Australian Government
Productivity Commission

Data Availability and Use

Productivity Commission
Inquiry Report

No. 82, 31 March 2017

ISSN 1447-1337 (online)
ISSN 1447-1329 (print)
ISBN 978-1-74037-617-4 (online)
ISBN 978-1-74037-616-7 (Print)

An appropriate reference for this publication is:

Productivity Commission 2017, *Data Availability and Use,* Report No. 82, 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).

Australian Government
Productivity Commission

Canberra Office
Level 2, 4 National Circuit
Barton ACT 2600
GPO Box 1428
Canberra City ACT
Telephone 02 6240 3200

Melbourne Office
Telephone 03 9653 2100

www.pc.gov.au

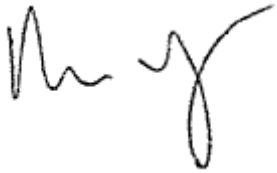
31 March 2017

The Hon Scott Morrison MP
Treasurer
Parliament House
CANBERRA ACT 2600

Dear Treasurer

In accordance with Section 11 of the *Productivity Commission Act 1998,* we have pleasure in submitting to you the Commission's final report into *Data Availability and Use.*

Yours sincerely



Peter Harris
Chair

Melinda Cilento
Commissioner

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]

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). Ms Cilento is also co-chair of the National Australia Bank's Advisory Council on Corporate Responsibility.

Contents

Acknowledgments

The Commissioners express their appreciation to the staff who worked on the Inquiry report and underlying analysis.

The Inquiry team was led by Rosalyn Bell and included Elise Whalan, Miriam Veisman-Apter, Greg Thompson, Joshua Runciman, Claire Prideaux, and Anthony Housego. Ian Moran, Kathryn Ovington (from Attorney-General's Department) and Gavin Walker (from CSIRO Data61) were also a part of the team for completion of the draft Report.

Abbreviations

ABS	Australian Bureau of Statistics
ACCC	Australian Competition and Consumer Commission
ACMA	Australian Communications and Media Authority
AGD	Attorney-General's Department
AGIMO	Australian Government Information Management Office
AGLDWG	Australian Government Linked Data Working Group
AIHW	Australian Institute of Health and Welfare
ALRC	Australian Law Reform Commission
ANAO	Australian National Audit Office
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
ARCA	Australian Retail Credit Association
ASAC	Australian Statistics Advisory Council
ASIC	Australian Securities and Investments Commission
ATO	Australian Tax Office
AURIN	Australian Urban Research Infrastructure Network
AUSTRAC	Australian Transactions Reporting and Analysis Centre
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
DSS	Department of Social Services
DTA	Digital Transformation Agency

DTO	Digital Transformation Office
FOI	Freedom of Information
GIF	Graphics Interchange Format
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
IP	Internet Protocol
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 has enabled a kaleidoscope of new business models, products and insights. Data frameworks and protections developed prior to sweeping digitisation need reform. This is a global phenomenon and Australia, to its detriment, is not yet participating.
- Improved data access and use can enable new products and services that transform everyday life, drive efficiency and safety, create productivity gains and allow better decision making.
- The substantive argument for 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 by both data custodians and users in existing data access processes and protections and numerous hurdles to sharing and releasing data are choking the use and value of Australia's data. In fact, improving trust community-wide is a key objective.
- Marginal changes to existing structures and legislation will not suffice. Recommended reforms are aimed at moving from a system based on risk aversion and avoidance, to one based on transparency and confidence in data processes, treating data as an asset and not a threat. Significant change is needed for Australia's open government agenda and the rights of consumers to data to catch up with achievements in competing economies.
  - At the centre of recommended reforms is a new Data Sharing and Release Act, and a National Data Custodian to guide and monitor new access and use arrangements, including proactively managing risks and broader ethical considerations around data use.
  - A new Comprehensive Right for consumers would give individuals and small/medium businesses opportunities for active use of their own data and represent fundamental reform to Australia's competition policy in a digital world. This right would create for consumers:
    - powers comparable to those in the Privacy Act to view, request edits or corrections, and be advised of the trade to third parties of consumer information held on them
    - a new right to have a machine-readable copy of their consumer data provided either to them or directly to a nominated third party, such as a new service provider.
- A key facet of the recommended reforms is the creation of a data sharing and release structure that indicates to all data custodians a strong and clear cultural shift towards better data use that can be dialled up for the sharing or release of higher-risk datasets.
  - 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. National Interest Datasets would be resourced by the Commonwealth as national assets.
  - A suite of Accredited Release Authorities would be sectoral hubs of expertise and enable the ongoing maintenance of, and streamlined access to, National Interest Datasets as well as to other datasets to be linked and shared or released.
  - A streamlining of ethics committee approval processes would provide more timely access to identifiable data for research and policy development purposes.
- Incremental costs of more open data access and use — including those associated with better risk management and alterations to business data systems — will exist but should be substantially outweighed by the opportunities presented.
- Governments that ignore potential gains through consumer data rights will make the task of garnering social licence needed for other data reforms more difficult. Decoupling elements of this Framework runs the risk of limiting benefits to, and support from, the wider public.

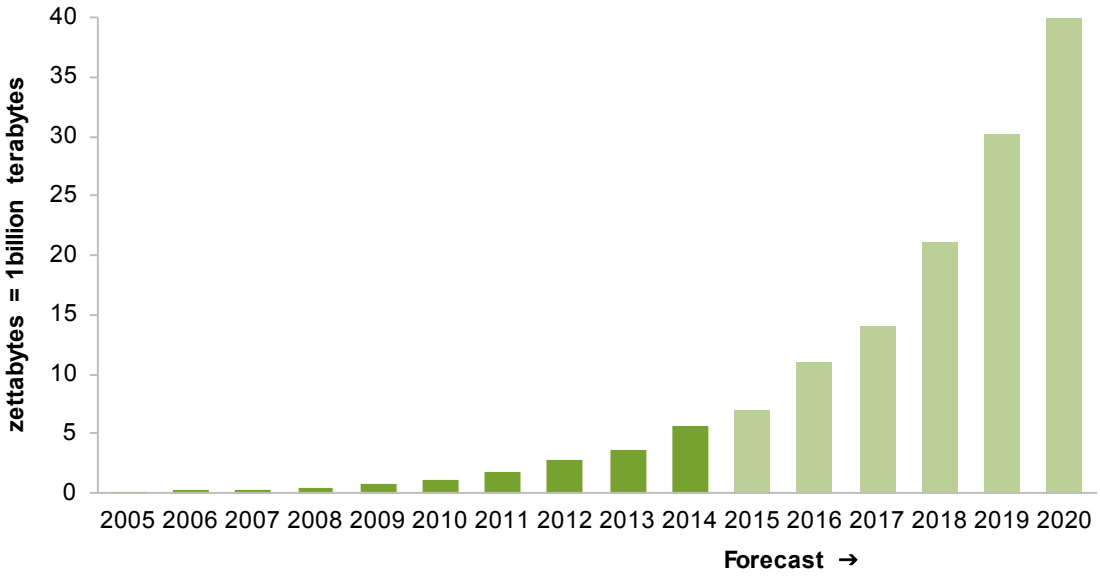
Overview

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, transformed 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 magnitude 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 may offer you a deal based on 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; as is the scope for Australia to forgo much of this value under the misconception that denial of access minimises 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. 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) or 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 from 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)
- data collected through personally controlled devices, such as smartphones and health monitors, that have increasing potential to assist both medical practitioners and patients.

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 powerful datasets, the equivalent of which capable, trusted researchers in other nations — the United States, New Zealand and the United Kingdom — can and do actively analyse to enable discovery and solutions to seemingly intractable problems. In that context, we fall short. Some of our best health researchers use UK health datasets, as ours are unavailable to them. Others wait up to eight years for access, in areas of life-saving significance.

Inquiry participants highlighted a range of health sector data that could underpin substantial long lasting benefits for the Australian community. We cannot afford to ignore these.

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 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:

> In a clinical sense, the lack of effective data sharing or data integration between different healthcare providers means that, in many cases, they are not in a position to deliver fully informed healthcare at an episodic level nor provide effective continuity-of-care to consumers. … [T]his represents a considerable waste of time and effort resulting in ongoing data quality issues. At a deeper level, it means that providers do not have access to a fully integrated patient history, with the potential for unnecessary duplication of services, particularly pathology and radiology services. This inability to see the 'big picture' of a patient's health is potentially dangerous, especially in an emergency setting….
>
> Administrative [health] data collections contain incomplete filtered and filleted data that limit their usefulness for planning and managing precision population health and monitoring the efficiency and quality of health services (Srinivasan et al. 2016, pp. 12, 21).

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, government agencies 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 (Research Australia, sub. 117). The researchers have been waiting for the linked data for more than eight years.

In the United Kingdom, 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', 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 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. The greater adoption of electronic health records in Australia (known as My Health Record) has the potential to enable more effective and holistic healthcare for patients who receive treatment from a range of healthcare providers. Pathology services are a case in point. 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). 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), so there are substantial gains in service efficiency and patient experiences 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. For example, 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.

The chequered history of electronic health records for all Australians (beginning in 2002) now has in its most recent version, My Health Record, trial outcomes that offer genuine hope for an effective nationwide rollout.

There is substantial potential for innovation in the use of data to improve individual health. Electronic health records 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).

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.

Risks from better data use are real but manageable

Allowing and enabling data more generally to be available and used 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. Public release of aggregated data on government regulatory activities, for example, may pose a very low risk of adverse consequences. Public 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 to 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 in some form 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 — and better managed than they are now. The likelihood of unintended or inappropriate data use 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. This does not mean every possible data release has a technical solution: wherever and while it is not possible to reduce risks to an acceptable level, the approach being advocated by the Commission would not support public release of the data.

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 are often unaware 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)

That privacy is often said to be a concern, and yet 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 and is collected in seemingly innocuous ways (mobile device location data being an example), 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. Often the amount of information that is required to be

provided far exceeds that necessary for delivery of the product or service that was initially sought. What you are consuming, how and when you are consuming it, is all being collected as data that is likely of considerably more value to the supplier than users may appreciate. This is not to condemn value-adding but rather to highlight the potential imbalance; and the risk to continuing consumer support that must accompany it.

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).

Increasing data use does not necessarily increase risk

In reality, most risks of data misuse arise not through controlled sharing of data or the public release of robustly de-identified data, but rather from dataset hacking made possible by poor or outdated data collection, storage and management practices, 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. If data collectors fail to remain vigilant and up-to-date in technology, the present generally benign environment for collectors (public or private) may alter. Regaining social licence in such circumstances may be very difficult.

