (SA), the *Health and Community Services Complaints Act 2004* (SA), the *Supported Residential Facilities Act 1992* (SA), the *Transplantation and Anatomy Act 1993* (SA), the SA Health Code of Fair Information Practice (2004) (SA), the SA Department of Premier and Cabinet Information Privacy Principles, PC012 (2013) (SA). the *Personal Information Protection Act 2004* (Tas), the *Health Records Act 2001* (Vic), the *Information Privacy Act 2000* (Vic), *Public Sector Management Act 1994* (WA), and the *State Records Act 2000* (WA).

It is worth noting that the National Ethics Application Form has been designed to enable researchers to complete research ethics proposals for submission to HRECs, and to assist HRECs to consistently and efficiently assess these proposals — it meets the requirements of relevant guidelines with the aim of increasing the efficiency and quality of the ethical review process for all parties involved. Ethics committee arrangements are discussed in appendix B.

Other requirements governing research

There are numerous other laws and rules that cover governance of human research. These include:

- State and territory specific policies and guidelines relating to health research — for instance, the WA Health Research Governance Policy and Procedures. Some health care providers also have their own guidelines and requirements (for instance, Mater Health Services requires all human research to undergo a research governance review in addition to ethical review).
- Intellectual property policies apply to research — for instance, the Intellectual Property Arising from Health Research Policy — NSW Department of Health, and the National Principles of Intellectual Property Management for Publicly Funded Research 2001 the Australian Research Council open publications policy, and the National Health and Medical Research Council open publications policy.
- Requirements surrounding adverse events and research monitoring, for instance, the Monitoring and reporting of safety for clinical trials involving therapeutic products, and the provisions of part 2 of the *Health Administration Act 1982* (NSW).
- Complaint handling — in addition to the Australian Code for the Responsible Conduct for Research and the National Statement on Ethical Conduct in Human Research, specific provisions apply such as the SA Health Research Governance Policy and the SA Consumer Feedback Management Policy Directive.
- Risk management — in addition to the Australian Code for the Responsible Conduct for Research and the National Statement on Ethical Conduct in Human Research, specific provisions apply such as the *Risk Management Policy* and *Health Service Directive Research Ethics and Governance* issued by Qld Health.

- Storage and retention of records — in addition to the Australian Code for the Responsible Conduct for Research, specific requirements apply such as, in New South Wales, the *State Records Act 1998* (NSW), and a number of guidelines including the General Retention and Disposal Authority: Public Health Services: Patient/Client Records; the General Retention and Disposal Authority: Public Health Services: Administrative Records; the Operations Manual: Human Research Ethics Committee Executive Officers; and the Operations Manual: Research Governance Officers.
- Ionising radiation — for instance, under the *Radiation Protection Act 2005* (Tas).
- Financial accountability — for instance, under the *Financial Management Act 2003* (NT).
- Working with children — for instance, under the *Working with Children (Criminal Record Checking) Act 2004* (Vic).
- Specific guidelines apply to research involving Aboriginal and Torres Strait Islanders – for instance, the *Values and Ethics: Guidelines for Ethical Conduct on Aboriginal and Torres Strait Islander Health Research*, and the *Statement on Consumer and Community Participation in Health and Medical Research*.
- Data linking and data management: for instance, the High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes — these are discussed in appendix B, and impose, among other things, a requirement to delete linked data at the end of a project. Also relevant are the Data Matching Program (Assistance and Tax) Guidelines 1997 (Cth), and use of health datasets for research purposes must comply with the Minimum Guidelines for Health Registers for Statistical and Research Purposes where relevant.

Finally, individual research bodies such as universities have their own internal policies and guidelines governing research. It is well beyond the scope of this appendix to detail them all here.

C.9 International frameworks

Internationally, legislative and policy frameworks for data collection, storage and disclosure vary substantially. However, Australia's frameworks share a number of similarities with those in some other Commonwealth countries, such as New Zealand and the United Kingdom. Below, legislation and policy relating to the collection, sharing and dissemination of information in New Zealand, the United Kingdom and the European Union is outlined.

New Zealand

Privacy legislation

In New Zealand, privacy is governed primarily by the *New Zealand Privacy Act 1993* (NZ Privacy Act), which regulates the collection, use, disclosure, storage and provision of access to personal information. Its application is broad, with exceptions limited to Members of Parliament, news media and courts and tribunals. Some key requirements of data holders within the NZ Privacy Act are that:

- collection of information is for a lawful purpose, necessary for that purpose and collection of information about the individual is from the individual
- access to personal information is provided on request to individuals to whom that data relates, with the ability to request record correction
- use of personal information collected for a specific purpose is limited to that purpose.

These are only some of the key principles within the privacy act, and there are a number of exceptions to each.

Variations to requirements of the NZ Privacy Act can be introduced by New Zealand's Privacy Commissioner. This involves issuing Codes of Practice which can alter the application of the NZ Privacy Act for specific sectors, including health, telecommunications and credit reporting. In addition, Part 9A of the Privacy Act contains an 'approved information sharing agreement' mechanism which allows New Zealand Government agencies and other entities to share personal information for service delivery purposes. Each data-sharing proposal is considered on its merits, requires transparency about proposed data-sharing activities and requires review by the Privacy Commissioner. The agreements have the status of legislative instruments requiring Cabinet approval.

In addition to the NZ Privacy Act, there are many examples of legislation relating to data collection, use and dissemination. For example, Statistics New Zealand's powers to collect and disseminate information are contained with the New Zealand *Statistics Act 1975*. The Act, among other things, specifies that the New Zealand Statistician may disclose individual information only if it is to be used for bona fide research or statistical purposes and the statistician is satisfied that the person has the necessary research experience, knowledge and skills.

Access to government information

New Zealand's right to access government records are contained within the New Zealand *Official Information Act 1982* (NZ OI Act). The express purpose of the NZ OI Act is to: increase the availability of official information; provide for proper access by each person to official information relating to that person; and protect official information to an extent consistent with the public interest.

The NZ OI Act specifies circumstances under which access may be withheld. Such circumstances include: prejudice of security or defence; endangering the safety of a person; and serious damage to the economy (among many others). It also requires that requests are passed onto appropriate organisation within 10 working days of their receipt and that a decisions on whether the request is granted be made available within 20 working days of the request being forwarded. The official information act allows agencies to charge applicants in accordance with the costs associated with accessing the data, and individuals have rights to correct erroneous information.

In addition to regulations relating to privacy and access to government information, key regulations in New Zealand's storage and access of government information are outlined in the *Public Records Act 2005,* which also specifies the extent of recordkeeping required relating to the affairs of central and local government. More specifically, it requires that every public office and local authority create and maintain full and accurate records of its affairs. These must be maintained in an accessible form. Moreover, unless specified otherwise, records classified as open access must be made available for inspection without charge and in a reasonable timeframe.

Open data policies

The New Zealand government initiated its Open Government Information and Data Programme in 2008. This program is hosted by the New Zealand Land Information Department, and led by the Open Government Data Chief Executives Governance Group and the Open Government Data Steering Group. In 2011, New Zealand released its *Declaration on Open and Transparent Government*, which focused on making publicly funded, high value data availability to the public. The documents holds that data held by the New Zealand government must be 'open, trusted, authoritative, well managed, readily available, without charge where possible and reusable, both legally and technically' (ICT NZ 2016).

United Kingdom

Privacy legislation

As in Australia, a complex array of statutory provisions and common law surrounds the sharing and release of personal data in the United Kingdom. The primary piece of legislation relating to privacy and data sharing in the United Kingdom is the *Data Protection Act 1998* (UK Data Protection Act) (table C.5).

The UK Data Protection Act transposes the *1995 General Data Protection Directive 95/46/EC* and regulates the collection, use, distribution and retention of personal data, where personal data is defined as that which relates to a living person and may allow the identification of that person. It places restrictions on the sharing of data mainly for

compliance or operation purposes, but also on the linkage of administrative datasets for research purposes. Broadly, the United Kingdom's data protection principles hold that personal data be: obtained only for lawful purposes; relevant and not excessive in relation to those purposes, accurate and up-to-date, kept for no longer than is necessary; protected from unauthorised or unlawful processing; and kept within the European Economic Area unless an adequate level of protections are ensured.

The UK Information Commissioner has issued a data-sharing code of practice under the UK Data Protection Act to clarify how the Act applies to data-sharing activities and to provide best practice data sharing. The Code is not legally binding in itself, though issued under the Act and approved by the relevant minister. However, courts, tribunals and the Commissioner must take the code into account when considering matters under the Act.

Other regulations relating to privacy include:

- The *Human Rights Act 1988* (UK), which gives effect to the European Convention on Human Rights and provides that person has the right to respect for his or her private and family life, home and correspondence.
- The *Privacy and Electronic Communications (EC Directive) Regulations 2003,* which implements the EU e-privacy Directive (Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector). Also known as 'the e-privacy Directive', it sets out rules for direct marketing, cookies or similar technologies and notification of data breaches.

Better use of data

As in Australia, data sharing in the United Kingdom is not only limited by the regulations outlined above, but also an array of domain-specific legislation. A proposal for new legislation to facilitate data sharing is currently under consideration, with the aim to simplify the complex legal landscape around data sharing for both research and administration purposes. More specifically, the suggested legislative changes include clauses to allow:

- public agencies to share personal data with other public agencies in specific contexts in order to improve the welfare of the individual in question
- public authorities to pilot projects that identify where individuals have debts with a number of public agencies, and then have a single interaction with them to help manage those debts
- access to civil registration data such as births, deaths and marriages to allow public authorities to prevent sending letters to people who have deceased and make it easier for citizens to interact with public services
- public authorities to pilot methods to spot conflicting information across different public services that could suggest patterns of fraud for further investigation by officials

- the Office for National Statistics to access detailed administrative data from across government and businesses to provide more accurate, frequent and timely statistics and to update how the census is managed, instead of relying on surveys
- the use of de-identified data to support accredited researchers to access and link data in secure facilities to carry out research for public benefit.

At the same time, key protective principles outlined in the consultation include ruling out the: building of new, large, and permanent databases, or collecting more data on citizens; indiscriminate sharing of data within Government; and amending or weakening of the Data Protection Act. Moreover the consultations suggests that the safeguards that apply to a public authority's data (such as Her Majesty's Revenue and Customs) continue to apply to the data once it is disclosed to another public authority.

A number of these proposals are contained within the *Digital Economy Bill 2016* (Cabinet Office (UK) 2016). Prior to the June 2016 referendum on membership of the European Union, the EU General Data Protection Regulation, or GDPR, agreed in April 2016, was due to come into force in the United Kingdom on May 2018 (European Commission 2016b) (box C.15). The result of the referendum now means that the Government needs to consider the impact on the GDPR (ICO (UK) 2016).

Access to government information

The United Kingdom was a number of years behind Australia in introducing legislation to allow access to government information. Its *Freedom of Information Act* 2000 (UK FOI Act) provides public access to information held by public authorities in England, Wales and Northern Ireland. The UK FOI Act covers public authorities (including government departments, executive agencies and some individually identified organisations), applies to all recorded information (ranging from emails and notes to CCTV footage) and is enforced by the Information Commissioner's Office. Unlike similar legislation in other countries (such as Australia and New Zealand), the UK FOI Act does not allow individuals to access information about themselves. This right is provided by the Data Protection Act 1998.

Additionally, access to government information within the United Kingdom is governed by the *Public Service Information Directive*. This directive is the transposition of the *European Directive 2003/98/EC on the re-use of public sector information*. The directive outlines: limits on charges associated with providing data; obligations on data holders to provide data in a timely, open and transparent manner; access application processes; and the prohibition of exclusive licenses.

Open data legislation or policies

The United Kingdom is a world leader on open data. It ranks first in the world on the World Wide Web Foundation's (2016) most recent ranking of open data progress (chapter 3). The

United Kingdom's push towards open data started with a letter by the UK Prime Minister in 2010 to department heads calling for increased availability of public sector information and identifying specific datasets for release, including: every item of central government and Quango spending over £25 000; publication of the names and salaries of all central government and Quango managers earning over £150 000 per year; salaries of the 35 000 most senior civil servants; and monthly online publication of local crime data.

In July 2011, the UK Prime Minister circulated a second letter on open data. It called for the open publication of data relating to the National Health Service, criminal justice, transport, government financial information and education (GOV.UK 2010). It also called for the improvement of data quality (including plain English descriptions and unique reference indicators) and the use of the Open Government License — a copyright license permitting anyone to copy, publish, distribute, transmit and adapt licensed public sector work.

Other key legislation, policies and practices

midata

As part of the United Kingdom's midata program, the *Enterprise and Regulatory Reform Act 2013* includes provisions giving Government the power to compel gas, electricity, mobile phone service or financial service business to provide customer data, on request, directly to a customer or to a third party person or business on the customer's behalf (sections 89-91). However, the UK Government has ultimately implemented the midata program by engaging with business on a voluntary basis (chapter 4). More recently, the UK Government introduced the *Digital Economy Bill* 2016 which builds on the midata legislation and makes it easier for consumers to change communications providers on request (clause 2 amending section 51(2) of the *Communications Act 2003*).

European Union

Privacy legislation

Until May 2016, privacy in the European Union was governed by the *European Parliament and Council Directive 95/46/EC of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data* [Official Journal L 281 of 23.11.1995] (EU 1995 Directive). The EU 1995 Directive set strict limits on the collection and use of personal data and required that each Member State set up an independent national body responsible for the supervision of any activity linked to the processing of personal data. Recently, *General Data Protection Regulation (EU)* 2016/679 and an accompanying Directive entered into force in May 2016 and will apply as an EU-wide law from May 2018 (box C.14).

Other relevant directives that apply in the European Union include:

- Privacy and Electronic Communications Directive (2002/58/EC) (also known as 'the e-privacy Directive') which sets out rules for direct marketing, cookies or similar technologies, notification of data breaches and data retention for the purposes of police surveillance.
- Regulation (EU) No 611/2013, which contains rules surrounding the notification of personal data breaches in the event that customers' personal data are lost, stolen or otherwise compromised.

Box C.14 The EU General Data Protection Regulation (GDPR)

The GDPR has significant implications for citizens and businesses in the EU. Its features include:

- *New right to be forgotten* — when an individual no longer wants his/her data to be processed, and provided that there are no legitimate grounds for retaining it, the data must be deleted.
- *New right to data portability* — intended to make it easier for individuals to transmit personal data between service providers
- *Stronger enforcement of the rules* — data protection authorities will be able to fine companies who do not comply with EU rules up to 4% of their global annual turnover
- *Strengthening the EU internal market* — companies will deal with one law and one single supervisory authority, not 28
- *Streamlining international transfers of personal data* — companies based outside of Europe will have to apply the same rules as European companies when they offer goods or services on the EU market
- *Setting global data protection standards* — companies and organisations must notify the national supervisory authority of data breaches that put individuals at risk and communicate to the data subject all high risk breaches as soon as possible so that users can take appropriate measures.

Sources: European Commission (2015, 2016a).

Access to government information

The European Directive on the reuse of public sector information (Directive 2003/98/EC) was introduced in 2003 and later revised in 2013. Compared with freedom of information regulations, the directive focusses on access to public sector information for economic benefit. It requires, among other things, that:

- conditions for reuse of information are non-discriminatory
- charges for reuse be limited to the marginal costs of providing that data, and that charges and other conditions for reuse be pre-established and published
- exclusive arrangements are prohibited
- requests for reuse are processed promptly (20 days for standard cases)
- licences do not unnecessarily restrict possibilities for reuse, nor are they used to restrict competition.

Table C.5 **Some key features of international privacy and data protection laws**

	Australia Privacy Act 1988	*New Zealand Privacy Act 1993*	*United Kingdom Data Protection Act 1998*	*European Union General Data Protection Regulation*	*OECD Privacy Guidelines (revised 2013)*[a]	*APEC Privacy Framework 2005*
Public/private sectors	Both	Both	Both	Both	Both	Both
Small business exemption	✓ (s6C, s6D)[b]	×	×	×	×	×
Openness principle / privacy policy	✓ (APP 1)	×	×	✓ (Articles 13, 14)	✓ (Para 12)	✓ (Principle 20)
Access and correction rights	✓ (APP 12, 13)	✓ (s6, Principles 6, 7)	✓ (s7, 14)	✓ (Articles 15, 16)	✓ (Para 13)	✓ (Principles 23-25)
Direct marketing rules	✓ (APP 7)	×	✓ (s11)	✓ (Articles 18, 21)	×	×
Data quality	✓ (APP 10)	✓ (s6, Principle 8)	✓ (Sch 1, Principle 4)	✓ (Article 5(1))	✓ (Para 8)	✓ (Principle 21)
Data security	✓ (APP 11)	✓ (s6, Principles 5, 9)	✓ (Sch 1, Principle 7)	✓ (Articles 5(2), 32)	✓ (Para 11)	✓ (Principles 14, 22, 26)
Data breach notification	Legislation expected 2016	Legislation expected 2017	✓ **c**	✓ (Articles 33, 34, 83)	✓ (Para 15(c))	**d**
Complaints handling mechanism	✓	✓	✓	✓ (Chapter VI, Articles 51-59)	✓ (Para 15(a))	✓ (Principle 31)
Enforcement (personal, regulator)	Regulator	Regulator and some personal	Personal and regulator	Personal and regulator	Regulator and some personal	Personal and/or regulator
Research using personal information (not de-identified)	✓ (s16B, 95A, 95AA) **e**	×	✓ Statistics or research – includes historical (s33)[f]	×	×	×

(Continued next page)

Table C5.5 (continued)

	Australia Privacy Act 1988	*New Zealand Privacy Act 1993*	*United Kingdom Data Protection Act 1998*	*European Union General Data Protection Regulation*	*OECD Privacy Guidelines (revised 2013)*[a]	*APEC Privacy Framework 2005*
De-identification rules	×	×	×	✓(Art 4(5), 32, 40)	×	×
Use of de-identified information - general	✓ (s6(1) [h]	✓	×	✓ (Art 4(5), 6(4), 11)	×	×
Use of de-identified information - research	✓ (s6(1) [24]	✓ (s6, Principle 3(4)(f)(ii)) [i]	✓ (s1)	✓ Only historical or scientific (Art 4(5), 9(2)(j), 89(1)) [j]	N/A	N/A

[a] Recommendation concerning Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data (2013) [C(80)58/FINAL, as amended on 11 July 2013 by C(2013)79]. [b] Note the exemption does not cover certain small businesses, for example those whose business involves the sale or purchase of personal information. [c] *Privacy and Electronic Communications (EC Directive) Regulations 2003* (UK), reg 5. [d] Likely to be included in the next revision of the Privacy Framework. [e] Personal information may be used only for health and medical research, in accordance with NHMRC guidelines approved by the OAIC, and where it is impracticable to obtain the individual's consent. [f] The results of the research or any resulting statistics are not to be made available in a form which identifies data subjects or any of them. [g] The information must be used in a manner that will ensure its confidentiality and the organisation must inform the Privacy Commissioner before it is used. [h] Personal information is 'de-identified' if it is longer about an identifiable individual or an individual who is reasonably identifiable (definition in s6(1)). De-identified information does not fall within the definition of 'personal information' and so is outside the scope of the Privacy Act. [i] The information must not be published in a form that could reasonably be expected to identify the individual. [j] Archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. [k] The information must be used in a manner that will ensure its confidentiality and the organisation must inform the Privacy Commissioner before it is used.

D Case Study: Health data

Key points

- Health data collection and use in Australia – by GPs, pharmacies, hospitals, and other healthcare providers — is scattered, unorganised and duplicative. There are substantial opportunities to make far greater use of the data collected, to the benefit of all Australians.
  - While broad health service indicators are readily available, access to underlying data is impeded by concerns about privacy and complex approval processes. Privacy controls that affect health information are more complex than for other types of data.
  - Linked datasets are particularly valuable for assessing the performance of the health system and in providing more integrated health services. However, the linking of datasets, particularly using data held by the Commonwealth, is not routinely carried out, and the process for obtaining linked datasets is complex, lengthy and expensive: only three bodies nationwide are accredited to link Commonwealth government health data. Legislative guidelines, mandating that datasets linking Medicare and PBS data must be destroyed after the completion of a project, further contribute to delays and expenses.
- eHealth systems can improve health data collection and transfer. However, as the Australian experience shows, rolling out such systems effectively cannot be done overnight.
  - Australia has been working towards the implementation of eHealth for a decade, and progress has been slow.
  - The central and final component of the plan is My Health Record, a centralised electronic health record-management system. It underwent major design changes in 2015-16 after negative feedback from users in the first years of the system's operation. The initial use of opt-in registration has been blamed for poor take-up rates, and opt-out registration is currently being trialled in parts of New South Wales and Queensland.
  - Other initiatives to improve particular aspects of individual health care with electronic means are also underway, and are at varying levels of completion.
- The technical inability of different parts of the health system to share information to improve patient care is an indication of how poor Australian health information systems can be.
  - IT system design and contracting place deliberate limits on interoperability. Some contract terms actively preclude proprietary systems exchanging data with other systems.
  - Health service providers face limited incentives in regard to interoperability and data transfer, and may have entrenched governance and service delivery models that do not place great emphasis on, or provide rewards for, data portability. Government procurement policies are similarly at fault.
  - The complexity of healthcare data means that standards development is a necessary part of any interoperability solution.
- In some areas, significant progress has been made. To continue this, government policies and practices must emphasise improved access to health data for both individuals and researchers, and improved data sharing between the participants in Australia's health system.

Health data includes a very diverse range of information (box D.1), collected in a wide variety of settings. GPs, specialists, allied health providers and pharmacists, as well as hospitals and other types of medical, diagnostic and pathology centres, collect this information. Data is also collected by various universities and research organisations through patient and practitioner surveys.[59] Other important sources of health information include population censuses and other surveys of the population, such as the ABS National Health Survey (ABS 2015).

Collected data is used either for direct patient care or for administrative purposes (such as receiving payments from Medicare). In addition, healthcare providers are required by law to provide information to certain registries, such as the cancer register and the immunisation register (AIHW 2016a).[60] Some of the data collected for administrative purposes feeds into the Commonwealth Department of Health's Medicare dataset (managed by the Department of Human Services) and/or into hospital datasets held by the state or territory Departments of Health.

Box D.1 What is health information?

The *Privacy Act 1988* (Cth) (OAIC nd) defines health information as:

(a) information or an opinion about:

  (i) the health, including an illness, disability or injury, (at any time) of an individual; or

  (ii) an individual's expressed wishes about the future provision of health services to the individual; or

  (iii) a health service provided, or to be provided, to an individual: that is also personal information;

(b) other personal information collected to provide, or in providing, a health service to an individual;

(c) other personal information collected in connection with the donation, or intended donation, by an individual of his or her body parts, organs or body substances;

(d) genetic information about an individual in a form that is, or could be, predictive of the health of the individual or a genetic relative of the individual.

In simpler terms, health information includes any information collected about a person's health (physical, mental or psychological) or disability, and any information collected in relation to a health service the person has received. Health information includes things such as: notes of symptoms, diagnosis, prognosis, treatment and prescriptions; medical history; test results and scans; appointment and billing details; dental records; individual healthcare identifiers; and genetic information. Information about a person's race, sexuality or religion can also be considered health information when it is collected by a health service provider (OAIC nd).

[59] Examples include the BEACH and MAGNET surveys and the Australian Rheumatology Association Database.

[60] Information is provided voluntarily to other registries, such as the National Joint Replacement registry.

This case study describes:

- the current landscape affecting health data (both in terms of policy and IT systems)
- the introduction of electronic health records in Australia and its effects on health data
- how health data is currently used to conduct research and inform policy
- ways to improve the availability and use of health data.

D.1 Health data – the policy and IT landscape

Despite the vast amounts of health data collected in Australia, current policy settings and the IT platforms used in the healthcare sector have caused the availability and use of health data to remain fraught with problems. Data flows across the health sector — and at times, within healthcare services — are inefficient (figure D.1). These issues affect both the provision of healthcare services to individuals and the ability of policy makers and researchers to understand and respond to public health trends (box D.2 presents the example of real time prescription monitoring, where the interplay between policy and technology issues can have substantial effects on individuals' wellbeing). This section looks at how such problems arise.

Policy frameworks for aggregate health data are fragmented

The institutional framework governing the collection of aggregate health data differs across states and territories and care settings. There is no single overarching framework for aggregating information about patients that is collected by GPs and specialists — the data available on primary and specialist care in Australia is mainly derived from administrative billing systems that reflect claims for Medicare reimbursements. Hospital data is collected and aggregated by States and Territories in accordance with various policies, which aim to support consistent data sharing between jurisdictions. However, significant data gaps still exist and effective data sharing is hampered by the multitude of data owners and custodians, and the inconsistencies in their authorising legislation and privacy regulation (or the absence of it) which makes approval processes for data sharing lengthy and complex, if they occur at all.

Hospital data

The health sector is one of only a few examples (along with early childhood and welfare) where jurisdictions have made a considerable effort to develop consistent policies on data collection and sharing. These policies apply primarily to data collected on the operation of hospitals (both public and private) and mental health services. However even in these areas, sharing is often not automatic and where it does occur, it is routinely or exclusively carried out by non-electronic means (and therefore involves delay).

Box D.2 A case study in inefficient handling of health data: real time prescription monitoring

In recent years, the adverse outcomes of prescription drug abuse, addictions and interactions have gained exposure as major public policy issues. All states and territories require pharmacists, drug wholesalers, and doctors to record controlled drug transactions in their own register; some jurisdictions also require pharmacists to directly report controlled drug dispensations (PGA 2015).

However, this reporting is manual and the information cannot be exchanged between participants in real-time. Similarly, Medicare operates a Prescription Shopping Information Service (PSIS), under which prescribers and dispensers can enquire (via website or telephone) whether MBS and PBS data indicate that a particular patient has been 'prescription shopping' (visiting many doctors to obtain more prescription drugs than they need). Much like the existing state and territory registers, the PSIS operates with a delay and relies on the prescribing doctor to detect drug-seeking behaviour (McDonald 2014a).

In 2008, Tasmania's Department of Health and Human Services (DHHS) received funding from the Commonwealth to develop and introduce a real-time controlled drug dispensation reporting system (McDonald 2012). The system was launched in 2011, and in 2012 the Commonwealth Government licensed the system for use in a national Electronic Recording and Reporting of Controlled Drugs (ERRCD) initiative (Department of Health and Ageing 2012), providing all states and territories with that licence. The 2010 Fifth Community Pharmacy Agreement (5CPA) between the Commonwealth Government and the Pharmacy Guild of Australia provided for the introduction of an ERRCD that could be used by both prescribers and dispensers (Dobbin 2014; Hore-Lacy 2007; Ogeil et al. 2016).

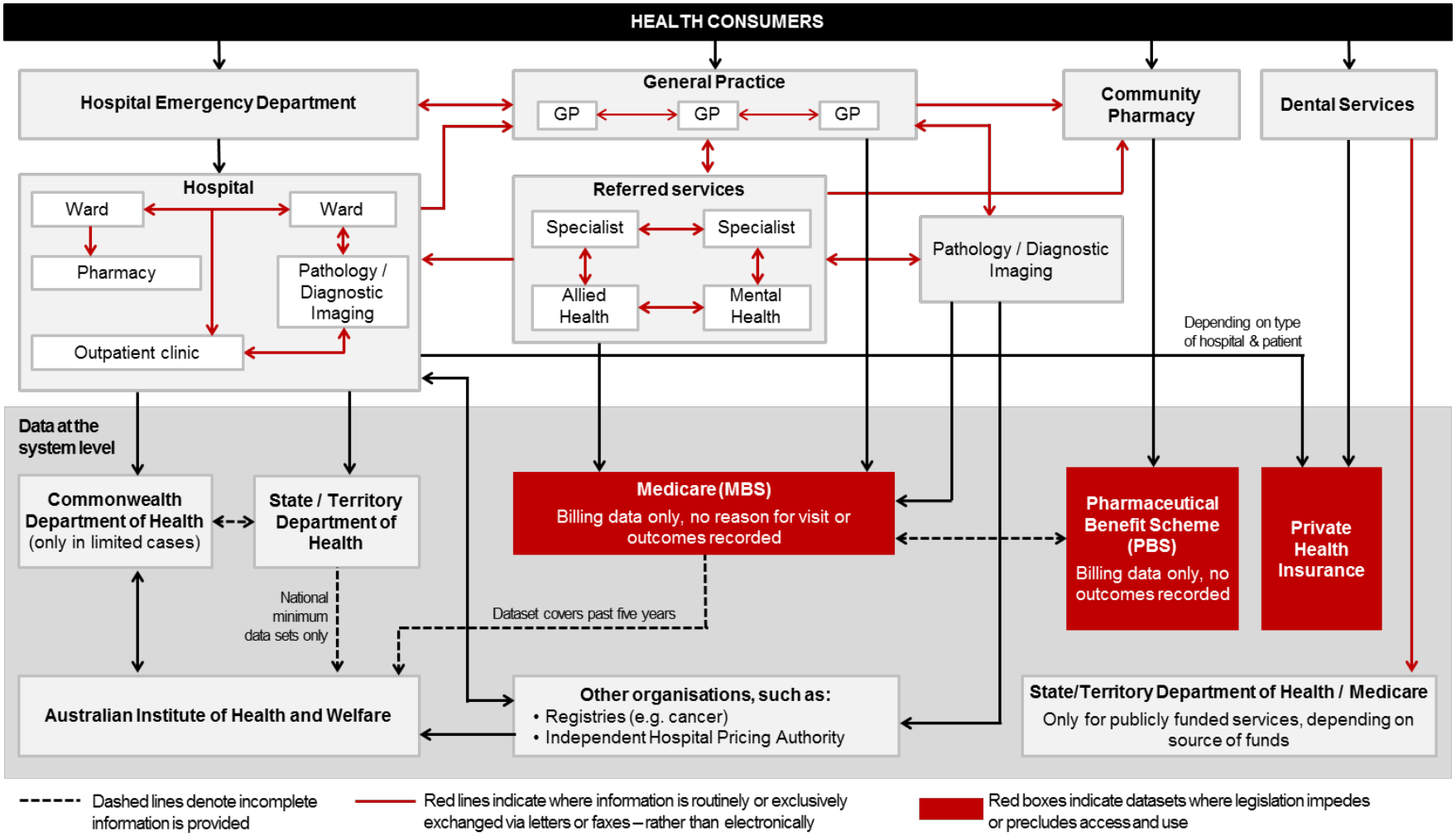
However, seven years after the formation of the 5CPA (Bruno 2013), and three years after the Medical Software Industry Association completed the development of nationally consistent specifications for reporting the dispensing of controlled drugs to state and territory health departments, the ERRCD has not achieved full roll-out in any state or territory apart from Tasmania (Pharmaceutical Society of Australia 2016).

The major reasons cited for this delay have been:

- a necessity for each state and territory to implement the system individually due to jurisdictional variations around drug regulations and classifications, and
- a need for most state and territory health departments to amend their jurisdiction's existing legislation relating to privacy and reporting requirements before the ERRCD system can be implemented (McDonald 2014b).

NSW Health appeared to be nearing an initial trial roll-out in late 2015 (McDonald 2015b), and achieved complete harmonisation of the state Poisons List schedules with the national Poisons Standard in August 2016; however, as of October 2016 there was no further information publicly available as to whether such testing had yet begun. Most recently, the Victorian Government allocated $30 million towards implementation of the ERRCD in the 2016-17 state budget (ABC 2016; Hennessy 2016).

Figure D.1 **Data flows in the Australian health sector**



The institutional arrangements for the collection and sharing of aggregate health data are described in the National Health Information Agreement (NHIA). First introduced in 1993 and most recently updated in 2013, the agreement is signed by the Commonwealth and all States and Territories, as well as the Australian Bureau of Statistics, the Australian Institute of Health and Welfare (AIHW), the National Health and Medical Research Council and other agencies.[61] The purpose of the agreement is to:

> …ensure the availability of nationally consistent high quality health information to support policy and program development, and improve the quality, efficiency, appropriateness, effectiveness and accountability of health services provided to individuals and populations. The Agreement promotes the efficient, secure, confidential and timely use of information across the complete lifecycle from development to use and supports reuse of information.…
>
> Nationally consistent health information also supports public discussion of health matters and research by health researchers and health professionals. The Agreement will therefore also improve opportunities for governments, health professionals, non-government organisations and consumer groups to share and use health information (COAG 2013, p. 6).

All signatories to the agreement have contributed to the development of the 'National Minimum Datasets', which include mandatory data collection and reporting at a national level. There are currently 16 National Minimum Datasets, covering a range of topics, such as government health expenditure, hospital admissions, public dental health, and mental health. In the case of datasets relating to hospitals, data is collected at each hospital from patient administrative and clinical record systems and regularly forwarded to the relevant state or territory health authority. State and territory health authorities provide the data to the AIHW for collation on an annual basis. The Independent Health Pricing Authority also collects data on hospital activity. Other types of administrative data are collected by State and Territory health authorities and provided to the AIHW (COAG 2013).

Under the NHIA, the AIHW was appointed as the body responsible for 'receiving, cleansing and disseminating information as a key national custodian of administrative health data collections and promoting national consistency of definitions and collections' (COAG 2013, p. 23). As part of this role, the AIHW uses established standards and methodologies to manage the data, including detailed metadata and consistent definitions of terms (AIHW 2007).[62] This distinguishes the health sector from many other parts of the economy and adds significant value to data collections. Extensive metadata and consistent definitions are vital in creating data linkages and enabling broader analysis of data.

While the AIHW is the data custodian, the ownership of the data remains with the original collecting jurisdiction, which can set publication conditions on the data collected. This

---

61 Numerous National Agreements have been signed between the all Australian jurisdictions, including the National Healthcare Agreement and the National Health Reform Agreement. The agreements include provisions for the collection and sharing of data between jurisdictions, covering a range of health and wellbeing topics (AIHW 2014).

62 The AIHW maintains and develops the National Health Data Dictionary and the Metadata Online Registry (METeOR), which are intended to improve the national consistency of data.

means that any data sharing or use requires the agreement of all contributing jurisdictions (COAG 2013). In effect, bureaucratic barriers and concerns about privacy prevent the use of existing health data collections to their full potential (section D.4).

For example, the Commission's Research Report on the Performance of Public and Private Hospital Systems concluded in 2009 that:

> The Commission encountered significant delays in accessing hospital related data beyond what could reasonably be expected to address privacy or confidentiality concerns.… The barriers to accessing hospital-related data are … wasteful because a substantial amount of information is currently collected at significant cost to governments and firms, and the potential broader public benefits from this are being unnecessarily curtailed (PC 2009, pp. 8–13).

As far as the Commission is aware, the situation has not changed in the past seven years.

Much of the AIHW-held data, while valuable and comprehensive, is not published openly; rather, AIHW more often publishes statistical overviews drawn from the raw data. Researchers may access some data, typically aggregated but sometimes at a unit record level, directly from the AIHW website. For data that is not published on the website, researchers must make a custom data request online, which can specify only one data collection, must be manually assessed by AIHW staff, and can take up to several months to be fulfilled. For linked datasets, researchers must additionally make an ethics approval request online, which attracts a fee of $600; data requests involving ethics approval are considered only on a quarterly basis by the AIHW Ethics Committee (AIHW 2016b).

Primary care data

The NHIA excludes most data relating to primary care (the treatment of non-admitted patients in the community, through GPs and other types of health care providers).[63] Therefore, unlike the data collected and managed by the AIHW, there is no single point of access for primary care data.

A number of organisations collect and publish data on primary care.

- Medicare Australia has a substantial administrative dataset, based on claims made by practitioners and patients. These statistics cover mainly the volume of services provided and the benefit paid for each Medicare item; there is no data on patient outcomes. High level summary statistics are available online.
- The Department of Health publishes statistical reports based on administrative data collected as part of the Pharmaceutical Benefit Scheme.
- Up until mid-2016, the National Health Performance Authority published reports on primary and hospital care, using a variety of existing data collections (NHPA 2015b).

63 Data on primary care provided to Indigenous people is included in the NHIA.

- There are a number of research bodies that have developed large datasets on primary care. For example, The General Practice Statistics and Classifications Centre is a collaboration between the University of Sydney and the AIHW. The Centre ran the 'Bettering the Evaluation and Care of Health' (BEACH) program for 18 years, surveying GPs and monitoring the characteristics of practitioners and patients, the reasons people seek medical care and the outcomes of GP consultations (FMRC 2016).

In recent months, there has been substantial change in the funding arrangements for primary care data collection. The National Health Performance Authority stopped operating on 30 June 2016. Its roles were transferred to the AIHW and the Australian Commission on Safety and Quality in Health Care. The BEACH program has been defunded and its data collection has ceased (historical data is still available for purchase by researchers), and funding for other bodies conducting research into primary health has also been cut (FMRC 2016; NHPA 2015a).

There has not been a clear directive from the Department of Health on the future of primary care research. The overarching Primary Health Care Research, Evaluation and Development Strategy was last updated in 2014 (Department of Health 2014). Stakeholders have voiced concern about reduced research capabilities, given recent funding decisions (Russell 2016).

IT systems limit the ability to share individuals' health information

The framework of information technologies and the related standards chosen to underpin individuals' medical records are also critical to improved data collection and access.

However, evidence provided to date to the Commission's inquiry suggests that in the Australian health system there is a diversity of IT platforms, and there are aspects of IT system design, procurement and contracting that significantly limit data sharing. This affects GPs and other small scale health service providers, hospitals and large scale administrative data collections. Health IT therefore remains a long way away from 'plug and play' solutions, one-to-many communication and real time exchange of data.

The systems and standards in place

Information technology is now widely used in the Australian health system. Specialised health IT systems are apparent within the current system in a range of contexts, including GP offices, hospitals and some specialist clinics.

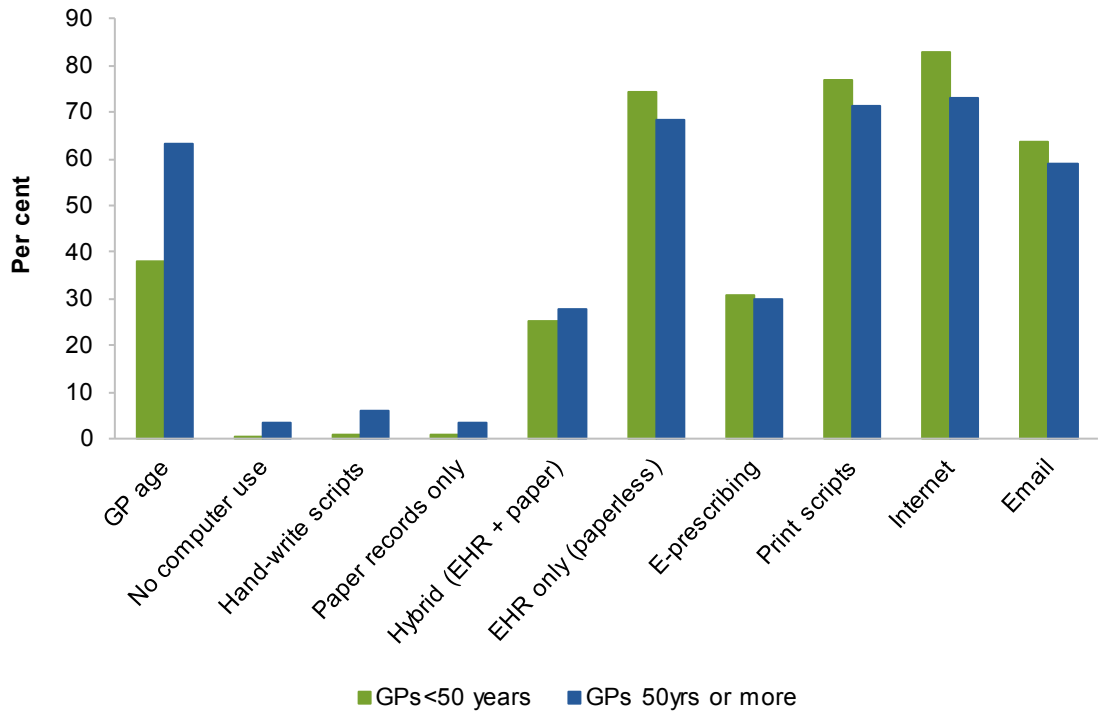
Current systems

In *GP offices* within Australia, there is now widespread use of IT overall and various IT platforms being used, although patterns of usage vary somewhat by age of practitioner (Henderson et al. 2014) (figure D.2). Recent survey work found that:

- only 4% of GPs did not use a computer at all for clinical purposes.
- 98% of GPs were producing prescriptions electronically (ePrescribing or printing scripts).
- 71% reported they used electronic medical records exclusively (that is, were paperless).
- 26% reported maintaining a hybrid record where some patient information is kept electronically and some on paper records. (Britt, Miller and Henderson 2015, p. 34)

This IT penetration stands in contrast with use by allied health providers, specialists and surgeons, with one estimate from 2012 that only around 37% of specialists and 22% of surgeons relied on computerised patient records (House of Representatives Standing Committee on Health 2016, p. 104).

Figure D.2 **IT use and the age of Australian GPs**



Source: (Henderson et al. 2014).

Among the different types of software used by GPs in Australia, two products, Best Practice and Medical Director, have a significant share of the market. The various systems used by GPs are based on inconsistent structures and standards, data elements and clinical

terminologies. This makes data transfers between doctors very difficult, and also creates challenges for data integration and linkage for research purposes. Unlike other types of medical software, medical records management systems are not regulated by the Therapeutic Goods Administration (Gordon, Miller and Britt nd).

In *hospitals*, both large generic IT systems, and more bespoke systems tailor-made for individual wards, hospitals or hospital groups, are apparent. With regard to the former more generic systems, there are large multinational providers, such as Joan Software, CERNER and EPIC, that provide comprehensive solutions for hospital IT within Australia and in many other countries. While these systems often have a generic structure, they will in many cases also have tailored elements designed to work with the particular structures and work flows that exist in the facilities where they are employed (McDonald 2015a).