Most researchers, ethics committees, public sector data custodians and large private sector data holders are strongly motivated to handle data in ways that retain community trust and their organisation's (or own personal) reputation. But where that motivation is combined with uncertainty about privacy and secrecy requirements, or an indifference to the reservations of those contributing data, then a disproportionate reporting of bad experiences can have a long lasting chilling effect on dataset sharing and release.

That most data breaches are inconsequential and go largely unnoticed is hardly the point. The tipping point is unknown, until too late.

Tightening privacy legislation or creating new offences cannot prevent human error; nor are they likely to dissuade criminal intent. Further, to the extent that such responses inhibit well-intentioned testing of the robustness of de-identification approaches and security measures, and inflame the risk aversion of data custodians, they would represent a setback in Australia's data sharing and release efforts.

Worst of all is inaction, a feature of many failed requests for data access that were described to this Inquiry. This not only denies Australia a raft of opportunities, it also takes no account of incentives — 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.

Greater use of data does not mean Australians should be put at greater risk of harm. A key issue in balancing access and trust is consideration of the level of data required for different uses and how to manage well any associated risks. 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 generally not currently necessary in order to obtain most 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, that also needs to be made more accessible for use and re-use.

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. Privacy has carved out a space, but privacy is only one aspect of data use, and a defensive one at that. Restrictions on use for data collections in the same field, even the same institutional setting, vary significantly. Uncertainty endorses inaction.

Yet 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 distributed widely in the private sector — will not diminish. It is a global movement and, to its detriment, Australia is not actively participating; and has remained nervous about making decisions.

Adding more ad hoc adjustments to existing structures and legislation will not suffice. 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 could be readily transferred across the economy, between sectors and across national boundaries with increasing ease. To ensure public awareness and social licence match these trends and that we make the most of them, 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. Culture shift requires both persistent change of attitude and permission to demonstrate that.
- There are key unanswered questions that go to the fundamental rights of individuals to control data held about them, and how individuals — as consumers — 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.

Broad criteria shaping the recommended 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; be able to adapt to higher risk; and preserve commercial incentives to collect, maintain and add value to data.

Recent progress in policy and practices around data management is acknowledged in the report: spatial data and population health have been strong for many years; linking tax and business performance data is nascent; and some State governments have been proactive. But overall, and despite positive sentiment for some time now, progress has to date been insufficient, given the broad effects possible across the social economy from the massive growth in digital data.

Thus the Commission has quite deliberately recommended the creation of a new, broad-reaching Framework that should, by design, be capable of enduring beyond current technologies, 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.

In fact, it is vital for Australia's data future that the risks of data handling are managed well. Businesses, as much as governments, rely on the willingness of the public — the source of so much of the data — to continue to trust data handling and use. Against the background of an ocean of personal data that is already public, there is now, and will be in the future, a need for continued community acceptance and trust in the handling of personal data by both governments and business. Social licence will develop if people:

- have a sound basis for believing in the integrity and accountability of entities (public and private) handling data
- feel they have some control over how their own data is used and by whom, and an inalienable ability to choose to experience some of the benefits of these uses themselves
- better understand the potential community-wide benefits of data use.

It can be difficult for a data holder to know if they have community support for use of data; but they will almost certainly know if they do not. Moreover, community acceptance of greater data use is not a one-off concept, nor is it enough that only a few better practice firms and agencies demonstrate a positive approach. Efforts of data holders to maintain community acceptance for the use of their data need to be ongoing and near universal.

Crucially, while the protections applying to personal information under the *Privacy Act 1988* (Cth) would remain, the recommended reforms would also take Australia beyond the stage of viewing data availability solely through a privacy lens. This recognises that there is much more than privacy at stake when it comes to data availability and use; this is not an Inquiry into privacy.

Although we would have preferred to find solutions that are non-regulatory, it is a clear conclusion of the Inquiry that legislative change is needed to implement the Commission's recommended reforms. This change primarily involves the creation of new Commonwealth legislation — a new Data Sharing and Release Act (DSR Act) — that would apply to all digital data.

By giving consumers new rights to use their digital data and data holders permission to be pro-active about data possibilities, the DSR Act creates a new lens through which to view data; the lens of a valuable asset being created and utilised, not merely a risk or an overhead.

The new Framework

We are recommending two facets to Australia's data Framework for the future:

1. a new right that enables both opportunities for active data use by consumers and fundamental reform in Australia's competition policy

2. a structure for data sharing and release that would allow access arrangements to be dialled up or down according to the different risks associated with different types of data, uses and use environments.

Each of these facets is discussed in detail below.

Giving individuals more control over their digital data

Australian consumers have little capacity to choose how digital data about them is used; and too often, organisations and governments make decisions (after complying with privacy principles) about the use of individuals' data on their behalf. 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.

The Commission is recommending that Australia's consumers — both individuals and small and medium sized businesses (SMEs) — be afforded a new Comprehensive Right to the use of their digital data. This Right would apply to digital data holdings only. The regulation would not require businesses with paper records to digitise these in order to supply consumer data. Businesses are increasingly holding data in digital formats and it is inevitable that the value of the Comprehensive Right for consumers would grow over time. And where businesses for their own purposes do convert legacy records to digital form, such information would become potential consumer data.

While the recommended new Right for consumers provides features that match those inherent in privacy provisions (ability to view your information and propose changes to it), it is not, nor is it intended to be, a replica of privacy law. Rather, the new Right is meant to lift up the opportunity for consumers and offer a *genuine* two-way street to support their continuing willingness to supply a crucial input to business, research and public policy — namely, their data (whether obtained directly or through other channels). Consumers would no longer be just a source of data, they would rank equally with the key data collectors — businesses and governments — in being able to trade and use their data.

The ability to require the transfer of your data from one data holder to an alternative party would offer consumers the opportunity to trade safely and conveniently on their data, as business and (increasingly) governments do. To date, Australian consumers have not had much of an opportunity. Other nations (United Kingdom, European Union, for example) are more active in this regard.

Apart from building social licence through greater opportunity to use data, this Right would afford individuals and SMEs more choice about the products and services they consume, and the providers of those, and be an avenue to improve market competitiveness and innovation. No longer will it be just the collectors of data that are able to determine its uses and realise its value.

The digital data of all businesses, not-for-profits, government agencies and government business enterprises (excluding data that is collected for security purposes) should be subject to this new Comprehensive Right when an individual or SME is seeking to transfer their consumer data.

A right that will regenerate markets, and widen service choices

The right for consumers to request that their data be transferred to a third party — be it an entity in the public or private sector or a not-for-profit — is very likely to reinvigorate competition policy.

Barriers to entry created by regulation or infrastructure choke points were the key focus of the original competition policy reforms more than twenty years ago, and analysis by the Productivity Commission showed how important they were to economic growth in Australia for many years thereafter. Today's barriers to competition include, high up on the

list, the availability of information — to consumers in order to know what is available today by way of new services; and to service providers, to identify interested consumers and improve the efficiency of operations. Information asymmetry has long been recognised in economics as a feature that weakens competitive markets.

Powerful as that may be, other benefits from consumer control of their data extend beyond competition between providers in a market. 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 would be to form a new customer relationship, or obtain a quote that may lead to one, at the consumer's discretion. While such trading in data is readily imagined (and seen to a limited degree via comparison web sites in a few service industries) it most likely would allow consumers to exercise real choice in ways yet unimagined, and play a role in the innovation of new products and services for them, a key feature of data use in recent years.

Under the Commission's recommended plan, the scope of digital data available to be accessed and transferred under this Right would desirably be developed and agreed by participants within each industry.

The overarching outcome to be legislated is that the scope of consumer data in an industry is that which is sufficient to generate a competitive offer for an individual's custom from another provider. In other words, the type of data held on an individual or SME that a competing or complementary service provider would themselves need, and *reasonably* expect to obtain, in seeking to provide a competitive offering.

At its broadest level, consumer data should include:

- personal information (as defined in the *Privacy Act 1988* (Cth)) that is in digital form
- files posted online by the consumer
- data created from consumers' online transactions, Internet-connected activity or digital devices
- data purchased or obtained from a third party that is about the identified consumer
- other data associated with transactions or activity that is held in digital form and relevant to the transfer of data to a nominated third party.

Data that is only *imputed* by a data holder to be about a consumer — that is, data that has been created by a data holder through the application of insights or analysis such that it cannot reasonably be considered the consumer's data — should be included in consumer data *only* with industry negotiated agreement. Illustratively, if an insurer had determined through its own analysis that people who drink a lot of milk and eat red meat are very good car insurance risks compared with those who buy petrol at night and drink spirits, we would not expect that information to be included in the data to be transferred, unless an industry agreed on its inclusion.

All data does, of course, have originating sources. But it is not always reasonable to trace the trail back to its source. When multiple data sources are transformed to an extent that it is merely probable, but not certain, that a characteristic is associated with an individual consumer, this data would most likely be proprietary information of the data holder entity or perhaps a data analytics supplier. Another party seeking access to it should invest accordingly, as others have done before them.

Various terms have been suggested to describe such data that might be exempted from consumer data (or included for that matter) through industry negotiated agreement. We favour imputed data as a general descriptor — imputed being a known term in both statistics and the law. Some have suggested an exemption for value added data. This is too broad: data that has been cleansed of errors, made better through simple statistical means such as aggregated or averaged for each consumer but remains unaltered, or made machine-readable could all be construed to be value added. Consumers would be unreasonably deprived under such a descriptor. This would hardly contribute to maintaining community support in a world of increasing data exploitation.

Others have suggested that consumer data could be limited to just the transactional data generated in the direct course of the relationship between the consumer and the data custodian. In a few industries, this may be sufficient to achieve a competitive outcome for a consumer; negotiation would demonstrate that under the recommended approach. But with the exhaustive depth and breadth of data collection today and the heightened relevance of knowing your customer to achieve the outcome specified in our approach, the likelihood of this seems limited.

Similarly, there may be some data over which an entity, other than the firm that is the subject of the consumer data request, holds an intellectual property right, even where data may still be identified with a consumer. While it is legally possible to require that such data be nevertheless provided to identified consumers, this might prove to be an extreme step.

The abuse of intellectual property rights to prevent consumers accessing their data would be troubling — a sister report to this one is currently under consideration by the government. But until the Comprehensive Right is in operation and evidence of abuse, if any, of intellectual property rights emerges, the Commission prefers to allow this to be excluded from consumer data.

Industry-agreed coverage of consumer data would be determined in a data-specification process, the outcome of which would be registered with the Australian Competition and Consumer Commission (ACCC), which may approve, reject or offer interim approval.

In the absence of industry agreement as to the composition of consumer data, the broadest level definition (discussed above) should be incorporated into a consumer's request for their data held by an entity. The ACCC would determine, through the presence or absence of a registered industry data-specification what level of access a consumer was entitled to, should a dispute arise.