The role of standards

While IT systems play an important part in facilitating (or blocking) data transfer, it is also important that once data is transferred across entities, it is able to be interpreted and used by the receiving party. What remains highly problematic in this regard are the use of agreed practices and terms for data interpretation (box D.3).

Box D.3 **Six requirements for health information exchange**

In a 2014 presentation, health interoperability expert Grahame Grieve outlined six main requirements for the exchange of health information:

- *Transmission of data*: a transmission channel between sources so meaningful symbols can be exchanged.
- *Common terminology*: a common set of terms with meanings that both parties understand.
- *Identification policies*: some way to identify instances of things that are being talked about.
- *Information structures*: a common method to assemble the terms or words into a coherent larger structure.
- *Behavioural models*: a conversation protocol about who says what when, and then what happens next.
- *Common understanding*: a common understanding of the context in which discussion is taking place.

Source: As quoted in McDonald (2015a).

There has been some recent progress in improving design and developing standards; however, this progress notwithstanding, significant challenges remain in this area, and there is little evidence in Australia of government efforts to address the need for interoperability through procurement processes.

Market practices, design elements and management practices limiting data quality and data sharing

Procurement and design

There is evidence that current market practices in health around systems procurement and contracting, in combination with proprietary systems design, are placing very significant limits on the improved use of data.

Where vendors of medical record-management systems demand hospitals sign 'no sharing' contracts before supplying the system, this can act as a hurdle to effective data linkage. Hospitals entering into such agreements may find that they cannot transfer data between two different systems without breaching contract. This can occur even when one vendor does not have an exclusive contract for an entire hospital, such that a hospital can be prevented from electronically sharing data between its wards.

Interoperability, across and even within systems offered by vendors, is also reportedly low. This means that the diverse range of IT products in use cannot operate in anything like a 'plug and play' way presently, and this can apply either within a given clinical setting or across settings. In many cases, this means hospitals and other health providers must devote significant resources in an attempt to bridge the gaps between systems (box D.4).

This problem is not confined to Australia but, rather, is apparent across the world. For example, US-based authors Cantwell and McDermott (2016, p. 1) state:

> Unfortunately, the vast majority of medical devices, electronic health records (EHRs), and other IT systems lack interoperability … Various systems and equipment typically are purchased from different manufacturers, and each comes with its own proprietary interface technology.

Broader systems governance is likely to be critical here. Some parties have suggested that in addition to supply side elements, such as inclusion by manufacturers of greater interoperability, demand side 'nudges' are needed. This could include, for example, the regulation of data management system procurement, whereby governments could mandate interoperability as part of their requirements for purchasing systems. The US Government worked with IT providers to promote interoperability as part of ongoing health reform, which puts a substantial emphasis on improving data access and use (box D.5).

Will technological innovations improve matters (and make current systems obsolete)?

While recent developments in systems design and standards will go some way towards solving the interoperability problem, it also appears likely that new technologies, being developed or enhanced at a rapid pace, may also provide a significant part of the eventual solution.

Box D.4 Some recent examples of interoperability problems in Australian health

Several recent examples indicate that interoperability problems, combined with existing management practices, continue to be a cause of concern.

A Parliamentary Committee in Western Australia reporting on problems encountered at the Fiona Stanley Hospital stated:

> The AMA reported that there had been significant difficulties with the implementation of the Intensive Care Unit's (ICU's) Clinical Information System (CIS), whilst the Health Services Union indicated that the ICU CIS was not compatible with the systems in use on the general wards. According to the HSU, this meant that patient's records must be printed and scanned when they transfer from the ICU to a general ward. The system does not currently provide the ability to export ICU medical records to BOSSnet, although this is an upgrade that is being considered. This manual paper-based process was confirmed to the Committee during its visit to the hospital.

The 2013 Ministerial Review of Victorian Health Sector Information and Communication Technology discussed an example of the results of poor interoperability:

> A similar issue arises when any electronic medical records (EMR) system receives a discharge notification in that it must, at that point, automatically discontinue the current medication chart and commence a new chart for the next admission. This means that when the patient is subsequently re-admitted from ward-based care into a different category, all previous medications will need to be reviewed and rewritten or copied across by the treating doctor or pharmacist. Clinicians have expressed concern to the panel regarding the handling by current EMR systems of these points of transition for care arrangements.
>
> The panel understands that concerned hospitals, the department, the clinical system vendor and other software vendors are in discussion regarding how best to improve the current processes by developing software solutions to decrease these 'points of hazard'.

Sources: Education and Health Standing Committee (Western Australia) 2015, p. 23; Department of Health & Human Services 2013.

Commentators such as Cook and Topol (2014) have emphasised the considerable potential of smartphones and other devices to enhance data transfer within medical contexts. While these devices may still require interoperability with larger systems, such as those in place within hospital settings, it does appear that they bring with them the prospect of rendering some parts of larger and less sophisticated networks as obsolete across time.

What also appears likely is that the more widespread and seamless transition of health data will still be accompanied by concerns around the maintenance of privacy regarding parts of an individual's health data. The recent case in the United Kingdom of the transfer of data between the NHS and Google DeepMind (Hodson 2016) illustrates some of the complexities around health data transfer and the variety of views around the net benefits of such practices.

Box D.5 US health policy – implementing data-driven reform

The US healthcare system is undergoing substantial changes, driven primarily by the introduction of the Affordable Care Act in 2010. The Act puts an emphasis on improving access to both personal health data and information about service providers, and encouraging sharing of data between healthcare providers.

The implementation of the Act, and other related legislation, includes policies that address several aspects of access to and use of data. These policies introduced financial incentives for providers to implement electronic health records, and share their information with other healthcare services. It also puts an emphasis on interoperability — unlike Australia, where each individual will have one eHealth record, in the US each healthcare provider maintains an electronic file about a patient. The systems managing these files use commons standards, so that information can be shared and aggregated (DHHS (US) 2016b). Examples of key policies include:

- promoting the implementation of electronic health records. In 2009, the Health Information Technology for Economic and Clinical Health (HITECH) Act introduced incentive payments to hospitals and other healthcare providers that implemented electronic health records. By 2015, nearly all US hospitals had implemented electronic health records, which also allow secure sharing of information (Henry et al. 2016).
- working with IT providers to improve interoperability. In 2016, the US Government announced it had reached an agreement with the major developers of IT systems for the healthcare industry to use agreed standards. These standards will allow healthcare providers to share individuals' health information and help consumers to access their health information and share it as they choose. This is intended to end the practice of information blocking, where fees and charges imposed by IT developers make it difficult for health service providers to access or share information (DHHS (US) 2016a).
- support data sharing between healthcare providers, and promote the collection of data on quality of health services. Providers can receive bonus payments from the government when they work together to share information and improve patients' health outcomes. At the same time, payments are reduced for providers who do not report data on quality improvement measures (CMS 2015, 2016).
- enhance individuals' access to their own health data. Beginning with the 'Blue Button Initiative', where the US Government allowed veterans to download their health record, health consumers in the US now have the right to access their personal health information and share it with a third party. In 2014, nearly all US hospitals allowed patients to view their records online, and over 60% offered them the option to download or transmit this information. Nearly half of all doctors also offered patients the ability to download their records (ONC 2015). The US Government is now supporting the development of applications that will allow patients to bring together health information held about them by different providers (HHS nd).

Summing up

As this section has discussed, there are many and varied operating systems observed across jurisdictions currently in Australia, and a diversity of procurement policies and practices. In many cases, the systems in place were implemented as part of separate purchasing

decisions, with little coordination. Providers of technological systems may also have a vested interest in limited or no commonality, and this has only served to further exacerbate the problems observed.

This present state of play ensures that primary care remains disconnected from hospitals and specialists, that hospital wards are often disconnected from each other, and that doctors are often not sharing (or able to share) sufficient information with nurses and other health professionals. It also has implications for the ability of policy makers to understand the needs of the health system, and the work of researchers using health data.

Interoperability is a critical aspect where, some encouraging recent developments notwithstanding, much remains to be done. In the present Australian system, interoperability of IT systems in health, and particularly of eHealth records, has been a persistent issue over time, and remains one of the key stumbling blocks to the introduction of many aspects of eHealth (see below). In many respects it appears that, while the transfer of health data across systems has improved greatly, interpretation of what such data actually means 'on arrival' also remains highly problematic.

D.2 The development of eHealth in Australia

Now that most healthcare providers engage in some form of electronic recordkeeping, the main focus of current eHealth policy in Australia is not the digitisation of recordkeeping systems themselves, but the creation of an Australia-wide database for electronic health records (EHR) that can be accessed by the patient and by any healthcare provider authorised by the patient.

A large variety of EHRs exist worldwide, operated by both the public and private sectors. Depending on the specific design, an EHR may include a range of data, such as demographics, medical history, medication and allergies, immunisation status, laboratory test results, and radiology images. In a highly linked system, such a record might incorporate billing or claiming data (such as Medicare data), pharmaceutical subsidy data, or geographical data (so that environmental features known to impact health, such as air pollution, might be factored into considerations of the patient's health).

Currently, Australia's centralised electronic health recordkeeping system is known as 'My Health Record' (MHR) — previously named the 'Personally Controlled Electronic Health Record' (PCEHR). The electronic health record, along with other elements of eHealth in Australia (box D.7), was designed by the National Electronic Health Transition Authority (NEHTA), in conjunction with the Australian Government.

Why use an electronic health record system at all?

The benefits of electronic health information management (or eHealth) have been widely recognised. Those benefits can be broadly categorised as: quicker, easier, cheaper access to

a patient's accurate medical history by healthcare providers and by the patient themselves; increased ability to transfer more of a patient's files between healthcare providers; and access to accurate population health data by policymakers (PC 2015). Other benefits, such as extending the ability to access specialised healthcare to remote residents, may also arise from an eHealth system (RACGP 2011).

An EHR can give healthcare providers better access to patient records

Given the specialisation of medicine, the localised electronic patient records (in separate IT systems acquired by each health professional for their practice) held by individual healthcare providers such as GPs are likely to be incomplete as they do not systematically include procedures, referrals, prescriptions and test results added by other health providers (such as hospitals or specialists) (Jolly 2011).

With a widely used centralised database, all healthcare providers could access a patient's medical history digitally and instantly (subject to the patient giving consent, and to their privacy settings, if these allow information to be hidden). Such a system would remove the need to rely on the patient to provide this information, either in paper form (which carries the risks of loss or damage) or by memorising it (which risks inaccuracy). This is particularly important in emergency situations. In Australia, AIHW research indicated that up to 18% of medical errors were due to inadequate patient information (AIHW 2002).

Depending on the doctor's choice of individual record system, the centralised electronic health record may not become the single, all-encompassing record for an individual's health events, that would enable health providers to function with the centralised record as their only source of information (Reeve, Hosking and Allinson 2013). Localised records may well continue to exist at each health provider organisation (DoH 2016d), particularly if the centralised record is limited in the information it can contain (as in the United Kingdom, box D.6) or if it allows information to be hidden (as in Australia).

An EHR can allow patients to access their own records

Currently, patients generally do not have access to their record outside of consultations with their doctor. Similarly, there is often no written record of test results, medical procedures or prescriptions that patients can share with other health professionals. Ideally, patients would be able to access their own medical history wherever they are, instantly, at no cost. Not only would this benefit the patient's knowledge of their conditions, but the patient could consequently provide this history to other healthcare organisations without incurring an expense or needing to wait.

Box D.6 The introduction of electronic health records in the UK

The United Kingdom's National Health Service (NHS) commenced moving towards a centralised electronic record in the early 2000s. Though the implementation of EHRs was marred by errors and stop-starts affecting much of the NHS National Program for Information Technology (Committee of Public Accounts 2013), by June 2015 more than 96% of the population had an electronic Summary Care Record (SCR) accessible by health service providers both public and private; furthermore, more than 97% of GPs could provide patients with the ability to access online the data held within their own SCR (Glick 2015). Government forecasts aim for complete coverage of the population – 'a paperless NHS' at the point of care – by 2020 (Parkin 2016).

The patient's consent is required for health practitioners to access an SCR, with exceptions provided for emergencies. A notable feature of the SCR is its brevity; the record contains only data deemed crucial to avoid potential treatment causing harm: current medications, allergies, and any previous adverse reactions to medication (NHS England 2016). Some jurisdictions, such as Cumbria and London, have begun to develop more detailed electronic data sharing arrangements at a local scale, with a focus on rapid transfer of patient records between hospitals and other urgent care providers (Healthcare Gateway 2013; London Connect 2013).

Data extraction and collation from NHS records has taken place on an increasing scale since 1989 when the Hospital Episode Statistics (HES) were launched (Presser et al. 2015). Currently, identifiable patient data is automatically extracted from hospitals into the Health and Social Care Information Centre's (HSCIC) 'safe haven' database, and is then used to generate aggregated statistics published by the NHS and disseminated to researchers.

An expansion of this approach for GP data, named care.data, was introduced in 2013. Care.data operated on an opt-out basis and involved individual patient data being uploaded from GP surgeries to the HSCIC database, where it was linked to HES data and could be disseminated. Aggregated data could be used by researchers or published, while de-identified individual data could only be made available to specified parties such as health providers and Public Health England (Presser et al. 2015). Issues with the program's impact on privacy – especially with regard to patients opting out of having their data collected – resulted in care.data being suspended in February 2014; it was intermittently recommenced and paused again in the intervening two years before being permanently cancelled in July 2016 on the back of a commissioned review of consent and opt-out models (Evenstad 2016). However, it is expected that data from GP surgeries will continue to be shared using other systems.

Further, because medical records are the property of the doctor, hospital or practice that created the documents, patients incur a cost to obtain a copy of their record or to have a copy provided to another healthcare organisation. Though at law this is limited to the 'reasonable expense' incurred by the practice in accessing, copying and providing those records to the patient (per Australian Privacy Principle 12.78–12.81), in practice there is enormous variation. While in some cases, copies are provided free of charge (for example, by public hospitals in Western Australia), there are also instances of very high fees. During the 2014-15 financial year, the Office of the Australian Information Commissioner (OAIC) received a complaint involving a medical centre charging $684 for a copy of a patient's file. After the OAIC ordered conciliation, this was reduced to $66, based on the actual costs incurred to produce the copy (OAIC 2015a).

An EHR can enable easier transfer of records between healthcare providers

At present, patient data does not tend to travel in a secure and systematic way between healthcare providers (see section on IT systems above). This is of particular concern to individuals suffering from chronic illness. The AIHW estimated in 2015 that roughly half of all Australians have a chronic disease, and about 20% have two or more. In such circumstances patients will often be treated by multiple specialists, and the absence of record centralisation can result in diagnostic duplication and confusion about the interaction of different treatments:

> [C]hronic illness requires close monitoring and ongoing management across an entire team of care professionals. … But healthcare providers largely operate in disconnected silos, hindering continuity of care. Doctors often do not know what medications and tests have been given to patients by other doctors, even when they are members of the same care team. It is even more difficult to bring relevant medical knowledge to the point of care, to create integrated care plans, to monitor a patient's progress against the care plan, or to alert care providers when a patient's condition requires intervention. (Georgeff 2007, pp. 6–7)

Georgeff (2007) cited figures estimating that improved information sharing and care plan management for sufferers of chronic disease would produce direct healthcare savings of $1.5 billion per year, based on 2007 levels of chronic disease prevalence.

Duplication of testing, which could be minimised through the effective use of EHRs, does not affect only patients with chronic health conditions. The National eHealth Strategy prepared for the Commonwealth Government in 2008 (Deloitte 2008) cited studies in hospital environments that found between 9% (CBO 2008) and 17% (Kwok and Jones 2005) of pathology and other tests were unnecessary duplicates. Further studies cited in the National eHealth Strategy found that duplicate test alerts could cut a hospital's absolute number of tests by up to 25% and reduce waiting time for radiology results by between 24 and 48% (Chaudhry et al. 2006).

With the use of an EHR that permits the digital transmission of vital signs, treatment information, diagnostics and pathology, the transfer of information provider-to-provider is faster, easier and less susceptible to mistakes — for example, the real-time transfer of a patient's entire medical file is possible when making a referral to a specialist.

An EHR can assist with the efficient collection of population health data

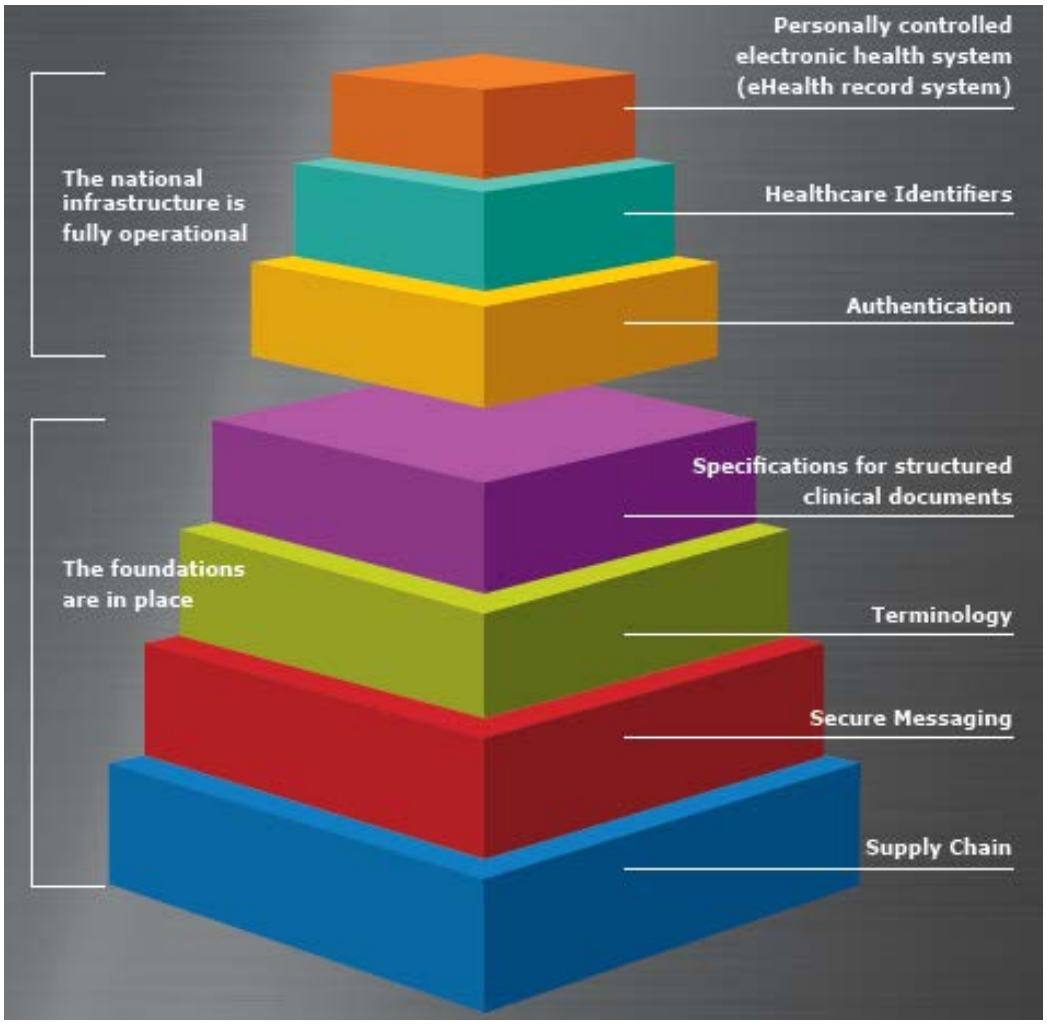
There are also potential benefits to public health research arising from universal electronic health records. For example, if clinical terminology is standardised across a jurisdiction and accurately coded into the EHR itself, the resultant statistics could provide a far more accurate picture of disease prevalence and treatment efficacy than either the Medicare database or GP reporting could deliver. For issues where population-wide statistics are especially important, such as immunisation and epidemiology, electronic health records should give governments more accurate information to inform policy decisions.

Major eHealth policy components in Australia

Commonwealth Government health policy had contemplated a centralised electronic healthcare record since the mid-1990s (Jolly 2011). Trials of the first EHR commenced in 2002; however, evaluation concluded that the system had been ineffective and the policy was scrapped in 2005 (Dearne 2010; Jolly 2011) everywhere but for the Northern Territory (eHealthNT 2011), which has retained its own EHR to the present day.

Between 2006 and 2012, a wide range of systems and standards were developed and implemented by NEHTA in collaboration with various Government agencies and industry bodies, in order to enable to eventual introduction of the EHR (figure D.3).

Figure D.3 **The building blocks of eHealth in Australia**



Source: NEHTA (nd).

NEHTA was created and jointly funded by the Australian Government and all State and Territory Governments in 2005, and operated until 2016, when it was disbanded. The 2013 Review of the PCEHR was highly critical of NEHTA, stating that the agency did not have the confidence of healthcare providers or consumers (Royle, Hambleton and Walduck 2013). NEHTA was replaced by the Australian Digital Health Agency (ADHA) in July 2016.

How to identify patients? The creation of Individual Healthcare Identifiers

A major milestone in creating the infrastructure for the EHR was the introduction of Individual Healthcare Identifiers in 2010.

Initial legislative development of a centralised electronic health record scheme (commencing in 2000) planned for Medicare numbers to be used as unique patient identifiers. However, various stakeholders opposed the scheme, citing concerns about privacy and critiquing the accuracy and integrity of the Medicare database (Chapman 2002). Audits of the Medicare Consumer Directory by the Australian National Audit Office (ANAO) in 2004 and 2014 lent weight to the claim that the database was not accurate enough to be used as a basis for identification. For example, the 2004 report concluded that a number of records were probably for people who were deceased, that data in some records indicated that a person had enrolled in Medicare before they were born, and that up to 500 people had duplicate entries (ANAO 2004). Similarly, the 2014 audit made reference to at least 18 000 possible duplicate entries, active records for customers without entitlements (which can result in payments to ineligible persons), records which had customer information inconsistently, inaccurately and incompletely recorded, and episodes of two different customers' records becoming intertwined by accident, giving rise to privacy and clinical safety risks (ANAO 2014).

As a consequence of the Medicare database's insufficient accuracy, Individual Healthcare Identifiers (IHIs) were launched in 2010, with the aim of creating a more accurate database of unique identifiers. The IHI Service is operated by Medicare and is designed to allocate a unique number to all Australians, foreigners seeking healthcare in Australia, and healthcare providers. While healthcare providers will attach a patient's IHI to their care record, and many hospitals use the IHI for patient wristbands to ensure correct identification and treatment, the IHI database itself does not contain clinical information – only identifying information such as name, date of birth, sex and addresses (Department of Health 2015b).

Before its implementation, the IHI Service was subject to three Privacy Impact Assessments over four years, which recommended several methods of achieving a level of privacy protection sufficient to comply with the *Privacy Act 1988*. The *Healthcare Identifiers Act 2010* therefore heavily restricts access to the IHI database and the use or disclosure of a person's IHI. Most patients today would probably not know their IHI.

Despite the stated intention for the IHI system to be more accurate than Medicare numbers, many IHIs are assigned on the basis of existing Medicare records. However, since the IHI

system is permitted to draw on multiple data sources, this is not the case for *all* IHIs — for example, veterans have their IHIs assigned on the basis of their Department of Veterans' Affairs numbers, and foreigners who do not have a Medicare number typically have their IHI assigned on the basis of their passport and/or visa documentation. While the IHI system design does include protocols for avoiding record duplication, given the shortcomings of the Medicare Consumer Directory discussed above, there is a possibility that the IHI database also contains some of the same flaws as the Medicare database.

Australia's first eHealth system was hampered by low community awareness

Australia's first nationwide electronic health record management system — the Personally Controlled Electronic Health Record (PCEHR) — was introduced by the Australian Government in July 2012 on an opt-in basis. The PCEHR was not designed to replace the existing records maintained by healthcare providers, but rather to be an additional, central repository for the most important information that all healthcare providers treating a patient could easily view online with the patient's consent (or without, in an emergency).

Implementation by healthcare providers was discretionary, though incentive payments were offered for GPs to procure PCEHR-compatible software (AMA 2012). Registration rates by GPs substantially exceeded policy targets in the first two years of the PCEHR's operation. However, only about 1.7 million people signed up by the end of the 2013-14 financial year. Several consumer surveys showed that there was very low awareness of the PCEHR in the community (Deloitte 2014; Partel 2015).

An evaluation of the program in 2013 concluded that there was 'overwhelming support' for continuing the implementation of a consistent electronic health record for all Australians, but that a major change in approach was needed (Royle, Hambleton and Walduck 2013, p. 13). Key recommendations included:

- Transition to an opt-out model for all Australians.
- Conduct an education campaign for individuals and clinicians about the impact of the change to an opt-out process, and the strength of security and privacy in the system.
- Establish a clinical systems capability group within the relevant department, to improve medical usability and work towards integration with all health systems and platforms.
- Alter the eHealth Practice Incentive Payment (ePIP) from a one-off registration payment – link ongoing ePIP funding to meaningful usage of My Health Record.

Re-creating eHealth: The introduction of My Health Record

In 2015, the Australian Government accepted many of the review's recommendations and promised a more user-friendly interface, better alignment with clinical workflows, and greater levels of training and support for healthcare providers (Ley 2015b). The *PCEHR Act 2012* was amended in November 2015 to become the *My Health Record Act 2012* and

reflect these major changes (Ley 2015a). Some technical features of the previous system were rolled into the new My Health Record (box D.7).

The most controversial of the recommendations – the move from opt-in to opt-out registration – was not unequivocally accepted, with the Government deciding to trial opt-out in selected areas of Australia before committing the entire country to the change.

Box D.7 My Health Record — putting patients in control

My Health Record (MHR) operates with a web browser-based interface for both patients and healthcare providers, found at www.myhealthrecord.gov.au. Some clinical software vendors have software that is conformant with My Health Record, meaning that healthcare providers can access a patient's My Health Record directly from their clinical software (NEHTA nd). The patient interface presents six sections for healthcare-related information: clinical records; prescription and dispensing records for medicines; childhood development; Medicare claims history; advance care planning and information added by the individual about allergies; adverse reactions, and current medications (Department of Health 2016c).

Despite the title, MHR does not necessarily give a patient access to their full health record. The most likely scenario is that some information is on MHR, but more is on each doctor's localised system, unless doctors choose to update patients' MHRs — not all clinical systems are able to automatically 'push' data into the MHR database. Some hospital data may be included in MHR but the failure to require interoperability among systems will generally mean that much is not.

Individuals are able to control which healthcare providers can access the information in their My Health Record and the content stored in their health record.[64] All documents and information stored on an individual's My Health Record can be hidden or completely removed, both of which prevent the information from being accessible by any users, even in an emergency. Hidden or deleted information and documents can also be rest ored by the individual. Individuals are not able to edit information except for the Personal Health section, even if they think that a clinical document may be incorrect, but are able to remove it (Department of Health 2016c).

This level of individuals' control over the information contained in their health record has been subject to much debate. Parties concerned with privacy place a very high value on the ability for individuals to control the presence of, or access to, information that they do not wish to be known by other healthcare providers (APF 2011; Australian Privacy Commissioner 2011). Meanwhile, there is concern that this precise feature of the record may result in healthcare providers relying on incomplete information, reducing its efficacy (AMA 2013; Jolly 2011). However, the National Health and Hospitals Reform Commission concluded that the personal control feature would, at worst, not render My Health Record any riskier than the status quo:

> [T]he concept of patients controlling access to their own health information may be confronting … [Patients] always had the right to choose whether or not to share some or all of their information with health professionals … (and some patients may choose to access different practitioners at different times because of the sensitivity of some health information) – this occurs regardless of whether we are living in an 'e-world' or relying on other forms of communication. (NHHRC 2009, p. 129)

64 Section 64 of the *My Health Record Act 2012* permits healthcare providers and the System Operator to collect, use and disclose information in an individual's My Health Record under certain emergency circumstances, if it is unreasonable or impracticable to obtain consent from the healthcare recipient or their authorised representative.

The privacy and permission issues arising as a result of an opt-out model were considered by the OAIC (2015b) in a submission on the amending legislation discussion paper. The OAIC stated that active and express consent was a crucial component of a recordkeeping system. So, while an opt-out model of permission was not necessarily incongruent with the Australian Privacy Principles, the model was to be implemented in the most privacy-enhancing way possible (OAIC 2015b, p. 4). This would include giving individuals maximum opportunity to exercise their right to opt-out, and ensuring existing personal controls over information within the record were not diminished.

A Privacy Impact Assessment for the opt-out trial was also conducted. In that report, the authors raised many of the same concerns as the OAIC; in particular, several recommendations focused on how the Government could maximise effective communication with participants of the opt-out trial (Minter Ellison 2015, pp. 92–96).

In early 2016, the Minister for Health announced plans for a trial to take place in Northern Queensland and the Nepean Blue Mountains region (Ley 2016). Communications regarding the trial took place from March, with records first created in June. The Explanatory Memorandum for the amending Act states that these trials may last for up to nine months (Ley 2015a, p. 11).

Stakeholders have indicated that there is still work to be done on the MHR system's clinical usability. In particular, there are differences in nomenclatures between doctors' preferred systems and the MHR system — the SNOMED-CT AU clinical terminology is apparently especially complicated, resulting in doctors retaining the use of various other clinical terminologies and coding systems for medical terms. Combined with a lack of full interoperability between some GP software and the MHR system (such that the doctor will not be automatically informed if the patient has an MHR, but will need to specifically search the MHR database for it), this may be impeding the creation of a single 'source of truth' patient record.

D.3 Using health data in research and policy development

The data collected by healthcare providers across Australia, and generated from administrative data sets such as Medicare, is used to produce myriad health indicators.[65] Policy development and evaluation is often based on such indicators.

Numerous organisations in many jurisdictions produce and publish hundreds of indicators that reflect various aspects of the Australian health system, as part of national agreement reporting (for example, the Productivity Commission's Report on Government Services

[65] A health indicator is defined as a 'key statistical measure selected to help describe (indicate) a situation concisely, to track change, progress and performance, and to act as a guide to decision making' (COAG 2013, p. 13).

publishes nearly 400 different indicators related to the health system each year). Measuring health system performance through indicators has a number of potential benefits:

- improving the accountability and transparency of service provision
- measuring the effectiveness of policies over time, and providing benchmarks for quality improvements
- offering the community information required to compare some aspects of service providers (for example, through myhospitals.gov.au) (AIHW 2014).

The extent to which these benefits are realised is questionable (Nous Group 2014). For performance indicators to be a valuable resource, they must be derived from data that is complete and up to date. This is not always the case — many types of health data are not collected, are inconsistent between jurisdictions, or are incomplete (PC 2015).

From the point of view of researchers and policy makers, it is the underlying datasets that are likely to yield much more insightful findings. These datasets can be used for:

- identifying the causes of disease, the prevalence of risk factors and identifying populations at risk;
- protecting public safety, especially with regard to infectious disease, but also in relation to prescription medicines, medical devices and environmental hazards;
- needs assessment, monitoring and evaluation of services, with a view to providing an optimum performance of healthcare systems; and
- improving the quality and safety of care in hospitals, practitioner's offices, clinics and other healthcare settings. (OECD 2013, p. 22)

Australia has some large scale health data collections, which have been used to answer important questions in different areas of medicine (box D.8). But researchers and policy makers are often constrained in accessing and using many administrative datasets, which can provide further insights if they were more widely accessible. A recent example is the Australian Atlas of Healthcare Variation, which was able for the first time to present variations in specific medical procedures across different parts of Australia. This analysis presented important findings, but it was limited by lack of data and restrictions on linkages (ACSQHC and NHPA 2015).

Box D.8 **Examples of large scale health data collections in Australia**

- The Busselton Health Study, one of the oldest of its kind in the world, commenced in 1966, when a local GP decided to collect detailed health information, including blood samples, from the entire population of Busselton WA. At the time, the town had about 6000 residents, and over 90% agreed to join the study. The data collection was repeated every three years until 1981. Since then, the residents in the area have continued to participate in smaller surveys, run through the Busselton Population Medical Research Institute (Busselton Population Medical Research Institute 2014).
- 45 and Up is a large-scale health study, involving over 250 000 people in NSW. The study, based at the Sax Institute, started recruiting participants in 2006 with the aim to create a comprehensive picture of health outcomes for people aged 45 and over. The data provided by participants is linked to administrative collections, such as cancer registries, through the NSW Centre for Health Record Linkage (45 and Up Study Collaborators 2008). The study aims to create a biobank, by inviting all current participants to provide a blood sample. To date, only about 1% participants were asked for blood samples, and response rates were relatively low (Banks et al. 2012; Sax Institute nd).

The Medicare dataset – an underutilised asset

The largest administrative dataset relating to primary care is the Medicare Consumer Directory, which contains all Medicare customer records. In 2014-15, there were 24.2 million people enrolled in Medicare (DHS 2015). Notwithstanding its data quality issues (discussed in the eHealth section above), the Medicare Consumer Directory is a very high value dataset that is underutilised (Centre for Big Data Research in Health, sub. 21, SSCH 2016).

In a recent Senate Inquiry into health policy, the Department of Health identified a long list of potential benefits from the use of Big Data (which would include the Medicare dataset):

- Better information to inform the government's policy decisions
- A clearer picture of the real experiences of patients as they engage with the health system
- A better understanding of what works, how well, for what cost, and in what circumstances
- Earlier detection of trends – both positive and negative
- Earlier detection of anomalous behaviour and deviations from expected results
- A more efficient health system, by supporting the most cost-effective treatments, strategies and interventions on broad-based independent evidence (SSCH 2016, pp. 23–24)

Privacy legislation seems to be a significant barrier to expanding the use of Medicare data, both for research and policy development. In fact, according to the Department of Health, there are cases where the government itself cannot use the data it collects:

> There are very strict guidelines under the National Health Act, the Health Insurance Act, the privacy guidelines and the Privacy Act. We also observe those provisions very strictly. Indeed,

> *sometimes those rules can limit our own potential to use data internally* (SSCH 2016, p. 46, emphasis added).

Health information, such as the data stored in the Medicare Consumer Directory, is subject to stronger privacy protections compared to other types of personal information. Under the *Privacy Act 1988* (Cth), health information is considered a particularly sensitive type of personal information and there are additional requirements for its protection. Organisations must have consent to collect health information, and to use it for secondary purposes (such as conducting research based on information collected by health practitioners in the course of treating their patients) (OAIC 2014).

However, the *Privacy Act 1988* (Cth) also authorises the National Health and Medical Research Council to issue guidelines for the 'use and disclosure of health information for the purposes of research, or the compilation or analysis of statistics, relevant to public health or public safety' (Privacy Act 1988, s. 95A). The guidelines acknowledge that:

> The individual's right to privacy is not an absolute right. In some circumstances, it must be weighed against the interests of others and against matters that benefit society as a whole. The conduct of research, and the compilation or analysis of statistics, relevant to public health or public safety and health service management fall within these circumstances. (NHMRC 2014, p. 2)

The guidelines empower human research ethics committees, which operate in many public and private research organisations, to consider whether the public interest in conducting research outweighs the public interest in the protection of privacy. In effect, once approved by a human research ethics committee, this allows researchers to access health data without seeking consent or using only de-identified data (NHMRC 2014).

In addition to the protections included in the Privacy Act, specific *Privacy Guidelines for Medicare and the Pharmaceutical Benefits Scheme (PBS)* are issued by the Privacy Commissioner under the *National Health Act 1953* (Cth). While linkages between the Medicare and PBS datasets are allowed in limited circumstances, any such linked data must be destroyed after use. The guidelines detail how Medicare information should be disclosed to the Department of Health, and how it can be used (Office of the Privacy Commissioner 2008). Health information is also covered by secrecy provisions that are contained in numerous acts relating to Medicare and other health data, including the *National Health Act 1953* and the *Health Insurance Act 1973* (ALRC 2010).[66]

Numerous stakeholders (including the Productivity Commission (2015)) have called on the Australian Government to review the Privacy Guidelines to allow linkages between Medicare and PBS data, most recently in the inquiry report released by the Senate Select Committee on Health (2016). The Acting Privacy Commissioner (sub. 200, p. 39) supported the calls for a review of the Privacy Guidelines:

---

66 Each Act uses different language in describing how information should be handled. According to DHS, this creates significant confusion (ALRC 2010).

> I am aware that some consider … the Guidelines, to be too restrictive and to not allow the disclosure and linkage of MBS and PBS data in ways that are needed for research and policy analysis activities. … Given these matters, together with the evolution of policy and research needs since these legislative provisions were originally enacted, further consideration of the operation of … the Guidelines may be warranted.
>
> My Office and the Department of Health are committed to working together to consider this further with the aim of improving access to de−identified MBS and PBS data, for the purpose of health policy evaluation and development (as well as research undertaken in the public interest).

Health data linkages

Linking different health datasets allows policy makers and researchers to trace individuals' outcomes across different healthcare settings. As such, it is a vital step in understanding public health outcomes and assessing the performance of healthcare systems (Oderkirk, Ronchi and Klazinga 2013).

In Australia, linked datasets are created for research purposes and there are linking bodies in all jurisdictions (for example, the Population Health Research Network, which brings together data from all states and territories; The Centre for Health Record Linkage, which uses data from New South Wales and the ACT; and SA-NT DataLink, which links data from South Australia and the Northern Territory, and the Data Linkage Branch in the WA Department of Health, discussed below). The AIHW, the ABS and the AIFS are the only bodies accredited to link data held by the Australian Government. These linking bodies all use standards and techniques that minimise the risk of re-identification from the linked data, and maintain privacy (see, for example, PHRN 2011).

However, the use of linked health datasets — while recognised by the NHIA as a core activity to be undertaken by governments — remains limited. There are a few reasons for this.

- The linkage of key datasets held by the Australian Government is limited by legislation, as well as inconsistent policies on data sharing (SSCH 2016). Linkages between the Medicare and PBS datasets are limited by the privacy guidelines described above.
- Linking datasets held by the Australian Government and State and Territory Governments (for example, a link between Medicare and hospital data) requires a complex approval process involving numerous data custodians and ethics committees, that can take a very long time to complete (SSCH 2016). Some jurisdictions have separate privacy legislation for health records, which needs to be considered (ALRC 2008).
- Once approved, researchers face a substantial waiting time to receive data. It can take years to receive the data, particularly where there are multiple data custodians and ethics committees that must grant access to the data. In addition, even once data is

made available, there is only limited linkage capacity and some researchers reported bottlenecks and long delays. This is partly due to the fact that there are only three linking agencies that are accredited to work with Commonwealth data. There are significant costs that researchers are required to pay in some instances (SSCH 2016).

- Linked datasets are normally destroyed after the project they were approved for is completed, particularly if they contain data from the Australian Government. This limits the opportunities to re-use and maximise the value gained from the data (SSCH 2016). The AIHW is attempting to negotiate a pilot project to create enduring linkage keys for national health data (such keys already exist for state and territory data) (AIHW 2015).

Recent times have seen some progress towards health data linkages. A fairly recent agreement between AIHW and the Department of Health will allow the AIHW to store Medicare enrolments data and a five-year dataset of Medicare and PBS claims. These datasets could be used in future linkage projects, and according to the AIHW, it will be able to offer 'more efficient and faster data linkage services to the research communit' (AIHW 2015, p. 3).

A further step towards increasing using Medicare and PBS data in linkage projects occurred in August 2016, when the Department of Health released a linkable, de-identified sample from the datasets on data.gov.au. The data was subsequently removed from the website following partial re-identification; however the Department stated that it will work towards making the data available again in the future, following further de-identification (Department of Health 2016b, 2016d). These changes may affect the usefulness of the data for researchers; nonetheless, the planned release of this sample file will be an important step towards improving access to health data held by the Commonwealth.

Western Australia — a leading example of data linkage

Data linkages using Australian health data were pioneered in Western Australia in the 1970s. By 1995, the University of Western Australia had secured funding from the WA Lotteries Commission to set up the WA Data Linkage Unit, which linked together 6.5 million records of births, deaths, hospital separations, and other health data. The State Health Department also joined the project, providing funding as well as opening up additional data sets to be linked (Holman et al. 2008). Currently, Data Linkage WA is able to create linkages between eight core data sets (seven health data sets, and the WA electoral roll), and over 20 other data sets, including geographical information, and data from other government agencies, such as the Department of Education, the Department of Housing and the Department of Corrective Services. The wide range of data sets has enabled researchers to understand individuals' pathways, and investigate the risk factors for delinquency in young people and better ways to identify children at risk of abuse and neglect, among many other topics (Data Linkage WA 2013, 2016).

Western Australia is currently reviewing its data linkage activities and capabilities, in response to concerns raised by linked data users about long wait times and high costs involved in accessing data (Department of the Premier and Cabinet 2016).

The data linkage work undertaken in Western Australia (and currently underway in other jurisdictions) demonstrates the benefits of such projects, including:

- enabling innovative and cost-effective research that contributes to medical and scientific knowledge as well as population health
- adding value to existing information assets, both by offering researchers a richer picture of the population, and by improving data quality, as linkages can uncover duplication and other errors in datasets.
- enhancing patient privacy in medical research. Linking datasets has removed the need for researchers to contact individuals and request further information required for their work. Instead, researchers receive the data they need without any personal identifiers. Therefore, the proportion of health research projects using named data in Western Australia has dropped considerably through the use of data linkages (Holman et al. 2008).

Data linkages in Western Australia and other jurisdictions are restricted to using state data only. The inability to link Commonwealth data, such as the Medicare data set, has often been cited as a barrier to further research (SSCH 2016). In the past, such linkages have occurred — between 2001 and 2012, Data Linkage WA worked with the Commonwealth Department of Health to link PBS, Medicare and aged care data to their state-based data holdings. Following a pilot project in 2001, which successfully linked the hospital, Medicare and PBS records of 148 000 patients, a Commonwealth-State agreement was signed in 2002, for a data linkage covering the entire WA population. Research using this data was successfully conducted from 2005, examining, for example, the use of GP services among people with mental health conditions, and potentially inappropriate medications given to the elderly (Holman 2014; Holman et al. 2008).