The joint nature of the data that is subject to the Right (that is, shared between the individual consumer and the businesses or agencies that hold the data) should ensure that incentives for data holders and collectors to collect data persist. Such entities would not be deprived of use of the asset, even as the consumer also uses it.

Transfers in action

The technological approach adopted to enable a transfer of data has received some attention.

We — and indeed, a majority of stakeholders responding to the Draft Report — are strongly against locking in a particular technological approach that all industries must adopt. In industries where frequent, real-time data transfers are needed and market participants have already made steps to enable this, transfer of data may be best achieved by the use of Application Programming Interfaces (APIs). In other industries where transfers reflect more one-off requests by consumers, alternative technology that enables the secure transfer of files may be a better approach.

Either way, standards around data formats and definitions would be necessary. We consider that participants in each industry, rather than governments, are best placed to develop these standards and determine them as part of the industry-agreed approach to transfer technology. And unlike our Draft Report, we no longer propose any additional right to opt out of data collection.

Knowing when your data has been sold

One of the most potentially pernicious practices with data is the onward trade or disclosure of data to third parties, leaving consumers unaware of who knows what about them. The damage is often not so much in monetary terms but in the feeling of exploitation. This has great capacity to undermine social licence over time, if misused. Around half of all Australians surveyed by Office of the Australian Information Commissioner (OAIC) have expressed concern about unknown organisations having obtained their personal information.

We do not propose that consumers be advised on each occasion data is traded or otherwise disclosed to a third party — the burden on businesses using contractors and outsourcing aspects of their operations could be enormous. Moreover, consumers in some areas could be inundated. But advising on which organisations data has been traded or disclosed to is a reasonable expectation of what is, after all, a joint right to data. You should surely be informed that something in which you now have a joint right is traded or disclosed to a third party.

Accordingly, entities should inform consumers about their data being traded or disclosed by including in their privacy policies, terms and conditions or on their websites, a list of

parties to whom consumer data has been traded or otherwise disclosed over the past 12 months. Such lists should easily accessible to consumers and updated in a timely manner.

Consumers may also be at risk of loss of data access on the wind up of a firm. In such circumstances, consumers should always be advised of who now holds their data if it is transferred (as an asset) by the insolvency practitioner; or dataset owner if the data is separately sold.

Costs, timeliness and transition

We recognise that there may be costs to business associated with their adherence to the Right. There are a number of aspects of the recommendation that seek to ensure these are manageable.

First, as noted above, it is expected that industry sectors themselves would determine the scope of data to be transferred, subject to approval by the ACCC.

Second, businesses and government data holders would be able to charge for costs reasonably incurred in transferring consumer data. We fully expect that there may be a tiered approach to such charges, namely that some digital data that is of high quality, readily available, and clearly identifiable with a particular individual (such as transactions data), should be made available at low or no cost and at relatively short notice. Data stored on different (yet still digital) systems, or that is of lesser quality may require additional effort to provide in a usable format and therefore could attract a higher charge and take longer. This would be for data holders themselves to determine *and explain.*

Our intention in recommending the creation of this Right is to enhance consumer outcomes, as a contribution to sustaining community support for the role data will play in the future. Business and governments as data holders would need to adjust to this Right. Neither should have interests in creating a process that was so costly as to prohibit its take up by most if not all consumers, as this would be counter to enhancing consumer outcomes and may eventually undermine the quality of data collections.

To make the process manageable, it is surely preferable to offer the parties affected in incurring expense the chance to meet the *intent* of the Right, namely enabling consumers to use their data. This is likely to involve degrees of iteration and transition. But the clear expectation is that there would be transparency on the part of businesses and agencies. Over time as systems evolve, the time taken and the cost involved should fall as these processes become part of each firm growing its business or government agency keeping faith with its clients, and while volume of data transferred might reasonably be expected to grow.

Similarly, it is expected that businesses and government data holders themselves would likely reap benefits from system transformation and better data management, such that all of the costs would not reasonably fall to consumers availing themselves of the Right.

Support for consumers in exercising their new Right

The ACCC would be the primary government entity charged with ensuring consumers are able to transfer their data and exercise their new rights. Specifically, any charges levied by data holders for access, editing, copying and/or transferring of data should be monitored, with the methodology used by a data holder recorded, transparent (such as on the data holder's web page) and reviewable on request by the ACCC.

While recourse for consumers not satisfied with the way their new Comprehensive Right can be exercised could primarily be through the ACCC, we recognise there are other bodies — industry-specific ombudsmen, State and Territory fair trading offices, and the OAIC — that may have industry-specific skills and knowledge to deal with particular complaints. There should be a 'no wrong door' approach to this. This means the key regulators need to implement systems that enable consumer concerns to be handled with efficacy — not leave the consumer straddling a regulator abyss.

While the changes proposed aim to enable consumers to exercise more control over the collection and use of their data, the onus remains on individuals to make responsible choices regarding to whom they provide personal information 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 (CCR).

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 is clear that a critical mass of accounts is not achievable on that basis. In the event that voluntary participation in the scheme remains below a critical mass of 40% in mid 2017, the Australian Treasury should proceed with developing draft legislation to mandate comprehensive credit reporting.

We note this is a date almost upon us, but those who have argued for more time could do their case greater service by noting that the industry has had since December 2015 to show clear movement, and a substantial notice period of CCR's imminence prior to that. Greater movement, if it occurs by December 2017, may convince the government to hold off proceeding to parliament. But the preparation of legislation to make participation mandatory cannot be a shock after such a time; and those with subsidiaries in New Zealand must surely have gained relevant experience in CCR participation from there.

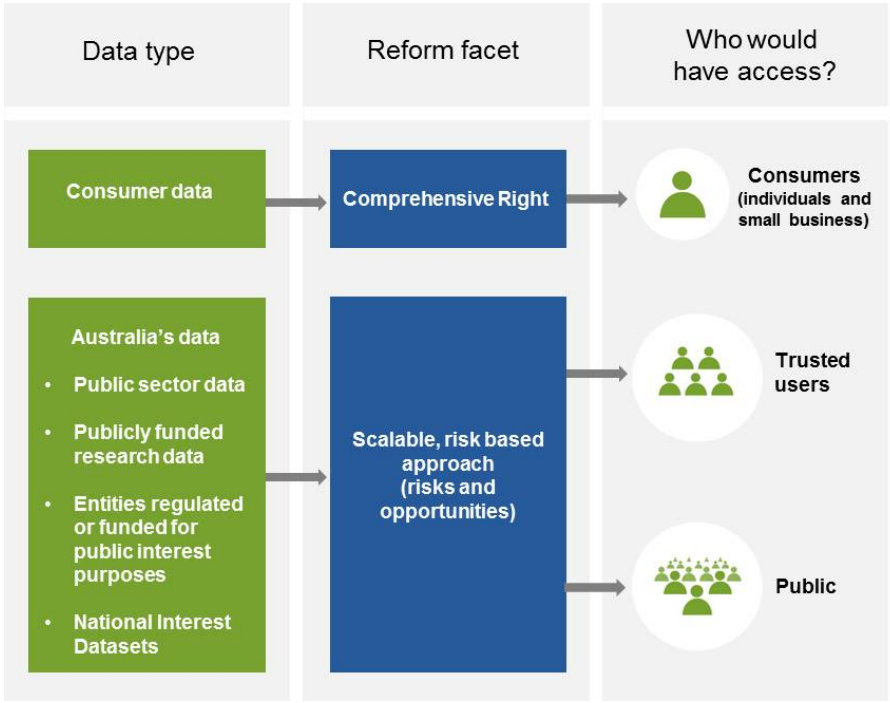
A structure to give substance to data aspirations

At both Commonwealth and State/Territory levels of government, there is an aspiration to improve the internal and external release and sharing of their data, with the objectives of creating opportunity for innovation by firms, lifting accountability and improving service delivery. New South Wales and South Australia have created legislative frameworks to put substance into aspiration. The Commonwealth has yet to do so.

Thus a substantive element of our reform Framework is a structure that provides institutional capability to seriously address these objectives and remove impediments to a consistent approach to handling data with varying levels of value and risk (figure 3).

Absent legislated permission to utilise the opportunity presented today by data analytics, aspiration will remain primarily a hope carried by well-intentioned but often under-equipped individuals.

Figure 3 **Framework of the recommended approach**



Institutions

In addition to focussing on institutional capability, our approach seeks to clearly define institutional accountability. The Commission has recommended both central agency responsibility for data in each jurisdiction and establishment of a new statutory office holder, the National Data Custodian (NDC). The NDC would be established as a statutory officer with a small advisory board to provide leadership and ongoing updating of the new

data Framework, including managing the broader ethical considerations around data use that are increasingly arising with the mass digitisation of data (box 2).

One of the NDC's primary activities authorised in the new DSR Act would be to accredit the processes and capabilities of a suite of Accredited Release Authorities (ARAs). The NDC would also publish guidance on data use — both opportunity and threat — for the benefit of ARAs and other data custodians, and to update continuously its accreditation processes. It would audit ARAs from time to time for compliance with both guidance material and the terms of Commonwealth funding.

ARAs would be sectoral hubs of expertise, independent entities that are funded by the Commonwealth (but may be based in a State or Territory, or even a not-for-profit entity that has a public interest focus). They would be tasked with:

- developing and integrating datasets drawn from across a sector (and jurisdictions) with high prospect to improve data use and release
- assisting a field of data custodians to improve the curation and quality of datasets to be released (including de-identifying or linking where necessary)
- facilitating timely updates and ongoing dataset maintenance
- approving trusted users of more sensitive data
- determining whether a dataset that they are responsible for should be shared or released.

Under this structure, access arrangements and safeguards could be dialled up or down by the ARA according to the nature of opportunities and risks likely. ARAs would have the capability necessary to make these judgements and apply them effectively. The extent to which use of a particular dataset(s) could be enabled to provide broad benefits beyond those currently derived would be a primary factor in the Commonwealth choosing to fund an ARA.

ARAs would, under governance arrangements agreed with cooperating dataset custodians, be responsible for whether a dataset is available for public release or limited to sharing with trusted users. Dataset custodians would retain their legislated responsibility for original data contributed to ARAs and so make the choice regarding its release or sharing in that *original* form; but where an ARA goes on to transform data (for example via integration, linkage, de-identification or use of Artificial Intelligence) the *transformed* dataset would become the primary responsibility of the ARA. The ongoing maintenance and updating of the transformed dataset would necessitate cooperation between ARAs and custodians of component datasets.

Box 2 Key participants in the new Framework

The National Data Custodian (NDC)

The new position of the NDC creates a role that would provide the focus of improved national effort to lift access to data and use of data analytics to contemporary standards and objectives. The NDC would have responsibility for enabling effective use of data; oversight, guidance and updating operations of the national data system; and be instrumental in assessing for designation possible National Interest Datasets (NIDs). The NDC would also accredit release authorities within the reformed national data system and have broad responsibility for determining which release authority is most suitable to handle the ongoing management of key datasets. The Office of the NDC would include a dedicated ethics adviser capable of providing advice and guidance on ethical issues related to data access and use. A small advisory board would support the NDC in technical matters of data access and use.