However, in 2009, the Commonwealth Department of Health raised concerns about the continuation of the data sharing arrangement with Western Australia, stating that it was 'unfunded and unsustainable in the longer term. [The then Secretary of the Department also] noted that data access arrangements were not being provided on an equitable basis with other jurisdictions' (Department of Health 2016a, p. 4). Funding for data sharing eventually stopped in February 2011 (Department of Health 2016a).

D.4 Improving the availability and use of health data

Enhancing the availability and use of health data requires substantial changes, both for individual healthcare providers, as well as the government agencies that develop policies for data access.

Data quality and the incentives faced by health service providers

The incentives faced by practitioners in the health system play a critical role in determining both the extent to which good quality health data is collected, and the degree to which it is shared.

In both hospital and general practice settings, there are limited incentives for the collection of good quality data. In many instances such collection is seen as additional to activities such as prescribing medications or performing procedures, rather than as a critical part of recording such activities for ongoing reference, and establishing their efficacy. There is often little premium placed on the accuracy of data, and poor mechanisms in place to encourage such accuracy. Data entry skills are also reportedly in short supply in many hospitals, particularly regarding the accurate collection of activity data (on which funding outcomes depend).

A further key factor that often acts as a blockage to data exchange is that hospitals and other health service providers have limited incentives to undertake such exchange. In many cases, providers face an array of governance and other requirements that actively prevent them from exchanging data. Providers in the health system can also have entrenched models of working that do not facilitate the greater use and exchange of data within their service delivery.

Changing the public sector's approach to data management

As with many other areas of the public sector, the availability and use of health data held by governments has been affected by a culture that prioritised the protection of data, over promoting its use to improve program design and service delivery. The Australian Government is encouraging departments to move towards an open data approach, which should improve availability and use:

> … [F]or many years there was that culture, 'We must absolutely protect this data at all costs.' But, of course, as techniques — computing and statistical techniques … — get more sophisticated there are more ways to 'perturb' the data … or to confidentialise the data so we can actually protect people's privacy and still be able to make information available for use by researchers (Alanna Foster, Department of Health, quoted in SSCH 2016, p. 46).

Cultural change is extensively covered in the main body of this report (see, for example, chapters 3 and 5).

And while an ingrained culture of absolutism around data protection is likely to be a substantial barrier to overcome, a number of other challenges also need to be considered.

- *IT and data management infrastructure*. Some of the health data collected in Australia is stored on proprietary systems, and there may be little interoperability and data sharing capability. This may affect both the accessibility of the data, and its quality (see section on IT systems above). For example, the Australian National Audit Office has found that the Australian Childhood Immunisation Register (ACIR), which includes

records for over 2.26 million children, is based on a number of different IT systems. As a result, data cleaning and matching activities need to be done manually in many cases (ANAO 2015). The ACIR was noted by a number of stakeholders as a high value dataset, as it is one of only three national immunisations registers in the world (ANAO 2015; SSCH 2016).

- *Data collection*. While a large volume of data is already collected across the health sector, some potentially valuable information is not available. Most commonly, researchers have raised the lack of data on quality and outcomes of care as a barrier to assessing the performance of the health system, as well as individual establishments and practitioners (ACSQHC and NHPA 2015; OECD 2015). Where data is collected, its processing can take a long time, which limits the relevance of the resulting dataset (OECD 2015).
- *Data that is collected but not used or published* (PC 2015). For example, unlike most OECD countries, Australia does not routinely use linked data to monitor the quality of its healthcare system (OECD 2013).
- *Data quality*. While the move to My Health Record has mostly been seen as a positive development in the context of data availability and use, there are concerns that the transition to electronic health records will have negative effects on the quality of the data available (for example, due to a lack of coded data or poorly coded data) (OECD 2013, 2015).

The Senate Select Committee on Health (2016), which examined the issues around improving access to health data has made a number of recommendations, including:

- reviewing the National Health Act 1953, and the Privacy Guidelines for the Medicare Benefits and Pharmaceutical Benefits Programs, with the aim of improving access to de-identified Medicare and PBS data
- streamlining the approval processes required to access health data for research purposes
- consider accrediting state-based linkage units to link Commonwealth and State data.

The Australian Government is yet to respond to the Senate's report. In response to questions raised by the Senate Committee, the Department of Health argued that existing data integration principles endorsed by the Portfolio Secretaries Board in 2010 allow for research projects that involve linking Commonwealth health data to take place (Department of Health 2016a). In 2015, the Department published its *Data Access and Release Policy*, which supports the release of data 'in an appropriately de-identified and confidentialised form unless there are compelling reasons to the contrary' (Department of Health 2015a).

According to the policy, the Department should grant structured access to data, by allowing researchers to use analytical tools to query the data in a controlled environment. This approach differs from the way data is supplied in many instances, either as open data available to download or as confidentialised unit record files that are prepared by the Department and given to researchers on disks or other type of media. The structured access

approach (box D.9) can give researchers more flexibility, both in how they access the data and the type of questions they can ask in the course of their research (Department of Health 2015a).

There are also numerous initiatives across jurisdictions to improve the access to and the use of health data. Some examples include:

- Health Stats NSW, a website designed to offer easy access to health data on the NSW population was first launched in 2012. It has since been updated, to include data from multiple sources as well as the ability to create tailored reports on various aspects of public health. Additional information on NSW public hospitals is available from the Bureau of Health Information (BIH 2016; NSW Government nd).
- The Australian eHealth Research Centre was established in 2003 as a joint venture between the Queensland Government and the CSIRO. The centre has since expanded, and currently has research activities nation-wide. Among its other activities, the centre develops a range of tools that enable better management of health data, including improving the terminology used in the development of My Health Record and enabling better data sharing between healthcare providers (AEHRC 2015).

Box D.9 Structured access models

Structured access models have already been used in numerous health research projects in Australia. As part of the Population Health Research Network, the Australian Government funded the establishment of SURE, a remote-access data research laboratory that enables structured access for researchers. SURE allows researchers to access and analyse secure and sensitive data sets, which are held on separate secure servers. This eliminates the risks involved in releasing the data, including lack of secure storage or data transfers.

In addition, before accessing any data researchers undergo a registration and training process, to ensure they handle the data appropriately. This is intended to further minimise risk, by ensuring only individuals with appropriate knowledge and training handle the data (a trusted user model) (Sax Institute nd).

Further changes are also likely to take place as private sector providers become more involved in managing health data. In May 2016, the Department of Health announced that Telstra Health will develop and operate the new National Cancer Screening Register. The new register will integrate information from nine separate cancer registries, and improve access to information for healthcare providers. The data included in the new register will continue to be owned by the Commonwealth (Department of Health 2016f).

The cooperation between public and private sector providers may contribute to more efficient data management, through competition and innovation. There are many private providers in the health information technology space offering different data management systems, and competing on innovative features that are designed to improve the provision of healthcare. However, once again the lack of interoperability between the different systems, which can make data sharing very difficult, is evident.

E Case Study: Financial data

Key points

- The financial sector, by nature resistant to disruption by technology, is undergoing a wave of innovation driven by digital technology and the expanded use of existing and new sources of data. New innovative businesses are capitalising on these developments ('fintech' businesses) along with incumbent businesses in the sector.
- Financial data of interest to this inquiry is that created by the interactions between finance sector businesses and their customers — that is, in the provision and consumption of financial products and services.
  - Finance sector businesses are using data and technology to expand their customer base, broaden their product offerings and improve the efficiency of their operations.
  - A significant amount of data is collected from finance businesses by government regulators — notably the Australian Prudential Regulatory Authority, the Australian Securities and Investments Commission and the Reserve Bank of Australia.
- There is scope for governments to adopt efficiency-enhancing measures in three areas:
  - ensuring the best opportunity for (retail) credit risk to be accurately priced in the market place by reducing information asymmetry between borrowers and lenders
  - minimising unintended adverse consequences of regulation
  - increasing the availability of the data when it is in the public interest.
- The move to comprehensive credit reporting in 2014 has the potential to help address information asymmetry between borrowers and lenders. While participation levels are below those of other OECD countries and there are clear public benefits associated with higher levels of participation.
- Market forces are gradually driving greater sharing and use of financial data.
  - For example, customers are seeking and, in some instances, gaining the cooperation of their financial institution to share data about them with their accountants through the use of data feeds. This is delivering increased efficiency for accountants and savings for customers.
  - Given these developments, caution is desirable in considering whether to mandate — by a preferred Application Program Interface (API) for example — third party access to financial data about customers, with the customer's consent. An alternative right to data portability is much broader than access mechanisms specifically related to financial data, and this is addressed in chapter 9.
- When deciding what datasets to make more widely available — for example, to the industry, researchers and the general public — regulators can face difficult decisions in weighing the public benefit against the 'commercial detriment' to the businesses the data is about.
  - APRA's approach — involving substantial consultation with interested parties and careful assessment of the costs and benefits — is worth applying more broadly.

The finance sector has historically been a sector of the economy resistant to disruption. Complex regulations (including capital requirements), economies of scale and consumer preferences for perceived safe and established brands have all contributed to high barriers to entry for newcomers, even those with innovative and potentially efficiency-enhancing business models.

However, the sector is experiencing a wave of innovation-driven disruption, one that has been enabled, to a significant extent, by digital technology and driven by innovative uses of new and existing data sources. Dietz et al. (2016) noted some of the changes occurring in the finance sector:

> [M]obile devices have begun to undercut the advantages of physical distribution that banks previously enjoyed. Smartphones enable a new payment paradigm as well as fully personalized customer services. In addition, there has been a massive increase in the availability of widely accessible, globally transparent data, coupled with a significant decrease in the cost of computing power. (p. 3)

This has coincided with the emergence of the 'fintech' sector, which is capitalising on these developments — prominent examples in Australia include RateSetter, SocietyOne and Tyro Payments. Incumbent firms are also developing new and innovative uses for data and creating or sponsoring tech and data driven hubs.

The economic and social benefits of such developments are potentially multifaceted. They include enhanced product design and pricing, improved consumer marketing, better-informed consumer decision making, improved credit-offering decisions by lenders and improved risk management by lenders after credit has been granted (Manyika et al. 2013).

The most recent comprehensive review of Australia's financial system (Murray et al. 2014) found that competition in the Australian financial system was 'generally adequate at present'. However, the review also noted that 'the high concentration and steadily increasing vertical integration in some sectors has the potential to limit the benefits of competition in the future' (p. 255).

This case study examines the ways in which the Australian Government could enable greater availability and more widespread use of finance sector data as a means to increase efficiency and competition in the sector. In so doing, it also examines policy developments overseas — such as the United Kingdom's Open Banking Standard and midata program — and assesses their potential for application in Australia.

E.1 Types and uses of financial data

Financial data can be characterised as information that is created in the provision and consumption of financial products and services, as well as data generated in the course of government regulation and supervision of the financial system.

It includes data held by:

- financial institutions such as banks, credit unions and building societies (for example, account-level transaction data and average balances across account portfolios)
- parties who facilitate transactions, such as security exchanges, brokers (and increasingly, technology firms such as PayPal) and superannuation firms
- credit bureaus — who gather data on credit histories, incomes and assets for individual consumers and groups of consumers to calculate credit scores
- third party developers and data services (data aggregators) that aggregate data about financial products and offer consumers comparison data about financial products and services
- regulators — in the course of their supervision of the financial sector, various regulators collect data about the businesses they are regulating (Manyika et al. 2013).

There are also new and emerging sources of data that are being used by some firms in the finance sector, such as data generated through social media and mobile phone apps. While some of this data might not fit a traditional view of what constitutes financial data, it is likely to become increasingly valuable to firms offering financial products and services, particularly as big data analytics becomes more widely accepted. For example, there are already examples of credit providers incorporating social media data into credit assessment processes (PwC 2015). The emergence of fintech firms has the potential to disrupt traditional banking models and lead to further evolution in how financial data is used (Accenture 2015c).

Customer data held by financial providers

A range of customer-specific data is collected by financial firms in the provision of financial products and services. Because customer-specific data could allow other parties to identify individuals, it is considered personal information and its use and disclosure is thus be regulated under the *Privacy Act 1998* (Cth) (the Privacy Act) (section E.2).

In some instances, this information is collected to comply with regulatory or legislative requirements. For example, under the *Anti-Money Laundering and Counter-Terrorism Financing Act 2006* (Cth), firms in the financial and gambling sectors, bullion dealers, and currency exchangers are required to collect and verify customer identification information and report on transactions on an ongoing basis (AUSTRAC 2014).

Customer-specific data currently held by banks and other firms includes:

- customer reference data such as individual's name, date of birth and address
- information that could provide insights about an individual's circumstances such as income, assets and liabilities, and life stage
- account-specific data such as balances, transaction history and payment history.

There are numerous ways in which financial institutions, fintech firms and other service providers can use customer-specific financial data. Historically, financial institutions have used traditional data management techniques to draw insights into the creditworthiness, preferences and needs of their customers (PwC 2013). With technology firms (such as Apple) and new fintech firms now operating in the traditional financial space, incumbent financial firms want to use data in new ways — with a commercial advantage in mind, unsurprisingly. Data analysis offers exceptional opportunities for new forms of targeted marketing (by suggesting new products based on an individual's circumstances), or providing forward-looking financial advice (Accenture 2015b).

The ways in which banks and other financial firms share customer data are also changing. Historically, one of the most notable ways in which financial services firms shared data in Australia has been through their participation in the credit reporting framework (section E.4). However, they have begun to provide data feeds that allow small to medium-sized enterprises (SMEs) and individuals to port their account and transaction data into accounting software packages (such as MYOB and Xero). The direct transfer of this data results in a number of efficiency gains, including greater accuracy of data and lower costs of preparing income tax returns — the latter being reflected in cheaper accounting services for taxpayers. It also opens up opportunities for third party 'data aggregator' firms — such as SSIS Data Services — to facilitate the transfer of data.

Such arrangements require the consent of the customer, via a 'general authority' form (to address privacy legislation requirements), the cooperation of the financial institution (driven by customer pressures and the incentive of some revenue in the form of fees) and the participation of accountants (who are attracted by the reduction in operating costs and increased potential to cross-sell advisory services).

One of the main barriers to such data transfer is the upfront costs to financial institutions of setting up the data feed facility. Nonetheless, the practice is spreading throughout the financial sector with, for example, each of the 'big four' banks and their subsidiaries participating. Another issue is ensuring that the data is provided in a usable format — although, here, all the major banks have adopted a widely used standard (the BAI2 standard).

In addition, various financial services firms have recently begun to collaborate with technology firms in the provision of financial products and services. A notable example is Apple's agreement with the ANZ to allow their customers to use Apple Pay.

Public sector data

The public sector information architecture of the Australian financial system is largely organised around four key elements:

- the payments system, as regulated by the Payment System Board of the Reserve Bank of Australia (RBA)

- the registry and licensing systems, as regulated by Australian Securities and Investments Commission (ASIC)
- the tax reporting and Self-Managed Superannuation Fund (SMSF) systems, as regulated by the Australian Taxation Office
- Approved Deposit-Taking Institutions, the general and life insurance sector and the non-SMSF superannuation funds, as regulated by the Australian Prudential Regulation Authority (APRA).

The RBA and Australian Bureau of Statistics also collect and publish a range of finance sector data. The Australian Transaction Reports and Analysis Centre — Australia's financial intelligence unit — has a regulatory responsibility for anti-money laundering and counter-terrorism financing investigations.

While some of the data collected by these agencies is released in statistical publications and annual reports, much of it is not publicly released.

Emerging sources of data

The rise of technology has resulted in the emergence of new sources of data (that is, the rise of big data). Moreover, in recent years the amount of information generated by individuals, and about individuals, has significantly expanded (Costa, Deb and Kubzansky 2016). The wide reach of social media, for example, and the large number of users has provided new sources of data for possible use by financial firms (PwC 2015).

The rapidly growing adoption of personal digital devices — such as laptops, mobile phones and wearables — has opened up access to new types of data. Browsing history from laptops and mobile phones can provide insights into the preferences of consumers, and location data from mobile phones and wearables can help to infer aspects of an individual's spending habits. Mobile phone usage patterns can also provide insights into an individual's financial situation — for example, an analysis undertaken in Ghana linked mobile phone usage and bank account balances, and found that individuals who heavily favour the use of SMS tended to have more frequent and higher-value banking transactions (Costa, Deb and Kubzansky 2016).

The emergence of third party payment services providers — such as PayPal and Square — has provided an alternative source of financial data to that collected and stored by banks (and other deposit-taking institutions) and credit card providers (Schaus 2015). Online marketplaces such as eBay and Etsy collect data — such as sales revenue and user ratings — that can be useful for evaluating the financial health of online businesses. Payment services providers, such as Alipay and PayPal, have begun using the data they collect from vendors as a basis for providing loans to those vendors ('secured' by future cash flows) (Shinal 2014; Taylor 2015). Online lender Mybank (partially owned by Alibaba) has taken this a step further, using data from Alipay to provide collateral-free loans to consumers (who use Alipay to purchase from online retail stores Taobao and TMall) (Horwitz 2015).

There are a myriad of ways in which these new data sources are being utilised within the financial sector. For example, Indian bank ICICI launched the functionality to allow its customers to transact on their accounts via Facebook, which involves allowing Facebook access to customer data (with the customer's approval) (ICICI Bank nd).

Credit providers use some of these new data sources to help them assess the credit risk of individuals and businesses. In particular, to improve credit assessment accuracy, some lenders have started to incorporate:

- data from professional social networks (such as LinkedIn) to verify information provided by an applicant, or to draw insights about employment stability
- data on an applicant's social media contacts — an applicant whose contacts are in stable employment and who are good credit risks strengthens the probability that the applicant is also a good risk
- behavioural data (Manyika et al. 2013; PwC 2015).

In the United States, companies such as MicroBilt Corp. are using histories of rent, utility, telecom and other types of bill payments to help assess the creditworthiness of individuals. Such information is beginning to be used by traditional credit bureaus as well:

- the telecommunications company Verizon reported the payment history of its landline customers to the credit bureau TransUnion
- Experian was the first credit-reporting agency to track tenants' on-time rent payments (Manyika et al. 2013).

General insurance providers can also benefit from emerging data sources. For example, big data can improve claims management and fraud detection by facilitating a shift away from a focus on claims to a focus on the individuals making claims. This could include using social media data to identify whether an individual's social media contacts have also made claims, and thus whether the individual is part of a network of individuals making fraudulent claims (Bharal and Halfon 2013). Richer data sources can also provide insurers with insights into customer sentiment and help identify likely customer behaviour (Bharal and Halfon 2013). This can be particularly valuable since insurance companies have limited opportunities for engagement with customers (typically at point of sale and when claims are made).

Furthermore, data from a range of sources, including geospatial, weather or traffic data can improve the ability of insurers to assess, influence and manage risk through the use of early warnings and 'close the loop' between risk estimation and claims (Bhargava 2013). The use of telematics devices — which record a car's movements in real time — also allows insurers to offer new products, such as pay as you go policies, or price an individual driver's risk on the basis of their driving behaviour, leading to lower premiums for careful drivers (Cognizant 2012).

E.2 Barriers to accessing financial data in Australia

Privacy requirements

The collection, use and disclosure of personal information in Australia is regulated by the Privacy Act. Specifically, the Australian Privacy Principles (APPs) establish how all government agencies and businesses (with an annual turnover of more than $3 million) handle, use and manage personal information (OAIC 2014a). Several of the principles directly limit the ways in which financial service providers use and disclose personal information.

For example, APP 6 states that personal information cannot be used or disclosed for a secondary purpose other than that for which it was collected in the first place, unless an exclusion applies:

- the individual has been informed about and consented to a secondary use
- the individual would reasonably expect that their personal information would be used for a secondary purpose *and* the use in particular is related to the primary use (and directly related in the case of sensitive information)
- the secondary use or disclosure is either required or allowed under an Australian law or court order (OAIC 2015).

The principle provides several examples of where an individual would reasonably expect that their personal information could be used for a secondary purpose, including that the entity has notified the individual of the *particular* secondary use (OAIC 2015). However, the onus is still on the entity in question to ensure that the secondary use is related to the primary reason for collecting data.

In practice, this means that when financial services firms are unable to obtain consent, they are limited in how they use and share data about their customers — uses that are not related to the primary purpose for collection would be in breach of the APPs. However, financial services firms would still be permitted to share data when it relates to the provision of financial services (which is the primary purpose for collecting the data). For example, Standard Chartered's Australian privacy policy lists a range of parties who it may share information with, including:

- solicitors, valuers and insurers (for credit products)
- information technology suppliers
- verification services
- organisations providing analysis and research
- cloud computing and data warehousing service providers.

The APPs (principle 7) also inhibit the use of personal information for the purposes of direct marketing, except where the individual would reasonably expect this (such as where

they have expressly consented) and has the right to opt out. Moreover, the principle gives individuals the right to request that an entity not use, or provide to third parties, their personal information for direct marketing purposes (OAIC 2015).

Sharing of data between financial services firms and overseas entities is also regulated by the APPs (principle 8). In particular, prior to disclosing information about an individual to an overseas entity, banks and other financial services firms must take reasonable steps to ensure that the entity will comply with the APPs, in addition to accepting responsibility for breaches by the overseas entity (OAIC 2015). However, the principles do not apply in situations where:

- the overseas entity is an office of the Australian financial services provider (as is the case for some banks in New Zealand)
- it is reasonable to expect that the overseas entity is subject to laws that have the effect of providing a similar degree of protections as the APPs *and* individuals are able to access mechanisms to enforce those protections
- the firm has informed the customer that APP 8 will not apply and the customer has provided consent (OAIC 2015).

Consumer Credit Reporting scheme restrictions on data sharing

Part IIIA of the Privacy Act provides a framework for the collection, disclosure and use of credit-related information (which is defined in part IIIA). It applies to all credit providers (regardless of whether they are subject to the APPs), and also modifies some of the APPs for those organisations to which the principles apply. For example, APPs 6, 7 and 8 (which relate to use and disclosure of personal information) are wholly superseded by the credit reporting provisions (OAIC 2014c). APPs related to the right to access and correct personal information contained in credit reports are also superseded (OAIC 2014c).

The credit reporting provisions facilitate the sharing of credit-related information between credit providers and credit reporting bodies.[67] In this sense, they permit a greater degree of sharing than that permitted under the APPs. In particular, credit providers (such as financial services firms and utility providers) and credit bureaus are permitted to share information related to:

- a credit provider having sought a credit report (from a credit bureau) in relation to an application for credit by an individual, and the amount of the credit sought
- an individual's current credit providers
- any credit defaults (which is the failure to meet legal repayment obligations) in the previous five years

[67] The credit reporting bodies in Australia are Veda Advantage, Dun and Bradstreet and Experian.

- a credit provider's opinion that an individual had committed a serious credit infringement (such as credit fraud)
- the type of credit account opened
- the date the account was opened
- the current limit of the account
- the date on which the account was closed (Veda nd).

In addition, credit providers that hold an Australian Credit Licence are permitted to share information related to payment history, including:

- whether the individual was meeting their payment obligations (at the end of each payment cycle) over the previous two years
- the number of repayment cycles the individual was in arrears (Veda nd).

While the recent reforms expanded the scope of information that can be shared, Australia's credit reporting system remains relatively narrow compared to those in Europe and the United States (ACCIS 2015; ARCA 2014).

In *addition* to the obligations imposed by the Privacy Act, deposit-taking institutions also have obligations related to confidentiality stemming from common law. Specifically, they have a duty to not disclose to a third party confidential information related to a customer's accounts including '… any information obtained as a consequence of the relationship between the customer and the bank' (McCoach and Landy 2014, p. 89). This duty is excepted only when the use and disclosure is:

- made with the customer's express or implied consent
- mandatory (or compulsory) under law
- necessary for the fulfilment of a public duty (such as in a time of war or emergency)
- in the interests of the institution, which occurs where disclosure is necessary to protect the legal rights of the financial institution — for example, when suing a customer to recover a debt, in which case prevention of disclosure would affect the institution's ability to enforce its rights (Chaikin 2011).

Commercial obstacles

There are several commercial factors that create disincentives for financial services firms to share data with other parties.

The first is that the customer data that financial services firms acquire in the course of their operations can give them a competitive advantage in developing, pricing and marketing financial products and services. For example, a bank that is assessing a credit application from an existing customer has access to a range of information on that customer that would not be available to a competitor, but which might be useful in assessing credit risk.

Behavioural analysis (such as spending patterns) might improve the accuracy of credit risk assessment (Capgemini 2014).

Customer preferences, and associated reputational risks, might also lead a financial services firm to limit when and how they share data with other parties. The Office of the Australian Information Commissioner conducts surveys on community attitudes to privacy, which consistently show that individuals view financial data as one of the most sensitive types of data (Office of the Australian Information Commissioner, sub. 200). This has not been lost on holders and users of financial data in Australia. As noted by Data Republic co-founder Paul McCartney:

> … banks are good at managing risk and governance around money. Now banks are realising all the information they have on all their customers is worth something. But they can't use it unless they apply the same security and risk management processes that they would to customer's money (Eyers 2016)

Other barriers created by market regulation

The licensing and regulatory requirements on financial services firms, aimed at maintaining the ongoing security and stability of the market, in turn limit the capacity of potential entrants to access data that may be necessary to enter the market. For example, innovation in services such as personal budgeting and product comparison may require access to individuals' account and transaction history, as well as data relating to fees and charges for different products and services. To obtain such data, fintech firms — such as data aggregators who draw together data from different accounts and financial services providers — are primarily resorting to so-called 'screen scraping' technology, which uses software to 'rip' data based on its known position on a webpage or on a statement (FinTech Australia, sub. 182, Australian Securities and Investments Commission, sub. 195). This requires an individual to provide their statements to the third party, manually enter their transaction into a portal managed by the third party, or grant the third party access to their online banking portal by providing their online banking credentials (that is, their username and password). The security risks associated with providing online banking credentials may provide a disincentive for customers to share their data.

What evidence is there that these barriers are reducing market efficiency?

Regulatory failure in data availability and use?

While many of the Murray Inquiry's conclusions and recommendations do not specifically refer to data, they clearly point to the potential for the regulatory framework to pose barriers to entry for firms that seek to use new, innovative business models. This may include firms that make innovative use of data and data analytics to provide targeted financial products and services.

Fintech businesses provide a new source of competition in the finance sector. The emergence of mobile devices, including smartphones, has enabled a new payment paradigm as well as fully personalized customer services. The rapid increase in the availability of widely accessible and globally transparent data, coupled with a significant decrease in the cost of computing power, have opened up opportunities for innovation. Fintech businesses have considerable potential to capitalise on these developments and, in so doing, increase the level of competition in markets that incumbent financial services firms have largely dominated. Some fintech firms specialise in data collection (2iQ Research) and credit scoring (ZestFinance). Others leverage large unstructured social media data sources to make better credit or insurance underwriting decisions (Wharton Finance 2016).

FinTech Australia, a Sydney-based fintech hub, noted that licensing requirements (to obtain an Australian Financial Services Licence) pose a substantial barrier to entry for early stage fintech businesses, citing four main reasons:

- Uncertainty — early stage fintech business models are fluid and frequently change during the development and testing stage, leading to uncertainty regarding the required authorisations and regulatory obligations.
- Lack of easy fit — some fintech business models do not fit neatly into existing authorisation categories, requiring substantial liaison with the regulator. It is 'inefficient and costly' for this work to be undertaken in relation to an immature business model that is likely to change.
- Time — timeframes for obtaining a licence range from two to six months.
- Cost — the complexity of the licensing regime in particular requires fintech businesses to retain external consultants and/or lawyers at costs ranging from $10 000 to over $200 000. Such costs may be beyond the financial capacity of many start-ups (FinTech Australia 2016).

New technologies and business models sometimes do not fit within existing regulatory requirements — for example, because certain actions are in contravention of the intended purpose of regulations or were simply not considered at the time the regulation was drafted. As noted by the Commission:

> Activities and behaviours of new business models can present real and complex regulatory issues, but governments and regulators should not act in a 'knee-jerk' fashion to tightly regulate or prescriptively enforce existing regulations. Such action could lead to poor regulatory outcomes that stifle innovation and limit the possible benefits from these new business models. (PC 2015, p. 211)

Regulators need to be mindful of the potential for existing regulatory frameworks to limit competition, particularly during periods when many new business models are emerging — such as those making innovative uses of data. In this regard, the Commission notes that the mandates of Australia's financial sector regulators contain an inconsistent approach to competition:

- The Australian Prudential Regulation Authority (APRA) is required to consider competition and contestability in its decisions, although its industry-specific frameworks (for example, across banking, general and life insurance, and superannuation) do not adopt a consistent approach to competition.
- ASIC lacks an explicit competition mandate (Murray et al. 2014).
- There is no current requirement for regulators to explain how they balance competition considerations with other regulatory objectives in reaching decisions (Murray et al. 2014).

Building consistent pro-competition provisions into the mandates of Australia's financial sector regulators would help to ensure that regulators keep a firm focus on market access, not least when reviewing and revising regulations (although competition objectives would still need to be balanced with systemic stability). ASIC's new 'regulatory sandbox' provisions could have the potential to help in this regard (by allowing fintech start-ups to test their ideas with customers without necessarily having to meet some licence requirements), but it is too early to assess how effective such arrangements have been with any certainty.

Are information asymmetry and adverse selection impeding efficiency?

The information asymmetry between lenders and borrowers has long impeded the efficiency of credit markets. All lenders face uncertainty about borrowers' creditworthiness — that is, the likelihood that a borrower will default on a loan. This uncertainty is compounded if lenders cannot observe some characteristics and actions of potential borrowers, particularly their current financial position and their loan repayment record.

When lenders are able to access comprehensive credit history information about borrowers, they are better equipped to allocate credit efficiently, and charge a borrower an interest rate that more closely reflects the risk involved in lending to that specific borrower. (This rate will typically be lower for a low-risk borrower and higher for a high-risk borrower.)

The Policy and Economic Research Council (2012), in conjunction with Dun and Bradstreet Australasia, undertook a study to estimate the effect on credit allocation stemming from inclusion of more information in credit reports. They found that comprehensive credit reporting increases the proportion of loans approved for a given target default rate — in other words, the presence of more information improved decisions around how to allocate credit.

In the absence of such information, low-risk borrowers effectively subsidise high-risk borrowers. This is the problem of adverse selection — a situation where high-risk borrowers find loans relatively cheap, low-risk borrowers find them relatively expensive, and the market becomes skewed in favour of high-risk borrowers (Turner and Varghese 2010).

Is data providing a source of market power?

In the process of lending and in the provision of other products (such as transaction accounts, savings accounts and wealth management services), financial services firms are able to gather quite a lot of information about their clients, including that related to their creditworthiness. A number of inquiry participants (such as Tyro Payments Limited, sub. 7 and FinTech Australia, sub. 182) have suggested that this provides lenders some degree of 'informational monopoly' about their clients, with the effect of reducing competition in the market for financial services.

The Murray Inquiry concluded that while the level of competition in Australia's banking sector was generally adequate, the concentrated nature of Australia's financial system had the potential to limit the benefits of competition into the future. The inquiry noted the potential for data to improve competition, particularly where it facilitates the emergence of services that compare products and services provided by a range of different businesses.

An alternative view, however, is that the growth in volume, variety and sources of data is helping to lower barriers to entry to new finance firms, particularly those that can make innovative use of these new sources and types of data.

On balance, it is likely that access to data provides some degree of competitive advantage for incumbents, at least in the short term. To the extent that governments can encourage data availability, there could be scope for increased competition in markets for financial services. That said, the costs of facilitating greater sharing of data would need to be carefully considered (section E.4).

Data release issues facing regulators

Regulators often face difficult decisions when considering whether or not to release data about the private sector. For instance, APRA's governing legislation requires it to weigh up the benefits to the public from disclosure of [the data] against any detriment to the commercial interests that the disclosure may cause' (box E.1).

This is often not a straightforward decision — and it becomes even more difficult if a regulator is required to weigh up the interests of all stakeholders, not least customers of finance sector businesses (such as credit card customers and small business customers). For example, while APRA's objectives are relatively narrow and largely limited to systemic stability, the finance sector has a broad range of stakeholders who may be affected — either directly or indirectly — by its decisions regarding data release.

Box E.1 Data release by APRA — the challenges of weighing up 'public benefit' against 'commercial detriment'

- APRA is a national statistical agency for the Australian financial sector. Using its powers under the *Financial Sector (Collection of Data) Act 2001* (Cth) (FSCOD Act), APRA has collected data for over a decade and, in some cases, made the data publicly accessible, primarily through its statistical publications.
- Under the FSCOD Act, APRA may determine that data submitted to it by firms under the Act to be non-confidential — and hence able to be made publicly available — if 'APRA considers that the benefits to the public from disclosure of [the data] outweighs any detriment to the commercial interests that the disclosure may cause'.
- APRA undertakes an extensive process — including the release of discussion papers, calls for submissions and direct consultation with stakeholders — to assist it to determine whether the benefits of public disclosure of certain data outweigh the costs.
- The benefits of public disclosure may include:
  - improved transparency and accountability of the finance sector
  - increased security for customers (such as life or general insurance policy holders)
  - increased quality of research and public discussion of policy issues
  - better-informed decision making by policy makers, other regulators, market analysts, researchers and managers
  - enhanced Australian observance of international standards.
- The costs may include:
  - detriment to the commercial interests of firms in the finance sector
  - erosion of competitive advantage of individual firms
  - reduction in individuals' privacy (although APRA is obliged to comply with the Privacy Act).
- The 'public benefits' are often difficult to determine because:
  - it can be difficult to know the potential benefits of data until it has been made available and used
  - stakeholders who could be interested in the data are diffuse, in some cases unaware of its existence and arguably not highly motivated to make a case for access to such data.
- By contrast, the 'commercial detriment' is easier to quantify, affects a relatively small number of stakeholders and is well articulated by those stakeholders.

Sources: APRA (2013, 2015).

E.3 What could governments do to increase data access and use in the financial sector?

Comprehensive credit reporting

Financial institutions in Australia have long participated in Australia's credit reporting system. Before 2014, the credit reporting regime limited the information that could be collected, used and disclosed by credit providers and credit reporting bureaus (CRBs) to so-called 'negative' information about an individual or company's credit delinquency (Veda nd) — such as defaults and late payments on loans. The majority of credit providers, including all of the major banks, participated in this system.

In 2012, the *Privacy Amendment (Enhancing Privacy Protection) Bill 2012* (Cth) was passed by both houses of Parliament and in March 2014, legislation from the Bill was enacted to allow credit providers to collect and share so-called comprehensive information, such as credit limits and repayment history. An industry-developed regulatory code for Australia's credit reporting system was approved by the Office of the Australian Information Commissioner in December 2014. Participation in the expanded system is voluntary, with information being shared on a reciprocal basis (participants have access only to the types of information that they themselves have shared) (ARCA nd). The Australian Retail Credit Association formalised this arrangement through the Principles of Reciprocity and Data Exchange, which were approved by the Australian Competition and Consumer Commission in December 2015.

To date, none of the major banks has begun sharing comprehensive credit reporting data publicly, although at least one is uploading this data to Veda — a credit bureau — privately. This has led to calls from some stakeholders — particularly from the fintech sector — for the Australian Government to make participation in the comprehensive scheme mandatory. Other parties, such as Financial Institutions and Management Advisory (FIMA) (sub. 73), have called for the mandating of partial comprehensive credit reporting (CCR) (as well as reporting of defaults).[68]

The Murray Inquiry noted that as participation levels and the amount of system-wide data in the CCR scheme grow, the net benefits for all participants in the scheme increase. It recommended that the Australian Government:

> Support industry efforts to expand credit data sharing under the new voluntary comprehensive credit reporting regime. If, over time, participation is inadequate, Government should consider legislating mandatory participation. (Recommendation 20, Murray et al. 2014, p. 190)

The incentives for an institution to participate in C can be mixed and quite complex. Participation depends on the perceived net benefits, which will differ between different classes of credit provider.

---

68 Partial CCR would not include repayment history information.

For a major institution with a relatively large customer base, early and full participation may provide, at least initially, relatively small benefits than to other participants — thus diluting their competitive advantage. For example:

> … Commonwealth Bank holds around 23 per cent of customer lending accounts in the banking sector. If they were to contribute full comprehensive data they would be providing more benefit in improved risk discriminatory power to other lenders relative to what they would get themselves. (Johnson 2013, p. 45)

However, non-participation is not without risks. Credit providers that do not participate are at risk of adverse selection with respect to potential new borrowers, a risk that becomes more acute as the level of industry participation increases. That is, while their competitors are benefiting from access to comprehensive information about the creditworthiness of potential borrowers, those outside the scheme may increasingly be approached for loans by relatively high-risk borrowers while still facing the traditional information barriers to assessing the creditworthiness of these potential customers.

The level of participation in CCR in Australia currently lags behind the OECD average. However, this at least partly reflects the fact that Australia's CCR system is relatively new. While the reforms to allow the expanded system were implemented in early 2014, the industry code was not approved until December 2014 and the Principles of Reciprocity and Data Exchange were not approved until December 2015.

Internationally, both mandatory and voluntary approaches have been adopted although most countries have maintained voluntary arrangements.

Arguments for mandatory scheme

There are compelling reasons to mandate participation in CCR.

- Additional availability of credit-related information would improve credit allocation and pricing, leading to benefits to (at least some) consumers who would be able to access cheaper loans, reflecting the lower default risk inferred from their credit history.
- Allowing smaller financial businesses and potential new entrants (as well as utilities providers) to have access to a large pool of customer data may help to facilitate their entry into the market and, in this way, could boost competition and innovation in the finance sector.
- A high level of participation might only be achieved by mandating participation in circumstances where large incumbent banks face significant disincentives to participate.
- If data collected and stored by credit providers is viewed as jointly owned with the customer then the customer should be allowed to share the data with third parties, including for the purposes of a credit assessment, regardless of whether their credit provider wishes to participate in CCR.

Arguments for continuation of voluntary scheme

Conversely, there are compelling reasons to retain CCR as a voluntary scheme.

- There is potential for unintended consequences for consumers if the credit reporting is mandated prematurely (which could ultimately harm consumers) (Financial Rights Legal Centre, sub. 107).
- Participation will likely increase over time, and it is still too early to conclude that voluntary participation levels in the comprehensive scheme will not rise to levels achieved in similar countries.
  - Across the OECD, Australia and France are the only two countries where participation levels are lagging. That said, Australian participation levels are consistent with the early levels of participation experienced in the United Kingdom. However, participation the United Kingdom (as well as in New Zealand) was encouraged by government — for example, it has been asserted to the Commission that those governments encouraged participation through an implied threat to mandate participation.
    - New Zealand moved to CCR two years before Australia and participation there has reached 50%. Veda estimates the overall size of the retail credit market in New Zealand to be 7.5 million open retail credit accounts, covering 2.7 million individuals or approximately two thirds of the New Zealand adult population.
- As participation and the level of system-wide data grow, net benefits increase for all CCR participants, providing an incentive for participation (such as the heightened risk of adverse selection facing non-participants).
  - Veda has estimated that the 'tipping point' for participation in Australia is below 50% (Murray et al. 2014, p. 192). As of March 2016, CCR data had been loaded on approximately 25% of all open retail accounts (Veda, sub. 163).
- Any legislation to compel mandatory participation is likely to require a fairly high level of prescription — for example, specifying the data that would be reported, how it would be reported (format) and to who it would be reported (for example, to all credit bureaus or just one). In rapidly evolving industry environments, highly prescriptive legislation can become dated quickly. Any legislation would also need to provide for policing and enforcement.
  - A voluntary scheme, by comparison, could evolve over time and settle on certain agreed arrangements over matters such as coverage, format and obligations, and in this way utilise the advantages that 'self-regulation' has in certain instances (compared to explicit government regulation).
    - Self-regulatory schemes tend to promote good practice and target specific problems within industries, impose lower compliance costs on business, and offer quick, low cost dispute resolution procedures. Effective self-regulation can also avoid the often overly prescriptive nature of regulation and allow industry

the flexibility to provide greater choice for consumers and to be more responsive to changing consumer expectations (The Treasury 2000).

  - In this regard, the Commission understands that a number of financial institutions are already sharing data outside of the scheme through mutually beneficial private arrangements.
  - Such arrangements could underpin the eventual widespread adoption of the scheme.
- Mandating the scheme could well slow progress being made by financial institutions towards participation in the scheme.
  - That is, they may be inclined to defer development of their internal systems until they see precisely what the requirements and specifications of a mandatory scheme will be.
  - In this regard, the Commission understand that at least some of the 'big four' banks are preparing their internal systems in order to be ready for CCR.
- Mandating CCR would impose costs on all finance sector businesses legally obliged to participate in the scheme — regardless of whether participation was in their commercial interest at the time.
- There is some evidence of data quality issues in the credit reporting system. Mandating participation in the expanded system — which involves much greater volumes of data — could compound data quality issues. It is important that sufficient time is given for credit providers to make necessary system changes and to undertake testing (Australian Retail Credit Association (ARCA), sub. 87; Dun and Bradstreet, sub. 135).

Broadening the scope of data collected under CCR

The CCR regimes in countries such as the United States and the United Kingdom allow for considerably more data fields to be collected and reported than in Australia. The Australian scheme could usefully be broadened to include the current balance of a credit contract, which would provide credit providers with visibility over levels of actual indebtedness, thus aiding their credit decisions and their responsible lending responsibilities.

Another possible extension would be to allow credit providers such as utilities and telecommunications businesses to provide and access repayment history information (which the current system prohibits). This would allow a way for some consumers, such as young adults who have not previously accessed credit from a financial services provider, to better demonstrate good credit behaviour. The Commission has heard views that data from such groups may have a lower degree of accuracy than data from other credit providers, and so would need to be used with caution.

There appears to be little doubt that additional sources of information have the potential to be valuable, and could lead to better outcomes for some consumers. Such value will, however, be influenced by the accuracy and completeness of the data.

Another means of broadening the scope of CCR is to mandate the use of small-medium enterprise (SME) credit data. The Murray Inquiry noted that such mandating would impose compliance costs on credit providers and may not have a significant impact on information asymmetries because:

> … the credit health of the business owner(s) as an individual remains the primary information source for credit decisions, rather than information about the SME itself. (Murray et al. 2014, p. 192)

However, some SMEs are now able to secure new sources of credit that are related to their business assets or cash flow rather than their personal circumstances. To the extent that this becomes more widespread, it may be worthwhile reviewing in the future the inclusion of SME credit data into Australia's credit reporting system.

Obstacles to greater participation

Several participants in the inquiry (such as the Australian Bankers' Association, sub. 93 and the Customer Owned Banking Association, sub. 132) noted that uncertainty surrounding the way in which CCR interacts with the hardship provisions of the *National Consumer Credit Protection Amendment Regulation 2012* was discouraging participation in the scheme. The Office of the Australian Information Commissioner (2014b) indicated that credit providers cannot disclose to a credit reporting body the existence of a hardship application but can disclose the termination or issuing of new credit associated with such applications. By withholding information on the existence of hardship situations, other credit providers have an incomplete and misleading picture of a borrower's capacity to repay credit.

Greater clarity on how the hardship provisions should interact with CCR could help pave the way for greater industry participation in the scheme. Alternatively, the inclusion of a hardship flag in credit reports could provide a more informative picture of an individual's credit worthiness. As this issue will need to be addressed at some point for the full benefits of CCR to be achieved, it would seem prudent for the Australian Government to attempt to resolve it as quickly as possible.

Customer-initiated access through the use of APIs

A number of submissions (for example, Tyro Payments Limited, sub. 7, FinTech Australia, sub. 182 and the Australian Securities and Investments Commission, sub. 195), and several other studies (such ODI and Fingleton (2014)), have noted the potential benefits that could arise from policies that allow customers to share their data with third parties, such as via Application Programming Interfaces (APIs). APIs are one way that financial bodies are able to share data (appendix B). Westpac Banking Corporation (sub. 197, p. 3) recommended that the Australian Government '… should require private and public sector organisations to provide individuals with access to a selection of information the organisations hold about them in a standardised and readily usable format.' It also

recommended that discussions should take place between governments and industry on specific mechanisms for enabling this access, including the use of APIs.

In many overseas markets, banks and other financial services providers have begun to implement APIs, without governmental support or encouragement, for a range of purposes (box E.2). In addition, some Australian financial services firms have already begun to implement APIs for commercial reasons (FinTech Australia, sub. 182).

Box E.2 Bank data sharing via APIs

Examples of how banks, internationally, are using APIs to allow third party access include:

- A range of banks in the United States have implemented, or are in the process of implementing, APIs to facilitate third party development of complementary apps (Macknight 2016). For example, Silicon Valley Bank has signalled its intent to give developers access to data and payments operations to facilitate integration with their own apps. The first step in this process will involve deploying open APIs to allow customers to direct the bank on how to handle payments on their behalf. Moreover, these developments open up the possibility for customers to share their data with other parties (Crosman 2015).
- In 2012, Credit Agricole (France) launched an online app store, CA store. APIs are used to facilitate customers accessing their data through the apps hosted on CA store, some of which also allow customers to share their data with other parties (Hoffman 2013; ODI and Fingleton Associates 2014).
- Banco Bilbao Vizcaya Argentaria (BBVA) (Spain) provides an 'API market' with a range of products, including:
  - Paystats, which provides access to aggregated card purchase data
  - BBVA Connect, which allows customers to authorise apps to access BBVA services on their behalf
  - BBVA Accounts, which allows pre-authorised users access key account data (BBVA nd).
- Fidor Bank (Germany) has also implemented an API which allows customers to authorise apps to:
  - access their bank account
  - see their transaction history
  - initiate various types of payments (Fidor nd).
- In China, Wechat Pay developed customised APIs to allow China Merchants Bank customers to link their credit card and Wechat accounts, providing functionality for the users to view information related to their credit accounts (such as transactions and credit limits) directly via Wechat (Sheng 2013).

The Open Banking Working Group, established by the Open Data Institute, has proposed several ways in which consumers could benefit if they were able to share their financial data with third parties via an API (ODI 2016).

- Given the complexity of financial products, it can be difficult for consumers to compare different products and to identify which product is most suitable for their

circumstances (ODI and Fingleton Associates 2014). However, there are third party comparator sites that are able to identify which account is most suitable for an individual, provided it has access to the individual's transaction history and data on the account's fees and charges.

- Consumers who are able to share their account balance and transaction history would be able to benefit from the use of personal financial management and budgeting tools (which can also pull in information from other financial products, such as credit cards).
- Transaction history is a powerful predictor of creditworthiness — being able to share this data in a streamlined manner could assist consumers to access credit from third parties at more competitive pricing, and could speed up the application process. It might also assist the third party credit provider in meeting their responsible lending obligations as well as regulatory obligations to identify credit applicants.
- The implementation of APIs could also allow SMEs to automatically import transaction data into accounting software packages, thereby eliminating a need to manually input transactions. In Australia, there are already options for business customers to import transaction data into their software accounts (via direct data feed and through businesses such as SSIS Data Services) (ANZ 2015).

There is also the potential for third-party providers to monitor an individual's account for fraudulent activity, particularly if access was granted over a range of accounts/products.

Broader potential benefits flowing from opening up access to customer data include:

- increased competition — for example, innovative lenders could use account transaction history to better understand an individual's credit risk, and thus offer more competitive loan pricing, without having to manually input transaction data into their credit assessment systems (which can be time and resource intensive) (ODI and Fingleton Associates 2014)
- greater innovation (Ley and Bailey 2016)
- enhanced consumer choice and protection by leveraging third party comparator and fraud monitoring services (ODI 2016)
- reduced transactions costs for SMEs in entering transaction data into accounting software (ODI 2016).

In addition, access to transaction data could allow third parties to draw insights that can be used as a basis for sophisticated marketing techniques, such as targeted marketing based on an individual's previous purchases.

Conversely, some inquiry participants pointed to risks that could arise from increased access to customer data. For example, the Commonwealth Bank of Australia (sub. 175, p. 2) suggested that measures to facilitate broad access to customer data could create 'privacy and security risks which customers may not be able to understand or control'.

Westpac Banking Corporation (sub. 197) highlighted a number of specific challenges related to privacy and security risks, including those related to:

- identity verification — where an individual directs a bank to share their data with a specified third party, it may be difficult for the bank to verify that the customer *should* share their data with the third party
- informed consent — consumers might not fully comprehend the type or amount of data to be shared (including the risk associated with different types of data), or the ways in which it could be used
- data governance — data holders lose the ability to control how data is used, and therefore prevent data misuse, once the data has been shared with third parties
- privacy policies — increasing the complexity of data access arrangements could create challenges for financial services firms seeking to balance disclosure requirements and the ease with which privacy policies can be understood
- data security — issues include different levels of security between financial services firms and third parties, identity fraud and ensuring that data is transferred securely.

Advantages of APIs over other methods for customers to share data with third parties

As noted earlier, at present, individuals who wish to share their financial data with a third party would need to provide copies of their account statements (in PDF or CSV format), or share their online banking credentials with the third party (section E.2). Both of these options have their shortcomings.

The first is that in handing over their banking credentials, consumers risk the possibility of fraudulent access to their accounts. The responsibilities of Australian Authorised Deposit-Taking Institutions and their customers in preventing fraudulent transactions are outlined in the ePayments Code (a voluntary code of conduct), which stipulates that participating financial services providers are liable for unauthorised transactions if the individual has not disclosed their online banking passcode to another party. Where customers do so, they could be liable for any losses incurred (Australian Securities and Investments Commission, sub. 195).[69]

The second is that the accuracy of screen scraping technology is affected by changes to the layout of webpages and statements. Fintech firms that rely on screen scraping technology could face additional costs to monitor webpages, and to update their screen scraping algorithms in responses to changes in the layout of websites and statements (ODI and Fingleton Associates 2014). At the time of publication, the Commission was not aware of

[69] The Australian Securities and Investments Commission also suggested that, providing security issues can be resolved, banking customers should not be disadvantaged if they use legitimate account aggregation services.

any financial services providers in Australia making customer data sharable through the use of APIs. There were, however, several businesses that had established data feeds from the major banks, and who facilitated individuals (and businesses) importing their transaction account data into selected accounting software packages.

Finally, even where consumers are able to download and provide their data to third parties in a machine-readable format, there is a risk that the data will have been altered by the customer. In other words, the current processes do not ensure data integrity, and could limit the usefulness of such data. Conversely, APIs can be designed to bypass the individual (consent notwithstanding) by directly connecting the financial services firm in question and the third party to whom the individual has granted access rights. Moreover, APIs can be used where the individual wishes to grant permission for a third party to initiate transactions on their behalf (as in the case in Europe with the approaching Payment Services Directive 2 reforms).

If the aforementioned data feeds provide the functionality for individuals to share their data with other third parties (such as fintech firms), in theory this could provide an alternative to implementing APIs.

In November 2011, the UK Government launched the midata program, with the aim of making it easier for customers to download data from service providers in four sectors — personal current (transaction) accounts, personal credit cards, energy and telecommunications (DBIS (UK) 2011, 2014). Participation in the program is voluntary. A review was undertaken in 2014, which found that most participating businesses provided functionality for consumers to download data, but only some did so in a format that was machine readable (DBIS (UK) 2014).

In 2014, the Open Data Institute and Fingleton Associates (2014), in a report prepared on behalf of the UK Government, found that giving customers the ability to share their banking data through APIs could improve competition and consumer choice. They also found that the costs to banks of implementing APIs was likely to be negligible, and subsequently recommended that banks implement APIs.

In response, the HM Treasury established the Open Banking Working Group in 2015 to develop a standardised approach to implementing bank APIs. Specifically, the Working Group's objective was '… to produce a detailed framework for how an Open Banking Standard could be designed and delivered, with a timetable for achieving this' (ODI 2016, p. 3). The Working Group made a number of recommendations, including that:

- an independent authority be established to oversee the development and deployment of the standard and to vet third parties seeking access to bank data (including publishing a white list of approved parties)
- access to data would be granted only with customer consent
- permission to both read and write data should be a feature of the standard.

Participation in the Open Banking Standard was on a voluntary basis.

In 2016, the Competition and Markets Authority (2016) found that the UK banking sector was not as competitive or innovative as it needed to be and announced that it would be implementing a range of remedies to improve the level of competition. This included mandating the development and implementation of an open API standard for banking by early 2018, with product reference data (such as fees and charges) made available in late 2017.

In addition, recent reforms in Europe, the Payments Services Directive 2 (PSD2), impose obligations on banks to build in APIs that facilitate read and write access to an individual's transaction accounts. In particular, the PSD2 requires banks to facilitate customers granting access rights (both read and write access) to third-party payment service providers and third-party account information service providers (who 'aggregate' into one place information from a range of accounts across different banking account providers) (Accenture 2015a).

The Commission is not aware of any other jurisdictions mandating the provision of open banking APIs, but notes that the Monetary Authority of Singapore has encouraged Singaporean banks to implement APIs (and is in the process of doing so to provide access to its own data), and that there are private-sector led efforts in Germany and the United States to set standards around banking APIs (FinTech Australia, sub. 182).

Disadvantages of the API approach

There are costs associated with the use of APIs, including infrastructure costs and potential security and operational risks. Mandating the use of APIs would also raise a number of issues.

There is the cost — to the bank or other financial enterprises that provides the data — of building the technical infrastructure required to facilitate the transfer of data to the third party. The Australian Bankers' Association (sub. 93) suggested that the cost of building APIs would be substantial. Conversely, the Open Data Institute and Fingleton Associates (2014), based on consultations with a number of organisations, estimated that capital costs of implementing APIs would be no more than £1 million, with smaller ongoing annual operating costs. They also found that uncertainty around technology and data standards, legal requirements and data security had the potential to significantly increase implementation costs.

To reduce this uncertainty, governments would be justified in playing a leading role to set standards, and thus might also face some costs. Regardless of who establishes data standards, they are an important pre-requisite for open APIs. Whether such standards should be mandated is a separate issue. On principle, governments would need to carefully consider whether to mandate, based on weighing up uncertain benefits with the costs that would be imposed on businesses not already using the particular standards in question.

There is also potentially a regulatory role for governments (such as approving third parties), that would have associated ongoing costs.

There would also be potential impacts on incentives, a point made by ANZ (sub. 64, p. 26):

> The effect of this [releasing customer data via an API] is that data custodians may not be able to control the terms on which data are used once released. … [S]uch usage could involve commercial activities detrimental to the data custodian's interests. Concern about this could limit the extent to which data custodians invest in data generation, protection and availability.

Aside from the costs of technical infrastructure, mandating the use of APIs would probably require resolution of matters surrounding data standards, responsibilities for data quality, data security and, not least, the overall costs and benefits. For example, would giving consumer the right to share their data also allow them to 'cherry pick' which transactions they share, or would there be rules in place to prohibit them from doing so? There would also be associated adjustment and compliance costs, such as technical infrastructure costs, and monitoring and enforcement costs.

In terms of data ownership, despite probable perceptions to the contrary, it is the Commission's understanding that customers generally do not own the data that financial services firms collect about them and their activities (such as account transactions). That said, ownership is most likely not essential in order to achieve improved data access — customers can already share their data with third parties, albeit in ways that are not necessarily ideal. And such processes can be substantially improved.

Weighing up the benefits and costs of mandating the use of APIs

Assessing the overall (community-wide) costs and benefits of mandating customer access to their financial data is difficult because of the large uncertainties involved.

The costs are probably easier to assess:

- building the technical infrastructure (compliance costs for financial services providers)
- ongoing compliance costs for financial services providers
- risks for security of customer data resulting from broader access to it.

The benefits are much more difficult to estimate with a high degree of accuracy because the demand for such data by consumers is unknown — that is, gauging how many consumers would seek data in an API format if it were available to them, and how much they would value it. Moreover, the benefits could evolve over time.

There are also outstanding questions about whether fintech firms would be more willing to use APIs (or data feeds) than the current screen scraping methods if they were charged an access fee. While APIs would be a more efficient mechanism for collecting data — they would eliminate the need to reconfigure screen scraping tools in response to changes in the layout of online banking webpages — current methods used to access data do not involve

access fees (but have other technology-related costs). If the Australian Government elected to mandate the use of APIs and allowed financial services firms to charge access fees (such as on a cost-recovery basis), it might also be necessary to consider whether there are sufficient incentives for fintech firms to adopt API-based approaches for accessing data.

Overseas initiatives, such as the Open Banking Standard and the Payment Services Directive 2, could provide insights into the potential costs and benefits of such a scheme in Australia, and it would be worthwhile for the Australian Government to monitor the outcomes of those initiatives. Notwithstanding the above discussion, however, the issue of a right to personal data portability (chapter 9) is much broader than just financial data portability.

Product reference information

In addition, the issue of access to publicly available general reference information about products and services, such as fees and charges, was raised by inquiry participants. This data is already published by Authorised Deposit-taking Institutions and credit providers to meet their disclosure obligations, typically on webpages and/or in Product Disclosure Statements. The issue, however, appears to be related to the *ease* with which third parties can collate this data.

Westpac Banking Corporation (sub. 197), for example, recommended that the Australian Government mandate the provision of such information in a standardised form to facilitate easier access (in addition to current disclosure requirements). It was also recommended that the mechanism for provision of this data be discussed further between industry and government, and that this should include the possible use of APIs.

ANZ (sub. 64) provided a counter viewpoint, arguing that recent changes to disclosure requirements by ASIC — intended to facilitate greater use of digital mediums — will likely push many financial services businesses towards providing this information through digital mediums.

The Commission considers that making such data available through APIs would almost certainly facilitate more efficient comparison of financial services and products, leading to greater competition and improved consumer outcomes. However, it is not clear how material these benefits would be for customers. Moreover, the costs of building APIs could be relatively large, and would have a relatively larger impact on smaller financial firms. That said, there could be less expensive technological solutions that would be fit for purpose — such as publication of this data in a CSV document (with an industry agreed layout/format).

In summary, given the potential benefits arising from easier access to product reference information, there is probably a strong case for financial businesses providing such information in digital formats, particularly if this could be done using relatively

inexpensive methods. The Australian Government could consider amending disclosure requirements to achieve this outcome.

Release of public data

As noted elsewhere in this report, governments around Australia hold multitudes of data, and as noted in this appendix, the availability of particular public sector data could improve the efficiency of financial markets. In this sense, governments can use the release of data as a lever to influence the operation of markets for financial services.

Participants to the inquiry identified several specific datasets that could be beneficial to particular parties, and to the operation of markets for financial services more broadly. For example, Veda (sub. 163) indicated that identity verification (of prospective customers) would be more straightforward if they had greater ability to access and use public data, such as electoral rolls (held by the Australian Electoral Commission).

As another example, the Australian Centre for Financial Studies (sub. 103) suggested that loan-level data on mortgages underlying mortgage-backed securities should be made available as it would be valuable for financial market participants (such as investors) as well as researchers.

E.4 Conclusions

The digital age — new data sources and increased technical capacity to analyse existing and new data — is transforming the financial sector, bringing with it innovation, competitive pressures, more efficient decision making by financial service providers and more empowered consumers.

Much of the change is being driven by market forces. New entrants with innovative business models are challenging incumbent firms, customers are demanding — and in some cases receiving — greater access to data about themselves, and third-party intermediaries are entering the market to provide new intermediary services between finance sector firms and their customers.

CCR has the potential to reduce one of the main sources of inefficiency in the financial sector — the information asymmetries between borrowers and lenders. Although levels of participation in CCR are not yet high, it is quite possible that participation will increase significantly over time, in line with the experience of other counties (such as New Zealand). The risk of adverse selection for non-participants is likely to be a driving force for participation.

However, even outside the CCR, innovative businesses are using non-traditional datasets (such as tenants' on-time rent payments and social media) and data analytics to overcome

these information asymmetries and, in so doing, are increasing access to credit (or more affordable credit) to those without an existing credit record, such as young adults.

There is little information on the number of customers that would be likely to utilise their transactions data — let alone the value they would place on it — if APIs were mandated for the financial sector. However, there would be implementation costs for financial service providers and these would be relatively more burdensome for smaller providers. Moreover, there would also be potentially greater risks for the security of customer data as a result of the broader access to it, and possibly reduced incentives for data custodians to invest in data generation and protection.

There is clear interest by finance sector businesses as well as researchers in gaining access to a wider range of public sector data than is currently accessible. The two main apparent barriers are privacy considerations and the potential for 'commercial detriment' to some financial service providers (particularly for data collected by finance sector regulators). APRA noted the difficulties in weighing public benefit and commercial detriment in determining what datasets to release, despite its thorough processes for informing such decisions. Explicit regulator mandates for increasing competition could help to address this situation.

F Case Study: Data from your Internet activities and intelligent devices

Key points

- Data that is generated by the use of social media, mobile devices such as phones and tablets, the Internet of Things and wearable devices such as smart watches and fitness trackers has emerged as a massive and important source of information on individuals and their activities.
  - Some of this data (such as the user's name and email address) is provided intentionally by users in their use of these products.
  - But much of the data derived from use of these products (such as the metadata behind photos posted online or the precise location of the user of an app) is collected either as a by-product of their use or in ways that would not be obvious to the product user.
- The terms of use for many social media sites give rights to users for the content that they generate. However, these rights usually do not include the right to exclusive use.
  - By using the services, users generally agree to give social media organisations an exclusive (with the exception of the user), royalty free licence to use the content.
- Rights to data generated by wearables, such as smart watches and fitness trackers, are similarly 'shared' between the individual wearing the device and the supplying business.
  - Fitness trackers have been used by health insurance companies as a means to obtain greater information about a customer's lifestyle, and to price policies accordingly. They have also been used overseas to support or defend legal action.
- Privacy laws in Australia apply to companies collecting data generated through the use of social media, mobile and intelligent devices where those companies have an 'Australian link' — such as where they carry on business in Australia and collect personal information from people who are physically in Australia.

Rapid expansion in Internet connectivity and the proliferation of sensor technology in consumer and business products and in infrastructure over the past 10 years in particular, have dramatically increased both the sources of data and the volumes that can be collected. This appendix describes some of these comparatively new sources of data and the issues associated with collecting and using data generated by the data.

F.1 Devices and social media — what data do they generate?

Devices and wearables that generate data

'Mobile devices' that generate data include mobile phones, tablets, laptops, and wearable technology (or wearables). Wearables can include clothing, jewellery or other accessories that incorporate electronic and computer technology. Traditional mobile phones and smartphones generate data such as call logs, text data, and location data. Smartphones also generate data via the use of applications (or 'apps') and financial payment mechanisms. Wearables can generate similar data, depending on their type, and particularly popular are fitness trackers, which capture an array of data about the activities of the wearer. Some of this data is provided deliberately by users, while other data is a by-product of the main activity for which the device is used.

Approximately 94% of the adult population in Australia use a mobile phone, and roughly three-quarters of these are smartphones ((ACMA 2015), Deloitte (2015)). With the second highest uptake of smartphones worldwide (behind South Korea) (Poushter 2016), Australians are potentially providing and generating more data about themselves via mobile devices than consumers elsewhere around the world.

Similar to smart phones and tablets, wearable technology (or wearables) generate data through their interaction with the Internet. Smart watches, such as the Apple Watch and devices using Google's Android Wear (a version of the Android operating system that works on smart watches), are wearable technology that can be used to operate a number of apps, and collect data generated by users. For example, smart watches can be used to receive notifications of incoming emails, text messages, and other communications such as Tweets. In the payments sphere, Visa has recently introduced payWave, which permits wearable users to make payments without requiring a physical card. MasterCard has similar technological capabilities offered under MasterCard Contactless. These devices use near field communication technology, enabling the device to act as a proxy for a card.

A recent survey of over 1000 people estimated that more than 2 million Australians possess a wearable device. Of the 944 000 wearables sold in Australia in the second half of 2015, roughly three-quarters were smart wristbands, such as the devices manufactured by Fitbit, Jawbone, and Garmin, rather than premium-priced smartwatches, such as the Apple Watch (Telsyte 2016). Worldwide, Fitbit is the most commonly sold wearable technology (IDC 2016).

The Internet of Things (IoT) comprises devices equipped with sensors used to collect data, which can then be reported, communicated to other devices in a network (for example, via a wireless network), or acted upon. As noted in chapter 3, smart electricity meters and refrigerators equipped with sensors that allow them to perform tasks such as monitor usage and identifies when food items have expired can be regarded as IoT devices. Products as diverse as aircraft engines and rail and road infrastructure are increasingly having IoT

technology built into them. A prominent application of the IoT has been in logistics management, where truck fleets have been fitted with devices that help minimise fuel consumption, and which improve safety by providing vehicle diagnostics. Wearables may also be classified as a category of IoT devices.

Types of data generated and collected

Since the advent of smartphones and tablets, the number of apps that can be used on such devices has rapidly increased. There are now a plethora of apps that facilitate the creation and sharing of numerous forms of data. Apps may generate or collect a wide variety of data, including on a person's location, address book and contacts, photographs, consumption and preferences, activities, and health. The information collected through apps is likely to be considerably more useful for secondary purposes than it once was — five years ago, games were the most popular app downloaded on smartphones; more recently its maps and navigation apps (table F.1). That the most popular apps in Australia relate to maps and navigation suggests that Australians are providing a considerable amount of personal location information to app owners.

Table F.1 **Popular categories of smartphone apps in Australia**
Apps used in preceding six months, percentage of respondents[a]

Type of app	*2011*	*2012*	*2013*	*2014*
Maps and navigation	55	**74**	**80**	**82**
News and weather	57	73	72	72
Games	**79**	74	64	66
Photos, videos and films	39	56	61	62
Instant messenger and social networking	46	27	52	61
Music	na	50	48	51
Search	26	53	45	43
Eating out	28	30	33	34
Shopping	30	35	34	34
Managing money	17	29	32	33
Travel	29	31	32	31
Health and wellbeing	23	23	28	27
Books	30	27	27	23
Time management	21	20	24	21
Education	21	18	17	15
Business	19	19	20	14

[a] A total of 81% of all survey respondents reported successfully downloading and installing apps on their mobile phone. The most popular category for each year is bolded. **na** Not available.

Source: Mackay (2014).

The motion sensors and GPS technology used in smartphones are also capable of capturing data that is indicative of the general health and wellbeing of the phone user. Apple Health and Google Fit apps are both capable of working in conjunction with other apps (including third party apps), as well as wearables (including Android Wear and Apple Watch), to record detailed health and fitness information such as steps taken or pulse levels. Apple Health also has a feature allowing people to set up a medical ID on their phone, which can be displayed in case of an emergency, and shares background medical information such as the presence of any medical conditions, allergies and reactions, and medications. Other apps that collect information on the health and fitness of the device user include Runtastic, DailyBurn, MapMyRun, Glow, and MapMyFitness. Fitness trackers are similarly capable of capturing an array of data on their wearers' activities, such as steps taken during the day, distance travelled, heart rate, and data relating to sleeping patterns.

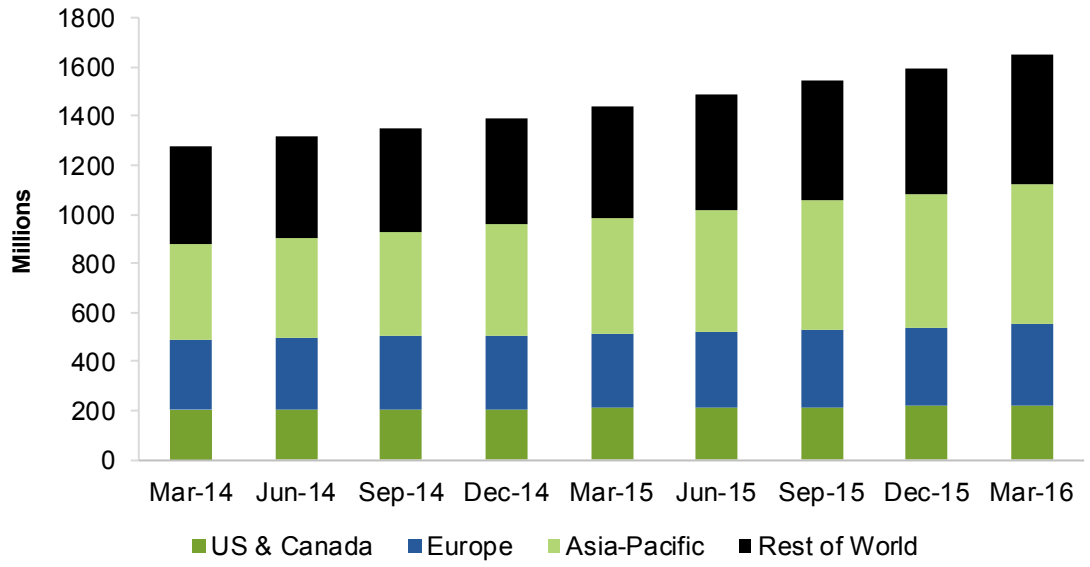
Data generated by devices on the IoT varies with the nature of the device. Typically (with the exception of wearables), data may reflect the activity, location, status (on or off) or performance of a device. Such data may reveal information about the individual person or business who owns the device. For example, a device in your car that conveys data to the manufacturer on the performance of your car as it is being driven is not particularly revealing of the characteristics of the driver (although your insurer may in some circumstances have a great deal of interest in it). If the device also reveals where your car is at every point in time, how many people are in the car, and the reaction speed of the driver, this does convey information that some may consider more personal.

Social media data

Social media are web-based environments that readily allow users to create, publish, and share content amongst each other. Around 68% of Australian consumers and 33% of businesses have a social media presence (Sensis 2015). Facebook is currently the most prominent social media company, with over 1.6 billion monthly active users (figure F.1), nearly 1.1 billion of whom use the service daily (Facebook 2016d).

In Australia, it is estimated that there are currently over 15 million monthly active Facebook users (Cowling 2016), with it being the most commonly used social networking site by both consumers and businesses (table F.2). Consumers use social media primarily to catch-up with family and friends; business use it primarily for two-way communication with clients and contacts (Sensis 2015).

Figure F.1 **Facebook monthly active users (global)**



Source: Facebook (2016c).

Table F.2 **Social networking sites used by Australian consumers and businesses**

2015, percentage of respondents[a]

Site	*Consumers*	*Small business*	*Medium businesses*	*Large businesses*
Facebook	93	94	89	89
LinkedIn	28	19	19	36
Instagram	26	11	13	11
Google+	23	9	4	20
Twitter	17	17	38	46
Pinterest	17	3	2	4
Blog	na	6	8	7
Snapchat	15	na	na	na
Tumblr	5	na	na	na
Vine	3	na	na	na
Yelp	3	na	na	na
Foursquare	1	na	na	na

[a] Survey of 800 consumers and 1100 businesses (of which 793 were small, 207 medium, and 100 large).

Source: Sensis (2015).

The range of data collected by social media, and Facebook in particular, is extensive. In its data policy, Facebook states that it may collect:

- information in or about the content you provide, such as the location of a photo or the date a file was created (that is, metadata)

- information about how you use Facebook services, such as the types of content you view or engage with or the frequency and duration of your activities
- contact information you provide if you upload, sync or import this information (such as an address book) from a device
- information about a purchase, transaction or donation made through Facebook, including payment information
- information from or about the computers, phones, or other devices where you install or access Facebook (depending on the permissions the you have granted), including specific geographic locations
- information about you from third-party websites, apps and partners of Facebook (Facebook 2015a, pp. 1–2)

All this data collected by Facebook is not kept in isolation — Facebook links data collected about you from different devices to form a more complete picture of you and your activities.

Twitter allows users to post short text messages, photos and short videos online and also collects data from users. Besides basic account and contact information, a user may provide Twitter with a short biography, location data, date of birth, and photograph. While the messages a person Tweets are classified by Twitter as public information, so too is the metadata provided with Tweets, the language and time zone associated with an account, and other information, such as the lists a user creates and the others they follow (Twitter 2016b).

At the end of March 2016, Twitter had 310 million monthly active users. Nearly 80% of all Twitter accounts originated from outside the United States (Twitter 2016a). CSIRO Data61 estimated that there are 2 to 3 million active Twitter users in Australia (pers. comm., 19 August 2016).

How obvious is the data collection?

There are three main avenues through which devices, apps and social networks collect data:

- Direct requests to the user.
- Indirectly from the user, as the device, app or service is being used.
- Indirectly from the user or their device, separate to the individual's use of the device, app or service.

In intentional provision, a user knowingly volunteers the data in question — for example, when a user signs up for an app, or for a service such as Facebook or Twitter, they will be required to provide basic personal information such as their name and email address.

Alternatively, the provision of user data could be inadvertent, perhaps because it is related to the nature of the product being used. For example, navigational apps will collect

location data from the device of a user — the user knows this, and the collection of the data is necessary for the service to be provided. As such, it is a byproduct of the ultimate objective of receiving navigational assistance. Similarly, the popular gaming app Pokemon GO collects location data from users' devices, which is necessary for participation in the game.

Other data may be provided to organisations such as device manufacturers or service providers through the use of devices, and without the full knowledge of consumers unless they have fully read the relevant privacy policy or terms of use. Evidence presented in chapter 5 suggests that not all consumers actually read privacy policies and terms of use. In many cases — such as where cookies are used — data collected may not be related to the primary purpose of the app. A number of providers of technology services, business and government websites and apps state in their privacy policies and terms of use that they collect data from devices via the use of cookies (box F.1).

Box F.1 **Cookies**

Cookies — also known as web cookies, browser cookies, or Internet cookies — are small text files that websites place on a device to store information about the user's browsing preferences. The files usually contain a unique user ID and the name of the site. The first time a user visits a website with cookies, a cookie is downloaded onto the user's device. The next time the user visits the website, the device checks to see if there is a relevant cookie, and then sends the information contained in the cookie back to the website. Cookies are used by a large number of modern websites.

Cookies can enhance a user's browsing experience by remembering a person's preferences, and can allow people to avoid signing into a site each time they visit. Cookies can be particularly useful for online shopping, as they can be used to target advertising. Cookies may be either session cookies, which are erased when a use closes their browser, or they may be persistent, remaining on a person's device for a pre-determined period of time. Cookies can also be first-party, in which case they are set by the web server of the visited page, or they may be third-party, and are stored by a different domain to the domain of the webpage visited.

However, cookies have been controversial due to their ability to allow website operators to track the browsing behaviour of users. Internet users have the option of blocking cookies, although this may reduce the functionality of some websites, and may prevent users from viewing some websites altogether.

In Europe, the ePrivacy directive requires prior informed consent for storage or for access to information stored on a user's device — essentially, website users must be asked if they agree to most cookies and similar technologies before a website starts to use them. Once a user has consented to cookies, and has been told what the cookies do and why, the process does not have to be repeated every time the same user visit the website.

Sources: BBC (2014); EC (2016); Microsoft (2016a).

Facebook is an example of a social media provider that makes significant use of cookies:

> We use cookies if you have a Facebook account, use the Facebook Services, including our website and apps (whether or not you are registered and logged in), or visit other websites and apps that use the Facebook Services (including the Like button or our advertising tools). (Facebook 2016a, p. 1)

Facebook also makes use of other tools such as plug-ins (box F.2) in conjunction with cookies. The use of these tools is in no way unique to Facebook; they are just more open and explicit about their approach.

Box F.2 Social plug-ins

Facebook lists its social plug-ins as the 'Like' button (which users can click on to share and connect with things from other websites that appeal to the user), the 'Share' button (which allows a user to write something about a link and post it to their Timeline), embedded posts (with which a user can add a public post to their blog or website), and their comments box (which allows a person to publicly comment on another website using their Facebook account).

Facebook also uses cookies, pixel tags (or 'pixels', since they are single pixel GIF images), device or other identifiers, and local storage in the course of offering its services. Pixels are small blocks of code on a website or app that allow the reading and placement of cookies, and the transmission of information to Facebook or its partners. This allows Facebook to receive information such as a device's IP address, the time of viewing of a pixel, the type of browser being used, and an identifier associated with the browser or device. Local storage enables a website or app to store and retrieve data on a computer, mobile phone, or other device.

Sources: Facebook (2016a, 2016b).

F.2 Uses of device and social media data

Social media

Use of social media data for targeted advertising

The primary commercial purpose of social media data is targeted advertising. The type of information collected on users, such as age, gender, location and interests enable marketers to better target their advertising to consumers.

Facebook generates its revenue by using the information collected from users via social media and tools such as plug-ins and cookies, to sell advertising placements to marketers. About 96% of Facebook's revenue is derived from advertising, with roughly half received in the United States and Canada and the remainder in the rest of the world (Facebook 2016d).

In discussing how this information is used and shared, Facebook explains:

> When we have location information, we use it tailor our Services for you and others, like helping you check-in and find local events or offers in your area or tell your friends that you are nearby. (Facebook 2015a, p. 3)
>
> … we use all of the information we have about you to show you relevant ads. We do not share information that personally identifies you … with advertising measurement or analytics partners unless you give us permission. We may provide these partners with information about the reach and effectiveness of their advertising without providing information that personally identifies you, or if we have aggregated the information so that it does not personally identify you. (Facebook 2015a, p. 5)

There are three key ways that Facebook data is used by marketers:

- *Custom audiences* are a group of existing customers that an advertising business can target. Specifically, custom audiences enable advertisers to use information collected outside of Facebook to target advertisements to individuals on Facebook. This can be done on the basis of a 'Customer List', 'Website Traffic', or 'App Activity'. An advertiser who chooses Customer List will be prompted to share a list of email addresses, phone numbers, and Facebook user IDs or mobile advertiser IDs with Facebook. Once the list has been created, advertisers are able to target or exclude individuals in their future Facebook advertising campaigns.

  Alternatively, the Website Traffic option requires advertisers to install a specialised 'Custom Audience Pixel' on its website. The pixel (box F.2) then enables an advertiser to automatically target advertisements to a Facebook user who visits the webpage where the pixel is installed. App Activity works in a similar fashion to Website Traffic, but allows advertisers to target their advertisements on the basis of actions Facebook users have or have not taken within an application. For example, advertisers can choose to target people who have used their app and loaded an item into their shopping cart, but not executed a purchase. Advertisers using the custom audience features can further choose to narrow down their audiences through using variables such as location, age, gender, and interests (Van Alsenoy et al. 2015).
- *Lookalike audiences* are Facebook users who are similar to those included in a custom audience. Lookalike audiences are created on the basis of common qualities with the custom audience such as demographics and interests. Facebook uses algorithms to identify a larger segment of Facebook users who are similar to those targeted in a specific custom audience (Van Alsenoy et al. 2015).
- *Atlas* was acquired by Facebook in 2013 and functions as an advertisement serving, management and measurement platform. By using cookies, Atlas endeavours to match individuals with devices. Atlas also uses a similar approach to custom audiences to try and link online advertising with offline purchasing behaviour (Van Alsenoy et al. 2015).

Community attitudes towards advertising based on social media are mixed. A minority of surveyed individuals are positive about targeted ads on social networks, most ignore some or all ads on social media sites, and some individuals indicated that they are put off by companies that advertise on social media (table F.3).

Table F.3 **Attitudes to advertisements on social networking sites**
2015, percentage share of respondents

Statement	*Agree*	*Neutral*	*Disagree*
I'm turned off by companies or brands that advertise on social network sites	31	32	37
I take no notice of the ads on social network sites	55	20	25
I ignore sponsored posts from businesses I do not follow	72	15	14
I sometimes click on ads I see on social network sites to find out more	42	13	46
I like sponsored posts from businesses I follow on social networks	32	26	43
I'm quite happy to see ads on social network sites	38	23	39

[a] Base is total number of social media users in survey = 539.

Source: Sensis (2015).

Similar to Facebook, Twitter receives the vast majority of its revenue from third party advertising on its platform. Indeed, the company received roughly 90% of all of its revenue in 2014 and 2015 from advertising. Advertising revenue is generated via the use of three promoted products: *promoted tweets* (which appear in a user's timeline or search results, and are based on Twitter's understanding of their interests), *promoted accounts* (appear in the same place and format as suggestions on which accounts a user should follow), and *promoted trends* (when a user clicks on one of these, search results for that trend are shown in a timeline and a Promoted Tweet created by advertisers is displayed to the user at the top of the search results).

Use of social media data for hiring decisions

Social media data can assist in the screening of job applicants. Some businesses and government agencies use the social media postings of job applicants to better assess the merit of potential employees. Indeed, in the United States, there have been instances of employers requesting job applicants' Facebook login and password details during interviews so that their posts can be read (Chang 2012).

A survey of more than 400 employers located in Australia and New Zealand found that 62% of employers used social media sites to check on prospective employees. When asked if they thought it was fair to use a candidate's social media postings to determine their suitability for a job, 50% of businesses said that it was not fair. In all, 25% of hiring managers surveyed said that they had rejected candidates based on their social media postings. Of these, a majority had done so because they believed the candidate did not suit the culture of the organisation, while others were rejected due to inappropriate comments or photographs, or due to inappropriate comments about their current or a previous employer (Robert Walters 2013).

Use of social media data by emergency services

Social media data has increasingly been employed by emergency services authorities in a number of countries to improve the effectiveness of public communication during emergency events.

For example, in May 2010, the Queensland Police Service began a trial use of social media accounts on Facebook, Twitter, and YouTube, with the aim of opening up a two-way conversation between the Police Service and the public, and developing an online community of followers prior to the occurrence of disasters. During significant flood events in December 2010 and January 2011, the Queensland Police Service streamlined their drafting, clearance, and release processes for information, and gravitated towards social media as the best means for reaching the public in the shortest possible timeframe (QPS nd).

Some emergency service providers have used social media as a way to exchange data with the public, providing opportunities to disseminate information on the extent to which locations are affected by disasters, and notifications to stay away from certain areas while emergency procedures are in place.

Facebook's Safety Check

In October 2014, Facebook introduced Safety Check, a tool designed for use in disasters and other emergency situations to share data about the location and safety of users. When the tool is activated, and if a person is in an affected area, they receive a Facebook notification asking if they are safe. If they are safe, they can elect an option to indicate this, and choose to publish a news feed story indicating their status. A person with friends in the disaster area will receive a notification about those friends (Gleit, Zeng and Cottle 2014).

Facebook submitted to this inquiry that it had activated Safety Check in disasters 20 times in 2016 alone. In 2015, following the earthquakes in Nepal, 8.5 million people were marked safe, and 150 million people were notified that a friend was safe. In all, Facebook claims that over one billion people have been reached by Safety Check following a crisis (sub. 172). Karsten and West (2016) argued that Safety Check has several advantages over traditional government response mechanisms to crises, including that the tool increases situational awareness and warns others to stay away from danger zones, provides an alternative to phone lines (which may be unreliable during a crisis) and also that the tool provides a means for global communication that does not rely on what may be chaotic media reporting.

Uses of wearables and mobile data

One of the main applications of fitness wearables data so far has been to refine health product and service offerings — health insurance in particular. Since fitness trackers can

compile data on metrics such as a person's heart rate and physical activity levels over time, it is conceivable that at some point they may be used regularly by health professionals, in conjunction with other data and health measurements (box F.3).