Accredited Release Authorities (ARAs)

ARAs would 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 is a working 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, approving 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 NIDs and other datasets. 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, to the NDC, government, and to the broader community of data custodians.

Trusted users

Trusted users would be accredited by ARAs to access data under its control or governance. They may include any individual working in an entity that has in place the necessary data governance requirements to safely handle the datasets in question and a signed legal undertaking that sets out safeguards for use and recognises relevant privacy requirements. Personnel in relevant Commonwealth and State/Territory government agencies and publicly funded research bodies should be eligible for accreditation as a trusted user. Private researchers should be able to achieve trusted user status, once their employer's commitment to the ARA's standards and approval requirements is evident.

Those wanting access to datasets maintained by ARAs would require the necessary governance structures and processes to address the risks associated with data use or release as a consequence of research findings, including access to secure computing infrastructure. Accreditation as a trusted user should enable a researcher to access a particular dataset or datasets on an ongoing basis for a program of work, rather than project by project.

There may remain some trusted users who access datasets held by data custodians under a variety of conditions set by each custodian. The introduction of ARAs would not alter these arrangements.

High value datasets

The Commission's Terms of Reference asked for advice on selecting high value datasets. The range of possible nominations for this is very broad and many submissions offered candidates. We have not, however, nominated any winners. Instead, informed by the breadth of advice and the impediments to maximising net benefit for Australians from further public investment in our data, we have designed processes to be applied to deliver the kind of outcome envisaged by recent governments' statements and our own assessment.

National Interest Datasets

Governments across Australia hold enormous amounts of data, but mostly lag behind other comparable economies in beneficially using data beyond the purposes for which it was initially collected, or allowing others access to do so. Australia's private sector data holders are often more innovative in their use of data. The public sector needs to spend the time and effort to identify strong candidates for the kind of investment via data analytics that the private sector is now applying. Absent this effort, the public sector will be denuded of core analytical capabilities in a digitally-driven future, and forgo the opportunities afforded by those capabilities.

The first step in this effort is to identify National Interest Datasets (NIDs).

From submissions in response to the Draft Report, it is evident that the concept is very positively viewed, albeit with uncertainty about the methodology of making the choices amongst many obviously very important datasets.

The simple summary of the process outlined below (and in more detail in the Report), concentrates on additionality. What additional community-wide benefit could be achieved by ensuring and consistently enabling wider sharing, use and possibly release of that data?

Designation as an NID would achieve wider sharing, use, and where appropriate, release by:

- providing sufficient and predictable funding for data curation and updating
- aggregating jurisdictionally-separated datasets
- providing a framework in which to better link datasets across sectors and fields of endeavour.

For instance, significant improvements could come from aggregating data across the States and Territories in health, education, social welfare, child support, aged care, and better linking them with elements of datasets from other fields — the population census, taxation, employment, business ownership, telecommunications, private health insurance or housing. This is not an exhaustive list. State datasets in justice, infrastructure, land use and

property ownership could be similar candidates for permanent linkages and curation by ARAs with a plan for delivering identified national benefits from this investment.

Such benefits would be identified on an ongoing basis by the NDC, after an initial set of NIDs is put forward with the DSR Bill.

Designating datasets as national interest collections would also signify their value as resources collected in the national interest. But this is not about labelling a collection as important in principle; its purpose must be additional national benefit in practice.

NIDs are thus likely to have the following characteristics:

- relatively few in number, certainly not in the thousands
- linked, integrated, transformed (for example, by de-identification or use of Artificial Intelligence) to suit a prior determined scope of outcomes
- offering clearly described public interest benefits of national application
- have been confirmed by public review (we are proposing consideration by a parliamentary committee).

For datasets designated as NIDs, new access and use arrangements devised under the DSR Act would take precedence over existing restrictions — Commonwealth or State — to access.

Impediments from jurisdictionally inconsistent access arrangements, confidentiality standards, collection purposes, obligations to seek approval for use and privacy rules would be replaced by a single legislated modern fit-for-purpose regime to apply when data is designated an NID and transferred to an ARA.

The extent to which this modern fit-for-purpose regime transforms each NID would be determined ultimately by the final legislated design and the restrictions in the head legislation of a data collection that is to be incorporated into an NID.

The DSR Act should draw upon all Commonwealth Constitutional powers — including of particular value, the telecommunications powers. This is likely to be effective with many datasets. Nevertheless, advice to us is that depending on the specific chosen datasets, some collection-specific regulation may also be required.

The intent is nevertheless that by the act of bringing forward datasets and arguing in public fora for their freedom to be used in the national interest, both the public and private sector (the latter in rare circumstances, exemplified by private education and health data) can create an asset and governance structure that each would be comfortable with, and which the Commonwealth could support in order to deliver the identified benefits.

We have recommended a parliamentary committee process be used to expose the NDC's analysis of national benefit to public comment. It is not often that the Executive willingly offers parliament the opportunity to critique its proposals but we see a number of benefits to this process.

First, the term national interest is an essentially arbitrary one. Rather than attempt to pin it down by definitional drafting, we propose that the comment of our national representatives is used. Second, the approach is consistent with seeking wherever possible to obtain social licence via public involvement. Third, it would allow for the NDC along with others in the community to continuously make known why better use of data is indeed in the national interest. This educational approach to community consultation would be far more effective than what is often simply box-ticking consultation. In other words, we recommend engaging in a full parliamentary scrutiny process prior to designating NIDs by disallowable instrument.

Datasets afforded NID status should be maintained as a national asset for a period of at least 10 years.

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

Datasets that provide additional opportunities for Australia-wide gains

Most datasets would not be NIDs. Beyond them, the ARA model is also intended to offer the opportunity for the much larger range of datasets in public hands to be better curated and managed for cooperative release by, or with the assistance of, ARAs. Thus, any dataset provided to an ARA under the DSR Act would also be subject to the single modernised legislative regime to enable improved data access. And Commonwealth data custodians, whether ARAs or not, should observe and apply the NDC's guidance on better data practice; and report on it annually.

Fee for service arrangements, such as that taken by the Australian Institute for Health and Welfare today, would support the resourcing required by ARAs.

The special case of higher risk data

Some data that identifies (or has the potential to identify) individual persons or businesses is already shared in a very limited way with policy developers and with researchers within government and/or the research community. This data is typically used for targeted program and product/service delivery, for research (such as rare medical conditions) where there are very small populations involved, and (in some limited cases) for regulatory compliance.

The current process requires, however, multiple approvals, delay and (often) work-arounds for all involved, even where there is great acceptance of the validity of the work. It surely can be improved, safely.

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 guideline and approval step may be reasonable in its own right, but collectively it is costly and can be self-defeating. We were advised of Australian researchers abandoning attempts to access Australian data in favour of UK, EU or US data.

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. Researchers in fields that require access to identified data already have strong professional reasons to avoid misusing it.

Simplified but still protected access to identifiable data would be enabled via ARAs. This would include all requests for access to identified data within ARA-managed datasets — whether NIDs or otherwise (the latter being where original data custodians have offered data to an ARA, thus allowing an ARA to create a transformed ARA-held dataset). Governance arrangements and technical capability to enable such access with safeguards would form a core part of the NDC's accreditation for that ARA.

Current duplicative efforts of ethics committees and lack of recognition of approvals granted should be addressed to streamline access to data by researchers. We are recommending reforms to both the registration processes for human research ethics committees and approval processes.

The capacity to use information without obtaining consent of individuals would be extended under the DSR Act to cover all public interest research, rather than be limited to health and medical research purposes as is the case at present. The NDC would be charged with resolving whether research in question was public interest or not.

Access would occur in a specified secure computing environment with output from the dataset reviewed (by an automated process, where possible) prior to project completion to ensure risk management approaches had been satisfactorily implemented.

Responsibility for appropriate use of datasets would rest with trusted users and their institutions, with incentives to maintain necessary safeguards, including assurance arrangements from institutions involved, up to and including financial bonds where the DSR Act may not have applicability (such as in the case of access by foreign institutions).

Making other publicly funded data readily available to all

Government understanding of the demand for data and its value is limited. Novel concepts and applications for data are arising continuously and anticipating how they might benefit the wider public interest is a continuous learning process.

While there may at times be good reasons for governments' inability to derive value from its data holdings — governments are not entrepreneurial nor would we necessarily want them to be — 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 non-sensitive datasets in fields where there are burgeoning opportunities and capability would be opened up and released, as resources and sectoral demand allow.

This includes data that, while it may identify individuals or businesses, 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. It also includes data that may identify and be used to evaluate the performance of publicly funded or regulated services.

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.

Not just an academic exercise

There is a need for the research community to also 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.

Leadership

New arrangements for data access will require, within Governments, strong and consistent leadership, from the Ministerial level as well as the upper echelons of the bureaucracy.

The application of the Framework and DSR Act will offer a clear sign of changed attitudes and permission to take a proactive approach to data use, but the first misstep will be a test. Custodians will not change culture without steady leadership in the face of short-term adversity. We are working to change attitudes and processes that have been entrenched for decades.

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

Establishment of the full Framework in this Report requires a firm implementation plan; and one that at the outset envisages implementation of all elements.

An underlying goal of the implementation plan should be to move forward in a way that retains and ultimately builds community confidence in what will be an all-encompassing data exchange between consumers, governments, researchers and business. While this Inquiry has created a new level of awareness of data opportunities amongst data holders and regulators, change of this order cannot be launched into the public arena without further planning and effective communication of intent and actions.

State and Territory agencies are an important part of the new Framework. It is essential that they be included in determining the details of its implementation — they would be beneficiaries of wider data accessibility and if they are allowed to realise these benefits, then would more likely also be contributors of datasets. They also hold much of the country's skills and expertise in some nationally-relevant sectoral data.

Apart from the new functions created — the Office of the National Data Custodian, a parliamentary committee (new or existing) for scrutiny of National Interest Datasets, and Accredited Release Authorities — the recommended reforms involve additional roles for several existing bodies. The ACCC would have additional functions in implementation of the Comprehensive Right for consumers; agencies with expertise in data integration and release, such as the Australian Institute of Health and Welfare, the Australian Bureau of Statistics, CSIRO Data61, and State linkage bodies, would have additional functions (either funded or charged) in implementation of the risk-based structure for data sharing and release.

The Australian Government should set an ambitious timeline for reform implementation, and move quickly where administrative change is all that is required to put substance into new policies (figure 4). We envisage that some of the recommendations can, and should, be implemented very quickly:

- allowing State-based linkage units to link Commonwealth data;
- abandoning the current policy obligation to destroy datasets and linkage keys on completion of research;
- publication of data registers for all public and publicly funded data holdings;
- administrative appointment of the National Data Custodian; and
- commencing the drafting and consultation processes for the new DSR Act.

To delay these would create a debilitating loss of policy momentum and forgo the possibility of early gains in community acceptance for reforms.

Beyond that *but still in the short term*: appointment of an advisory board and ethics advisor to the Office of the National Data Custodian; reform of human research ethics committees processes around registration and approval recognition; development of processes for the accreditation of ARAs and nomination of National Interest Datasets; and establishment of an early set of priority projects to reach agreement on industry standards for transfer of consumer data under the Comprehensive Right; should be pursued relatively quickly.

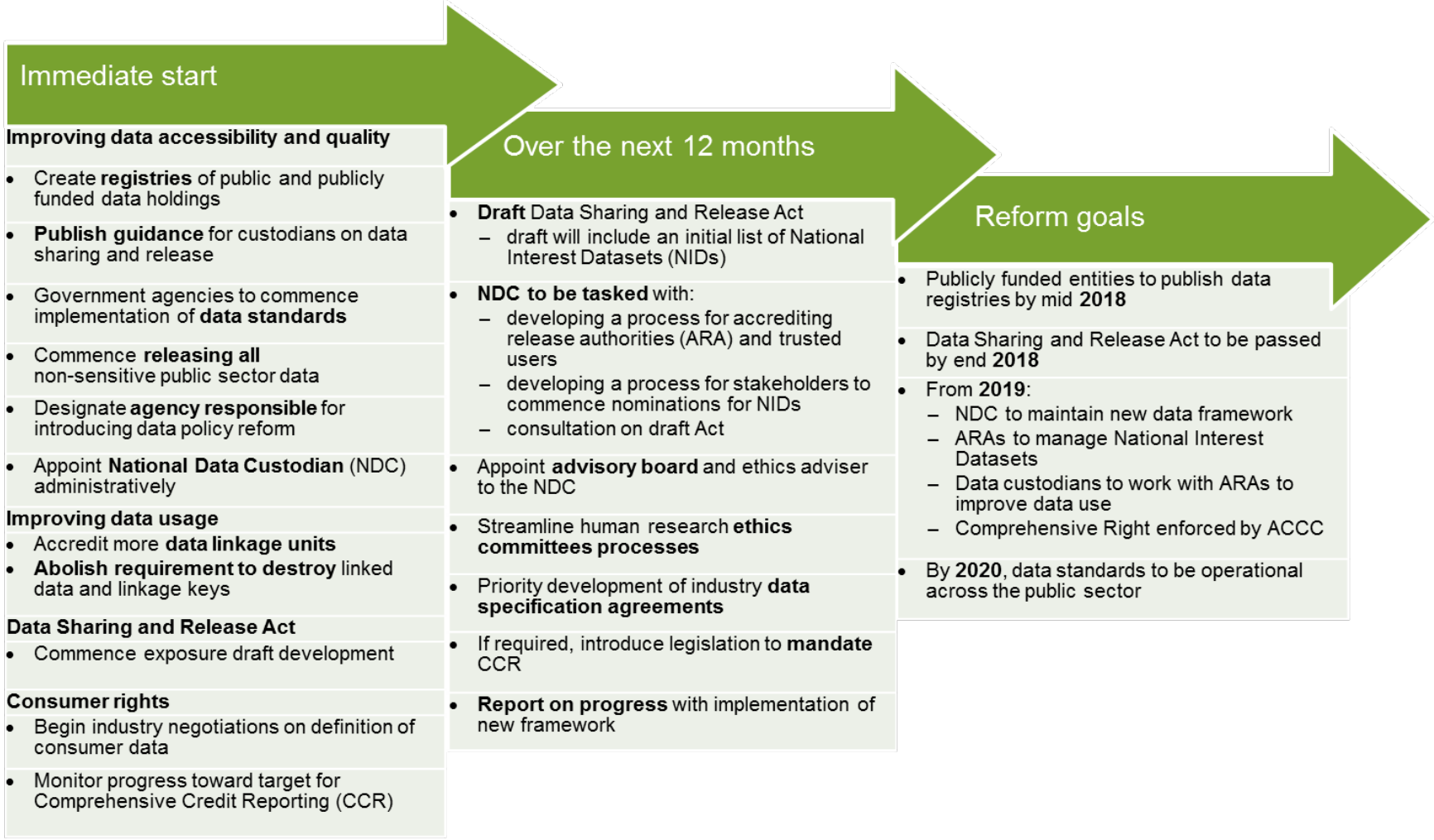
The new data Framework is intended to alter approaches to data holdings across public and private sector entities, including research bodies and not-for-profits. As noted earlier, we see no basis for distinguishing between these entities in their attitudes to ensuring data is more accessible.

The Commission cautions against any temptation to decouple parts of the reform Framework. In a project that aims to create new opportunity for both public and private benefit, each element supports the others. While gains from reforming public sector data access are likely to be substantial and benefit the public over time, governments tempted to ignore consumers' data rights, for example, would make their life difficult in trying to get the social licence needed for data reforms more generally.

It is only by allowing individuals and SMEs as consumers to share benefits of greater data use directly that governments and business would garner the community approval necessary for ongoing data collection and use.

Other developed countries have been making long term commitments necessary to engender community acceptance and reap benefits from the opportunities that mass digitisation of data now presents. It is time for Australia to reach further.

Figure 4 **Implementation timeline for the Commission's key recommendations**



Findings and recommendations

The current data environment

FINDING 1.1

Australia's provision of open access to public sector data is below comparable countries with similar governance structures, including the United States, the United Kingdom and New Zealand.

While there remains considerable scope to improve the range of datasets published (and, correspondingly, the diversity of agencies and research bodies publicly releasing data), poor formatting and the lack of frequency with which data is publicly updated are reducing data usability.

FINDING 2.1

The benefits from greater access to data would be widespread, but consumers, in particular, have much to gain, collectively, from action on Australia's data sharing and release arrangements.

FINDING 3.1

Individuals are likely to be more willing to allow data about themselves to be used by private and public organisations, provided they understand why and how the data is being used, can see tangible benefits, and have control over who the data is shared with.

FINDING 3.2

A wide range of more than 500 secrecy 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 would no longer be fit for purpose.

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

FINDING 3.3

Data integration in some jurisdictions (particularly Western Australia and New South Wales) has progressed in some fields, but highlights a lack of action in equivalent fields at both Commonwealth and State level, and reveals the large unmet potential in data integration opportunities.

FINDING 3.4

The boundaries of personal information are constantly shifting in response to technological advances and new digital products, along with community expectations.

The legal definition of personal information, contained in the *Privacy Act 1988* (Cth), has always had an element of uncertainty, and is managed by guidelines. In the face of rapid changes in sources and types of data, outcome-focused data definitions remain essential. But practical guidance (that data custodians and users can rely on) is required on what sorts of data are covered by the definitions.

FINDING 3.5

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, and often perverse incentives, greatly inhibits data discovery, analysis and use.

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.

FINDING 3.6

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

Breaches of personal data, often compounded by individuals' unwary approach to offering data, are largely dominated by malicious database hacking or criminal activity. By comparison, breaches due to sharing or release are rare.

A new Framework for sharing and release

FINDING 4.1

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.

FINDING 4.2

Community trust and acceptance will be vital for the implementation of any reforms to Australia's data infrastructure. These can be built through enhancement of consumer rights, genuine safeguards, transparency, and effective management of risk.

A Comprehensive Right for Consumers

RECOMMENDATION 5.1

Consumer data must be provided on request to consumers or directly to a designated third party in order to exercise a number of rights, summarised as the Comprehensive Right to access and use digital data. This Comprehensive Right would enable consumers to:

- share in perpetuity joint access to and use of their consumer data with the data holder
- receive a copy of their consumer data
- request edits or corrections to it for reasons of accuracy
- be informed of the trade or other disclosure of consumer data to third parties
- direct data holders to transfer data in machine-readable form, either to the individual or to a nominated third party.

Where a transfer is requested outside of an industry (such as from a medical service provider to an insurance provider) and the agreed scope of consumer data is different in the source industry and the destination industry, the scope that applies would be that of the data sender.

RECOMMENDATION 5.2

The Australian Government should introduce an outcome-based definition of consumer data that is, as an overarching objective, data that is sufficient to enable the provision of a competing or complementary service or product for a consumer.

In the relevant service or product context, consumer data is digital data, provided in machine-readable format, that is:

- held by a product or service provider, and
- identified with a consumer, and
- associated with a product or service provided to that consumer.

Participants in an industry should determine the scope of consumer data relevant to their industry (where an industry in this context would be determined by a broad description of the service). This should be in the form of a data-specification agreement.

Data-specification agreements should also articulate: transfer mechanisms, and security of data, to ensure that data use is practical and robust to technology updates; and the requirements necessary to authenticate a consumer request prior to any transfer.

These agreements should be registered with the ACCC, which may offer interim approval where an agreement has been reached but other industry agreements have been prioritised for approval.

In the absence of such agreement, consumer data must be in machine-readable form and include all of:

- personal information, as defined in the *Privacy Act 1988* (Cth), that is in digital form
- information posted online by the consumer
- data created from consumers' online transactions, Internet-connected activity, or digital devices
- data purchased or obtained from a third party that is about the identified consumer
- other data associated with transactions or activity that is relevant to the transfer of data to a nominated third party.

Data that is solely imputed by a data holder to be about a consumer may only be included with industry-negotiated agreement. Data that is collected for security purposes or is subject to intellectual property rights would be excluded from consumer data.

A consumer for the purposes of consumer data should include a natural person and an ABN holder with a turnover of less than $3m pa in the most recent financial year.

Data that is not able to be re-identified to a consumer in the normal course of business within a data holder should not be considered consumer data.

The definition should be included in a new Act for data sharing and release (Recommendation 8.1). Given the need for consumer data to have broad applicability, the outer boundary definition and reference to ACCC registered industry-specific definitions should also be included within the *Acts Interpretation Act 1901* (Cth). Consequential amendments to other legislation in the future would ensure harmonisation across federal laws.

RECOMMENDATION 5.3

All holders of consumer data should include in their privacy policies, terms and conditions, or on their websites a list of parties to whom consumer data has been traded or otherwise disclosed over the past 12 months.

On the windup of an entity that holds consumer data, consumers should be informed if data to which they hold a joint right has been traded or transferred to another entity. For businesses entering formal insolvency processes, insolvency practitioners should ensure consumers have been informed. For businesses closing but not in insolvency proceedings, the entity acquiring consumer data should inform consumers of this fact and give them the opportunity for data collection to cease.