Box F.3 Using fitness tracker data in health services

In one case in the United States, a man presented to a hospital emergency department following a seizure. The patient was diagnosed with an abnormal heart rhythm. However, as the patient did not display symptoms, it was not possible from medical assessment alone to determine an onset time for his abnormal heart rhythm. The patient happened to be wearing a Fitbit that was synchronised with an application on his smart phone, and which recorded his pulse rate. Medical staff accessed the application, and were able to determine that the patient's arrhythmia had begun three hours earlier, allowing treatment via electrical cardioversion. The medical professionals involved in the case concluded:

> To date, activity trackers have been used medically only to encourage or monitor patient activity, particularly in conjunction with weight loss programs. To our knowledge, this is the first report to use information in an activity tracker-smartphone system to assist in specific medical decision making. The increased use of these devices has the potential to provide clinicians with objective clinical information before the actual patient encounter. (Rudner et al. 2016, p. 3)

However, some have questioned whether the data generated by such devices has wider applicability, for instance, use by health professionals. Rosenblum (2015) quoted one doctor in the United States as saying:

> I'm an oncologist, and I have these patients who are proto 'quantified self' kinds of people … They come in with these very large Excel spreadsheets, with all this information — I have no idea what to do with that. (p. 1)

Rosenblum also pointed out that, in the United States, fitness trackers have not been clinically validated to perform at the same standards for reliability that the Food and Drug Administration uses for medical devices, such as devices that measure blood pressure. However, at least one wearable device on the market aims to be a medical quality device. Embrace, manufactured by Empatica, sends out an alert when an unusual event, such as a seizure, happens to the wearer. It goes via the smartphone to a person's caregiver, roommates or parents, enabling those people to check on the person.

Some doctors have also argued that wearable technology has the ability to improve the efficiency with which healthcare is delivered in the future. Comite (2015), an endocrinologist, argued that data from fitness trackers has the ability to help in the delivery of precision medicine, by providing a detailed log of various aspects of a person's health over time. In turn, this could help cut down on unnecessary office visits and testing, and permit a clearer observation of breakdowns in the body's systems before they become readily apparent.

Use of data in health insurance

In December 2014, US insurance start-up Oscar offered members a free wearable fitness tracker — members could sync their number of steps with Oscar's app to get credit for their activity. Upon reaching their daily goal, members would receive a credit in Amazon gift card vouchers (Oscar 2014). In April 2015, US insurance company John Hancock announced a program under which new policyholders would receive a health review with

personalised health goals and be provided with a Fitbit to keep track of their progress. When a policyholder completed health related activities, they receive points toward rewards and discounts at selected retailers of up to 15% of their premium (John Hancock 2015).

In Australia from March 2016, Qantas and insurer nib have partnered to provide health insurance (Qantas Assure) for Qantas frequent flyers. The product allows Qantas frequent flyers who take out a Qantas Assure policy and use the associated app with a wearable device (such as a Jawbone or Apple Watch) to receive points for meeting fitness challenges. There are approximately 11 million Qantas frequent flyers, and Qantas has announced that they are targeting a 2–3% share (on a revenue basis) of the Australian private health insurance market in the first five years of the life of the product (Qantas and nib 2015).

Use of data for health management and research

Other potential benefits of the data generated by wearable technology include proactive health engagement and a focus on preventative measures, leading to earlier rectification of health problems. The data generated at an individual level could also be integrated to reduce inefficiencies in the healthcare system, such as unnecessary laboratory tests, and to simplify the management of chronic conditions such as diabetes and heart disease. Integrated data sets may also assist in the identification of population-level health issues earlier and more efficiently than clinical trials (Patterson 2013).

The health and fitness data collected via other (nonwearable) mobile devices is being used for health research and to improve health outcomes. In March 2015, Apple announced the introduction of ResearchKit in the United States, an open-source software framework for medical and health research, in which participants can chose to voluntarily share medical and health information for research purposes. In-built iPhone features, such as the gyroscope, GPS, and accelerometer are used to collect data on such variables as activity levels and motor function. The primary objective of ResearchKit is to make it easier for medical researchers to obtain data by enrolling more study participants via iPhones and other devices. To some extent this removes the need for study participants to be located close to the research group or a researcher and enables greater pooling of data from various locations (adding more texture to medical datasets). Six months after its launch, over 100 000 people were using ResearchKit apps (McGarry 2015). A number of apps have been developed using ResearchKit to obtain data.

mPower is an app to assist with the management of Parkinson's disease, designed by Sage Bionetworks and the University of Rochester Medical Centre. The app uses features of the iPhone, such as the touch screen, GPS, and motion sensors to capture real-time data on dexterity, gait and balance. Tracking disease symptoms and how these vary with time and medication supports research into the disease and its treatment. The app has more than 12 000 registered users across the United States, and more than 9500 users have consented to have their information shared for research. In March 2016, Sage Bionetworks (which

uses ResearchKit but also hosts data for other mobile health projects) released a large tranche of data consisting of millions of data points for researchers to use (URMC 2016).

The American Sleep Apnea Association and IBM have jointly created an app called SleepHealth, with the objective of creating the world's largest longitudinal dataset on healthy and unhealthy sleepers, with the aim of publication as an open dataset to be shared with other researchers. The app makes considerable use of the sensors in the Apple Watch to detect movements, orientation, and heart rate during sleep. It is hoped that after several years of data collection, the resulting research will lead to the development of personalised and public health interventions for the treatment of sleep disorders (IBM 2016).

23andMe, a genomics and biotechnology company helps individuals to identify (for a fee) whether they have susceptibilities to any genetic conditions. Users can choose to make their data available to researchers, and the company claims that, on average, a user who shares their data contributes to over 230 studies (23andMe 2016).

CareKit (like ResearchKit) provides a platform from which app developers can create new offerings. However (unlike ResearchKit) the primary function of CareKit is to assist patients to manage their medical conditions. While relatively new (the platform started in the United States in March 2016), one app that has already been launched on the platform is a diabetes management app called One Drop. Users of the app can record their blood glucose levels, foods consumed (and have their carbohydrate intake measured), activity levels, and medications taken. If needed, this data can then be shared instantaneously with a person's care team (IDS 2016).

Use of data in legal cases

There have been examples of fitness trackers being used to provide evidence in court cases. For instance, in a personal injury case in Canada in 2014, a plaintiff's attorney used Fitbit data to support his argument that his client, who was a personal trainer in peak physical shape prior to a car accident, had suffered a significant decline in physical activity (Patton, Wetmore and Magill 2016).

Uses of Internet of Things data by governments and product manufacturers

Government uses

Governments have begun exploring applications of the IoT and implementing this technology. As observed by Meyers, Niech and Eggers (2015), the IoT has potential uses by governments to collect better data on how effectively public programs and policies are addressing their objectives, in addition to assisting governments to deliver services based on real-time or near real-time data.

One of the early areas of government activity in the IoT space has been in the development of 'smart cities'. This has focused on elements such as improving traffic flows, providing services such as street lighting, water management and waste management. For instance, the City of Melbourne is piloting a network of 50 rubbish bins equipped with sensors that report to rubbish truck operators when the bins reach 70% capacity (Gutierrez 2016). The ACT government is similarly trialling public rubbish bins that provide real-time data on fullness levels to waste collectors (ACT Territory and Municipal Services 2016).

The United States Government has explored applications of the IoT in a number of areas, including national defence. 'Network centric warfare' permits the US military to provide a shared awareness of the battlefield for its forces — for instance, military bases collect a variety of data using connected devices including cameras, infrared sensors and chemical detectors. Data is also collected using drones, surveillance satellites, and ship and ground stations, as well as by military personnel. Data can be used for purposes such as helping ground forces navigate unfamiliar terrain, and to be aware of the presence of threats and targets within striking range. Defence contractors are also undertaking research into 'smart skin', which covers the fuselage of aircraft with thousands of sensors, enabling a wide range of data to be transmitted from the aircraft in real time (Castro, New and McQuinn 2016).

Intelligent Transport Systems (ITS) have been another area of government use of IoT capabilities. ITS technology allows vehicles and infrastructure to transfer data across systems, with the objective of improving safety, productivity and environmental performance. For example, Infrastructure Australia (2016) noted that the installation of electronic signs and additional CCTV cameras on a section of the Monash Freeway in Melbourne allowed 16 to 19% more people to travel in each lane of the freeway, equivalent to an additional 0.5 to 0.8 lanes, but delivered for substantially lower cost than the construction of a new lane.

Uses by product manufacturers

An increasing number of product manufacturers are using IoT capabilities to collect data on product quality on production lines, and for monitoring products for signs of required maintenance prior to breakdowns. General Electric, for example, has embraced the IoT as part of its commercial strategy, embedding IoT capabilities in its products and using these capabilities to provide and facilitate services after the manufacture of industrial goods.

Some pharmaceuticals manufacturers have begun using optical sensors to collect data continuously on product quality (PwC 2015). Prior to the advent of optical sensors, quality inspection typically had to occur via random sampling, a less comprehensive inspection method.

IoT technology has also been used to monitor the safety of manufacturing processes. Spanish manufacturing company Polibol, which specialises in the manufacture of printed coils and aluminium laminated plastics used for flexible packaging, has used IoT to collect

data on environmental variables and critical processes. The company uses sensor network technology in its plants to collect data and monitor the air temperature around printing machines and pipes, light intensity, and C02 concentration in areas occupied by workers in real time. Sensors are also used to measure volatile organic compound readings, which allows plant operators to ensure that solvents retained in ink or adhesive that come into contact with food remain below the minimum levels of tolerance permitted (LCD 2015).

Harley-Davidson has installed software in its motorcycle plant in York, Pennsylvania that keeps data on how well production equipment is working. Software automatically adjusts machinery if sensors detect that a variable such as humidity or fan speed have deviated from the acceptable range (Lopez Research 2014).

Rolls-Royce and Microsoft announced an agreement in 2016 for the utilisation of IoT capabilities in aircraft engines. Using technology developed by Microsoft, Rolls-Royce's aircraft engines will have sensors placed inside them to collect data on variables such as engine health and fuel usage, enabling for easier detection of operational anomalies and trends. This will allow for earlier detection of potential problems (with the scope to reduce flight delays) and improved fuel efficiency (Microsoft 2016b).

IoT in manufacturing has the ability to deliver benefits to the consumers of manufactured products, through the collection of data to improve product safety and design, in addition to helping manufacturers increase the efficiency of their production processes.

F.3 Issues in data use and collection

Rights to social media data and commercial terms

Social media allows users to generate content, but then requires it to be posted on a platform owned and maintained by another organisation, raising the matter of ownership of the data generated by users. Social media companies often state in their user conditions or statement of rights and responsibilities that content and information posted and created by users belongs to users. However, companies also state that they then have a licence to use such content in a manner consistent with their commercial interests. Facebook's statement of rights and responsibilities specifies:

> You own all of the content and information you post on Facebook, and you can control how it is shared through your privacy and application settings. In addition:
>
> 1. For content that is covered by intellectual property rights, like photos and videos (IP content), you specifically give us the following permission, subject to your privacy and application settings: **you grant us a non-exclusive, transferrable, sub-licensable, royalty-free, worldwide license to use any IP content that you post on or in connection with Facebook** [emphasis added] (IP License). This IP License ends when you delete your IP content or your account unless your content has been shared with others, and they have not deleted it.

2. When you delete IP content, it is deleted in a manner similar to emptying the recycle bin on a computer. However, **you understand that removed content may persist in backup copies for a reasonable period of time** [emphasis added] (but will not be available to others).
3. When you use an application, the application may ask for your permission to access your content and information as well as content and information that others have shared with you
4. When you publish content or information using the Public setting, it means that you are allowing everyone, including people off of Facebook, to access and use that information, and to associate it with you (i.e. your name and profile picture) … (Facebook 2015b)

Furthermore:

> Our goal is to deliver advertising and other commercial or sponsored content that is valuable to our users and advertisers. In order to help us do that, you agree to the following:
>
> 1. You give us permission to use your name, profile picture, content, and information in connection with commercial, sponsored, or related content (such as a brand you like) served or enhanced by us. This means, for example, that **you permit a business or other entity to pay us to display your name and/or profile picture with your content or information, without any compensation to you** [emphasis added] …
> 2. We do not give your content or information to advertisers without your consent.
> 3. **You understand that we may not always identify paid communications and services as such** [emphasis added]. (Facebook 2015b)

Therefore, although Facebook users retain intellectual property rights to their content, users agree to Facebook having considerable scope to use that content for commercial purposes, including in advertising, but not disclosing that such communication is paid for. Similarly, Twitter's terms of service specify:

> You retain your rights to any Content you submit, post or display on or through the Services. By submitting, posting or displaying Content on or through the Services, you grant us a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display, and distribute such Content in any and all media or distribution methods (now known or later developed) …
>
> You agree that this license includes the right for Twitter to provide, promote, and improve the Services and to make Content submitted to or through the Services available to other companies, organizations or individuals who partner with Twitter … Such additional uses by Twitter, or other companies, organizations or individuals who partner with Twitter, may be made with no compensation paid to you with respect to the Content that you submit, post, transmit or otherwise make available through the Services. (Twitter 2016c, p. 4)

YouTube's terms of service specify:

> … you retain all of your ownership rights in your Content. However, by submitting Content to YouTube, you hereby grant YouTube a worldwide, non-exclusive, royalty-free, sublicenseable and transferable license to use, reproduce, distribute, prepare derivative works of, display, publish, adapt, make available online or electronically transmit, and perform the Content in

> connection with the Service and YouTube's (and its successors' and affiliates') business, including without limitation for promoting and redistributing part or all of the Service (and derivative works thereof) in any media formats and through any media channels … YouTube may retain, but not display, distribute, or perform, server copies of your videos that have been removed or deleted. The above licenses granted by you in user comments you submit are perpetual and irrevocable. (YouTube 2010, p. 1)

Hence, the standard for terms of service agreements in social media is generally that users retain intellectual property rights over content they have created, but that by using the services of a particular social media network, users agree that their content and data can be used in a wide variety of applications. This use will be to further the commercial interests of the platform in question, and in some cases, associated third parties, usually without compensation to the creator of the content[70]. People may be attracted to platforms such as Facebook because they are free to use, have a critical mass of participants, and can be used across Internet service providers, and on numerous devices. This may be traded against other features of services, such loss of autonomy of content and implications for privacy.

Privacy

Concerns have been raised about the adequacy of privacy protections relating to apps, IoT devices, social media, and wearables. For instance, in wearables technology, Patterson (2013) observed that data flows on an increasing scale to social networks and community groups, and from businesses to associates, insurers and employers, and data brokers. Coupled with data mining techniques, she argued that there may be risks to users of events such as employment and insurance discrimination and unwanted marketing.

[70] One notable exception to this is YouTube. Those who submit content may receive a share of advertising revenue generated when users of the site view that person's content (Google 2016).

Specific privacy risks associated with fitness tracking technology identified by Patterson (2013) arise from:

- ubiquitous monitoring — users are encouraged to treat every moment of their lives as an opportunity to log information about their physical self, resulting in complete profiles of the user's behavioural patterns over periods of weeks, months, and longer.
- granular information collection — many fitness trackers and associated apps are able to capture a large variety of detailed information about the demography, physiology, and behavioural attitudes of an individual. This results in fitness technology and app companies holding troves of highly personal health data that is potentially of value to business associations, insurance companies, and employers.
- de-contextualised information flows — health information flows using fitness tracking devices and related apps are multi-directional, multi-purpose, and not subject to the well-established norms that apply for health professionals, such as general practitioners and surgeons. Fitbit users, for instance, can easily make large swaths of personal health information available to friends or the general public by adjusting their privacy settings.
- insufficient disclosures — users are unlikely to know of the extent to which their data can be shared with third parties because companies do not provide complete descriptions of data flows in privacy policies and because privacy settings are not sufficiently mapped to collection settings. Users may also not be aware of the total amount of information they are sharing over time, and with whom, since access authorisations may be granted incrementally, and users may forget to revoke authorisation once they have moved on.
- security risks — a 2013 review of 43 health and wellness apps by the Privacy Rights Clearinghouse in the United States revealed that many apps send personally identifiable and other sensitive information to third parties on an unencrypted basis. Companies were found to have failed to take precautions with data such as never sending user information in clear text, and failing to fully anonymise data shared with third party analytics services.
- erosion of social norms — 'unravelling' is a term used to describe the phenomenon by which the disclosure of personal information for economic gain becomes so common, inexpensive, and easy that those who do not disclose will be assumed to be withholding negative information. For example, a driver who agrees to have a tracking device placed on their car may get a lower insurance premium, while a driver who refuses to install such a device might be assumed to be hiding unsafe driving practices, leading to a higher insurance premium.

In a series of interviews with a very small sample of 21 Fitbit users, Patterson (2013) found that users were vulnerable to persistent health tracking due to the fact that the majority seldom removed their devices. Users also tended to underestimate the amount of information they shared with Fitibt and other health tracking devices, and lacked the necessary tools to objectively evaluate the data flow practices of Fitbit and third parties.

In a Canadian study of the privacy practices and terms of service of nine fitness tracking companies, the authors argued that the companies gave themselves very broad rights to use, and in some cases, sell, the fitness data of customers. When examining the ease with which consumers could access their own data as permitted under Canadian law, the researchers found only six companies of the nine actually responded to consumer requests for access, and these six responses showed varying levels of regard for security and identity verification approaches taken, the level of detail of responses to questions, and how much raw personal information was actually provided to users (Hilts, Parsons and Knockel 2016).

Further, these researchers found a number of issues that could compromise the security of user data. All fitness trackers studied, with the exception of the Apple Watch, were susceptible to Bluetooth address surveillance (which can allow for particular devices to be recognised — this was also found by Cyr et al. (2014)). Garmin, Withings, and Bellabeat failed to use transit-level data security for at least one data transmission, leaving user data exposed. In addition, one of the applications of the Jawbone fitness tracker transmitted information on the user's precise location, for reasons not made apparent to the user (Hilts, Parsons and Knockel 2016).

Similar conclusions were found in an analysis of Fitbit by a group of researchers at the Massachusetts Institute of Technology. Echoing the concerns of Hills, Parsons and Knockel (2016), Cyr et al. (2014) stated:

> … companies like Fitbit do not provide any control of the data, upload it to their personal cloud, and force the user to pay a subscription fee in order to get further analysis. In the end, the user is given very little indication of what data the device or its associated applications are able to collect. Historically, the Fitbit has had numerous security vulnerabilities, some leading to awkward disclosure of data, security bungles with communication between the device and the web server, and a myriad of issues relating to the device itself. (pp. 1–-2)

The 'awkward disclosure of data' noted by Cyr et al. (2014) refers to the fact that in 2011, it was discovered that users who had recorded their sexual activity using Fitbit could have that information found online in Google search results of Fitbit profiles. The reason for this was that Fitbit user accounts were set to 'public access' by default, allowing user profiles to be found and viewed using search engines. Fitbit responded by hiding user activity records and removing sexual activity as an option to be recorded in a user's activity log, as well as changing access settings to be private by default (Marshall 2015).

As noted, there is a wide array of applications to which IoT devices can be put, meaning that these devices collect large quantities of data, some of which may be personal in its nature. As a consequence, privacy is an important consideration in the design and use of IoT devices.

The Federal Trade Commission (FTC 2015) pointed out that some of the privacy risks associated with IoT devices are not necessarily any different to those relevant to personal data transmitted over the Internet and mobile phones — for example, data collected may include financial data, and precise geo-location. However, given the sheer volume of data

that IoT devices may generate and the range of consumer and business products and infrastructure facilities with IoT potential, rich datasets could be created. While these datasets could be used to the overall benefit of the community, there is also a risk that they may be breached or misused. Another possible security risk noted by the Federal Trade Commission was the possibility of eavesdropping by device manufacturers or intruders, citing an example of researchers in Germany who used unencrypted smart meter data to determine the television program that an individual was watching (FTC 2015).

The Office of the Privacy Commissioner of Canada argued that as IoT devices are often designed to operate quietly as part of the environment, people may have difficulty determining precisely what type of data and how much data these devices are collecting. That is, questions are raised as to how transparent the collection of data is, and how individuals can give meaningful consent to the collection of data, especially in the setting of a person's home (OPCC 2016).

In social media, there has been some concern, primarily in Europe, about the data collection and usage practices of Facebook. These concerns have primarily related to the tracking of Internet browsing activity via cookies and plug-ins (box F.4).

Facebook uses cookies and plug-ins for similar purposes in Australia as it does overseas. However, the practice appears to have not caused the same degree of concern in Australia as in Europe, where data protection laws are generally seen to be amongst the most stringent in the world.

There are no specific laws governing social media in Australia. However, provisions of existing laws cover several aspects of social media use. For example, consumer protection laws prohibit businesses from making false, misleading, or deceptive claims in social media as in other media. The tort of defamation applies to electronic communications, thus including various forms of unstructured data, including social media.

There is a lack of clarity however, about whether Australia's *Privacy Act* 1988 (Cth) applies to social media organisations. The Office of the Australian Information Commissioner (OAIC) states that to be covered by the Privacy Act, an organisation must have an Australian link:

> A number of factors will determine whether an organisation has an Australian link, including whether it has a presence in Australia and whether it carries on business in Australia. If the social networking site is based in another country and does not have a presence in Australia, then you may not have privacy rights under Australian law when you use the site. (OAIC nd, p. 1)

The question of whether the Privacy Act applies to a given social media organisation is therefore dependent on the specific facts of the circumstances in question.

Box F.4 Investigation of Facebook practices in Belgium and France

The Belgian Privacy Commission was prompted to initiate an investigation of Facebook's new terms of use, introduced in January 2015, following multiple queries from concerned Facebook users, the media, Belgian Parliament, and the Secretary of State for Privacy. Van Elsenoy et al. (2015) assessed Facebook's new data policy in the following terms:

> Much of the DUP [Data Use Policy] consists of hypothetical and vague language rather than clear statements regarding the actual use of data. Moreover, the choices Facebook offers to its users are limited. For many data uses, the only choice for users is to simply 'take-it-or-leave-it' … Facebook leverages its dominant position on the online social network market to legitimise the tracking of individuals' behaviour across services and devices.
>
> … It is impossible to add any information that may not later be re-used for targeting … Users are even more disempowered because they are unaware exactly how their data is used for advertising purposes. Furthermore, they are left in the dark about their appearance in promotional content. (p. 11)

In response to the research carried out by Van Alsenoy et al. (2015) and Acar et al. (2015), the Belgian Privacy Commission considered that with respect to tracking:

> … Facebook is thus in a unique position, since it can easily link its users' surfing behaviour to their real identity, social network interactions and sensitive data such as medical information and religious, sexual and political preferences. This implies that Facebook tracking is more intrusive compared to most of the other cases of so-called 'third party tracking'. (CPP 2015, p. 17)

The Commission consequently recommended that Facebook:

- provide full transparency about the use of cookies, specifying the content and purpose of each individual cookie
- refrain from systematically placing long-life and unique identifier cookies with non-users of Facebook
- refrain from collecting and using the data of Facebook users by means of cookies and social plug-ins, except when (and only to the extent that) it is strictly necessary for a service explicitly requested by the user or unless unambiguous and specific consent is obtained (CPP 2015).

Besides Belgium, Germany, Italy, Spain and France have also investigated Facebook's practices. In 2015, the French data protection authority, the CNIL, undertook operations to verify whether Facebook was acting in accordance with the French Data Protection Act of 1978.

A number of the CNIL's findings were similar to those of the Belgian academics and authorities who investigated Facebook's activities. For example, the CNIL also observed that visiting a third party website containing Facebook plug-ins led to the enabling of cookies and collection of data on the browsing activity of Internet users who did not have a Facebook account, without informing them, and without their consent.

Third party data collection

As noted above, organisations involved in providing social media platforms and apps may supply data in some situations to third parties, depending on the specific business model adopted.

Privacy policies and terms of service provide information on the types of third parties with whom organisations may exchange data, and how that data are likely to be used.

Navmii, a navigation app, stipulates in its privacy policy that it shares data with data processors, other companies as well as third party business partners of Navmii. The third parties with whom Navmii shares data may, in turn, share data with their own third parties for the purposes of improving their products, advertising, and carrying out other activities that are disclosed with a user or to which they consent (Navmii 2016).

Similarly, in terms of wearables, Fitbit (2014) outlines the uses to which it puts data:

- Height, weight, gender and age is used to estimate the number of calories you burn.
- Contact information is used to send you account modifications, allow other Fitbit users to add you as a friend, and to inform you about new features or products we think you would be interested in.
- Your data is used for research to understand and improve Fitbit products and services.
- Logs and other data are used to troubleshoot Fitbit services; detect and protect against error …
- De-identified data that does not identify you may be used to inform the health community about trends; for marketing and promotional use; or for sale to interested audiences … (p. 1)

Fitbit further states that it may share personally identifiable information with companies that are contractually engaged to provide services such as order fulfilment. It may also share data if doing so is necessary to comply with a law, regulation, or valid legal process, or if it is judged necessary in connection with the sale, merger, bankruptcy, sale of assets or reorganisation of the Fitbit company. The company also states that:

> Fitbit may share or sell aggregated, de-identified data that does not identify you with partners and the public in a variety of ways, such as by providing research reports about health and fitness or in services provided under our Premium membership. When we provide this information, we take legal and technical measures to ensure that the data does not identify you and cannot be associated back to you. (Fitbit 2014, p. 1)

Garmin states that it uses personal information for communications, customer support, purchases, promotions, providing posts for discussion, and for company use. The latter category comprises the use of information for internal statistical, marketing, and operational purposes, including for the purposes of generating sales reports and understanding user demographics and trends. Garmin may also share the personal information of customers with its business affiliates (Garmin 2016).

Whether or not users know that data about them may be shared with third parties depends on the degree to which consumers read privacy policies and terms and conditions. However, privacy policies do not necessarily outline precisely with whom data may be shared, meaning that even those consumers who make an effort to read privacy policy may

not necessarily be fully informed about all of the parties with whom a service or device provider exchanges data.

A related issue is how the sharing of data with third parties affects the user about whom the data relates. In some instances, third party sharing may not affect the individual at all, where third parties provide support (such as technical assistance and monitoring) for the activities of the primary organisation. In other cases, the data obtained by third parties is used for advertising — some consumers may view this as a nuisance, but a necessary 'cost' of using the service.

At the more serious end of the spectrum however, are situations in which third parties possess data protection standards that are different to those of the initial data collector, potentially leading to security breaches. Organisations that share data with third parties do have options to mitigate these risks, such as by conducting risk assessments prior to sharing data.

Stopping data collection and the right to delete

Typically, an individual can prevent an organisation from collecting data about them by electing not to use the service or device in question. This may not necessarily be the case for some social media companies. In their analysis of Facebook practices in Belgium, Acar et al. (2015) revealed that even for those who did not possess a Facebook account, visiting the Facebook.com domain resulted in the site setting a number of cookies on the user's device. Further, Acar et al. (2015) found that the same cookies that were activated for logged out users also applied to users who had deactivated their accounts, enabling Facebook to continue to collect information about both sets of users' Internet browsing activity.

In the European Union, there is a so-called 'right to be forgotten'. Specifically, individuals have the right to request that search engines remove links containing personal information, where that information is inaccurate, inadequate, irrelevant, or excessive for the purposes of the data processing in question. Under the current European Union law, the right to be forgotten is not an absolute right, requiring a case-by-case assessment balanced against other fundamental rights, such as freedom of expression (EC nd).

The principle underpinning the right to be forgotten is derived from the European Union's 1995 Data Protection Directive, which permits an individual to ask that personal data be deleted once it is no longer necessary. The European Union's new Data Protection Directive, due to be implemented in member countries by May 2018 — includes a 'right to erasure'. Under the right to erasure, an individual will have the ability to request the erasure of any links to, copy or replication of the data in question, provided that:

- the data are no longer necessary in relation to the purposes for which they were collected
- the individual withdraws consent or the relevant storage period has expired

- the individual objects to the processing of data under relevant provisions of European Union law
- the data was unlawfully processed (EC nd).

The European Union Committee of the UK House of Lords (HoL EUC 2014) argued that the very expression 'right to be forgotten' was misleading, as information cannot be deliberately forgotten — at best, the Committee argued, information can be made less readily accessible. In a world in which information is readily accessible on the Internet and easily shared, permanently deleting data can prove difficult. For example, a person wishing to delete their Facebook account can delete the messages, status updates and photos they have posted. However, information that other Facebook users have shared about a person is not part of that person's account, and they cannot delete it (Facebook 2015a).

The Committee also pointed out that the right may actually have the opposite of the desired effect by raising awareness of the information that the subject wishes to be forgotten. The Committee further considered the right in the context of the technological change afforded by the Internet:

> In the early 1990s, when the World Wide Web was in its infancy and Google was not even in gestation, it may have seemed reasonable to include in the first EU data protection legislation a right for the data subject labelled 'The right to be forgotten'. Developments in the subsequent twenty years have made clear that the right is as elusive as the name is misleading. (HoL EUC 2014, p. 21)

Unlike Europe, Australia does not have an explicit 'right to be forgotten' or 'right to erasure' in law. Despite this, the Australian Privacy Principles (APPs) do impose obligations on APP entities to ensure the quality and accuracy of information they hold. APP 10 requires an entity to take reasonable steps to ensure that the information it collects is accurate, up-to-date, and complete, as well as being relevant to the purpose of the use or disclosure. APP 11 stipulates that if an entity holds personal information about an individual and no longer needs that information for the purposes covered by the APPs, that information should be destroyed or de-identified. Further, APP 13 specifies that an entity that is required to adhere to the APPs must correct personal information they hold if an individual requests the entity to correct the information, or if the entity is satisfied that the information is inaccurate, out-of-date, incomplete, irrelevant or misleading. The entity must also notify third parties of corrections to the information (unless impractical or unlawful to do so) (OAIC 2014).

In its report on invasions of privacy in the digital era, the Australian Law Reform Commission (ALRC 2014) noted that, in relation to interferences with the privacy of an individual, the Information Commissioner has the power to make a declaration that a respondent must not repeat the conduct in question or take specified steps to ensure that such conduct is not repeated or discontinued. Such declarations may require the respondent to delete, remove, or de-identify personal information.

On the prospect of a 'regulator take-down mechanism' for personal information that individuals wished to have removed, the Australian Law Reform Commission concluded that such a system 'may have an undesirably chilling effect on online freedom of expression, and any such power would need to balance the interests of the complainant against the interests of the party in publishing the material and broader public interests' (ALRC 2014, p. 313). The Australian Law Reform Commission considered that the availability of declarations could provide a suitable mechanism for removing information, while avoiding the chilling effect that could result from a take-down mechanism (ALRC 2014).

The Australian Law Reform Commission also noted that, given the ease with which information can be proliferated on the Internet, a take-down mechanism may be ineffective. Furthermore, information may be difficult to remove or delete where the respondent is located overseas. Despite this, the Australian Law Reform Commission argued that the possibility of a take-down mechanism having a limited effect in some cases was not in itself a reason not to make the mechanism available in those cases in which it would be effective (ALRC 2014).

References

23andMe 2016, *Research*, https://www.23andme.com/en-int/research/ (accessed 3 June 2016).

ABC (Australian Broadcasting Corporation) 2016, *'Prescription shopping' crackdown to monitor Victorians buying drugs*, ABC News Online, http://www.abc.net.au/news/2016-04-25/prescription-shopping-the-focus-of-30-million-dollar-crackdown/7355002 (accessed 31 May 2016).

ABR (Australian Business Register) 2014, *Australian Reporting Dictionary*, https://abr.gov.au/Media-centre/Publications/Connecting-business-and-government/Australian-Reporting-Dictionary/ (accessed 26 July 2016).

—— 2015, *Report of the Australian Business Registrar 2014–15*, Australian Government, Canberra.

ABRDRSC (Australian Business Roundtable for Disaster Resilience and Safer Communities) 2014, *Building an Open Platform for Natural Disaster Resilience Decisions*, Research Report.

ABS (Australian Bureau of Statistics) 2009, *ABS Data Quality Framework*, 4 May, Cat. 1520.0, Canberra.

—— 2012, *National Early Childhood Education and Care Collection: Data Collection Guide, 2011*, Cat. 4240.0.55.002, Canberra.

—— 2013, *Essential Statistical Assets for Australia*, Cat. 1395.0, Canberra.

—— 2015a, *National Agricultural Statistics Review – Final Report*, Cat. 7105.0.55.004, Canberra.

—— 2015b, *National Health Survey: First Results, 2014–15*, Cat. 4364.0.55.001, Canberra.

—— 2015c, *TableBuilder User Manual*, http://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder%20manual%20contents?opendocument&navpos=240 (accessed 22 July 2016).

—— 2016a, *About the Australian Bureau of Statistics*, http://www.abs.gov.au/about?OpenDocument&ref=topBar (accessed 20 October 2016).

—— 2016b, *Household Use of Information Technology, Australia, 2014–15*, Cat. 8146.0, Canberra.

—— 2016c, *Internet Activity, Australia, June 2016*, Cat. 8153.0, Canberra.

—— 2016d, *Organisations approved to use CURF Microdata*, http://www.abs.gov.au/websitedbs/d3310114.nsf/home/organisations+approved+to+use+curf+microdata (accessed 13 October 2016).

—— 2016e, *Personal Fraud, Australia, 2014–15*, Cat. 4582.0, Canberra.

—— 2016f, *Residential Property Price Indexes: Eight Capital Cities, Mar 2016*, Cat. 6416.0, Canberra.

—— 2016g, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*.

—— 2016h, *Submission to the Productivity Commission Inquiry into the Regulation of Australian Agriculture*.

Acar, G., Van Alsenoy, B., Piessens, F., Diaz, C. and Preneel, B. 2015, *Facebook Tracking Through Social Plug-ins*, version 1.1, KU Leuven, Leuven, Belgium.

ACCC (Australian Competition and Consumer Commission) 2014a, *Australian Competition & Consumer Commission Submission to the Competition Policy Review – Response to the Draft Report*, Australian Government.

—— 2014b, *The comparator website industry in Australia*, Australian Government.

—— 2015, *Application for authorisation lodged by Australian Retail Credit Association Ltd in respect of the Principles of Reciprocity and Data Exchange*, Draft determination, A91482, Australian Government.

—— 2016, *New Car Retailing Industry – a market study by the ACCC*, Issues Paper, Australian Government.

Accenture 2015a, *Revised Payment Services Directive (PSD2) Everyday Payments — Accenture*, https://www.accenture.com/au-en/insight-everyday-payments-europe (accessed 8 July 2016).

—— 2015b, *The Digital Disruption in Banking: Demons, demands and dividends*, https://www.accenture.com/au-en/~/media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Industries_5/Accenture-2014-NA-Consumer-Digital-Banking-Survey.pdf (accessed 18 May 2016).

—— 2015c, *The Future of FinTech and Banking: Digitally disrupted or reimagined?*, https://www.accenture.com/au-en/insight-future-fintech-banking (accessed 3 October 2016).

—— and General Electric 2014, *Industrial Internet Insights Report for 2015*, http://www.geautomation.com/content/industrial-internet-insights-report-pn (accessed 20 June 2016).

ACCIS (Association of Consumer Credit Information Suppliers) 2015, *ACCIS 2015 Survey of Members: An Analysis of Credit Reporting in Europe*, Brussels.

ACIL Tasman 2008, *The Value of Spatial Information: The impact of modern spatial information technologies on the Australian economy*, Spatial Information Systems; ANZLIC — the Spatial Information Council, Sydney.

ACMA (Australian Communications and Media Authority) 2012, *Location services, personal information and indentity: Exploratory community research*,

http://www.acma.gov.au/theACMA/Library/researchacma/Research-reports/here-there-and-everywhere-consumer-behaviour-and-location-services (accessed 4 July 2016).

—— 2013a, *Privacy and personal data*, Occasional Paper #4, Emerging issues in media and communications, http://www.acma.gov.au/theACMA/About/The-ACMA-story/Connected-regulation/privacy-and-digital-data-emerging-issues (accessed 4 July 2016).

—— 2013b, *Sharing digital identity*, November, Digital footprints and identities research - Short report 2.

—— 2015a, *Communications report 2014–15*, Australian Government.

—— 2015b, *Industry code development*, http://www.acma.gov.au/Industry/Telco/Infrastructure/The-NBN-and-industry/industry-code-development (accessed 16 September 2016).

—— 2015c, *Register of telco industry codes & standards*, http://www.acma.gov.au/theACMA/Library/Corporate-library/Forms-and-registers/register-of-telecommunications-industry-codes-and-standards (accessed 16 September 2016).

—— 2016, *Service provider regulation*, http://www.acma.gov.au/sitecore/content/Home/Industry/Telco/Carriers-and-service-providers/Licensing/telecommunications-carrier-and-service-provider-regulation-fact-sheet (accessed 16 September 2016).

ACORN (Australian Cubercrime Online Reporting Network) 2015, *Attacks on computer systems*, https://www.acorn.gov.au/learn-about-cybercrime/attacks-computer-systems (accessed 10 October 2016).

Acquisti, A. 2010, *The Economics of Personal Data and the Economics of Privacy*, Background Paper #3, OECD Roundtable on The Economics of Personal Data and Privacy: 30 Years After the OECD Privacy Guidelines, OECD Publishing.

——, Taylor, C. and Wagman, L. 2016, 'The Economics of Privacy', *Journal of Economic Literature*, vol. 54, no. 2, pp. 442–492.

ACSQHC and NHPA (Australian Commission on Safety and Quality in Health Care and National Health Performance Authority) 2015, *Australian Atlas of Healthcare Variation*, http://www.safetyandquality.gov.au/atlas/ (accessed 18 May 2016).

ACT Government 2015, *Proactive Release of Data (Open Data) Policy*, December, Canberra.

Acxiom 2016, *Accurately Identify Relevant Audiences for All of Your Media Campaigns*, http://www.acxiom.com.au/data-packages/ (accessed 12 July 2016).

Adams, C. and Allen, J. 2013, 'Data custodians and decision-making: A right of access to government-held databases for research?', presented at 2013 AIAL National Administrative Law Conference, Canberra, 19 July.

—— and Lee-Jones, K. 2016, *A study into the legislative — and related key policy and operational — frameworks for sharing information relating to child sexual abuse in institutional contexts*, Report for the Royal Commission into Institutional Responses to

Child Sexual Abuse, May, Macquarie University, http://www.childabuseroyalcommission.gov.au/policy-and-research/our-research/published-research/legislative-and-related-frameworks-for-information (accessed 12 October 2016).

ADHA (Australian Digital Health Agency) 2016a, *Legislation: Changes to legislation*, https://myhealthrecord.gov.au/internet/mhr/publishing.nsf/Content/legislation (accessed 20 October 2016).

—— 2016b, *My Health Record Statistics – at 11 September 2016*, Australian Government Department of Health, https://myhealthrecord.gov.au/internet/mhr/publishing.nsf/Content/news-002, (accessed 22 September 2016).

ADLS (Administrative Data Liaison Service) 2012, *Section 33 of the Data Protection Act — A practical note for researchers*, http://www.adls.ac.uk/wp-content/uploads/Section-33-of-the-DPA-a-practical-note-for-researchers.pdf (accessed 28 September 2016).

Administrative Data Taskforce 2012, *Improving Access for Research and Policy*, https://www.gov.uk/government/publications/administrative-data-taskforce-report-government-response (accessed 16 June 2016).

AEC (Australian Electoral Commission) 2016, *Supply of elector information for use in medical research*, http://www.aec.gov.au/Enrolling_to_vote/About_Electoral_Roll/medical_research.htm (accessed 8 August 2016).

AEHRC (The Australian EHealth Research Centre) 2015, *Making health records more accessible*, https://aehrc.com/research/case-studies/making-health-records-more-accessible/ (accessed 10 June 2016).

AEMC (Australian Energy Market Commission) 2012, *Power of choice review — giving consumers options in the way they use electricity*, Final report, Sydney.

—— 2014, *Customer access to information about their energy consumption*, Information release, Sydney.

—— 2015, *Metering Data Provision Procedures*, Final report and determination, Sydney.

Attorney-General's Department nd, *Document Verification Service*, Australian Government, Canberra, https://www.ag.gov.au/rightsandprotections/identitysecurity/pages/documentverificationservice.aspx (accessed 20 October 2016).

—— 2012, *Protective Security Policy Framework - Information security management guidelines: Management of aggregated information*, Australian Government, Canberra.

—— 2015a, *Data Retention*, https://www.ag.gov.au/dataretention (accessed 1 March 2016).

—— 2015b, *Discussion paper — Mandatory data breach notification*, https://www.ag.gov.au/consultations/pages/serious-data-breach-notification.aspx (accessed 22 July 2016).

—— 2016a, *Amendment to the Privacy Act to further protect de-identified data*, https://www.attorneygeneral.gov.au/Mediareleases/Pages/2016/ThirdQuarter/Amendmen

t-to-the-Privacy-Act-to-further-protect-de-identified-data.aspx (accessed 29 September 2016).

—— 2016b, *Australian Government Information Security*, https://www.protectivesecurity.gov.au/informationsecurity/Pages/default.aspx (accessed 18 July 2016).

—— 2016c, *Serious data breach notification*, https://www.ag.gov.au/consultations/pages/serious-data-breach-notification.aspx (accessed 21 July 2016).

AGIMO (Australian Government Information Management Office) 2006, *Australian Government Information Interoperability Framework*, April, http://www.finance.gov.au/archive/policy-guides-procurement/interoperability-frameworks/information-interoperability-framework/ (accessed 3 March 2016).

AIFS (Australian Institute of Family Studies) 2016a, *Mandatory reporting of child abuse and neglect*, May, Australian Government, Canberra.

—— 2016b, *What we do*, Text, Australian Institute of Family Studies, https://aifs.gov.au/about-us/what-we-do (accessed 20 October 2016).

AIHW (Australian Institute of Health and Welfare) 2002, *Australia's Health 2002*, Cat. AUS 25, Australian Government, Canberra.

—— 2007, *A guide to data development*, Cat. HWI 94, Australian Government, Canberra.

—— 2010, *Guidelines for the formulation of good data standards*, National Data Development and Standards Unit, Australian Government, Canberra.

—— 2014, *Australia's Health 2014*, Cat. AUS 178, Australian Government, Canberra.

—— 2015, *Submission to the Senate Select Committee on Health*, Australian Parliament House, http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Health/Health/Submissions (accessed 26 May 2016).

—— 2016a, *Cancer registration in Australia*, http://www.aihw.gov.au/cancer-registration-in-australia/ (accessed 9 June 2016).

—— 2016b, *Custom Data Request Service*, http://www.aihw.gov.au/custom-data-request-

—— 2016c, *About*, http://www.aihw.gov.au/about/ (accessed 22 September 2016).

Al-Erini, K. and Tang, J. 2014, *News Feed FYI: Click-baiting*, Facebook, http://newsroom.fb.com/news/2014/08/news-feed-fyi-click-baiting/ (accessed 30 August 2016).

Allianz Australia 2016, *Privacy Policy*, https://www.allianz.com.au/about-us/privacy/ (accessed 15 June 2016).

ALRC (Australian Law Reform Commission) 2003, *Essentially Yours: The Protection of Human Genetic Information in Australia*, 96, Australian Government, Sydney.

—— 2008, *For Your Information: Australian Privacy Law and Practice*, 108, Australian Government, Sydney, http://www.alrc.gov.au/publications/report-108 (accessed 8 April 2016).

—— 2010a, *Family Violence - A National Legal Response (ALRC Report 114)*, 11 November, http://www.alrc.gov.au/publications/family-violence-national-legal-response-alrc-report-114 (accessed 13 October 2016).

—— 2010b, *Secrecy Laws and Open Government in Australia*, 112, Australian Government, Sydney, http://www.alrc.gov.au/publications/report-112 (accessed 20 April 2016).

—— 2014a, *Serious Invasions of Privacy in the Digital Era - Discussion Paper*, Discussion Paper, March, DP80, Australian Government, Sydney.

—— 2014b, *Serious Invasions of Privacy in the Digital Era - Final Report*, September, ALRC Report 123, Australian Government, Sydney, https://www.alrc.gov.au/publications/serious-invasions-privacy-digital-era-alrc-report-123 (accessed 17 June 2016).

AMA (Australian Medical Association) 2012, *Media Release: Government delivers PCEHR incentives for doctors*, https://ama.com.au/media/government-delivers-pcehr-incentives-doctors (accessed 1 June 2016).

—— 2013, *Submission to the Minister for Health's Review of the PCEHR*.

ANAO (Australian National Audit Office) 2004, *Integrity of Medicare Customer Data*, Audit Report 24 of 2004–05, Australian Government, Canberra.

—— 2014, *Integrity of Medicare Customer Data*, Audit Report 27 of 2013–14, Australian Government, Canberra.

—— 2015a, *Administration of the Australian Childhood Immunisation Register*, Audit Report 46 of 2014–2015, Australian Government, Canberra.

—— 2015b, *Administration of the Imported Food Inspection Scheme*, Audit Report 49 of 2014–2015, Australian Government, Canberra.

Anderson, B., Argent, R., Comeadow, S. and Barlow, A. 2010, 'Harmonising Australia's water information: Reflection on first steps', presented at the Australian Hydrographers' Assoication Conference 2010, pp. 1–19.

Andrews, D. 2016, *Media Release: New Digital Start For Victorian Government*, Premier of Victoria, http://www.premier.vic.gov.au/new-digital-start-for-victorian-government/ (accessed 25 July 2016).

ANDS (Australian National Data Service) 2016a, *Submission to the Productivity Commission Draft Report on Intellectual Property Arrangements*, http://www.pc.gov.au/inquiries/completed/intellectual-property/submissions (accessed 6 June 2016).

—— nd, *Data storage*, Australian National Data Service, http://www.ands.org.au/guides/data-storage (accessed 5 October 2016).

—— 2016b, *About us*, http://www.ands.org.au/about-us/what-we-do (accessed 16 October 2016).

Anonalytix nd, *Privacy-neutral Analytics*, http://www.anonalytix.com/ (accessed 20 October 2016).

ANZ nd, *ANZ Business Insights*, https://anzbusinessinsights.com/ (accessed 15 August 2016).

—— 2015, *Activate Bank Feeds — ANZ Internet Banking Help*, http://www.anz.com/internet-banking/help/update-details/bank-feeds/ (accessed 5 July 2016).

—— 2016, *ANZ Privacy Policy*, June, ANZ, Melbourne.

APF (Australian Privacy Foundation) 2011, *The PCEHR: Checklist of Privacy Concerns (Discussion Draft)*, https://www.privacy.org.au/Papers/PCEHR-Privacy-110215.pdf (accessed 1 June 2016).

Apple 2016, *Privacy*, http://www.apple.com/au/privacy/approach-to-privacy/ (accessed 1 June 2016).

APRA (Australian Prudential Regulation Authority) nd, *How to apply for an ADI authority*, http://www.apra.gov.au/adi/Pages/how-to-apply-for-an-ADI-authority.aspx (accessed 3 October 2016).

—— 2013, *Confidentiality of general insurance data and changes to general insurance statistical publications*, Discussion Paper, Australian Government, Sydney.

—— 2015, *Confidentiality of General Insurance data (letter to industry)*, http://www.apra.gov.au/CrossIndustry/Documents/150622-LTI-Public-disclosure-for-prudential-purposes-for-insurers-June-2015.pdf (accessed 4 October 2016).

APSC (Australian Public Service Commission) 2012, *Capability review: Department of Human Services*, Australian Government, Canberra, http://www.apsc.gov.au/publications-and-media/current-publications/dhs (accessed 14 September 2016).

ARCA (Australian Retail Credit Association) 2014, *Additional Submission to the Australian Financial System Inquiry*, http://fsi.gov.au/files/2015/03/arca.pdf (accessed 19 May 2016).

—— nd, *Principles of Reciprocity and Data Exchange*, http://www.arca.asn.au/focus/principles-of-reciprocity-data-exchange-prde.html (accessed 4 February 2016).

ARC (Australian Research Council) 2016, *State of Australian University Research: Volume 1*, Excellence in Research for Australia Reports, Australian Government, Canberra.

Archer, P., Bargiotti, L., De Keyzer, M., Goedertier, S., Loutas, N. and Van Geel, F. 2014, *Report on high-value datsets from EU institutions*, SC17DI06692, Interoperability Solutions for European Public Administrations, European Commission, Brussels.

ASD (Australian Signals Directorate) 2016, *ISM – Information Security Manual*, http://www.asd.gov.au/infosec/ism/index.htm (accessed 18 July 2016).

ASIC (Australian Securities and Investments Commission) 2012, *Media Release: ASIC warns comparison websites*, 12–304, http://asic.gov.au/about-asic/media-centre/find-a-

media-release/2012-releases/12-304mr-asic-warns-comparison-websites/ (accessed 24 February 2016).

ATO (Australian Taxation Office) 2016, *2013–14 Individual Sample File — Basic user info about sample unit record file*, https://data.gov.au/dataset/taxation-statistics-individual-sample-files/resource/39678fd4-26b3-41fc-8a61-b63c029fa93a (accessed 22 July 2016).

Audit Office of New South Wales 2015, *New South Wales Auditor-General's Report*, Financial Audit — Transport, Volume 6, Sydney.

AusGOAL 2011a, *AusGOAL — Australian Governments Open Access and Licensing Framework*, http://www.ausgoal.gov.au/ (accessed 25 July 2016).

—— *What Is AusGOAL?*, http://www.ausgoal.gov.au/overview (accessed 12 August 2016).

AUSTRAC (Australian Transaction Reports and Analysis Centre) 2014, *About the Anti-Money Laundering and Counter-Terrorism Financing Act 2006*, http://www.austrac.gov.au/businesses/legislation/amlctf-act (accessed 17 May 2016).

—— 2016, *AUSTRAC: Australia's financial intelligence unit*, http://www.austrac.gov.au/about-us/intelligence (accessed 1 September 2016).

Australian Data Archive 2016, *Submission to the National Research Infrastructure Roadmap Capability Issues Paper*, Department of Education (Australian Government), https://submissions.education.gov.au/Forms/National-Research-Infrastructure-Capability-Issues-Paper-Submissions/Documents/Australian%20Data%20Archive.pdf (accessed 15 October 2016).

Australian Farm Institute 2016, *The Implications of Digital Agriculture and Big Data for Australian Agriculture*, Sydney.

Australian Government 2012, *Explanatory Memorandum — Privacy Amendment (Enhancing Privacy Protection) Bill 2012*, Canberra.

—— 2015, *Preparing the National Action Plan*, Open Government Partnership – Australia, http://ogpau.govspace.gov.au/national-action-plan/ (accessed 16 September 2016).

—— 2016a, *Data.gov.au Site Statistics*, http://data.gov.au/stats#res-by-org (accessed 27 September 2016).

—— 2016b, *OGP Updates*, Open Government Partnership – Australia, https://ogpau.govspace.gov.au/category/updates/ (accessed 4 October 2016).

Australian Human Rights Commission 2006, *How Are Human Rights Protected in Australian Law?*, https://www.humanrights.gov.au/how-are-human-rights-protected-australian-law (accessed 11 August 2016).

Australian Policy Online 2016, *Submission into the National Research Infrastructure Roadmap Capability Issues Paper*, Department of Education (Australian Government), https://submissions.education.gov.au/Forms/National-Research-Infrastructure-

Capability-Issues-Paper-Submissions/Documents/Australian%20Policy%20Online.pdf (accessed 17 August 2016).

Baker, M. 2015, 'First results from psychology's largest reproducibility test', *Nature*, http://www.nature.com/doifinder/10.1038/nature.2015.17433 (accessed 19 October 2016).

Banks, E., Herbert, N., Mather, T., Rogers, K. and Jorm, L. 2012, 'Characteristics of Australian cohort study participants who do and do not take up an additional invitation to join a long-term biobank: The 45 and Up Study', *BMC Research Notes*, vol. 5, pp. 655–660.

——, Redman, S., Jorm, L. and Armstrong, B. 2008, 'Cohort profile: the 45 and up study', *International Journal of Epidemiology*, vol. 37, no. 5, pp. 941–947.

Bapat, A. 2013, 'The new right to data portability', *Privacy and Data Protection*, vol. 13, no. 3, pp. 3–4.

Barth-Jones, D. 2012, 'The "Re-Identification" of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now', *SSRN Electronic Journal*.

BBC (British Broadcasting Corporation) 2014, *What are cookies?*, http://www.bbc.co.uk /webwise/guides/about-cookies (accessed 25 July 2016).

BBVA nd, *BBVA APImarket*, https://www.bbvaapimarket.com/web/api_market/ (accessed 8 July 2016).

Beagrie, N., Lavoie, B. and Woollard, M. 2010, *Keeping research data safe (Phase 2)*, http://www.webarchive.org.uk/wayback/archive/20140614192110/http://www.jisc.ac.uk/publications/reports/2010/keepingresearchdatasafe2.aspx#downloads (accessed 19 July 2016).

Beckett, L. 2013, *Big Data Brokers: They Know Everything About You And Sell It To The Highest Bidder*, Gizmodo, http://www.gizmodo.com.au/2013/03/big-data-brokers-they-know-everything-about-you-and-sell-it-to-the-highest-bidder/ (accessed 1 August 2016).

Beef Central 2016a, *AgData: Putting farm data to profitable use*, Beef Central, http://www.beefcentral.com/news/agdata-putting-farm-data-to-profitable-use/ (accessed 13 July 2016).

—— 2016b, *Big data: What is it, and what does it mean for cattle?*, Beef Central, http://www.beefcentral.com/news/big-data-what-is-it-and-what-does-it-mean-for-cattle/ (accessed 13 July 2016).

Belcher, B. 2015, *The Independent Review of Whole-of-Government Internal Regulation*, Australian Government, Canberra, http://www.finance.gov.au/publications /reducingredtape/ (accessed 4 July 2016).

Ben-Shahar, O. and Schneider, C.E. 2011, 'The Failure of Mandated Disclosure', *University of Pennsylvania Law Review*, vol. 159, no. 3, pp. 647–749.

Bernal, P. 2014, *Internet Privacy Rights: rights to protect autonomy*, Cambridge University Press, United Kingdom.

Bharal, P. and Halfon, A. 2013, *Making Sense of Big Data in Insurance*, http://www.marklogic.com/resources/making-sense-of-big-data-in-insurance/ (accessed 27 June 2016).

Bhargava, A. 2013, *White Paper: A Dozen Ways Insurers Can Leverage Big Data for Business Value*, http://www.tcs.com/resources/white_papers/Pages/Business-Value-Big-Data-Insurers.aspx (accessed 27 June 2016).

Bickers, P., Hopkins-Burns, V., Bennett, A. and Namay, R. 2015, 'Information sharing by government agencies: The effect on the integrity of the tax system', *eJournal of Tax Research*, vol. 13, no. 1, pp. 183–201.

BIH (Bureau of Health Information (NSW)) 2016, *Bureau of Health Information — Home*, http://www.bhi.nsw.gov.au/ (accessed 10 June 2016).

Bioregional Assessments 2016, *Data register for the Galilee subregion*, http://www.bioregionalassessments.gov.au/assessments/16-data-register-galilee-subregion (accessed 11 July 2016).

Bitcoin 2016, *Open source P2P money*, https://bitcoin.org/en/ (accessed 16 October 2016).

bitglass 2016, *Healthcare Breach Report 2016*, https://pages.bitglass.com/BR-Healthcare-Breach-Report-2016_PDF.html (accessed 9 September 2016).

BITRE (Bureau of Infrastructure, Transport and Regional Economics) 2015, *Infrastructure benchmarking report*, Australian Government, Canberra, https://bitre.gov.au/publications/2015/cr_003.aspx (accessed 13 September 2016).

Bligh, A. 2012, *The Queensland Government's 'open data' revolution*, Cabinet Release, October, Queensland Government, Brisbane.

Boiten, E. 2016, *Care.data has been scrapped, but your health data could still be shared*, The Conversation, http://theconversation.com/care-data-has-been-scrapped-but-your-health-data-could-still-be-shared-62181 (accessed 10 October 2016).

Boulding, W. and Christen, M. 2001, 'First-Mover Disadvantage', *Harvard Business Review*, vol. 86, no. 12, pp. 20–21.

Box, P., Simons, B., Cox, S. and Maguire, S. 2015, *A Data Specification Framework for the Foundation Spatial Data Framework*, Digital Productivity Flagship — Report for the Department of Communications (Australian Government), CSIRO, Canberra.

Brandimarte, L., Acquisti, A. and Loewenstein, G. 2010, 'Misplaced Confidences: Privacy and the Control Paradox', presented at the *Ninth Annual Workshop on the Economics of Information Security (WEIS)*, Harvard University, June.

Braucher, J. 2006, 'New Basics: Twelve Principles for Fair Commerce in Mass-Market Software and Other Digital Products', in Winn, J.K. (ed), *Consumer Protection in the Age of the Information Economy*, Ashgate Publishing, Farnham, England, pp. 177–204.

Britt, H., Miller, G. and Henderson, J. 2015, *General practice activity in Australia 2014-15*, http://ses.library.usyd.edu.au/bitstream/2123/13765/4/9781743324530_ONLINE.pdf (accessed 9 May 2016).

Brodaric, B. and Gahegan, M. 2006, 'Representing geoscientific knowledge in cyberinfrastructure: Some challenges, approaches, and implementations.', *Geological Society of America Special Papers*, vol. 397, pp. 1–20.

Brody, H. 2013, *Making Your Website Scrape-Proof: How to Thwart Content Thieves*, Speed Awareness Month, http://www.speedawarenessmonth.com/making-your-website-scrape-proof-how-to-thwart-content-thieves/ (accessed 30 August 2016).

Bruce, D. and Bruce, J. 2015, *Transformation Index Monitor: Baseline Report*, Digital Transformation Office (Australian Government), Canberra.

Bruno, R. 2013, 'Real time monitoring of opioid prescriptions: DORA and her big brother ERRCD', presented at 2013 National Drug Trends Conference, Melbourne.

Bupa Australia 2015, *Information Handling Policy*, www.bupa.com.au/staticfiles/BupaP3/pdfs/BUPA_Info_Handling_Policy.pdf (accessed 30 June 2016).

Bureau of Communications Research 2016, *Open Government Data and Why It Matters: A Critical Review of Studies on the Economic Impact of Open Government Data*, Australian Government, Canberra.

Bushfire CRC 2014, *Fire Behaviour Model Enhancement*, http://www.bushfirecrc.com/projects/2-2/enhancement-fire-behaviour-models (accessed 9 July 2016).

Busselton Population Medical Research Institute 2014, 'Busselton Health Study — History', http://bpmri.org.au/about-us/history/busselton-health-study-history.html (accessed 6 September 2016).

Butler, K., Winkworth, G., McArthur, M. and Smyth, J. 2010, *Experiences and Aspirations of Younger Mothers*, Institute of Child Protection Studies — Australian Catholic University.

Cabinet Office (UK) 2013, *G8 Open Data Charter and Technical Annex UK*, UK Government, London, https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex (accessed 9 June 2016).

—— 2016, *Better use of data in government: Consultation paper*, UK Government, London, https://www.gov.uk/government/consultations/better-use-of-data-in-government (accessed 16 June 2016).

Caldicott, F. 2016, *Review of Data Security,Consent and Opt-Outs*, https://www.gov.uk/government/publications/review-of-data-security-consent-and-opt-outs (accessed 1 August 2016).

Canning, S. 2016, *Facebook and Quantium sign deal to measure advertising impact on store sales*, Mumbrella, https://mumbrella.com.au/facebook-quantium-sign-deal-372301 (accessed 9 July 2016).

Cantwell, E. and McDermott, K. 2016, *Making Technology Talk: How Interoperability Can Improve Care, Drive Efficiency, and Reduce Waste*,

http://medicalinteroperability.org/wp-content/uploads/2016/04/Making-Technology-Talk_HFM-reprint_May2016.pdf (accessed 11 May 2016).

Cao, L., Hosking, A., Kouparitsas, M. and Mullaly, D. 2015, *Understanding The Economy-wide Efficiency and Incidence of Major Australian Taxes*, The Treasury (Australian Government), http://www.treasury.gov.au/PublicationsAndMedia/Publications/2015/working-paper-2015-01 (accessed 27 September 2016).

Capgemini 2013, *The Open Data Economy — Unlocking Economic Value by Opening Government and Public Data*, https://www.capgemini.com/resources/the-open-data-economy-unlocking-economic-value-by-opening-government-and-public-data (accessed 13 July 2016).

—— 2014, *Big Data Alchemy: How Can Banks Maximize the Value of their Customer Data?*, https://www.capgemini.com/resources/big-data-customer-analytics-in-banks (accessed 22 June 2016).

Card, D., Chetty, R., Feldstein, M. and Saez, E. 2010, *Expanding access to administrative data for research in the United States*, White Paper, Future Research in the Social, Behavioral & Economic Sciences, National Science Foundation.

Carrasco, M. 2015, *myGov beats customer fatigue with 'tell us once' promise*, The Mandarin, http://www.themandarin.com.au/33039-mygov-juggernaut-tackles-customer-fatigue/ (accessed 3 October 2016).

Carter, P., Laurie, G. and Dixon-Woods, M. 2015, 'The social licence for research: why care.data ran into trouble', *Journal of Medical Ethics — Online First*, pp. 1–6.

Castro, D., New, J. and McQuinn, A. 2016, *How Is the Federal Government Using the Internet of Things?*, Center for Data Innovation, Information Technology and Innovation Foundation, Washington, D.C., https://itif.org/publications/2016/07/25/how-federal-government-using-internet-things (accessed 4 July 2016).

Cavoukian, A. and Castro, D. 2014, *Big Data and Innovation, Setting the Record Straight: De-identification Does Work*, Office of the Information and Privacy Commissioner, Ontario.

CBA (Commonwealth Bank of Australia) nd, *When we may send your information overseas*, Sydney, https://www.commbank.com.au/content/dam/commbank/security-privacy/country-list.pdf (accessed 19 June 2016).

—— 2014, *Privacy Policy*, Sydney, https://www.commbank.com.au/security-privacy/general-security/privacy-policy-html-version.html (accessed 19 June 2016).

CBO (Congressional Budget Organization (US)) 2008, *Evidence on the Costs and Benefits of Health Information Technology*, United States Congress, Washington, D.C.

CCAAC (Commonwealth Consumer Affairs Advisory Council) 2012, *Sharing of repair information in the automotive industry*, Australian Government, Canberra.

CDC (Centres for Disease Control and Protection) 2016, *Flu Activity & Surveillance*, http://www.cdc.gov/flu/weekly/fluactivitysurv.htm (accessed 19 October 2016).

Chaikin, D. 2011, 'Adapting the qualifications to the banker's common law duty of confidentiality to fight transnational crime', *Sydney Law Review*, vol. 33, no. 2, pp. 265–294.

Chang, A. and Li, P. 2015, 'Is economics research replicable? Sixty published papers from thirteen journals say "usually not"', presented at the *Finance and Economics Discussion Series 2015-083*.

Chang, K.K. 2012, 'All up in your Facebook: using social media to screen job applicants', *New England Law Review On Remand*, vol. 47, no. 1, pp. 1–13.

Chapman, S. 2002, *Medicare numbers debatable identifiers in medical database scheme*, Computerworld, http://www.computerworld.com.au/article/34557/medicare_numbers_debatable_identifiers_medical_database_scheme/ (accessed 20 May 2016).

Chaudhry, B., Wang, J., Wu, S. and Maglione, M. 2006, 'Systematic review: impact of health information technology on quality, efficiency, and costs of medical care', *Annals of Internal Medicine*, vol. 144, no. 10, pp. 742–752.

CHeReL (Centre for Health Record Linkage) 2016a, *Completed projects*, http://www.cherel.org.au/completed-projects (accessed 27 September 2016).

—— 2016b, *Master Linkage Key (MLK)*, http://www.cherel.org.au/master-linkage-key (accessed 25 July 2016).

CHOICE 2014, *Submission to the Competition Policy Review Issues Paper*, https://www.choice.com.au/consumer-advocacy/policy-submissions (accessed 20 October 2016).

Christensson, P. 2016, *API (Application Program Interface) Definition*, Tech Terms, http://techterms.com/definition/api (accessed 27 September 2016).

Chui, M., Farrell, D. and Jackson, K. 2014, *How government can promote open data*, http://www.mckinsey.com/industries/public-sector/our-insights/how-government-can-promote-open-data (accessed 18 July 2016).

City of Melbourne nd, *Urban Forest Strategy*, http://www.melbourne.vic.gov.au/community/parks-open-spaces/urban-forest/pages/urban-forest-strategy.aspx (accessed 18 July 2016).

Clayton Utz 2015, *Metadata can be personal information under privacy laws, says Privacy Commissioner - Knowledge*, https://www.claytonutz.com/knowledge/2015/may/metadata-can-be-personal-information-under-privacy-laws-says-privacy-commissioner (accessed 17 October 2016).

CLEDS (Vic) (Commissioner for Law Enforcement Data Security (Vic)) 2013, *Social Media and Law Enforcement*, July, Victorian Government, Melbourne.

CLIR (Council on Library and Information Resources) 2014, *Data curation*, https://www.clir.org/initiatives-partnerships/data-curation (accessed 21 October 2016).

CMA (Competition and Markets Authority) 2016a, *Energy Market Investigation*, Final report, June, London.

—— 2016b, *Retail Banking Market Investigation*, Final report, August, London.

—— 2016c, *Retail Banking Market Investigation: Overview*, https://www.gov.uk/government/publications/retail-banking-market-investigation-overview (accessed 23 September 2016).

CMS (Centres for Medicare and Medicaid Services) 2015, *Accountable Care Organizations (ACO)*, https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ACO/index.html?redirect=/ACO/ (accessed 10 June 2016).

CMS (Centres for Medicare and Medicaid Services) 2016, *Physician Quality Reporting System*, https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html?redirect=/pqri/ (accessed 10 June 2016).

CMTEDD (ACT) (Chief Minister, Treasury and Economic Development Directorate (ACT)) 2016, *Smart Parking*, Smart Parking, http://www.cmd.act.gov.au/smartparking/home (accessed 9 August 2016).

COAG (Council of Australian Governments) 2013, *National Health Information Agreement*, http://www.coaghealthcouncil.gov.au/Publications/Reports/ArtMID/514/ArticleID/44/Revised-National-Health-Information-Agreement-September-2013 (accessed 12 May 2016).

Cognizant 2012, *The New Auto Insurance Ecosystem*, http://www.the-digital-insurer.com/the-new-auto-insurance-ecosystem-telematics-mobility-and-the-connected-car-cognizant-report-by-aala-santhosh-reddy/ (accessed 22 September 2016).

Colley, A. 2016, *NBN still grappling with inaccurate housing data*, iTnews, http://www.itnews.com.au/news/nbn-still-grappling-with-inaccurate-housing-data-417164 (accessed 10 August 2016).

Comite, F. 2015, *Some doctors DO want your Fitbit data*, 3 September, http://venturebeat.com/2015/09/03/some-doctors-do-want-your-fitbit-data/ (accessed 25 May 2016).

Community Insight nd, *About*, https://www.communityinsight.org/about/ (accessed 2 September 2016).

Confluence nd, *About Phishing Attacks*, https://confluence.biola.edu/display/itservices/About+Phishing+Attacks (accessed 20 October 2016).

Cook, R. and Topol, E. 2014, 'How Digital Medicine Will Soon Save Your Life', *Wall Street Journal*, 21 February, http://www.wsj.com/articles/SB10001424052702303973704579351080028045594 (accessed 13 May 2016).

CoreLogic RP Data 2014, *About us*, http://www.corelogic.com.au/about-us/ (accessed 18 July 2016).

Costa, A., Deb, A. and Kubzansky, M. 2016, *Big Data, Small Credit*, https://www.omidyar.com/insights/big-data-small-credit (accessed 25 May 2016).

Cowan, P. 2016, *Former NAB exec to lead billion-dollar Centrelink IT overhaul*, iTnews, http://www.itnews.com.au/news/former-nab-exec-to-lead-billion-dollar-centrelink-it-overhaul-414365 (accessed 14 September 2016).

Cowling, D. 2016a, *Facebook Reaches 15 Million Australian Users*, 3 February, http://www.socialmedianews.com.au/facebook-reaches-15-million-australian-users/ (accessed 17 May 2016).

—— 2016b, *Social Media Statistics Australia – August 2016*, SocialMediaNews.com.au, http://www.socialmedianews.com.au/social-media-statistics-australia-august-2016/ (accessed 16 September 2016).

CPDC (Vic) (Commisioner for Privacy and Data Protection (Vic)) 2014, *Privacy by Design: Effective Privacy Management in the Victorian Public Sector*, https://www.cpdp.vic.gov.au/images/content/pdf/CPDP_Privacy_by_Design_Background_paper_Oct_2014.pdf (accessed 26 August 2016).

CPP (Commission for the Protection of Privacy) 2015, *Recommendation no. 04/2015 of 13 May 2015*, 13 May, Brussels.

CPSIC (Cross Portfolio Statistical Integration Committee) 2010, *High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes*, http://www.nss.gov.au/nss/home.NSF/pages/High+Level+Principles+for+Data+Integration+-+Content?OpenDocument (accessed 3 March 2016).

Creative Commons Australia 2010, *About the Licences*, Creative Commons Australia, http://creativecommons.org.au/learn/licences/ (accessed 11 October 2016).

CrimTrac 2015, *Annual Report 2014-2015*, https://www.crimtrac.gov.au/publications/crimtrac-annual-report-2014-15 (accessed 2 March 2016).

Crosman, P. 2015, *Fintech Glasnost: Why US Banks Are Opening Up APIs to Outsiders*, American Banker, http://www.americanbanker.com/news/bank-technology/fintech-glasnost-why-us-banks-are-opening-up-apis-to-outsiders-1075284-1.html (accessed 26 September 2016).

CSIRO nd, *The Provenance Management System*, https://confluence.csiro.au/public/PROMS (accessed 20 October 2016).

—— 2016a, *Data61 and Treasury to examine blockchain technology potential*, http://www.csiro.au/en/News/News-releases/2016/Data61-and-Treasury-to-examine-blockchain-technology-potential (accessed 17 October 2016).

—— 2016b, *Privacy: Keeping data confidential*, Confidential Computing – Insights from data without seeing the data, http://www.csiro.au/en/Research/D61/Areas/Cybersecurity/Privacy/Confidential-computing (accessed 17 October 2016).

Cyr, B., Horn, W., Miao, D. and Specter, M. 2014, *Security Analysis of Wearable Fitness Devices (Fitbit)*, Working paper, MIT, Cambridge, Massachusetts.

Daityari, S. 2014, *Using .htaccess to Prevent Web Scraping*, Sitepoint, https://www.sitepoint.com/using-htaccess-prevent-web-scraping/ (accessed 30 August 2016).

Daraganova, G., Mullan, K. and Edwards, B. 2014, *Attendance in Primary School: Factors and Consequences*, SSRN Scholarly Paper, 1 October, ID 2519883, Social Science Research Network, Rochester, NY, http://papers.ssrn.com/abstract=2519883 (accessed 27 September 2016).

Data61 2016, 'Automating Data Integration with Machine Learning', https://a.confui.com/public/conferences/575eb5b495c1bfbaaf00001f/topics/57dce7ed8d3ed20182000035/slides (accessed 1 October 2016).

Data Linkage WA 2013, *Developmental Pathways Project*, http://www.datalinkage-wa.org.au/projects/developmental-pathways-project (accessed 10 June 2016).

—— 2016a, *Data Collections*, http://www.datalinkage-wa.org/data-collections (accessed 10 June 2016).

—— 2016b, *Frequently Asked Questions*, http://www.datalinkage-wa.org.au/about-us/faq#46 (accessed 25 July 2016).

—— 2016c, *Glossary*, http://www.datalinkage-wa.org.au/data-linkage/glossary (accessed 21 October 2016).

Data.gov.au 2016, *Bioregional Assessment areas v03*, dataset, http://data.gov.au/dataset/96dbf469-5463-4f4d-8fad-4214c97e5aac (accessed 20 October 2016).

Davenport, T.H. and Dyché, J. 2013, *Big Data in Big Companies*, http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf (accessed 19 July 2016).

Davis, K. 2015, *Privatising Public Information: The Sale of the ASIC Business Registers*, http://www.australiancentre.com.au/News/privatising-public-information-sale-asic-business-registers (accessed 26 February 2016).

DBIS (UK) (Department for Business, Innovation & Skills (UK)) 2011, *The midata vision of consumer empowerment*, https://www.gov.uk/government/news/the-midata-vision-of-consumer-empowerment (accessed 10 March 2016).

—— 2012, *Impact Assessment for midata*, UK Government, London.

—— 2014, *Review of the Midata Voluntary Programme*, UK Government, London, UK, https://www.gov.uk/government/publications/midata-voluntary-programme-review (accessed 10 March 2016).

DDS IT Security nd, *Top 10 BlackHat Hacking Techniques*, http://ddsitsecurity.in/top-10-blackhat-hacking-techniques/ (accessed 20 October 2016).

Dearne, K. 2010, *Labor unveils e-health records trial sites*, The Australian, http://www.theaustralian.com.au/business/technology/labor-launches-e-health-records-trials/story-fn4htb9o-1225906463440 (accessed 24 May 2016).

DEDJTR (Department of Economic Development, Jobs, Transport and Resources) 2015, *About smart meters*, http://www.smartmeters.vic.gov.au/about-smart-meters/reports-and-consultations/lockstep-dpi-ami-pia-report/part-3 (accessed 14 July 2016).

—— 2016, *Smart Meter Compatible Web Portals*, State Government of Victoria, http://www.smartmeters.vic.gov.au/interactive-devices/web-portals (accessed 29 March 2016).

Deloitte 2008, *National eHealth Strategy*, National eHealth and Information Principal Committee.

—— 2013, *Market Assessment of Public Sector Information*, May, https://www.gov.uk/government/publications/public-sector-information-market-assessment (accessed 18 July 2016).

—— 2014, *Report to the Commonwealth Department of Health on the Public Consultation into the Implementation of the Recommendations of the Review of the PCEHR*, Department of Health, Canberra.

—— 2015, *Mobile Consumer Survey 2015 - the Australian Cut*, Deloitte Touche Tohmatsu.

Department of Education (NT) 2016, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*.

Department of Finance 2014a, *Australian Government Cloud Computing Policy*, http://www.finance.gov.au/cloud/ (accessed 23 March 2016).

—— 2014b, *Australian Government Cost Recovery Guidelines, Resource Management Guide No. 304 – Third Edition*, https://www.finance.gov.au/sites/default/files/australian-government-cost-recovery-guidelines.pdf (accessed 1 October 2016).

—— 2015, *Charging for Data Services Information Sheet*, https://www.finance.gov.au/sites/default/files/charging-for-data-services.docx (accessed 1 October 2016).

—— 2016a, *Australian Government ICT Trends Report 2014-15*, Australian Government, Canberra.

—— 2016b, *Australian Securities and Investments Commission (ASIC) Registry – FAQs*, https://finance.gov.au/procurement/scoping-studies/asic-faqs/ (accessed 4 July 2016).

—— 2016c, *Fedlink*, Text, http://www.finance.gov.au/collaboration-services-skills/fedlink/ (accessed 5 October 2016).

—— 2016d, *Historical Australian Government Contract Data*, data.gov.au.

Department of Health 2014, *Primary Health Care Research, Evaluation and Development (PHCRED) Strategy*, http://www.health.gov.au/internet/main/publishing.nsf/Content/pcd-programs-phcred (accessed 9 June 2016).

—— 2015a, *Data Access and Release Policy*, http://www.health.gov.au/internet/main/publishing.nsf/Content/Data-Access-Release-Policy (accessed 10 June 2016).

—— 2015b, *Healthcare Identifiers Service – Frequently Asked Questions*, http://www.health.gov.au/internet/main/publishing.nsf/Content/pacd-ehealth-consultation-faqs (accessed 23 May 2016).

—— 2016a, *Answer to Questions on Notice — Senate Select Committee on Health*, http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Health/Health/Additional_Documents (accessed 30 May 2016).

—— 2016b, *Data update*, http://www.health.gov.au/internet/main/publishing.nsf/Content/mr-yr16-dept-dept005.htm (accessed 29 September 2016).

—— 2016c, *Frequently Asked Questions: Managing your My Health Record*, My Health Record, https://myhealthrecord.gov.au/internet/mhr/publishing.nsf/Content/find-out-more?OpenDocument&cat=Managing%20your%20My%20Health%20Record (accessed 14 June 2016).

—— 2016d, *Linkable de-identified 10% sample of Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Schedule (PBS)*, https://data.gov.au/dataset/mbs-sample-10pct-1984-gz (accessed 2 August 2016).

—— 2016e, *My Health Record — What you should know before you get one*, https://myhealthrecord.gov.au/internet/mhr/publishing.nsf/Content/my-health-record-what-you-should-know (accessed 1 April 2016).

—— 2016f, *National Cancer Screening Register*, http://www.health.gov.au/internet/main/publishing.nsf/Content/mr-yr16-dept-dept002.htm (accessed 10 June 2016).

Department of Health (WA) 2012, *Human Research Ethics Committee Standard Operating Procedures*, http://www.health.wa.gov.au/healthdata/HREC/index.cfm (accessed 9 September 2016).

—— 2013, *Confidentiality Agreement for Researchers*, http://www.health.wa.gov.au/healthdata/HREC/index.cfm (accessed 9 September 2016).

—— 2016, *WA Public Hospital Activity: ED*, Department of Health, http://www.health.wa.gov.au/emergencyactivity/edsv/ (accessed 24 August 2016).

Desai, T., Ritchie, F. and Welpton, R. 2016, 'Five Safes: design data access for research', presented at the *Economics Working Paper Series*, University of the West of England.

Dewey, C. 2016, *98 personal data points that Facebook uses to target ads to you*, The Washington Post, https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/ (accessed 5 October 2016).

DFD (Department of Finance and Deregulation) 2013, *Big Data Strategy: Issues Paper*, Australian Government, Canberra.

DFSI (NSW) (Department of Finance, Services and Innovation (NSW)) nd, *Web Services: Land and Property Information*, Spatial Services, http://spatialservices.finance.nsw.gov.au/mapping_and_imagery/lpi_web_services (accessed 16 October 2016).

—— 2016a, *NSW Data Analytics Centre*, https://www.finance.nsw.gov.au/nsw-data-analytics-centre (accessed 25 July 2016).

—— 2016b, *NSW Government Open Data Policy*, NSW Government.

DHA (Department of Health and Ageing) 2012, *Fact Sheet: Electronic Recording and Reporting of Controlled Drugs*, Australian Government.

DHHS (US) (Department of Health and Human Services (US)) nd, *Consumer Health Data Aggregator Challenge*, https://www.challenge.gov/challenge/consumer-health-data-aggregator-challenge/ (accessed 10 June 2016).

—— 2013, *What & Why of Usability*, Usability.Gov, https://www.usability.gov/ (accessed 6 October 2016).

—— 2014, *What is HIE?*, US Government, Washington, D.C., https://www.healthit.gov/providers-professionals/health-information-exchange/what-hie (accessed 31 August 2016).

—— 2016a, *Breaches Affecting 500 or More Individuals*, https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf (accessed 5 September 2016).

—— 2016b, *HHS announces major commitments from healthcare industry to make electronic health records work better for patients and providers*, http://www.hhs.gov/about/news/2016/02/29/hhs-announces-major-commitments-healthcare-industry-make-electronic-health-records-work-better.html (accessed 10 June 2016).

—— 2016c, *Unlocking data*, http://www.hhs.gov/healthcare/delivery-system-reform/unlocking-data/index.html (accessed 10 June 2016).

DHHS (Vic) (Department of Health and Human Services (Vic)) 2013, *Ministerial Review of Victorian Health Sector Information and Communication Technology*, https://www2.health.vic.gov.au:443/about/publications/researchandreports/ministerial-review-of-victorian-health-sector-information-and-communication-technology (accessed 20 May 2016).

DHS (Department of Human Services) 2015, *Annual Report 2014-15*, https://www.humanservices.gov.au/corporate/annual-reports/annual-report-2014-15 (accessed 10 June 2016).

—— 2016, *Public Key Infrastructure*, https://www.humanservices.gov.au/health-professionals/services/medicare/public-key-infrastructure (accessed 10 October 2016).

DHS (VIC) (Department of Human Services (Vic)) 2007, *Providing Support to Vulnerable Children and Families*, Victorian Government, Melbourne.

Dietz, M., Khanna, S., Olanrewaju, T. and Rajgopal, K. 2016, *Cutting Through the FinTech Noise: Markers of Success, Imperatives For Banks*, February, http://www.mckinsey.com/industries/financial-services/our-insights/cutting-through-the-noise-around-financial-technology (accessed 3 October 2016).

DIIS (Department of Industry, Innovation and Science) 2015, *Innovation and Science Policy Report*, September, http://www.industry.gov.au/innovation/reportsandstudies/Pages/InnovationPolicyReport.aspx (accessed 20 July 2016).

Dobbin, M. 2014, 'Pharmaceutical drug misuse in Australia', *Australian Prescriber*, vol. 37, no. 3, pp. 79–81.

Doran, M. 2016, *Welfare overhaul will 'crack the back' of long-term reliance*, Text, ABC News, http://www.abc.net.au/news/2016-07-25/welfare-system-overhaul-to-eliminate-long-term-dependency/7657270 (accessed 13 October 2016).

DPC (SA) (Department of Premier and Cabinet (SA)) 2013, *Open Data Declaration*, September, Adelaide, http://digital.sa.gov.au/resources/topic/open-data/open-data-declaration (accessed 20 October 2016).

DPC (Tas) (Department of Premier and Cabinet (Tas)) 2015, *Stats Matter: A long-term strategy to build Tasmanian Government statistical assets and capability*, Tasmanian Government, Hobart.

DPC (WA) (Department of Premier and Cabinet (WA)) 2016, *Data Linkage Review*, https://www.dpc.wa.gov.au/Consultation/Pages/Data-Linkage-Review.aspx (accessed 30 May 2016).

DPMC (Department of Prime Minister and Cabinet) 2015, *Public Sector Data Management*, July, Australian Government, Canberra, https://www.dpmc.gov.au/sites/default/files/publications/public_sector_data_mgt_project.pdf (accessed 16 September 2016).

—— 2016a, *Data Skills and Capability in the Australian Public Service*, Australian Government, Canberra.

—— 2016b, *Guidance on Data Sharing for Australian Government Entities*, Australian Government.

—— 2016c, *Public Sector Data Management: Implementation Report*, Australian Government, Canberra.

—— 2016d, *Smart Cities Plan*, 29 April, Australian Government, Canberra.

——, DoC and NICTAL (Department of Prime Minister and Cabinet, Department of Communications and National ICT Australia Limited) 2015, *The National Map*, About, http://www.nationalmap.gov.au/about.html (accessed 16 October 2016).

DTF (Vic) Department of Treasury and Finance (Vic) 2012, *DataVic access policy*, http://www.dtf.vic.gov.au/Publications/Victoria-Economy-publications/IP-and-DataVic/DataVic-Access-Policy (accessed 28 September 2016).

—— 2015, *DataVic Access Policy Guidelines*, http://www.dtf.vic.gov.au/Publications/Victoria-Economy-publications/IP-and-DataVic/DataVic-Access-Policy-Guidelines (accessed 9 June 2016).

DTO (Digital Transformation Office) 2015a, *Announcing our work programme*, 14 October, Australian Government, Canberra.

—— 2016, *Digital Identity - early days in the Discovery process*, https://www.dto.gov.au/blog/digital-identity-early-days-in-the-discovery-process/ (accessed 10 October 2016).

Dunn, H.L. 1946, *Vital Statistics Analytical Report No. 2*, Dominion Bureau of Statistics, Ottawa.

Dwork, C. 2006, 'Differential Privacy', vol 4052, presented at 33rd International Colloquium on Automata, Languages and Programming, part II, Microsoft Research, pp. 1–12.

EA (Energy Australia) 2015, *Privacy Policy*, 1 June, https://www.energyaustralia.com.au/privacy (accessed 15 June 2016).

EDIC (Vic) (Joint Committee on Economic Development and Infrastructure Committee (Victoria)) 2009, *Inquiry into Improving Access to Victorian Public Sector Information and Data*, Parliamentary Paper, no. 198, Parliament of Victoria, June, Victorian Government, Melbourne.