RECOMMENDATION 5.4

The Australian Government should provide for broad oversight and complaints handling functions relating to the use of the Comprehensive Right. Accordingly, the Australian Competition and Consumer Commission (ACCC) should be resourced to undertake the following additional responsibilities:

- approving and registering industry data-specification agreements and standards
- handling complaints in relation to a data holder's failure to meet the terms of the Comprehensive Right, including in regard to the scope of consumer data
- educating consumers (in conjunction with State And Territory fair trading offices) on their rights and responsibilities under the Comprehensive Right
- assessing the validity, when requested or at their discretion, of charges levied by data holders for application of the Comprehensive Right.

The Office of the Australian Information Commissioner and industry ombudsmen should, in order to ensure a 'no wrong door' approach to handling consumer engagement, coordinate with the ACCC on the receipt and handling of consumer complaints on data access and use.

RECOMMENDATION 5.5

The Australian Government should adopt a minimum target for voluntary participation in Comprehensive Credit Reporting of 40% of all active credit accounts, provided by Australian Securities and Investments Commission (ASIC)-licensed credit providers, for which comprehensive data is supplied to the credit bureaux in public mode.

If this target is not achieved by 30 June 2017, the Government should circulate draft legislation by 31 December 2017, to impose mandatory participation in Comprehensive Credit Reporting (including the reporting of repayment history) by ASIC-licensed credit providers in 2018.

The Office of the Australian Information Commissioner and ASIC should consult with other regulators, industry groups and consumer advocates to collaboratively consider whether there is a need for a hardship flag in credit reporting.

The Department of the Treasury should be given responsibility for monitoring and publicly reporting on a regular basis on participation in Comprehensive Credit Reporting.

A risk-based approach to data sharing and release

RECOMMENDATION 6.1

As an immediate objective, all Australian governments should direct the early release of all non-sensitive publicly funded datasets — whether held by a government agency or other body receiving public funding for data collection activities.

A realistic assessment of the risks attached to public release of identifiable information that is already public (in a less accessible form) should be undertaken by all governments, with the intention of releasing low risk data, and mitigating risks where possible to enable far greater public release of data, including that which could be used for program or agency performance management purposes.

Agencies should report annually on the proportions of their datasets made publicly available, shared, and not available for release.

RECOMMENDATION 6.2

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 (Recommendation 6.6) to allow them to link Australian Government data.

RECOMMENDATION 6.3

All Australian governments entering into contracts with the private sector that involve the creation of datasets in the course of delivering public services should assess the strategic significance and public interest value of the data as part of the contracting process.

Where data is assessed to be valuable, governments should retain the right to access or purchase that data in machine-readable form and to subsequently apply any analysis and release strategy that is in the public interest.

The Australian Government Department of Finance should modify template contracts to, by default, vest access and purchase rights in governments, and avoid the need for negotiating separate rights in each contract. State and Territory governments should adopt a similar approach.

RECOMMENDATION 6.4

Publicly funded entities, including all Australian Government agencies, should create comprehensive, easy to access registers of data, including metadata and linked datasets, that they fund or hold. These registers should be published on data.gov.au. Where datasets are held or funded but are not available for access or release, the register should indicate this and the reasons why this is so.

States and Territories should create an equivalent model for their agencies where such registers do not exist. These should, in turn, be linked to data.gov.au.

A reasonable timeframe in which to achieve this is within one year (by March 2018).

RECOMMENDATION 6.5

In determining datasets for public release, a central government agency in each jurisdiction with overarching policy responsibility for data should offer a public process whereby datasets or combinations of datasets can be nominated, with a public interest case made, for release.

A list of requested datasets, and decisions regarding dataset release or otherwise, should be transparent and published online — in the Commonwealth's case, on data.gov.au.

RECOMMENDATION 6.6

The Australian Government should establish an Office of the National Data Custodian (NDC) to take overall responsibility for the implementation of data management policy, in consultation with all levels of Government.

The Office of the NDC should have responsibility for:

- broad oversight and ongoing monitoring of and public reporting on Australia's national data system and the operation of the new Data Sharing and Release Act (recommendation 8.1)
- preliminary assessments for, and recommending designation of, National Interest Datasets (recommendation 7.1)
- accrediting release authorities, be party to determining a funding agreement for Accredited Release Authority (ARA) activities, and promoting cooperation between ARAs
- managing complaints about ARA processes
- providing practical guidance material for ARAs and data custodians on matters such as risk management, data curation and metadata, data security, data de-identification and trusted user models
- advising on ethics and emerging risks and opportunities in data use.

The Office of the NDC should include a small advisory board, comprising members with technical skills related to the NDC's activities, and a dedicated ethics adviser.

The NDC role should be filled administratively by the end of 2017 to be operational by the time that new draft legislation for data access is completed for public consultation (Recommendation 10.2).

RECOMMENDATION 6.7

The National Data Custodian should streamline approval processes for access to data by:

- issuing clear guidance to all Australian Government data custodians on their rights and responsibilities, ensuring that requests for access to data they hold are dealt with in a timely and efficient manner and are consistent with the risk management approach to be adopted by Accredited Release Authorities (ARAs)
- requiring that these data custodians report annually on their handling of requests for data access, including requests from ARAs.

State and Territory governments may opt in to these approaches to enable use of data for jurisdictional comparisons and cross-jurisdictional research.

RECOMMENDATION 6.8

Selected public sector and public interest entities should be accredited as release authorities. Accreditation should be determined based on sectoral expertise, capability, governance structures, and include consultation throughout the relevant sector.

Accredited Release Authorities (ARAs) would be responsible for:

- deciding (in consultation with original data custodians) whether a dataset is available for public release or limited sharing with trusted users
- collating, curating, linking and ensuring the timely updating of National Interest Datasets and other datasets
- offering advice, services and assistance on matters such as dataset curation, de-identification and linking
- providing risk-based access to trusted users.

ARAs should be fully operational from the beginning of 2019.

RECOMMENDATION 6.9

All Accredited Release Authorities must have and publish formal risk management processes to effectively assess and manage the risks associated with sharing and release of data under their control.

Standardised, access-friendly Data Sharing Agreements should be implemented with external data providers and users to formalise the activities that can take place with identifiable and de-identified data.

Risk management processes should be regularly reviewed and revised to account for new and emerging risks.

RECOMMENDATION 6.10

Funding of Accredited Release Authorities (ARAs), for the purposes of data management, curation, storage and access should be set via a funding agreement with the National Data Custodian.

ARAs should have the power to charge fees sufficient to recoup costs where ARAs undertake requested work beyond that envisaged in their funding arrangement with the National Data Custodian.

In assessing the scope to undertake such activities, ARAs must ensure they do not detract from their primary focus on the public benefits of enabling greater access to, and use of, data (which is the basis for their accreditation and funding).

RECOMMENDATION 6.11

The Office of the National Data Custodian should be afforded the power to require an audit of a data custodian's de-identification processes and issue assurance of de-identification practices used.

RECOMMENDATION 6.12

Accredited Release Authorities (ARAs) should be given responsibility to grant, on a continuing program-wide basis, data access to trusted users from a range of potential entities that:

- 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, and
- have a signed legal undertaking that sets out safeguards for data use and recognises relevant privacy requirements.

In assessing trusted user access, the ARAs should accept existing current approvals of the trusted user's work environment.

Trusted user status for use of identifiable data would cease for that user when they leave the approved environment, when a program is completed, or if a data breach or mishandling occurs in that same environment and/or program.

RECOMMENDATION 6.13

Accredited Release Authorities (ARAs) and data custodians should be required to refer suspected and actual violations of data use conditions that have system-wide implications to the National Data Custodian.

Clarification should be issued detailing how this process would interact with the *Privacy Amendment (Notifiable Data Breaches) Act 2017* (Cth).

RECOMMENDATION 6.14

Progress by individual research institutions receiving Australian Government funding in making their unique research data and metadata widely available to others should be openly published by those institutions, with reference to past performance.

All bodies channelling public funds for research, such as the National Health and Medical Research Council and Australian Research Council, should similarly require in future funding agreements with research applicants that data and metadata is to be publicly available, and publish the results of progress on this for their funded projects.

On completion of projects, research institutions should include in their reports details of when and how other researchers can access the project's data and metadata.

RECOMMENDATION 6.15

Processes for obtaining approval from human research ethics committees (HRECs) should be streamlined.

To achieve this in the health sector:

- All HRECs should be required to register with the National Health and Medical Research Council (NHMRC). The NHMRC should receive funding to expand its current registration process, to include audits of registered HRECs.
- To maintain their registration, HRECs must implement efficient and timely approval processes, which ensure projects are not unduly delayed. The time taken to consider and review projects should be reported to the NHMRC, and included in the annual report on HREC activity.
- As a condition for registration, all HRECs and the institutions they operate in would be required to accept approvals issued by certified HRECs for multi-site projects, without additional reviews. The Australian Health Ethics Committee should develop uniform review processes to be used by certified HRECs.
- The Council of Australian Governments' Health Council should sign an intergovernmental agreement that extends the existing National Mutual Acceptance Scheme to all jurisdictions, including the Commonwealth, and all types of projects. As part of this agreement, all jurisdictions should also implement streamlined governance approvals.

RECOMMENDATION 6.16

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 in the legislative package that implements these reforms to apply to all research that is determined by the National Data Custodian to be in the public interest.

RECOMMENDATION 6.17

The Australian Government should abolish its requirement to destroy linked datasets and statistical linkage keys at the completion of researchers' data integration projects. Where an Accredited Release Authority is undertaking multiple linkage projects, it should work towards creating enduring linkage systems to increase the efficiency of linkage processes.

Data custodians should be advised as part of early implementation of this reform package to 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 made available to other dataset users.

National Interest Datasets

RECOMMENDATION 7.1

The Australian Government, in consultation with State and Territory governments, should establish a process whereby public (and in some exceptional cases, private) datasets are nominated and designated as National Interest Datasets (NIDs).

This process should be public, driven by the National Data Custodian, and involve:

- The National Data Custodian accepting nominations for NIDs, assessing their public interest merits and, after consideration by the Government, referring selected nominations to a public scrutiny process. Designation would occur via a disallowable instrument on the recommendation of the National Data Custodian.
- The establishment of a parliamentary committee, or addition of such a role to the work of an existing parliamentary committee, to conduct public scrutiny of nominations for NIDs.

The process of nomination should be open to the States and Territories in order to cover linked datasets.

This process should be in place by the end of 2018, as part of the legislative package to implement these reforms.

RECOMMENDATION 7.2

In considering nominations for National Interest Datasets (NIDs), the National Data Custodian's public interest test should establish that through sharing or release, the designation of a dataset would be likely to generate significant *additional* community-wide net benefits beyond those obtained by the original data holder.