Education and Health Standing Committee (Western Australia) 2015, *Managing the transition? The report of the inquiry into the transition and operation of services at Fiona Stanley Hospital*, November, 6, http://www.parliament.wa.gov.au/parliament/commit.nsf/(Report+Lookup+by+Com+ID)/70864F6AC389DCFA48257F08002B75F9/$file/151120+Final+Version+post-adoption+Signature+PDF+Cropped.pdf (accessed 17 May 2016).

eHealthNT 2011, *The Origins of eHealthNT*, eHealthNT, http://ehealthnt.nt.gov.au/About_Us/Origins_of_eHealthNT/index.aspx (accessed 24 May 2016).

El Emam, K., Jonker, E., Arbuckle, A. and Malin, B. 2011, 'A Systematic Review of Re-Identification Attacks on Health Data', *PLos ONE*, vol. 6, no. 12, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0028071 (accessed 21 July 2016).

Elliss-Brookes, L., McPhail, S., Ives, A., Greenslade, M., Shelton, J., Hiom, S. and Richards, M. 2012, 'Routes to diagnosis for cancer – determining the patient journey using multiple routine data sets', *British Journal of Cancer*, vol. 107, no. 8, pp. 1220–1226.

EMC Corporation 2014, *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*, http://www.emc.com/leadership/digital-universe/2014iview/index.htm (accessed 14 July 2016).

Emergency Alert nd, *National Emergency Alert Warning System – Home*, http://www.emergencyalert.gov.au/ (accessed 23 August 2016).

European Commission 2011, *Pricing of Public Sector Information Study — Models of Supply and Charging .for Public Sector Information.*

—— 2015, *European Commission - PRESS RELEASES - Press release - Questions and Answers - Data protection reform*, http://europa.eu/rapid/press-release_MEMO-15-6385_en.htm (accessed 27 September 2016).

—— 2016a, *Cookies*, 21 September, Information Providers Guide: the EU internet handbook, http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm (accessed 25 July 2016).

—— 2016b, *Fact Sheet on the 'Right to be Forgotten' ruling*, http://ec.europa.eu/mwg-internal/de5fs23hu73ds/progress?id=SKvNgnJOI179RJ1hwX5mIKM7NMmYzwUXl58aAFH5Ixo, (accessed 22 July 2016).

—— 2016c, *How does the data protection reform strengthen citizens' rights?*, http://ec.europa.eu/mwg-internal/de5fs23hu73ds/progress?id=h2JQ2KdRHZvEqnXx25j6wZXCpKwxYs8aNw7Wv_Al6sU, (accessed 6 July 2016).

—— 2016d, *Reform of EU data Protection Rules*, http://ec.europa.eu/justice/data-protection/reform/index_en.htm (accessed 18 February 2016).

—— nd, *About INSPIRE*, Infrastructure for Spatial Information in the European Community, http://inspire.ec.europa.eu/index.cfm/pageid/48 (accessed 16 September 2016).

Evenstad, L. 2016, *NHS England scraps controversial Care.data programme*, ComputerWeekly, http://www.computerweekly.com/news/450299728/Caldicott-review-recommends-eight-point-consent-model-for-patient-data-sharing (accessed 1 September 2016).

Experian nd, *Experian's combined Facebook and e-mail marketing activity drives an ROI increase of 350% for a luxury retailer*, Experian Australia.

Eyers, J. 2016, 'Banks planning big data deals to target customers', *Australian Financial Review*, 22 May, http://www.afr.com/technology/banks-planning-big-data-deals-to-target-customers-20160521-gp0pnq (accessed 22 June 2016).

Facebook 2015a, *Data Policy*, 30 January, https://www.facebook.com/policy.php (accessed 18 May 2016).

—— 2015b, *Terms of Service*, https://www.facebook.com/terms (accessed 29 April 2016).

—— 2016a, *About Facebook Adverts*, https://www.facebook.com/ads/about/?entry_product=ad_preferences (accessed 5 October 2016).

—— 2016b, *About Social Plugins*, https://www.facebook.com/help/443483272359009/ (accessed 1 October 2016).

—— 2016c, *Cookies & Other Storage Technologies*, 26 May, https://www.facebook.com/policies/cookies/# (accessed 30 May 2016).

—— 2016d, *Facebook Annual Report 2015*, April, Facebook, California.

—— 2016e, *Facebook Q1 2016 results*, http://investor.fb.com/results.cfm (accessed 17 May 2016).

—— 2016f, *Suicide Prevention*, https://www.facebook.com/help/suicideprevention (accessed 30 August 2016).

Faculty of Education — University of Western Australia 2016, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*, http://www.pc.gov.au/__data/assets/pdf_file/0006/199356/sub010-education-evidence.pdf (accessed 16 June 2016).

FCAI, AADA, AAA, AMIF and AAAA (Federal Chamber of Automotive Industries, Australian Automotive Dealer Association, Australian Automobile Association, Australian Motor Industry Federation and Australian Automotive Aftermarket Association) 2014, *Agreement on access to service and repair information for motor vehicles*, December.

Fidor nd, *Fidor API Reference*, http://docs.fidor.de/#introduction (accessed 8 July 2016).

Fielding 2000, *Architectural Styles and the Design of Network-based Software Architectures*, University of California, Irvine.

FinTech Australia 2016, *Priorities for Reform of the Australian Financial Services Industry (Prepared for the Australian Government Treasury)*.

Fitbit 2014, *Fitbit Privacy Policy*, https://www.fitbit.com/au/privacy#PrivacyPolicy (accessed 5 April 2016).

FMRC (Family Medicine Research Centre) 2016, *Bettering the Evaluation and Care of Health (BEACH)*, http://sydney.edu.au/medicine/fmrc/beach/index.php (accessed 9 June 2016).

FOS (Financial Ombudsman Service) 2016, *Annual Review 2015-16*, September, Melbourne.

Frontier Economics 2008, *Economic Study of the Consumer Benefits of eBay*, London.

FTC (Federal Trade Commission) 2015, *Internet of things - Privacy & Security in a Connected World*, FTC Staff Report, January.

Gardiner, B. 2015, *Offshore sensors record NSW Coast wave data*, CIO, http://www.cio.com.au/article/575781/offshore-sensors-record-nsw-coast-wave-data/ (accessed 13 October 2016).

Garmin 2016, *Privacy Statement*, 11 January, http://www.garmin.com/en-AU/legal/privacy-statement (accessed 23 May 2016).

Gartner 2015, *Gartner Says 6.4 Billion Connected*, http://www.gartner.com/newsroom/id/3165317 (accessed 22 September 2016).

Garvey, G., Percival, N., Izquierdo, L., Moodie, D. and Moore, S. 2016, 'Big data in an indigenous health context: opportunities and obstacles', *CancerForum*, vol. 40, no. 2, pp. 93–97.

Georgeff, M. 2007, *E-Health and the Transformation of Healthcare*, Australian Centre for Health Research, Melbourne.

Geoscience Australia 2014, 'Natural Hazard Impact Assessment at Geoscience Australia', http://www.sydneycoastalcouncils.com.au/sites/default/files/coovermar_ga_slides.pdf (accessed 16 September 2016).

Gershon, P. 2008, *Review of the Australian Government's Use of Information and Communication Technology*, Department of Finance and Deregulation (Australian Government), Canberra.

Gleit, N., Zeng, S. and Cottle, P. 2014, *Introducing Safety Check*, http://newsroom.fb.com/news/2014/10/introducing-safety-check/ (accessed 18 August 2016).

Glick, B. 2015, *NHS England GPs offer online services to 97% of patients*, ComputerWeekly, http://www.computerweekly.com/news/4500246514/NHS-England-GPs-offer-online-services-to-97-of-patients (accessed 3 June 2016).

Google nd, *Google Flu Trends*, https://www.google.org/flutrends/about/ (accessed 30 August 2016).

—— 2016a, *European privacy requests for search removals*, Google Transparency Report, https://www.google.com/transparencyreport/removals/europeprivacy/?hl=en (accessed 12 October 2016).

—— 2016b, *Helping emergency services find you when you need it most*, 25 July.

—— 2016c, *YouTube partner earnings overview*, https://support.google.com/youtube/answer/72902?hl=en (accessed 20 May 2016).

Gordon, J., Miller, G. and Britt, H. nd, *Reality check - reliable national data from general practice electronic health records*, Deeble Institute Issues Brief, https://ahha.asn.au/system/files/docs/publications/deeble_institue_issues_brief_no_18.pdf (accessed 25 July 2016).

Governance Institute of Australia 2015, *Public Display of Personal Information of Officeholders*, http://www.governanceinstitute.com.au/advocacy-research/submissions/2015/ (accessed 4 July 2016).

Government 2.0 Taskforce 2009, *Engage: Getting on with Government 2.0*, Department of Finance and Deregulation, http://www.finance.gov.au/archive/publications/gov20taskforcereport/ (accessed 7 April 2016).

Government of South Australia 2016a, *Alert South Australia - About*, Alert SA, https://www.alert.sa.gov.au/ (accessed 24 August 2016).

—— 2016b, *Public Sector (Data Sharing) Bill 2016*, 146.

GOV.UK 2010, *Letter to government departments on opening up data*, https://www.gov.uk/government/news/letter-to-government-departments-on-opening-up-data (accessed 16 June 2016).

Graham, D. 2016, *Discounts for data*, Choice, https://www.choice.com.au/shopping/consumer-rights-and-advice/your-rights/articles/loyalty-program-data-collection (accessed 7 July 2016).

GrainGrowers 2016, *Submission to the Productivity Commission Inquiry into the Regulation of Australian Agriculture*.

GSA (General Services Administration) nd, *Introduction to APIs in government*, http://18f.github.io/API-All-the-X/pages/introduction_to_APIs_in_government (accessed 21 October 2016).

Gutierrez, P. 2016, *Melbourne reveals its smart city ambitions*, http://www.iothub.com.au/news/melbourne-reveals-its-smart-city-ambitions-418252 (accessed 12 September 2016).

Hack Canada nd, *Canada's Big Brother: HRDC and The Longitudinal Labour Force File*, https://www.hackcanada.com/canadian/freedom/canadasbigbrother2000.html (accessed 15 October 2016).

Harper, I., Anderson, P., McCluskey, S. and O'Bryan, M. 2015, *Competition Policy Review Final Report*, Australian Government, Canberra.

Harrison, P., Hill, L. and Gray, C. 2016, *Confident, but Confounded - Consumer Comprehension of Telecommunications Agreements*, September, Deakin University and the Australian Communications Consumer Action Network, Sydney.

Hawke, A. 2013, *Review of Freedom of Information Laws*, August, https://www.ag.gov.au/consultations/pages/reviewoffoilaws.aspx (accessed 15 September 2016).

Head, A. 12 2016 at 12:15 2016, *Woodside retains corporate memory using cognitive computing*, Financial Review, http://www.afr.com/news/special-reports/the-cognitive-era/woodside-retains-corporate-memory-using-cognitive-computing-20160711-gq3d0u (accessed 20 September 2016).

Headd, M.J. 2016, *Open Data Guide*, http://opendata.guide/ (accessed 9 June 2016).

Health& 2016, *Health&*, https://healthand.com/au/ (accessed 23 September 2016).

Healthcare Gateway 2013, *Urgent care record sharing in Cumbria*, Healthcare Gateway, http://www.healthcaregateway.co.uk/case-studies/urgent-care-record-sharing-in-cumbria (accessed 3 June 2016).

Heath, T. and Bizer, C. 2011, 'Linked Data: Evolving the Web into a Global Data Space', in Hendler, J. and Harmelen, F. van (eds), *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1st edn, Morgan & Claypool.

Henderson, J., Pollack, A., Gordon, J. and Miller, G. 2014, 'Technology in practice – GP computer use by age', *Australian Family Physician*, vol. 43, pp. 831–831.

Hennessy, J. 2016, *Media Release: Real-time prescription monitoring will save lives*, Department of Health and Human Services (Vic).

Henry, J., Pylypchuk, Y., Searcy, T. and Patel, V. 2016, *Adoption of Electronic Health Record Systems among US Non-Federal Acute Care Hospitals: 2008-2015*, ONC Data Brief, 35, Office of the National Coordinator for Health Information Technology.

Heydon, G. and Zeichner, F. 2015, *Enabling the Internet of Things for Australia*, October, Communications Alliance.

Hill, K. 2012, 'How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did', *Forbes*, 16 February, http://www.forbes.com/sites/kashmirhill/2012/02/

16/how-target-figured-out-a- teengirl- was-pregnant-before-her-father-did/ (accessed 19 September 2016).

Hilts, A., Parsons, C. and Knockel, J. 2016, *Every Step You Fake - A Comparative Analysis of Fitness Tracker Privacy and Security*, Open Effect, Toronto.

Hodson, H. 2016, 'Did Google's NHS patient data deal need ethical approval?', *New Scientist*, https://www.newscientist.com/article/2088056-did-googles-nhs-patient-data-deal-need-ethical-approval/ (accessed 6 June 2016).

Hoffman, K.E. 2013, *Open API for Bank Apps: Can Credit Agricole's Model Work Here?*, American Banker Magazine, http://www.americanbanker.com/magazine/123_8/open-api-for-bank-apps-can-credit-agricoles-model-work-1060535-1.html (accessed 8 July 2016).

HoL EUC (House of Lords European Union Committee) 2014, *EU Data Protection law: a 'right to be forgotten'?*, 2nd Report of Session 2014-15, House of Lords, London.

Holman, D. 2014, 'Health, Political Arithmetic and Public Accountability: Bringing Down the Great Cth-State Data Divide', presented at the *Valedictory Lecture of the Chair in Public Health*, UWA, 29 July, http://www.aph.gov.au/DocumentStore.ashx?id=0d007e55-0cfb-4542-be07-046532649219 (accessed 30 May 2016).

——, Bass, J., Rosman, D., Smith, M., Semmens, J., Glasson, E., Brook, E., Trutwein, B., Rouse, L., Watson, C., de Klerk, N. and Stanley, F. 2008, 'A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system', *Australian Health Review*, vol. 32, no. 4, pp. 766–777.

Holmes, C.E. 2012, *Queensland Floods Commission of Inquiry*, Final Report, March, Brisbane.

Hore-Lacy, D. 2007, *Findings on death of case number 3236 of 2005*, Coroner's Written Findings, 28 November, Coroners Court of Victoria, Melbourne.

Horwitz, J. 2015, *Alibaba's customers can now get a loan based on their online shopping history*, *Quartz*, 25 June, http://qz.com/436889/alibabas-customers-can-now-get-a-loan-based-on-their-online-shopping-history/ (accessed 22 September 2016).

Houghton, J. 2011, *Costs and Benefits of Data Provision: Report to the Australian National Data Service*, September, Victoria University.

House of Commons Committee of Public Accounts 2013, *The dismantled National Programme for IT in the NHS*, 19, Session 2013-14, United Kingdom Parliament, London.

House of Representatives Standing Committee on Health 2016, *Report on the Inquiry into Chronic Disease Prevention and Management in Primary Health Care*, http://www.aph.gov.au/Parliamentary_Business/Committees/House/Health/Chronic_Disease/Report (accessed 9 May 2016).

Hunn, D. 2016, *Oil companies joining open source world by sharing data*, *Fuel Fix*, 25 August, http://fuelfix.com/blog/2016/08/25/oil-companies-joining-open-source-world-by-sharing-data/ (accessed 12 October 2016).

Huq, N. 2015, *Follow the Data: Analyzing Breaches by Industry*, TrendLAbs Research Paper, http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp-analyzing-breaches-by-industry.pdf (accessed 20 July 2016).

Hutchinson, J. 2012, *NBN slows over inaccurate address data*, iTnews, http://www.itnews.com.au/news/nbn-slows-over-inaccurate-address-data-302137 (accessed 10 August 2016).

Iansiti, M. and Lakhani, K.R. 2014, *Digital Ubiquity: How Connections, Sensors, and Data Are Revolutionizing Business*, November, https://hbr.org/2014/11/digital-ubiquity-how-connections-sensors-and-data-are-revolutionizing-business (accessed 30 August 2016).

IBM 2016, *American Sleep Apnea Association and IBM Launch Patient-led Sleep Study App; First ResearchKit App on Watson Health Cloud*, 2 March, https://www-03.ibm.com/press/us/en/pressrelease/49275.wss (accessed 27 April 2016).

ICA (Insurance Council of Australia) 2013, *Premiums explained*, Understand Insurance, http://understandinsurance.com.au/premiums-explained (accessed 6 July 2016).

ICICI Bank nd, *ICICI Bank Pockets on Facebook*, http://www.icicibank.com/Personal-Banking/insta-banking/internet-banking/pockets-on-facebook/index.page (accessed 5 July 2016).

ICO (UK) (UK Information Commissioner's Office) 2014, *Deleting Personal Data, Guide for organisations*, London.

—— 2016, *GDPR is still relevant for UK*, 7 July, https://iconewsblog.wordpress.com/2016/07/07/gdpr-still-relevant-for-the-uk/ (accessed 20 October 2016).

ICT NZ 2016, *Declaration on Open and Transparent Government*, 20 September, NZ Government, Wellington, NZ.

IDC (International Data Corporation) 2016, *IDC Forecasts Worldwide Shipments of Wearables to Surpass 200 Million in 2019, Driven by Strong Smartwatch Growth and the Emergence of Smarter Watches*, https://www.idc.com/getdoc.jsp?containerId=prUS41100116 (accessed 30 August 2016).

IDS (Informed Data Systems) 2016, *One Drop*, http://onedrop.today/ (accessed 3 June 2016).

Infrastructure Australia 2016, *Australian Infrastructure Plan*, February, Canberra.

Instagram 2016, *Privacy Policy*, https://www.instagram.com/about/legal/privacy/ (accessed 3 August 2016).

ITS International 2010, *Detection analysis technology successfully predicts traffic flows*, http://www.itsinternational.com/categories/detection-monitoring-machine-vision/features/detection-analysis-technology-successfully-predicts-traffic-flows/ (accessed 15 July 2016).

Janssen, K. and Kronenburg, T. 2012, *Open Data Standardization before publication?*, 2012/12, European Public Sector Information Platform,

https://www.europeandataportal.eu/sites/default/files/library/201212_open_data_and_standardization.pdf (accessed 11 March 2016).

Jawbone 2014, *Up Privacy Policy*, 16 December, https://jawbone.com/up/privacy (accessed 11 July 2016).

Jeseke, M., Grüner, M. and Weiß, F. 2013, *Big Data in Logistics*, December, http://www.dhl.com/en/about_us/logistics_insights/dhl_trend_research/bigdata.html#.WAgNxZj5j1s (accessed 19 July 2016).

Johansson 2016, *Vectors of trust*, https://tools.ietf.org/html/draft-richer-vectors-of-trust-03 (accessed 10 October 2016).

John Hancock 2015, *John Hancock Introduces a Whole New Approach to Life Insurance in the US. That Rewards Customers for Healthy Living*, 8 April, http://www.johnhancock.com/about/news_details.php?fn=apr0815-text&yr=2015 (accessed 23 May 2016).

Johnson, S. 2013, 'Consumer lending: Implications of new comprehensive credit reporting', *JASSA*, no. 3, p. 44.

Jolly, R. 2011, *The eHealth revolution - easier said than done*, Research Papers 2011-12, 3, Social Policy Section, Department of Parliamentary Services, Canberra.

Joskow, P.L. and Schmalensee, R. 1986, 'Incentive Regulation For Electric Utilities', *Yale Journal on Regulation*, vol. 4, no. 1, pp. 1–49.

Juniper Research 2016, *'Internet of Things' Connected Devices to Almost Triple to Over 38 Billion Units by 2020*, http://www.juniperresearch.com/press/press-releases/iot-connected-devices-to-triple-to-38-bn-by-2020 (accessed 22 September 2016).

Kantor, L. and Bhunia, P. 2016, *High Value Open Data in Australia — Quality, Availability and Use*, http://www.opengovasia.com/articles/7126-exclusive--dealing-with-data-in-australia---availability-accessibility-and-use (accessed 20 October 2016).

Karsten, J. and West, D.M. 2016, *Are you safe? Facebook's Safety Check and the future of emergency management*, https://www.brookings.edu/blog/techtank/2016/08/31/are-you-safe-facebooks-safety-check-and-the-future-of-emergency-management/?utm_campaign=Brookings+Brief&utm_source=hs_email&utm_medium=email&utm_content=33790332 (accessed 5 September 2016).

Kennedy, S. 2011, 'Canberra kicks off', *The Australian*, http://www.theaustralian.com.au/business/technology/canberra-kicks-off-tell-us-once-pilot/story-fn4htb9o-1226067685097 (accessed 13 October 2016).

King, T., Brankovic, L. and Gillard, P. 2012, 'Perspectives of Australian adults about protecting the privacy of their health information in statistical databases', *International Journal of Medical Informatics*, vol. 81, no. 1, pp. 279–289.

Kitchin, R. 2014, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, Sage, London.

Knapton, S. 2016, 'How the NHS got it so wrong with care.data', *The Telegraph*, 7 July, http://www.telegraph.co.uk/science/2016/07/07/how-the-nhs-got-it-so-wrong-with-caredata/ (accessed 23 September 2016).

Kwok, J. and Jones, B. 2005, 'Unnecessary Repeat Requesting of Tests: An Audit in a Government Hospital Immunology Laboratory', *Journal of Clinical Pathology*, vol. 58, no. 5, pp. 457–462.

Land and Property Information 2016, *Information brokers*, Land titles, http://www.lpi.nsw.gov.au/land_titles/information_brokers (accessed 16 October 2016).

Landgraf, M. 2016, *Comprehensive Credit Reporting*, Dun & Bradstreet, Auckland, New Zealand.

Lane, J., Stodden, V., Bender, S. and Nissenbaum, H. 2014, *Privacy, Big Data, and the Public Good*, Cambridge University Press.

Larson, D. 2014, *Data Analysis - Structured vs. Unstructured Data*, Presentation to Texas Digital Government Summit, 16 June.

Lateral Economics 2014, *Open for Business: How Open Data Can Help Achieve the G20 Growth Target*, June, Melbourne.

Lawson, C. 2008, 'Compulsory Licensing under the Patents Act 1990 (Cth) to remedy anticompetitive conduct under the Trade Practices Act 1974 (Cth)', *Australian Business Law Review*, vol. 36, no. 5, pp. 369–383.

LCD (Libelium Comunicaciones Distribuidas) 2015, *Smart Factory: Reducing Maintenance Costs and Ensuring Quality in the Manufacturing Process*, http://www.libelium.com/smart-factory-reducing-maintenance-costs-ensuring-quality-manufacturing-process/ (accessed 12 September 2016).

LCSCSI (Legislative Council Standing Committee on Social Issues (NSW)) 2015, *Service Coordination in Communities with High Social Needs*, https://www.parliament.nsw.gov.au/committees/inquiries/Pages/inquiry-details.aspx?pk=1746#tab-reports (accessed 16 September 2016).

Leadbetter 2014, *Facebook Messenger is a Blessing in Disguise*, *Lucas Leadbetter*, 14 August, https://lucasleadbetter.com/2014/08/14/facebook-messenger-is-a-blessing-in-disguise/ (accessed 22 September 2016).

Ley, S. 2015a, *Health Legislation Amendment (eHealth) Bill 2015 - Explanatory Memorandum*.

—— 2015b, *Media Release: Patients to get new myHealth Record: $485m 'rescue' package to reboot Labor's e-health failures*, Department of Health.

—— 2016, *Media Release: My Health Record gets one million more reasons to sign up*, Department of Health.

—— and Bailey, S. 2016, *PSD2 opens the door to new market entrants*, http://www2.deloitte.com/lu/en/pages/banking-and-securities/articles/psd2-new-market-entrants.html (accessed 29 June 2016).

Lindell, Y. and Pinkas, B. 2008, *Secure Multiparty Computation for Privacy Preserving Data Mining*.

Link Labs 2015, *What is M2M?*, 10 November, http://www.link-labs.com/what-is-m2m/ (accessed 13 July 2016).

Loff, B., Campbell, E., Glass, D., Kelsall, H., Slegers, C., Zion, D., Brown, N. and Fritschi, L. 2013, 'Access to the Commonwealth electoral roll for medical research', *The Medical Journal of Australia*, vol. 199, no. 2, pp. 128–130.

London Connect 2013, *Fact Sheet: Sharing your health and social care information*, London Health Improvement Board.

Lopez Research 2014, *Building Smarter Manufacturing With the Internet of Things (IoT)*, January, San Francisco.

Loshin, D. 2001, *Enterprise Knowledge Management — The Data Quality Approach*, Morgan Kaufmann, Academic Press, London.

Ludwig, J. 2009, *Enhancing National Privacy Protection: Australian Government First Stage Response to the ALRC Report 108: For Your Information: Australian Privacy Law and Practice*, October, Australian Government, Canberra.

Lyman, P. and Varian, H.R. 2003, *How Much Information?*, http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/ (accessed 31 August 2016).

Ma, W. 2013, *Woolworths: No ads, just data*, 5 September, AdNews, http://www.adnews.com.au/adnews/woolworths-no-ads-just-data (accessed 7 July 2016).

Mackay, M.M. 2014, *The Australian Mobile Phone Lifestyle Index*, December, AIMIA, Adelaide.

Macknight, J. 2016, *Is It API or Die for Banks?*, Xignite, http://resources.xignite.com/h/i/217848582-is-it-api-or-die-for-banks (accessed 26 September 2016).

Magistrates' Court of Victoria 2015, *Submission to the Royal Commission into Family Violence*, http://www.rcfv.com.au/Report-Recommendations (accessed 4 August 2016).

Mandel, R. and Noyes, E. 2013, 'Beyond the NWS: Inside the Thriving Private Weather Forecasting Industry', *Weatherwise*, Jan/Feb.

Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S. and Almasi Doshi, E. 2013, *Open Data: Unlocking Innovation and Performance with Liquid Information*, October, http://www.mckinsey.com/mwg-internal/de5fs23hu73ds/progress?id=_mStkMtliFGGMdwRSmbnml_wIYqHYaE2UleH_jZLOS8, (accessed 4 July 2016).

Marshall, G. 2015, *The story of Fitbit: How a wooden box became a $4 billion company*, 30 December, http://www.wareable.com/fitbit/youre-fitbit-and-you-know-it-how-a-wooden-box-became-a-dollar-4-billion-company (accessed 25 May 2016).

Martin, C. 2014, 'Barriers to the Open Government Data Agenda: Taking a Multi-Level Perspective', *Policy & Internet*, vol. 6, no. 3, pp. 217–240.

Maus, G. 2015, *How corporate data brokers sell your life, and why you should be concerned*, The Stack, https://thestack.com/security/2015/08/24/how-corporate-data-brokers-sell-your-life-and-why-you-should-be-concerned/ (accessed 3 August 2016).

McCoach, L. and Landy, D. 2014, 'Australia', *The Banking Regulation Review*, Fifth Edition, Law Business Research.

McColl, R. 2010, *Freedom of Information — A New Paradigm (The 2010 Whitmore Lecture)*, Council of Australasian Tribunals (NSW Chapter).

McDonald, K. 2012, *Real-time access to controlled drugs data from July*, Pulse+IT, http://www.pulseitmagazine.com.au/australian-ehealth/857-real-time-access-to-controlled-drugs-data-from-july (accessed 30 May 2016).

—— 2014a, *RACGP calls for urgent national roll-out of ERRCD*, Pulse+IT, http://www.pulseitmagazine.com.au/index.php?option=com_content&view=article&id=2198:racgp-calls-for-urgent-national-roll-out-of-errcd (accessed 31 May 2016).

—— 2014b, *Regulations and rail gauge problems holding up ERRCD roll-out*, Pulse+IT, http://www.pulseitmagazine.com.au/australian-ehealth/1808-regulations-and-rail-gauge-problems-holding-up-errcd-roll-out (accessed 30 May 2016).

—— 2015a, 'Digital chart forms an intuitive record for Calvary Bethlehem', *Pulse+IT*, http://www.pulseitmagazine.com.au/index.php?option=com_content&view=article&id=2303:digital-chart-forms-an-intuitive-record-for-calvary-bethlehem-hospital&catid=16:australian-ehealth&Itemid=328 (accessed 20 October 2016).

—— 2015b, *NSW takes the lead on real-time prescription drug monitoring*, Pulse+IT, http://www.pulseitmagazine.com.au/news/australian-ehealth/2496-nsw-takes-the-lead-on-real-time-prescription-drug-monitoring (accessed 31 May 2016).

McGarry, C. 2015, *ResearchKit at 6 months: 100,000 people now using medical apps*, 15 October, Macworld, http://www.macworld.com/article/2993838/ios/researchkit-at-6-months-100-000-people-now-using-medical-apps.html (accessed 2 June 2016).

McKinney, J., Guidoin, S. and Marczak, P. 2015, *Gaps and opportunities for standardization in OGP members' open data catalogs*, Open Data Working Group of the Open Government Partnership, http://www.opengovpartnership.org/resources-2 (accessed 11 March 2016).

McLeod, K., Templeton, R., Ball, C., Tumen, S., Crichton, S. and Dixon, S. 2015, *Using Integrated Administrative Data to Identify Youth Who Are at Risk of Poor Outcomes as Adults*, December, http://www.treasury.govt.nz/publications/research-policy/ap/2015/15-02 (accessed 14 July 2016).

Medibank Private 2015, *Medibank Privacy Policy*, May, Melbourne.

Melbourne Institute nd, *Organisations with Organisational Deed of Licence*, https://www.melbourneinstitute.com/hilda/data/organisational_licences.html (accessed 13 October 2016).

Mell, P. and Grance, T. 2011, 'The NIST Definition of Cloud Computing', *NIST Special Publication*, no. 800–145.

Menzies Foundation 2013, *Public Support for Data-based Research To Improve Health*, A discussion paper based on the proceedings of a Menzies Foundation Workshop 16th August, 2013, https://www.lowitja.org.au/sites/default/files/docs/10-Menzies-Foundation-Public-support-data-based-research.pdf (accessed 21 July 2016).

Meyer, M., Niech, C. and Eggers, W.D. 2015, *Anticipate, sense, and respond: Connected government and the Internet of Things*, Deloitte University Press, http://dupress.deloitte.com/dup-us-en/focus/internet-of-things/iot-in-government.html (accessed 12 September 2016).

Mickelburough, P. 2015, 'Hundreds of personal myki files released to police investigating crimes and missing persons cases', *Herald Sun*, 6 September, http://www.heraldsun.com.au/news/law-order/hundreds-of-personal-myki-files-released-to-police-investigating-crimes-and-missing-persons-cases/news-story/4134400f50942effe31d52c10becbed4 (accessed 10 August 2016).

Microsoft 2016a, *Delete and manage cookies*, https://support.microsoft.com/en-au/help/17442/windows-internet-explorer-delete-manage-cookies (accessed 25 July 2016).

—— 2016b, *Rolls-Royce agrees deal with Microsoft*, https://news.microsoft.com/en-gb/2016/04/26/rolls-royce-agrees-deal-with-microsoft/#sm.0001ioe8an 8f0fnwyfm2cvyrpix0x (accessed 12 September 2016).

MinterEllison 2015, *Privacy Impact Assessment Report: PCEHR System Opt-Out Model*, Department of Health, Canberra.

—— nd, *Ferry System Contract*, Execution version, MinterEllison, Sydney.

Mitchell, R., Cameron, C., McClure, R. and Williamson, A. 2015, 'Data linkage capabilities in Australia: practical issues identified by a Population Health Research Network "Proof of Concept project"', *Australian and New Zealand Journal of Public Health*, vol. 39, no. 4, pp. 319–325.

Mobbs, J.D. 2001, 'Crimtrac - Technology and Detection', presented at 4th National Outlook Symposium on Crime in Australia, Canberra, http://www.aic.gov.au/media_library/conferences/outlook4/mobbs.pdf (accessed 16 August 2016).

Moore, G. 2012, *Thoughts from the week #1*, Twitter post, 12 August, https://twitter.com/geoffreyamoore/status/234839087566163968 (accessed 20 September 2016).

Morey, T., Forbath, T. and Schoop, A. 2015, *Customer Data: Designing for Transparency and Trust*, Harvard Business Review, https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust (accessed 4 June 2016).

Munro, D. 2015a, 'Data Breaches In Healthcare Totaled Over 112 Million Records In 2015', *Forbes*, 31 December, http://www.forbes.com/sites/danmunro/2015/12/31/data-breaches-in-healthcare-total-over-112-million-records-in-2015/#316b59867fd5 (accessed 5 September 2016).

Munro, K. 2015b, 'Opal card data surrendered to police and immigration authorities', *The Sydney Morning Herald*, 22 May, http://www.smh.com.au/nsw/opal-card-data-surrendered-to-police-and-immigration-authorities-20150521-gh76wn.html (accessed 10 August 2016).

Murray, D., Davis, K., Hewson, C. and McNamee, B. 2014, *Financial System Inquiry: Final Report*, Canberra.

Murray, P. 2012, 'Congestion pricing for roads: An overview of current best practice, and the economic and transport benefits for government', *Public Infrastructure Bulletin*, vol. 1, no. 8.

Murray Irrigation 2014, *2014 Statutory Review of the Water Act 2007 — Submission to the Independent Expert Panel*, Deniliquin, New South Wales.

NAA (National Archives of Australia) nd, *Normal administrative practice*, http://www.naa.gov.au/records-management/agency/keep-destroy-transfer/nap/index.aspx (accessed 20 October 2016).

—— 2014, *API Documentation*, https://api.naa.gov.au/Help (accessed 16 October 2016).

—— 2015a, *Digital Continuity 2020 Policy*, http://www.naa.gov.au/records-management/digital-transition-and-digital-continuity/digital-continuity-2020/index.aspx (accessed 17 March 2016).

—— 2015b, *What we keep: Principles for selecting the Australian Government's national archives*, http://www.naa.gov.au/records-management/publications/what-we-keep/index.aspx (accessed 20 October 2016).

—— 2016, *Access to Records Under the Archives Act — Fact Sheet #10*, National Archives of Australia, http://www.naa.gov.au/collection/fact-sheets/fs10.aspx (accessed 27 September 2016).

NAB (National Australia Bank) nd, *National Australia Bank Privacy Policy*, http://www.nab.com.au/common/privacy-policy (accessed 4 July 2016).

Naone, E. 2011, *Homomorphic Encryption*, MIT Technology Review, http://www2.technologyreview.com/news/423683/homomorphic-encryption/ (accessed 20 October 2016).

Narayanan, A., Huey, J. and Felten, E. 2015, 'A Precautionary Approach to Big Data Privacy', *Data protection on the move*, Springer, pp. 357–385, http://randomwalker.info/publications/precautionary.pdf (accessed 5 September 2016).

National Research Infrastructure Council 2010, *A process to identify and prioritise Australia's Landmark Research Infrastructure needs*, Discussion Paper, June.

National Weather Service 1991, *Policy Statement on the Weather Service - Private Sector Roles*, http://www.nws.noaa.gov/im/fedreg.htm (accessed 20 October 2016).

Navmii 2016, *Navmii, Navmii World & Navfree – Privacy Policy*, http://navmii.com/navfree-privacy-policy/ (accessed 5 September 2016).

NCOA (National Commission of Audit) 2014, *Towards responsible government - reports of the National Commission of Audit*, 31 March.

NCRIS (National Collaborative Research Infrastructure Strategy) 2014, *The Australian Research Data Infrastructure Strategy*, Australian Government.

NEHTA (National Electronic Health Transition Authority) 2016, *Features of the My Health Record system: Clinical Documents*, https://www.nehta.gov.au/get-started-with-digital-health/what-is-digital-health/features-of-the-my-health-record-system/clinical-documents (accessed 1 June 2016).

—— nd, *My Health Record system and Healthcare Identifiers (HI)*, http://www.nehta.gov.au/get-started-with-digital-health/what-is-digital-health/features-of-the-my-health-record-system/my-health-record-system-healthcare-identifiers (accessed 1 June 2016).

Nemschoff, M. 2014, *A Quick Guide to Structured and Unstructured Data*, 28 June, http://www.smartdatacollective.com/michelenemschoff/206391/quick-guide-structured-and-unstructured-data (accessed 10 May 2015).

New Zealand Data Futures Forum 2014, *Harnessing the economic and social power of data*, https://www.nzdatafutures.org.nz/sites/default/files/NZDFF_harness-the-power.pdf (accessed 1 June 2016).

New Zealand Government 2016, *New Zealand Government Open Access and Licensing framework (Version 2)*, https://www.ict.govt.nz/guidance-and-resources/open-government/new-zealand-government-open-access-and-licensing-nzgoal-framework/nzgoal2/ (accessed 20 October 2016).

New Zealand Law Commission 2012, *The Public's Right to Know: Review of the Official Information Legislation*, June, Law Commission Report 125, http://r125.publications.lawcom.govt.nz/ (accessed 7 June 2016).

NHHRC (National Health and Hospitals Reform Commission) 2009, *A Healthier Future For All Australians*, Final Report, NHHRC, Canberra.

NHMRC, ARC and AVCC (National Health and Medical Research Council, Australian Research Council and Australian Vice Chancellors' Committee) 2007, *National Statement on Ethical Conduct in Human Research*, https://www.nhmrc.gov.au/print/book/export/html/51613 (accessed 20 October 2016).

——, —— and UA (National Health and Medical Research Council, Australian Research Council and Universities Australia) 2007, *Australian Code for the Responsible Conduct*

of Research, https://www.nhmrc.gov.au/guidelines-publications/r39 (accessed 20 October 2016).

NHMRC (National Health and Medical Research Council) 2011, *Research Governance Handbook: Guidance for the National Approach to Single Ethical Review*, Australian Government, Canberra.

—— 2012, *The National Certification Scheme of Institutional Processes related to the Ethical Review of Multi-Centre Research*, https://hrep.nhmrc.gov.au/certification? (accessed 18 July 2016).

—— 2014, *Guidelines approved under Section 95A of the Privacy Act 1988*, https://www.nhmrc.gov.au/guidelines-publications/pr2 (accessed 1 June 2016).

—— 2015, *NHMRC Funding Rules 2015*, Australian Government, Canberra.

—— 2016, *Human Research Ethics Committees*, https://hrep.nhmrc.gov.au/certification/hrecs (accessed 18 July 2016).

—— nd, *NHMRC Statement on Data Sharing*, http://www.nhmrc.gov.au/grants-funding/policy/nhmrc-statement-data-sharing (accessed 4 April 2016).

NHPA (National Health Performance Authority) 2015a, *About us*, http://www.nhpa.gov.au/internet/nhpa/publishing.nsf/Content/About-us (accessed 10 June 2016).

—— 2015b, *National Health Performance Authority - Data Plans*, http://www.nhpa.gov.au/internet/nhpa/publishing.nsf/Content/Data-Plans (accessed 9 June 2016).

NHS England (National Health Service England) 2016, *Your health and care records*, NHS Choices, http://www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Pages/overview.aspx (accessed 3 June 2016).

nib health funds 2016, *nib privacy policy*, https://www.nib.com.au/legal/privacy-policy (accessed 30 June 2016).

Nielsen, J. 2007, *Do Government Agencies and Non-Profits Get ROI From Usability?*, Nielsen Norman Group, https://www.nngroup.com/articles/government-non-profits-usability-roi/ (accessed 7 October 2016).

—— and Nicol, D. 2008, 'Whither patent use without authorisation in Australia?', *Federal Law Review*, vol. 36, no. 3, p. 331.

Nike 2016, *Nike Sustainable Business Report*, http://www.nikeresponsibility.com/report/ (accessed 9 March 2016).

Nott, G. 2016, *Integrators sought for $1 billion welfare payment system transformation*, CIO, http://www.cio.com.au/article/604419/integrators-sought-1-billion-welfare-payment-system-transformation/ (accessed 14 September 2016).

Nous Group 2014, *Perspectives on the Use of Performance Frameworks in the Australian Federation*, Research commissioned by the COAG Reform Council, April, Melbourne.

NSS (National Statistical Service) nd, *Statistical Infrastructure*, http://www.nss.gov.au/nss/home.nsf/NSS/35BFD39E0E2A8597CA25763F000B622C?opendocument (accessed 13 September 2016a).

—— nd, *What is statistical data integration?*, A guide for data integration projects involving Commonwealth data for statistical and research purposes.

—— 2010, *High level principles for data integration involving Commonwealth data for statistical and research purposes*.

—— 2013a, *Australian Government Statistical Forum October Meeting*, http://nss.gov.au/nss/home.NSF/533222ebfd5ac03aca25711000044c9e/b691218a6fd3e55fca257af700076681/$FILE/ATTLC9F4.pdf/AGSF%2030%20October%202013%20-%20Summary%20Record%20Final.pdf (accessed 1 June 2016).

—— 2013b, *Risk Assessment Guidelines (December 2013)*, http://www.nss.gov.au/nss/home.NSF/pages/Data%20integration%20projects%20-%20how%20to%20determine%20the%20risk%20level%20-%20Risk%20assessment%20guidelines (accessed 24 August 2016).

—— 2015, *Rights, responsibilities and roles of data custodians*, http://www.nss.gov.au/nss/home.NSF/533222ebfd5ac03aca25711000044c9e/59fd060543b4e9e0ca257a4e001eacfe/$FILE/Rights,%20responsibilities%20and%20roles%20of%20data%20custodians_Dec2013.pdf (accessed 18 July 2016).

—— 2016a, *A Guide to Data Integration Projects Involving Commonwealth Data for Statistical and Research Purposes*, A Guide to Data Integration Projects Involving Commonwealth Data for Statistical and Research Purposes, http://statistical-data-integration.govspace.gov.au/topics/applying-the-separation-principle/ (accessed 19 September 2016).

—— 2016b, *Accreditation*, A Guide for Data Integration Projects Involving Commonwealth Data for Statistical and Research Purposes, https://statistical-data-integration.govspace.gov.au/topics/accreditation/ (accessed 10 August 2016).

—— 2016c, *Information and communication technology security*, A Guide for Data Integration Projects Involving Commonwealth Data for Statistical and Research Purposes, https://statistical-data-integration.govspace.gov.au/topics/secure-data-management/information-and-communication-technology-security/ (accessed 25 July 2016).

—— 2016d, *Public Register of Data Integration Projects*, http://www.nss.gov.au/nss/home.NSF/pages/Data+Integration+Find+A+Project?OpenDocument (accessed 2 June 2016).

—— 2016e, *Scope of the Commonwealth Arrangements*, https://statistical-data-integration.govspace.gov.au/topics/scope-of-the-commonwealth-arrangements/ (accessed 20 October 2016).

NSW Government nd, *About the Information Asset Register*, http://data.nsw.gov.au/iar/pages/about-the-iar (accessed 20 October 2016).

—— 2015, *NSW ICT Strategy: NSW Data Analytics Centre*, NSW Department of Finance, https://www.finance.nsw.gov.au/ict/nsw-data-analytics-centre (accessed 15 September 2016).

—— 2016a, *Health Stats NSW*, http://www.healthstats.nsw.gov.au/ (accessed 10 June 2016).

—— 2016b, *Real-time data - Overview*, Office of Water, http://www.water.nsw.gov.au/realtime-data (accessed 24 August 2016).

—— 2016c, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*.

NZ Data Futures Partnership 2015, *Who we are*, http://datafutures.co.nz/who-we-are/ (accessed 1 August 2016).

OAIC (Office of the Australian Information Commissioner) 2011a, *Freedom of information – The information publication scheme for Australian Government agencies*, https://www.oaic.gov.au/freedom-of-information/foi-resources/foi-fact-sheets/foi-fact-sheet-4-information-publication-scheme (accessed 6 April 2016).

—— 2011b, *Submission to the Department of Health and Ageing in response to the PCEHR System: Legislation Issues Paper*.

—— 2013a, *Community Attitudes to Privacy Report: 2013*, Canberra.

—— 2013b, *Open Public Sector Information: From Principles to Practice*, Office of the Australian Information Commissioner, https://www.oaic.gov.au/information-policy/information-policy-resources/open-public-sector-information-from-principles-to-practice (accessed 3 March 2016).

—— 2014a, *Australian Privacy Principles*, Privacy fact sheet 17, January, Canberra.

—— 2014b, *Data breach notification guide: A guide to handling personal information security breaches*, August, Canberra.

—— 2014c, *Privacy business resource 4: De-identification of data and information*.

—— 2014d, *Privacy fact sheet 38: Hardship assistance and your credit report*, https://www.oaic.gov.au/individuals/privacy-fact-sheets/credit-reporting/privacy-fact-sheet-38-hardship-assistance-and-your-credit-report (accessed 4 October 2016).

—— 2014e, *Privacy fact sheet 40: Credit providers, the APPs and your credit report*, https://www.oaic.gov.au/individuals/privacy-fact-sheets/credit-reporting/privacy-fact-sheet-40-credit-providers-the-apps-and-your-credit-report (accessed 21 June 2016).

—— 2015a, *Annual Report 2014-15*, Office of the Australian Information Commissioner, ——

—— 2015b, *Australian Privacy Principles guidelines*, 1 April, https://www.oaic.gov.au/resources/agencies-and-organisations/app-guidelines/APP_guidelines_complete_version_1_April_2015.pdf (accessed 21 June 2016).

—— 2015c, *Submission to the Department of Health: Electronic health records and healthcare identifiers*.

—— 2016a, *Australian Privacy Commissioner's investigation into published MBS and PBS data sets*, https://www.oaic.gov.au/media-and-speeches/statements/australian-privacy-commissioner-s-investigation-into-published-mbs-and-pbs-data-sets (accessed 29 September 2016).

—— 2016b, *Determinations*, https://www.oaic.gov.au/privacy-law/determinations/ (accessed 16 October 2016).

—— 2016c, *Does my business have privacy obligations in relation to consumer credit reporting under the Privacy Act?*, https://www.oaic.gov.au/agencies-and-organisations/faqs-for-agencies-orgs/businesses/does-my-business-have-privacy-obligations-in-relation-to-consumer-credit-reporting-under-the-privacy-act (accessed 5 July 2016).

—— 2016d, *Information policy agency resource 1: De-identification of data and information*, https://www.oaic.gov.au/information-policy/information-policy-resources/information-policy-agency-resource-1-de-identification-of-data-and-information (accessed 22 July 2016).

—— 2016e, *Loyalty program assessment: flybuys*, Summary report, July.

—— 2016f, *Loyalty program assessment: Woolworths Rewards*, Summary report, July.

—— 2016g, *Privacy Act*, https://www.oaic.gov.au/privacy-law/privacy-act/ (accessed 31 March 2016).

—— 2016h, *Privacy business resource 15: Keeping records of disclosures under the Telecommunications Act 1997*, February, Office of the Australian Information Commissioner, Sydney.

—— 2016i, *Privacy management framework: enabling compliance and encouraging good practice*, https://www.oaic.gov.au/agencies-and-organisations/guides/privacy-management-framework (accessed 15 September 2016).

—— 2016j, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*, http://www.pc.gov.au/inquiries/current/education-—— nd, *What is health information?*, https://www.oaic.gov.au/individuals/faqs-for-individuals/health/what-is-health-information (accessed 23 September 2016).

Obar, J.A. and Oeldorf-Hirsch, A. 2016, *The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of*, Working paper, 24 August, York University.

Oderkirk, J., Ronchi, E. and Klazinga, N. 2013, 'International comparisons of health system performance among OECD countries: Opportunities and data privacy protection challenges', *Health Policy*, vol. 112, no. 1, pp. 9–18.

ODI and Fingleton Associates (Open Data Institute and Fingleton Associates) 2014, *Data Sharing and Open Data for Banks*, A report for HM Treasury and Cabinet Office, UK Government, London, UK.

—— 2016, *The Open Banking Standard*, https://theodi.org/open-banking-standard (accessed 17 May 2016).

OECD (Organisation for Economic Cooperation and Development) 2008, *Council Recommendation on Enhanced Access and More Effective Use of Public Sector Information*, Seoul.

—— 2013, *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges*, OECD Health Policy Studies, OECD Publishing, http://dx.doi.org/10.1787/9789264193505-en (accessed 12 May 2016).

—— 2014, *OECD Public Governance Reviews Open Government in Latin America*, OECD Publishing.

—— 2015a, *Assessing government initiatives on public sector information: A review of the OECD Council Recommendation*, OECD Digital Economy Papers, http://dx.doi.org/10.1787/5js04dr9l47j-en (accessed 4 May 2016).

—— 2015b, *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris.

—— 2015c, *OECD Reviews of Health Care Quality: Australia 2015*, OECD Publishing, http://dx.doi.org/10.1787/9789264233836-en (accessed 18 May 2016).

—— 2015d, 'Open government data', *Government at a Glance*, Organisation for Economic Co-operation and Development, pp. 150–151, http://www.oecd-ilibrary.org/content/chapter/gov_glance-2015-48-en (accessed 27 September 2016).

OEL (Origin Energy Limited) 2016, *Privacy Policy*, https://www.originenergy.com.au/privacy/privacy-policy.html (accessed 22 June 2016).

Office of the Press Secretary (US) 2012, *Memorandum for the Heads of Executive Departments and Agencies: Building a 21st Century Digital Government*, The White House.

OGCIO (WA) (Office of the Government Chief Information Officer (WA)) 2015, *Whole of Government Open Data Policy*, http://gcio.wa.gov.au/2015/07/03/whole-of-government-open-data-policy/ (accessed 20 October 2016).

Ogeil, R., Heilbronn, C., Lloyd, B. and Lubman, D. 2016, 'Prescription drug monitoring in Australia - capacity and coverage issues', *Medical Journal of Australia*, vol. 204, no. 4, pp. 148–149.

OGP (Open Government Partnership) 2015, *About the Open Government Partnership*, Open Government Partnership, http://www.opengovpartnership.org/about (accessed 16 September 2016).

OGTR (Office of the Gene Technology Regulator) 2014, *Confidential Commercial Information*, http://www.ogtr.gov.au/internet/ogtr/publishing.nsf/content/cci (accessed 4 July 2016).

OIC (Qld) (Office of the Information Commissioner - QLD) 2012, *IP addresses, Google Analytics and the privacy principles*, https://www.oic.qld.gov.au/guidelines/for-government/guidelines-privacy-principles/transferring-personal-information-out-of-australia/ip-addresses,-google-analytics-and-the-privacy-principles (accessed 4 July 2016).

—— 2014, *Privacy and Mobile Apps*, 13 February, https://www.oic.qld.gov.au/guidelines/for-government/guidelines-privacy-principles/applying-the-privacy-principles/privacy-and-mobile-apps (accessed 6 June 2016).

OIRA (US) (Office of Information and Regulatory Affairs (US)) 2011, *Memorandum 8*, September, The White House, Washington, D.C.

O'Keefe, C.M. and Rubin, D.B. 2015, 'Individual privacy versus public good: protecting confidentiality in health research', *Statistics in Medicine*, vol. 34, no. 23, pp. 3081–3103.

Olszewski, P. and Xie, L. 2006, 'Modelling the effects of road pricing on traffic in Singapore', *Transportation Research Part A: Policy and Practice*, vol. 39, no. 7–9, pp. 755–772.

ONC (Office of the National Coordinator for Health Information Technology) 2015, *A Majority of Providers Provide Online Access to Health Information*, https://www.healthit.gov/newsroom/majority-providers-provide-online-access-health-information (accessed 10 June 2016).

OPC (NZ) (Office of the Privacy Commissioner of New Zealand) 2013, *Limits on use of personal information*, https://www.privacy.org.nz/the-privacy-act-and-codes/privacy-principles/limits-on-use-of-personal-information-principle-ten/ (accessed 28 September 2016).

OPC (Office of the Privacy Commissioner) 2008, *Privacy Guidelines for the Medicare Benefits and Pharmaceutical Benefits Programs*, https://www.legislation.gov.au/Details/F2008L00706 (accessed 10 June 2016).

OPCC (Office of the Privacy Commissioner of Canada) 2014, *Data Brokers - A Look at the Canadian and American Landscape*, Research report, September, Gatineau, Qubec.

OPCC (Office of the Privacy Commissioner of Canada) 2016, *The Internet of Things - An introduction to privacy issues with a focus on the retail and home environments*, Policy and Research Group research paper, February, Gatineau, Qubec.

OpenAustralia nd, *About us*, http://www.openaustralia.org.au/about/ (accessed 18 July 2016).

Oscar 2014, *It Pays to Walk: Oscar Rewards Members for Staying Active*, 23 December, http://blog.hioscar.com/post/105971652148/it-pays-to-walk-oscar-rewards-members-for-staying (accessed 23 May 2016).

Parkin, E. 2016, *A paperless NHS: electronic health records*, Briefing Paper 07572, 25 April, United Kingdom House of Commons, London.

Parkopedia nd, *About Parkopedia*, http://au.parkopedia.com/about-us/ (accessed 19 July 2016).

Partel, K. 2015, *Toward better implementation: Australia's My Health Record*, Issues Brief, 30 October, 13, Deeble Institute for Health Policy Research, Canberra.

Patterson, H. 2013, *Contextual Expectations of Privacy in Self-Generated Health Information Flows*, The 41st Research Conference on Communication, Information and Internet Policy, TPRC 41, New York University.

Patton, L.P., Wetmore, S.E. and Magill, C.T. 2016, 'How Wearable Fitness Devices Could Impact Personal Injury Litigation in South Carolina', *South Carolina Lawyer*, January, pp. 44–48.

Pawsey, M. 2015, *The Agriculture of Things: Submission to the Senate Standing Committee on Agriculture and Industry's Inquiry into Agricultural Innovation*, Parliament of Australia, Canberra.

PC (Productivity Commission) 2001, *Cost Recovery by Government Agencies*, Report 15, Canberra.

—— 2006, *Standard Setting and Laboratory Accreditation — Research Report*, Australian Government, Canberra.

—— 2007, *Annual Review of Regulatory Burdens on Business: Primary Sector*, Research Report, Australian Government, Canberra.

—— 2009a, *Annual Review of Regulatory Burdens on Business: Social and Economic Infrastructure Services*, Research report, Australian Government, Canberra.

—— 2009b, *Performance of Public and Private Hospital Systems*, Research report, Canberra.

—— 2010, *Gambling*, Report no. 50, Canberra.

—— 2011a, *Caring for Older Australians*, Report no. 53, Canberra.

—— 2011b, *Disability Care and Support*, Report no. 54, Canberra.

—— 2012, *Benchmarking in Federal Systems*, Roundtable Proceedings, Productivity Commission, Canberra.

—— 2013a, *Annual Report 2012-13*, Annual Report Series, Canberra.

—— 2013b, *Compulsory Licensing of Patents*, Inquiry Report no. 61, Canberra.

—— 2013c, *Mineral and Energy Resource Exploration*, Australian Government, Canberra.

—— 2013d, *National Access Regime*, Australian Government, Canberra.

—— 2014a, *Childcare and Early Childhood Learning*, Inquiry Report no. 73, Canberra.

—— 2014b, *Natural Disaster Funding Arrangements*, Australian Government, Canberra.

—— 2014c, *Public Infrastructure*, Inquiry Report no. 71, Canberra.

—— 2015a, *Business Set-up, Transfer and Closure*, Inquiry Report no. 75, Canberra.

—— 2015b, *Efficiency in Health*, Commission Research Paper, Canberra.

—— 2015c, *Housing Assistance and Employment in Australia*, Commission Research Paper, Canberra.

—— 2016a, *Digital Disruption: What do governments need to do?*, Commission Research Paper, Canberra.

—— 2016b, *National Education Evidence Base: Draft Report*, Canberra.

—— 2016c, *Regulation of Australian Agriculture: Draft Report*, Canberra.

PCU (Pulse Credit Union) nd, *Pulse Credit Union Limited Privacy Policy*, http://www.mucu.com.au/about/privacy.html (accessed 4 July 2016).

Pearce, R. 2016, *How data analytics could change NSW*, Computerworld, http://www.computerworld.com.au/article/594426/how-data-analytics-could-reshape-nsw/ (accessed 15 September 2016).

PERC (Policy and Economic Research Council) 2012, *Credit Impacts of More Comprehensive Credit Reporting in Australia and New Zealand*, http://www.perc.net/publications/credit-impacts-comprehensive-credit-reporting-australia-new-zealand/ (accessed 8 April 2016).

Pettifer, R. E. W. 2015, 'The development of the commercial weather services market in Europe: 1970–2012', *Meteorological Applications*, vol. 22, pp. 419–424.

PGA (Pharmacy Guild of Australia) 2015, *Fact Sheet: Electronic Recording and Reporting of Controlled Drugs (ERRCD)*, https://www.guild.org.au/issues-resources/ehealth/electronic-recording-and-reporting-of-controlled-drugs-(errcd) (accessed 30 May 2016).

PHRN (Population Health Research Network) nd, *About us*, http://www.phrn.org.au/about-us/ (accessed 20 October 2016).

—— 2011a, *Application process*, http://www.phrn.org.au/for-researchers/data-access/application-process/ (accessed 7 September 2016).

—— 2011b, *How Is Data Linked*, http://www.phrn.org.au/about-us/data-linkage/how-is-data-linked/ (accessed 10 June 2016).

—— 2011c, *Linkage and Security*, About Us, http://www.phrn.org.au/about-us/data-linkage/linkage-and-security/ (accessed 16 October 2016).

—— 2016a, *Overview*, http://www.phrn.org.au/about-us/overview/ (accessed 20 October 2016).

—— 2016b, *Population Health Research Network Response to Medical Research Future Fund consultation for the development of the Australian Medican Research and Innovation Strategy and related Priorities*, http://www.phrn.org.au/media/80968/phrn_mrff-priorities-submission-_v10.pdf (accessed 20 October 2016).

—— 2016c, *Submission to National Research Infrastructure Roadmap Capability Issues Paper*, https://submissions.education.gov.au/Forms/National-Research-Infrastructure-Capability-Issues-Paper-Submissions/Documents/Population%20Health%20Research%20Network.pdf (accessed 15 October 2016).

—— 2016d, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*, http://www.pc.gov.au/inquiries/current/education-evidence/submissions (accessed 5 July 2016).

Pinterest 2016, *Privacy Policy*, https://about.pinterest.com/en/privacy-policy (accessed 8 July 2016).

Pollock, R. 2008, *The Economics of Public Sector Information*, University of Cambridge.

Poloni, G.D. 2014, *ASKAP telescope takes shape in WA outback*, ABC News, http://www.abc.net.au/news/2014-10-11/australian-square-kilometre-array-telescope-project-takes-shape/5805918 (accessed 15 September 2016).

Ponemon Institute 2016, *2016 Cost of Data Breach Study: Australia*, June, http://www-03.ibm.com/security/au/data-breach/index.html (accessed 18 July 2016).

Porter, M. 2016, 'GE and the turning point for Boston', *Boston Globe*, 20 January, https://www.bostonglobe.com/opinion/editorials/2016/01/20/and-turning-point-for-boston/WUBLLEidDbaEsqHwFmxN5H/story.html (accessed 30 August 2016).

—— and Heppelmann, J. 2014, *How Smart, Connected Products Are Transforming Competition*, Harvard Business Review, https://hbr.org/2014/11/how-smart-connected-products-are-transforming-competition (accessed 1 September 2016).

Poushter, J. 2016, *Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies*, Pew Research Center, http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ (accessed 13 October 2016).

Poynter, K. 2008, *Review of information security at HM Revenue and Customs - Final Report*, June, http://roselabs.nl/files/audit_reports/PwC_-_HM_Revenue_and_Customs.pdf (accessed 20 September 2016).

Presser, L., Hruskova, M., Rowbottom H and Kancir, J. 2015, 'Care.data and access to UK health records: patient privacy and public trust', *Technology Science*, no. 2015-08-11.

Privacy Committee Of South Australia 2015, *Privacy and Open Data Guideline*, http://www.archives.sa.gov.au/sites/default/files/20150121%20Privacy%20and%20Open%20Data%20Guideline%20Final%20V1.1_Copy.pdf (accessed 20 July 2016).

PSA (Pharmaceutical Society of Australia) 2016, *Real-time recording and reporting of drugs of dependence: Position statement*, https://www.psa.org.au/policies/position-statement-real-time-recording-and-reporting-of-drugs-of-dependence (accessed 20 October 2016).

PTV (Public Transport Victoria) 2014, *myki Privacy Policy*, https://ptv.vic.gov.au/assets/PTV/PTV%20docs/Privacy/PTV-Policy-myki-Privacy-Policy-September-2014.pdf (accessed 10 August 2016).

PwC (PricewaterhouseCoopers) 2013, *Where have you been all my life? How the financial services industry can unlock the value in Big Data*, PwC, http://www.pwc.com/us/en/financial-services/publications/viewpoints/unlocking-big-data-value.html (accessed 18 May 2016).

—— 2014, *Deciding with data: How data-driven innovation is fuelling Australia's economic growth*, http://www.pwc.com.au/consulting/assets/publications/data-drive-innovation-sep14.pdf (accessed 4 April 2016).

—— 2015a, *Is it time for consumer lending to go social? How to strengthen underwriting and grow your customer base with social media data*, PwC,

http://www.pwc.com/us/en/consumer-finance/publications/social-media-in-credit-underwriting-process.html (accessed 27 June 2016).

—— 2015b, *The extra mile: Risk, regulatory, and compliance data drive business value*, April.

—— 2015c, *The Internet of Things: what it means for US manufacturing*, February.

Qantas and nib 2015, *Qantas and nib to create a more rewarding health insurance experience*, Media release, 23 November, Sydney.

QBE Insurance 2016, *Insurance Box*, https://www.qbe.com.au/personal/quote/vehicle/insurance-box (accessed 5 September 2016).

QPS (Queensland Police Service) nd, *Disaster Management and Social Media - a case study*, Media and Public Affairs Branch, Brisbane.

Quantium 2016, *Quantium*, https://www.quantium.com/ (accessed 3 August 2016).

Queensland Government 2015, *1996-2000—Bowen tide gauge archived interval recordings*, Queensland Government Data, https://data.qld.gov.au/dataset/bowen-tide-gauge-archived-interval-recordings/resource/dd6d88b4-ae2b-42de-916a-b1417feea14e (accessed 16 October 2016).

—— 2016, *Water Monitoring Information Portal: Home*, Water Monitoring Information Portal, https://water-monitoring.information.qld.gov.au/ (accessed 24 August 2016).

Quisquater, J.-J., Guillou, L.C. and Berson, T.A. 1990, 'How to Explain Zero-Knowledge Protocols to Your Children', presented at the *Proceedings of the 9th Annual International Cryptology Conference on Advances in Cryptology*, pp. 628–631.

RACGP (Royal Australian College of General Practitioners) 2011, 'E-health', *The RACGP Curriculum for Australian General Practice 2011*, RACGP, Melbourne, pp. 439–450.

Rao, L. 2015, *How Palantir Is Helping Corporations Make Sense Of Big Data*, Fortune, http://fortune.com/2015/12/01/palantir-big-data/ (accessed 5 October 2016).

Reeve, J., Hosking, R. and Allinson, Y. 2013, 'Personal electronic health records: the start of a journey', *Australian Prescriber*, vol. 36, no. 3, pp. 70–73.

Research Australia 2016, *Australia Speaks! Research Australia Opinion Polling 2016*, http://researchaustralia.org/reports/public-opinion-polling/ (accessed 20 September 2016).

Research Data Services 2016, *Submission to National Research Infrastructure Roadmap Capability Issues Paper*, https://submissions.education.gov.au/Forms/National-Research-Infrastructure-Capability-Issues-Paper-Submissions/Documents/Research%20Data%20Services.pdf (accessed 15 October 2016).

Retailer and Supplier Roundtable Ltd 2014, *Paddock to Plate: Reform of Regulations to Enhance Competitiveness, Increase Productivity and Decrease the Cost of Doing Business*, Sydney.

Ritchie, F. and Welpton, R. 2011, 'Sharing risks, sharing benefits: Data as a public good', https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2011/21_Ritchie-Welpton.pdf (accessed 13 July 2016).

—— and —— 2014, *Addressing the human factor in data access: incentive compatibility, legitimacy and cost-effectiveness in public data resources*, Economics Working Paper Series, 1413, University of the West of England, http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/Economics%20Papers%202014/1413.pdf (accessed 13 July 2016).

Robert Walters 2013, *Understanding the role of social media to complement attraction strategies*, Robert Walters Whitepaper, Sydney.

Rogers, S. 2013, *Meet the man who turned David Cameron onto open data*, The Guardian, http://www.theguardian.com/news/datablog/2013/apr/30/rohan-silva-interview-david-cameron-open-data (accessed 15 June 2016).

Rosenblum, A. 2015, *Your Doctor Doesn't Want to Hear About Your Fitness-Tracker Data*, https://www.technologyreview.com/s/543716/your-doctor-doesnt-want-to-hear-about-your-fitness-tracker-data/ (accessed 23 May 2016).

Royle, R., Hambleton, S. and Walduck, A. 2013, *Final Review of the Personally Controlled Electronic Health Record*, Report to the Minister for Health, Department of Health, Canberra.

Rubinsztein-Dunlop, S. 2014, 'Warnings follow big supermarket moves into banking', Australian Broadcasting Corporation, *7:30*, transcript, http://www.abc.net.au/7.30/content/2014/s4062642.htm (accessed 7 July 2016).

Rudner, J., McDougall, C., Sailam, V., Smith, M. and Sacchetti, A. 2016, 'Interrogation of Patient Smartphone Activity Tracker to Assist Arrhythmia Management', *Annals of Emergency Medicine*, vol. 68, no. 3, pp. 292–294.

Russell, G. 2016, 'Australia's primary health care research needs an urgent check', *Sydney Morning Herald*, 18 April, http://www.smh.com.au/comment/australias-primary-health-care-research-needs-an-urgent-check-20160418-go8xig.html (accessed 9 June 2016).

SA NT DataLink 2015, *Submission to Inquiry into and Report on Health Policy, Administration and Expenditure*, http://www.aph.gov.au/DocumentStore.ashx?id=e86cc309-a6c7-4c5b-ab1e-cf704e7d9c8e&subId=407273 (accessed 6 June 2016).

—— 2016a, *List of Completed SA NT DataLink Projects by Financial Year*, Darwin.

—— 2016b, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*, http://www.pc.gov.au/__data/assets/pdf_file/0010/199711/sub057-education-evidence.pdf (accessed 1 June 2016).

—— 2016c, *Supporting health, social and economic research, education and policy in South Australia and the Northern Territory*, https://www.santdatalink.org.au/ (accessed 20 October 2016).

SA Premier 2016, *Parliament urged to deal swiftly with child protection reform package*, http://www.premier.sa.gov.au/index.php/john-rau-news-releases/1167-parliament-urged-to-deal-swiftly-with-child-protection-reform-package (accessed 27 September 2016).

Saiyid, A. 2016, *Real-Time Water Monitoring Data Challenging for Regulators*, Bloomberg Bureau of National Affairs, http://www.bna.com/realtime-water-monitoring-n73014446663/ (accessed 24 August 2016).

Samsung Electronics 2016, *Samsung Introduces an Entirely New Category in Refrigeration as Part of Kitchen Appliance Lineup at CES 2016*, https://news.samsung.com/global/samsung-introduces-an-entirely-new-category-in-refrigeration-as-part-of-kitchen-appliance-lineup-at-2016-ces (accessed 2 September 2016).

SANS Institute 2016, *Layered Security: Why It Works*, https://www.sans.org/reading-room/whitepapers/.../layered-security-works-34805 (accessed 1 August 2016).

Sax Institute nd, *The 45 And Up Study Policy On Collection Of Biological Specimens In Sub-Studies*, https://www.saxinstitute.org.au/wp-content/uploads/Policy-on-Collection-of-Biological-Samples-in-Sub-Studies.pdf (accessed 6 September 2016).

—— 2016a, *SURE*, http://www.saxinstitute.org.au/our-work/sure/ (accessed 10 June 2016).

—— 2016b, *SURE: The Secure Unified Research Environment fact sheet*, https://www.saxinstitute.org.au/news/sure-fact-sheet/ (accessed 20 October 2016).

Schaus, P. 2015, *Will online lenders disrupt small business banking?*, Banking Exchange, http://www.bankingexchange.com/community-banking/viewpoints/item/5795-will-online-lenders-disrupt-small-business-banking (accessed 3 October 2016).

Schrier, B. 2014, 'Government Open Data: Benefits, Strategies, and Use', *University of Washington Evans School Review*, vol. 4, pp. 12–27.

SCRGSP (Steering Committee for the Review of Government Service Provision) 2016, *Report on Government Services 2016*, Productivity Commission, Canberra.

SEEK nd, *Real company reviews from real employees*, SEEK.com.au, https://www.seek.com.au/companies/ (accessed 19 July 2016).

SELC (Senate Economics Legislation Committee) 2006, *Provisions of the Intellectual Property Laws Amendment Bill 2006*, Australian Government, Canberra.

Select Committee on Health 2016, *Big health data: Australia's big potential*.

Senate Committee on Commerce, Science, and Transportation (US) 2013, *A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes*, United States Congress, Washington, D.C., https://www.commerce.senate.gov/public/index.cfm/2013/12/data-brokers-report (accessed 4 July 2016).

Sensis 2015, *Senis Social Media Report May 2015*, May, Melbourne.

Sharwood, S. 2012, *PSMA defends data quality after NBNCo criticism*, The Register, http://www.theregister.co.uk/2012/08/08/psama_defends_gnaf/ (accessed 3 October 2016).

Shaw, D.R. 2014, *Personal Big Data - Is there a missing third party in our emerging Big Data society?*, White paper, March.

Sheng, C. 2013, *China Merchants Bank Tries Deeper Integration with WeChat*, TechNode, http://technode.com/2013/06/18/china-merchants-bank-tries-deeper-integration-with-wechat/ (accessed 26 September 2016).

Shinal, J. 2014, *Amazon, Alipay, PayPal are quietly becoming big lenders*, USA TODAY, http://www.usatoday.com/story/tech/columnist/shinal/2014/09/04/alipay-paypal-amazon-online-payments/14993343/ (accessed 22 September 2016).

SIIAA (Spatial Information Industry Action Agenda) 2001, *Positioning for Growth - Technical Report*, Australian Government.

Simons Institute for the Theory of Computing 2013, *Using Data-Oblivious Algorithms for Private Cloud Storage Access*, https://simons.berkeley.edu/talks/michael-goodrich-2013-10-24 (accessed 1 October 2016).

Singapore LTA (Singapore Land Transport Authority) 2008, *LTA Employs Innovation in Traffic Forecasting*, https://www.lta.gov.sg/apps/news/page.aspx?c=2&id=1998 (accessed 14 July 2016).

Singtel Optus 2016, *Privacy Policy*, https://www.optus.com.au/about/legal/privacy (accessed 30 June 2016).

Smith, J., Tennison, J., Wells, P., Fawcett, J. and Harrison, S. 2016, *Applying blockchain technology in global data infrastructure*, Open Data Institute, https://www.scribd.com/document_downloads/direct/315354748?extension=pdf&ft=1476661381&lt=1476664991&source=embed&uahk=hiq42iTsmm5QyscZW7Drd8tutG8 (accessed 20 October 2016).

Smith, R. and Hutchings, A. 2014, *Identity crime and misuse in Australia: Results of the 2013 online survey*, AIC Reports - Research and Public Policy Series, 128, Australian Institute of Criminology.

Snapchat 2016, *Privacy Policy*, https://www.snapchat.com/privacy (accessed 3 August 2016).

Solove, D.J. 2013, 'Introduction: Privacy Self-Management and the Consent Dilemma', *Harvard Law Review*, vol. 126, pp. 1880–1903.

SSCH (Senate Select Committee on Health) 2015, *Official Committee Hansard — Health* —— 2016, *Big health data: Australia's big potential*, Sixth interim report, May, Australian Government, http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Health/Health/Sixth_Interim_Report (accessed 17 May 2016).

Stanley, F. 2010, 'Privacy or public good? Why not obtaining consent may be best practice', *Significance*, vol. 7, no. 2, pp. 72–75.

State Records of South Australia nd, *Privacy law in South Australia | State Records of South Australia*, https://www.archives.sa.gov.au/content/privacy-law-sa (accessed 11 October 2016).

Statistics New Zealand 2016, *Integrated Data Infrastructure*, http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx (accessed 27 September 2016).

Stewart, D. 2015, 'Assessing Access to Information in Australia: The impact of freedom of information laws on the scrutiny ans operation of the Commonwealth government', *New Accountabilities, New Challenges*, ANZSOG ANU Press, Canberra, pp. 79–152.

Stiglitz, J.E., Orszag, P.R. and Orszag, J.M. 2000, *The Role of Government in a Digital Age*, Computer and Communications Industry Association.

Suncorp Group nd, *Suncorp Group Privacy Policy*, Suncorp Group, Brisbane.

SysNucleus nd, *What is Web Scraping?*, WebHarvy, https://www.webharvy.com/articles/what-is-web-scraping.html (accessed 25 July 2016).

Tasmanian Government 2016a, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*, http://www.pc.gov.au/inquiries/current/education-evidence/submissions (accessed 6 July 2016).

—— 2016b, *Tasmanian Government Open Data policy*, http://www.egovernment.tas.gov.au/stats_matter/open_data/tasmanian_government_open_data_policy (accessed 20 October 2016).

Tauberer, J. 2014, *The 8 Principles of Open Government Data*, https://opengovdata.org/ (accessed 27 September 2016).

Tay, L. 2012, *Immigration targets 'problem travellers' with analytics*, iTnews, http://www.itnews.com.au/news/immigration-targets-problem-travellers-with-analytics-321562 (accessed 27 September 2016).

Taylor, H. 2015, *PayPal has lent more than $1 billion to small biz*, CNBC, http://www.cnbc.com/2015/10/27/paypal-ceo-announces-working-capital-loans-have-crossed-1-billion.html (accessed 22 September 2016).

Technology Transactions 2013, *Woolworths Limited acquires a non-controlling 50% interest in Sydney customer data analytics firm Quantium for A$20m*, 1 May, http://tmt-transactions.com/woolworths-limited-acquires-a-non-controlling-50-interest-in-sydney-customer-data-analytics-firm-quantium-for-a20m/ (accessed 6 July 2016).

Telstra 2015, *Privacy Statement (Including Credit Reporting Policy)*, September, Telstra.

Telsyte 2016, *Smartphone Sales Down as Price Rises and Less Upgrades Impact Maturing Australian Market*, http://www.telsyte.com.au/announcements/2016/3/15/lwyakigaympj35g2khr66j9lwl5rr1 (accessed 5 September 2016).

The Benevolent Society 2015, *Submission to Inquiry into Service Coordination in Communities with High Social Needs*, https://www.parliament.nsw.gov.au/committees/

DBAssets/InquirySubmission/Summary/50884/013%20The%20Benevolent%20Society%20.pdf (accessed 6 May 2016).

The Data Scraping Group nd, *Data Scraping*, http://www.data-scraping.com.au/ (accessed 25 July 2016).

Directivity, Citrus and First Point Research 2015, *For love or money*, http://www.theloyaltypoint.com.au/ (accessed 25 July 2016).

The Treasury 2000, *Industry Self-Regulation in Consumer Markets*, August, Australian Government, http://archive.treasury.gov.au/documents/1131/HTML/docshell.asp?URL=01_prelims.asp (accessed 4 October 2016).

The University of Sydney 2016, *Submission to the Productivity Commission Draft Report on Intellectual Property Arrangements*, http://www.pc.gov.au/__data/assets/pdf_file/0004/194863/sub104-intellectual-property.pdf (accessed 5 June 2016).

TIO (Telecommunications Industry Ombudsman) 2016, *Complaint statistics January-March 2016*, http://www.tio.com.au/publications/news/complaint-statistics-january-march-2016 (accessed 29 September 2016).

Tudge, A. 2016, *Media Release: Digital transformation to enhance welfare payment system*, Department of Human Services, https://www.mhs.gov.au/media-releases/2016-08-01-digital-transformation-enhance-welfare-payment-system-0 (accessed 14 September 2016).

Turnbull, M. 2015, *Australian Government Public Data Policy Statement*, Department of Prime Minister and Cabinet, https://www.dpmc.gov.au/sites/default/files/publications/aust_govt_public_data_policy_statement_1.pdf (accessed 1 May 2016).

Turner, M.A. and Varghese, R. 2010, 'The Economic Consequences of Consumer Credit Information Sharing : Efficiency , Inclusion , and Privacy', *ResearchGate*, https://www.researchgate.net/publication/215991947_The_Economic_Consequences_of_Consumer_Credit_Information_Sharing_Efficiency_Inclusion_and_Privacy (accessed 4 October 2016).

Twitter 2016a, *Company*, https://about.twitter.com/company (accessed 18 May 2016).

—— 2016b, *Twitter Privacy Policy*, https://twitter.com/privacy (accessed 5 September 2016).

—— 2016c, *Twitter Terms of Service*, 27 January, Twitter, San Francisco.

UK Parliament 2016, *Digital Economy Bill 2016–2017*, http://services.parliament.uk/bills/2016-17/digitaleconomy.html (accessed 20 October 2016).

UN ECE (United Nations Economic Commission for Europe) 2015, *Big Data*, http://www1.unece.org/stat/platform/display/msis/Big+Data (accessed 1 August 2016).

UNSW Social Policy Research Centre 2015, *Opportunities for information sharing*, https://www.sprc.unsw.edu.au/media/SPRCFile/SPRC_Report__Opportunities_for_Information_Sharing.pdf (accessed 6 May 2016).

URMC (University of Rochester Medical Center) 2016, *Parkinson's App Celebrates Milestone, Featured by Apple*, https://www.urmc.rochester.edu/news/story/4528/parkinsons-app-celebrates-milestone-featured-by-apple.aspx (accessed 1 June 2016).

VAGO (Victorian Auditor-General's Office) 2015a, *Access to Public Sector Information*, 2015–16:20, Melbourne, http://www.audit.vic.gov.au/reports_and_publications/latest_reports/2015-16/20151210-access-to-information.aspx (accessed 23 March 2016).

—— 2015b, *Audit of Access to Information*, http://www.audit.vic.gov.au/reports_and_publications/latest_reports/2015-16/20151210-access-to-information.aspx (accessed 28 September 2016).

—— 2015c, *Financial Systems Controls Report: Information Technology 2014–15*, Melbourne.

—— 2015d, *Operational Effectiveness of the myki Ticketing System*, PP 40, Melbourne.

—— 2015e, *Realising the Benefits of Smart Meters*, Victorian Auditor-General's Report, 2015–16:8, Melbourne.

Van Alsenoy, B., Verdoodt, V., Heyman, R., Ausloos, J., Wauters, E. and Acar, G. 2015, *From social media service to advertising network - A critical analysis of Facebook's Revised Policies and Terms*, Draft, 25 August, v1.3.

VCAA (Victorian Curriculum and Assessment Authority) 2016, *National Assessment Program - Literacy and Numeracy Testing (NAPLAN)*, http://www.vcaa.vic.edu.au/Pages/prep10/naplan/index.aspx (accessed 13 June 2016).

Veda nd, *Comprehensive credit reporting — Your Credit and Identity*, http://www.veda.com.au/yourcreditandidentity/comprehensive-credit-reporting (accessed 7 April 2016).

Verizon 2016, *2016 Data Breach Investigations Report*, http://www.verizonenterprise.com/verizon-insights-lab/dbir/2016/ (accessed 5 September 2016).

VHA (Vodafone Hutchison Australia) 2016, *Privacy Policy*, http://www.vodafone.com.au/about/legal/privacy (accessed 5 July 2016).

Victorian Government 2016, *Information Technology Strategy for the Victorian Government: 2016-2020*, Department of Premier and Cabinet (Vic).

—— 2009, *Own Motion Investigation into the Department of Human Services child protection program*, Melbourne, http://www.parliament.vic.gov.au/papers/govpub/VPARL2006-10No253.pdf (accessed 5 June 2016).

W3C 2013, *An Overview of the PROV Family of Documents*, W3C Working Group Note, 30 April.

WA Land Information Authority 2015, *Providing access to WA government data*, http://data.wa.gov.au/open-data-policy (accessed 10 October 2016).

WA Ombudsman 2013, *Guidelines for the Management of Personal Information*, May, WA Government, Perth.

Waller, 2013 Matt and Boccasam, P.V. 2013, *How Sharing Data Drives Supply Chain Innovation*, Industry Week, http://www.industryweek.com/supplier-relationships/how-sharing-data-drives-supply-chain-innovation (accessed 6 October 2016).

Warkentin, M., Gefen, D., Pavlou, P. and Rose, G. 2002, 'Encouraging Citizen Adoption of e-Government by Building Trust', *Electronic Markets*, vol. 12, no. 3, pp. 157–162.

Warren, D. and Brandis, L.D. 1890, 'The Right to Privacy', *Harvard Law Review*, vol. 4, no. 5, pp. 193–220.

Waugh, P. 2013, *New data.gov.au – now live on CKAN*, Text, https://www.finance.gov.au/blog/2013/07/17/new-datagovau-%E2%80%93-now-live-ckan/ (accessed 27 September 2016).

WEF (World Economic Forum) 2011, *Personal Data: The Emergence of a New Asset Class*, https://www.weforum.org/reports/personal-data-emergence-new-asset-class/ (accessed 4 July 2016).

Weiss, P. 2010, *Borders in Cyberspace: Conflicting Government Information Policies and their Economic Impacts*, in Fitzgerald, B., ed., Access to Public Sector Information: Law, Technology & Policy, Volume 2, Sydney University Press.

West, M. 2016, *Investigation: ASIC fees highest in world, even before data sale*, Michael West, http://www.michaelwest.com.au/asic-fees-highest-in-world/ (accessed 20 October 2016).

Western Australian Land Information System nd, *Location Information Access Framework*, http://www.walis.wa.gov.au/projects/location-information-access-framework-liaf (accessed 10 October 2016a).

—— nd, *Location Information Strategy for WA*, http://www.walis.wa.gov.au/projects/location-information-strategy-for-wa (accessed 10 October 2016b).

Westpac 2014, *Westpac Privacy Policy*, 12 March, Sydney.

WhatsApp 2012, *WhatsApp Legal Info*, https://www.whatsapp.com/legal/ (accessed 3 August 2016).

Williams, M. 2013, *IAG's $1.85bn Wesfarmers underwriting bet*, Asia-Pacific Banking & Finance, http://www.australianbankingfinance.com/insurance/iag-s--1-85bn-wesfarmers-underwriting-bet/ (accessed 5 January 2016).

Winn, J.K. and Wrathall, J.R. 2000, 'Who Owns the Customer? The Emerging Law of Commercial Transactions in Electronic Customer Data', *Business Lawyer*, vol. 56, pp. 213–233.

Wood, J. 2008, *Report of the Special Commission of Inquiry into Child Protection Services in NSW*, November, NSW Government, Sydney.

Woolworths 2016, *Woolworths Group Privacy Policy*, February, https://www.woolworths.com.au/Shop/Discover/about-us/privacy-policy (accessed 22 June 2016).

WWWF (World Wide Web Foundation) 2015a, *Open Data Barometer Global Report*, 3rd edn, http://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-GlobalReport.pdf (accessed 1 July 2016).

—— 2015b, *Open Data Barometer Methodology - v1.0*, opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-Methodology.pdf (accessed 1 August 2016).

Yaraghi, N. 2016, *Hackers, phishers, and disappearing thumb drives: Lessons learned from major health care data breaches*, Center for Technology Innovation at Brookings, https://www.brookings.edu/research/hackers-phishers-and-disappearing-thumb-drives-lessons-learned-from-major-health-care-data-breaches/ (accessed 9 September 2016).

Yarnell, M. 2016, 'Data Sharing - The key to a better society?', *Australasian Lawyer*, http://www.australasianlawyer.com.au/country-editor/new-zealand/data-sharing--the-key-to-a-better-society-215842.aspx (accessed 1 June 2016).

YouTube 2010, *Terms of Service*, 9 June, https://www.youtube.com/static?gl=AU&template=terms (accessed 18 May 2016).