Once designated, NIDs that contain non-sensitive data should be made available for immediate release.

NIDs that include data on individuals would be available to trusted users only in a manner that reflects the accreditation processes of the relevant Accredited Release Authority, as established and updated by the National Data Custodian, to respect privacy and confidentiality.

Where data from the private and/or not-for-profit sectors is recommended to be included in a NID, the analysis prior to designation should specifically note the ways the designation addresses genuine commercial sensitivity associated with the information and costs (including those related to ongoing dataset maintenance).

RECOMMENDATION 7.3

Trusted users should be accredited by the relevant Accredited Release Authority (ARA) for access to those National Interest Datasets (NIDs) that are not publicly released, under processes accredited and updated as needed by the National Data Custodian.

Trusted users should be personnel from a range of potential entities that:

- 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, and
- have a signed legal undertaking that sets out safeguards for data use and recognises relevant privacy requirements.

The default position should be that after applicants and their institution establish capability to respect the processes and obligations of the ARA's accredited standard, an individual researcher from one of these organisations would be readily approved for access.

For trusted users of NIDs, this status should provide an ongoing access arrangement to specified unreleased datasets that would only cease on completion of a researcher's engagement with their relevant institution, or a loss of trust in the user or their organisation (via processes also established in accreditation of the ARA by the National Data Custodian).

RECOMMENDATION 7.4

The Australian Government should make provision, in select circumstances as approved by the funding Minister, for the National Data Custodian to pay for access or linkage to private sector datasets (Recommendation 9.4).

Equally, the National Data Custodian may consider applying charges for access to National Interest Datasets where this would not be inconsistent with the public interest purpose of the National Interest Dataset.

It is expected this would not be a common occurrence, in either case.

Legislative reform

RECOMMENDATION 8.1

New Commonwealth legislation — the Data Sharing and Release Act — should be passed drawing on the full range of Commonwealth powers to regulate digital data, in order to authorise the better sharing and release of data.

The new Act should also establish the Comprehensive Right of consumers to access their data from government and private data holders alike, for the purposes of improving the services that are offered to them by alternative providers.

RECOMMENDATION 8.2

The Data Sharing and Release Act should establish the risk-based approach to data sharing and release and accompanying institutional frameworks.

- All non-sensitive data held by agencies and Accredited Release Authorities (ARAs) should be explicitly presumed to be made public, consistent with the Australian Government's Public Data Policy Statement.
- Data custodians and ARAs would be authorised to provide sensitive data to trusted users in a secure environment, with de-identification where necessary for risk management of the data.
- The National Data Custodian should have the authority to issue guidance on how the risks of *all* sharing of identifiable data between entities should be managed. This guidance should be updated where it judges the risks have shifted.

RECOMMENDATION 8.3

The Data Sharing and Release Act (DSR Act) would, where possible, override secrecy provisions or restrictions on use that prevent original custodians actively providing access to data to other public sector data custodians and Accredited Release Authorities (ARAs).

Access should be governed by Data Sharing Agreements that embed the trusted user principles, actively assist data sharing and create clarity of understanding amongst the parties. The National Data Custodian (NDC) should issue a model Data Sharing Agreement early in its life, and update it from time to time.

The DSR Act should establish modern, clear and supportive standards — the new 'rules of the game' — for data sharing and release. The Commonwealth Privacy Act would continue to apply, as well as any residual obligations emanating from the original data custodian's legislation.

Existing protections would remain on datasets that do not utilise the DSR Act, in order to ensure there is no gap between the accountability obligations on original public sector data custodians and the ARA.

In limited exceptional circumstances as the DSR Act transitions to becoming nationally effective, it may be necessary to provide access to data shared under the new Act to a party that has yet to adopt its provisions. The NDC should be provided with the power to use a disallowable instrument to allow access or sharing for such transitional purposes.

RECOMMENDATION 8.4

The Australian Government's Protective Security Policy Framework (and equivalent State and Territory policies) should be amended to recognise that the risk and therefore the classification needed for data can be reduced by:

- transforming a dataset, for example through de-identification, such that the risks of misuse on dataset release are reduced
- only making the transformed data available to trusted researchers in a secure computing environment, with usage monitored and output checked for disclosiveness.

This would align the Protective Security Policy Framework with the current legal environment.

The Australian Government should consider doing this as part of its response to the Belcher Review.

RECOMMENDATION 8.5

Legislative reform to implement the Commission's recommendations would need to be undertaken in two parts, moving forward together:

- the first part is the passage of the Data Sharing and Release Act (DSR Act) itself, that authorises to the greatest extent practical in a single statute, the sharing and release of data for the purposes of the Act and removes existing Commonwealth and State restrictions on integrating, linking and research uses of datasets by Accredited Release Authorities
- the second part is a further legislative amendment process that may be necessary, depending on the particular characteristics of, for example, National Interest Datasets, in order to address residual restrictions on the use of specific datasets that were not able to be effected by the DSR Act itself.

The National Data Custodian should be asked to identify residual legislative restrictions that need removal in its consideration of National Interest Datasets.

RECOMMENDATION 8.6

The Data Sharing and Release Act (DSR Act) should have national reach — to create a simplified and transparent one-stop location for a national framework for data volunteered, declared or acquired for inclusion under the DSR Act.

The Act should allow for the acquisition of private datasets via disallowable instruments as part of the process of creating National Interest Datasets (NIDs). Acquisition should only occur on just terms after parliamentary scrutiny determines the benefits are demonstrable.

An initial set of NIDs should be identified by the National Data Custodian to accompany the DSR Bill, following processes to establish additionality and public interest.

The DSR Act should apply Commonwealth privacy legislation to datasets managed by Accredited Release Authorities where feasible. It should be drafted with reference to (and with the intention of being consistent with) the *Data Sharing (Government Sector) Act 2015* (NSW) and the *Public Sector (Data Sharing) Act 2016* (SA) to the extent possible.

RECOMMENDATION 8.7

The Australian Competition and Consumer Commission (ACCC) and the Office of the Australian Information Commissioner should enter into working arrangements with each other, industry ombudsmen and other relevant bodies at all levels of government to support a 'no wrong door' approach to how individuals (including small businesses) pursue complaints or queries regarding their rights as consumers to data held on them.

Where an industry data-specification agreement (Recommendation 5.2) seeks to use a recognised industry ombudsman to address consumer complaints, this should be considered by the ACCC as part of its acceptance or rejection of a proposed industry agreement.

Data transformation and pricing

FINDING 9.1

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 would 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.

RECOMMENDATION 9.1

The emphasis for government agencies in handling data should be on making data available at a 'fit for release' standard in a timely manner. Beyond this, agencies should only transform data beyond the basic level if there is a clearly identified public interest purpose or legislative requirement for the agency to undertake additional transformation, or:

- the agency can perform the transformation 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.

RECOMMENDATION 9.2

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

RECOMMENDATION 9.3

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

Where data has been transformed, the transformed dataset may be priced above the marginal cost of release. Data custodians should experiment with low prices initially to gauge the price sensitivity of demand, with a view to sustaining lower prices if demand proves to be reasonably price sensitive.

RECOMMENDATION 9.4

Funding should be provided to agencies for the curation and release of those datasets determined through the central data agencies' public request process (Recommendation 6.5) to be of high value with a strong public interest case for their release. This funding should be limited and supplemental in nature, payable only in the event that agencies make the datasets available through public release.

Funding would also be required for the Office of the National Data Custodian, for functions undertaken by Accredited Release Authorities and, in some cases, for the purchase and ongoing maintenance of National Interest Datasets. Additional responsibilities required of the Australian Competition and Consumer Commission in regard to the Comprehensive Right should also be resourced.

Aside from these purposes, no additional supplementary funding appears warranted for agencies' activities related to their data holdings as a consequence of this report.

Implementation of the new Framework

RECOMMENDATION 10.1

The Australian Government should engage actively with the community on matters related to data availability and use.

At a minimum, the National Data Custodian should regularly convene forums for consultation, to ensure community concerns about increased use of data are addressed.

RECOMMENDATION 10.2

The Australian Government should set an ambitious — but realistic — timeline for implementation of the Commission's recommended reforms.

A set of actions in this Report can be completed in 2017, to ensure they deliver benefits to the community in the short term.

Passage of the Data Sharing and Release Act and supporting Part 2 amendments for an initial suite of National Interest Datasets should be in place by the end of 2018.

A central agency with data responsibility should actively support the progress made against the implementation plan until the Office of the National Data Custodian is legislatively established.

Once established, the National Data Custodian should assume responsibility for monitoring and evaluating the effects of the new data Framework, reporting annually on progress and with a formal evaluation after three years' experience of the Framework's reforms.

RECOMMENDATION 10.3

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
- deal effectively with 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 would likely be most commonly used), including where relevant and authorised, for machine-to-machine interaction.

Policy documents outlining the standards and how they would 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 to improve the quality of their data holdings.

RECOMMENDATION 10.4

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

In the event that cooperative 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.

1 Australia's data landscape

Key points

- Data refers simply to a collection of material, which can include characters, text, words, numbers, pictures, sound or video. Data may be stored digitally or in hard copy formats, with digitisation enabling data to be copied, stored and transferred rapidly.
- 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, a source of considerable potential value, is being collected ubiquitously.
- The extraordinary capabilities of data analytics and the increasing ability to link previously separate datasets are compounding the usefulness of new data sources, offering important opportunities for better-informed decision making by individuals, businesses and governments, and for research breakthroughs.
- The frameworks and protections for data collection and access, developed prior to sweeping digitisation, require serious re-examination. As one example only, privacy law is neither the only lens, nor even the best, through which to view the use of an asset such as data.
- A shift to viewing data as an opportunity, not necessarily a threat, is a global phenomenon.
- There can be many different competing interests in a particular dataset, including: the subject of the data (such as an individual, who is often also the source); 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 both legally and practically, as shown by a recent Federal Court ruling. The readiness of individuals to share information about themselves on social media, and other avenues such as loyalty programs, may indicate that social appetite for some types of data use are changing.
- A common misperception is that privacy laws — or, indeed, the privacy policies of individual organisations — give individuals ownership over data created by or about them.
  - In Australia, no one 'owns' data, although copyright law may apply in limited circumstances. Privacy legislation, the primary generic tool offering individuals some control, regulates how personal information is collected, used and disclosed.
  - In a world increasingly making use of the data *of others*, the primary unaddressed question should be: for how long will the public — the source of most of this information — trust a structure in which their actual rights are mainly limited to privacy?
- An enormous range of information is collected by governments, researchers and businesses about individuals and their activities, institutional and economic structures, and the built and natural environments. However, there is less publication — or controlled sharing — of this information than would help achieve widespread benefits for the community.
- This Report offers guidance on how governments may generate community acceptance of the processes, costs and risks associated with enhanced data use, and to do so where benefits may be most evident.

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. The Productivity Commission too has repeatedly 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 2013a).[1]

The Australian Government 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.

The Commission published the Draft Report of this Inquiry in November 2016. In their responses to the Draft, many stakeholders raised the need for additional detail and guidance on the implementation of the Commission's recommended reforms.

The factors currently restricting data availability and use in Australia are summarised in chapter 3, while the potential benefits of opening up data are discussed in chapter 2. The majority of this Final Report discusses the structure and goals of the reforms and their implementation in detail (chapters 4–10). In the interest of focusing on matters essential to the outcomes favoured by the Report, some information from the Draft Report is not repeated in full in this Report. It nevertheless remains of relevance unless corrected or withdrawn in this Report.

Categories of data used in the report

Hierarchies of data

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* refers to 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 may have little, if any, inherent meaning.

[1] Other Commission reports that have raised the need for better availability and use of data include *Gambling* (PC 2010), *Disability Care and Support* (PC 2011b), *Caring for Older Australians* (PC 2011a), *Annual Report 2012-13* (PC 2013a), *Childcare and Early Childhood Learning* (PC 2014) and *Housing Assistance and Employment in Australia* (PC 2015).

- 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.
- *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, each 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).

Sectoral groupings of data

The terms of reference for this Inquiry include examination of data collected, stored and used in both the public and private sectors, as well as research data, which is spread across both sectors.

Public sector data is defined in this Report to be data held by government agencies — at all levels of government — and other entities that are publicly funded. This includes data held by Government Business Enterprises and bodies such as universities and research institutes.

It is important to note that while security data (such as that pertaining to military organisations, much police investigatory work, or secret intelligence services) is certainly public sector data, there are typically national security or other compelling public interest considerations that tell against its release, even in the absence of privacy concerns, and as such we do not consider security data within the scope of this Inquiry.

For completeness, however, we note that the Australian Government is also considering changes to the regime governing such data, and the Protective Security Policy Framework, as part of its implementation of the Belcher Review recommendations — though no specific reforms have yet been confirmed (AGD 2017b; Belcher 2015, pp. 136–149; Department of Finance 2015).

We have also not considered cabinet-in-confidence information (at either a Commonwealth or State and Territory level) to be within the scope of this Inquiry, given that the continued confidence of this information is (generally) important to the effective functioning of Australian governments.

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 much of this data has significant commercial value and sensitivity, and that there are a range of commercial incentives to consider when contemplating any increase in access to such data.

Identifiability and sensitivity of data

Data collected directly from individual people or businesses, or that results from their activities, is typically able to identify those people or businesses — in other words, it is identifiable information.

Information that identifies, or could help to identify, an individual person, is termed personal information. Common examples of personal information include a person's name, address, contact details, signature, place of employment, work role, bank account details and vehicle registration (among many others: see box 1.1).

Some personal information is considered sensitive — that is, it has the potential to harm an individual either physically, financially or emotionally if mishandled. For individuals, sensitive information might include information or opinions about characteristics such as a person's: racial or ethnic origin; religious/philosophical/political beliefs or affiliations; membership of professional or trade associations/unions; sexual orientation; family or relationships; criminal record or other interactions with the justice system; or health, genetics and biometrics (ss. 6, 6FA *Privacy Act 1988* (Cth)).

For businesses, identifiable information can be sensitive if it reveals commercially confidential information that might, for example, cause reputational damage, void contracts or give a firm's competitors an informational advantage in the market. There is also an array of information that, while it might identify an individual business, is not sensitive. Examples include a business name and address, and material published in business annual reports.

While data that does not identify individuals or businesses can also be sensitive (examples include the locations of endangered species), typically data on the natural and built environment, and data on institutional and economic structures, is typically not sensitive.

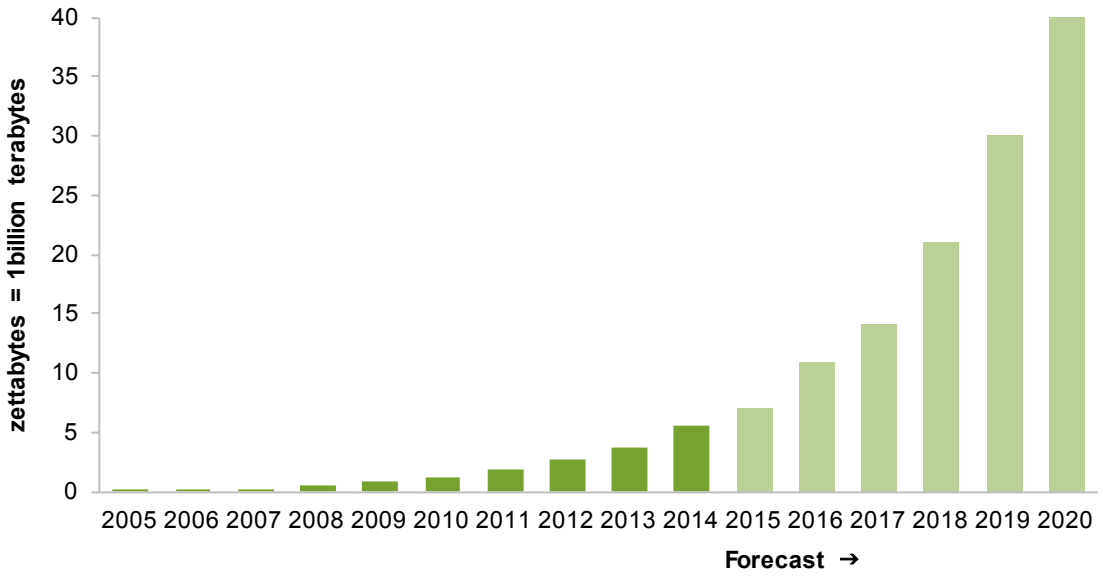
Given the focus in this Inquiry on enabling data to be more widely available, approaches to reducing the identifiability and sensitivity of data are of interest (and are discussed further in section 1.6). Data aggregation and perturbation are such approaches (appendix C), as is de-identification of sensitive identifiable information. De-identified data refers to information that previously identified individual people or businesses, but has had certain variables removed or encrypted to suppress its identifiability. Therefore, in its existing de-identified state, it cannot be used to identify individual people or firms; however, it may still carry the potential to be re-identified if it can be decrypted or matched against other datasets containing identifying information (chapter 3, chapter 4, chapter 6).

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 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; Lyman and Varian 2003). Structured data — 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: United Nations Economic Commission for Europe (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 a range of information and communication technologies. In particular, over the last two decades, massive amounts of data have been generated from three emerging sources:

- social media posts, video and audio files, and emails
- mobile devices, such as mobile phones and fitness trackers

- physical objects (apart from computers, mobile phones or tablets) embedded with sensors — the 'Internet of Things' refers to the computerisation and inter-connectivity (via the Internet) of 'things' as varied as buildings, cars, aeroplanes, traffic lights, dishwashers, toothbrushes and dog collars (appendix G).

At the same time there has been enormous growth in connectivity within and between these data sources, such that personal data on individuals may be 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.

Growth in data generation and use has gone hand-in-hand with increasing levels of concern about the privacy of personal information

Australians voluntarily provide a lot of personal information (box 1.1) to various organisations, particularly online. 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 examples of Australians' widespread provision of personal information include:

- 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 social media site, 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.

Box 1.1 'Personal information' is a moving target

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.

The exact meaning of the phrase 'about an individual' is somewhat unclear, but was recently considered by the Full Court of the Federal Court in *Privacy Commissioner v Telstra Corporation Limited* [2017] FCAFC 4. This case queried whether all information pertaining to a person's mobile phone usage (such as the network geolocation data generated when a phone periodically 'pings' off a mobile/cell tower, or when the user commences an outgoing call or text message) automatically constitutes information about the person — and therefore must be provided to the person under the Privacy Act — or, rather, whether it is about the service delivered by Telstra, and therefore not subject to any rights under the Privacy Act.

The Administrative Appeals Tribunal found, and the Federal Court did not dispute, that some network data is not 'about an individual' but rather is about the telecommunication company's attempt to provide a service to the individual. More broadly, this finding reflects that not all data generated by a person's use of a product or service will constitute personal information, even if the person may be reasonably identifiable by cross-matching the data with other information. As such, individuals will not automatically be entitled to access all of the information generated through their use of a product or service. The individual must be 'the subject matter' of the data.

However, the Full Court of the Federal Court noted that information and opinions can, in theory, have multiple subject matters. As such, any legal action turning on the definition of personal information in the near future may see some quite specific developments in case law around the Privacy Act.

Source: Hamilton and O'Dowd (2017); Johnston (2017)

However, despite an apparent willingness to share about themselves in a wide range of contexts, individuals have voiced concerns about the privacy and security of their personal information — indeed, the Australian Information and Privacy Commissioner recently noted that both of these factors are on the rise:

> People are becoming more aware of just how much information they are giving [online]. We see the conundrum: people are aware they are providing a lot of information, but they are still concerned about what's happening to it … [and] want to find out what their rights are. (Pilgrim, as quoted in White 2017).

Previously, the Office of the Australian Information Commissioner (OAIC) had found that 74% of survey respondents were more concerned about the privacy of their personal information online than they had been five years previously (OAIC 2013).

Community responses to recent events have also demonstrated strongly that some Australians are apprehensive about what data is held on them, by whom, what is done with that data, and which interests might be served by data use or disclosure (see, for instance, Chirgwin (2017)). The best-known examples include the Australian Bureau of Statistics' retention of names and addresses from the 2016 Census of Population and Housing, the passage of the *Telecommunications (Interception and Access) Amendment (Data*

Retention) Act 2015 (Cth), and various data breaches. There are also indications that community members are concerned about what could conceivably be done with their data in the future — for instance, the Attorney-General's Department recently undertook consultation on allowing parties to civil lawsuits to access telecommunications metadata, and received several hundred submissions expressing concerns about the risks to privacy such a change could engender (AGD 2017a; Cooper 2017). Recent legislative activity, including the passage of the *Privacy Amendment (Notifiable Data Breaches) Act 2017* (Cth) and the introduction of the *Privacy Amendment (Re-identification Offence) Bill 2016* (Cth), has attempted to address some of the risks that fuel these concerns.

Data digitisation and the growing power of data analytics

Digitisation — the conversion of hard copy data into electronic formats, 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 (Shaw 2017). 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 fundamentally related. Data enables the production of digital products and services but it is also an output from the use of these products and services. New digital services, including numerous online consumer services — such as Amazon, Netflix, Airbnb and Uber — have been highly disruptive to the industries they have entered. The financial sector, typically conservative in the face of technological disruption, is also experiencing a wave of innovation driven by digital advances and the use of existing and new data sources. Innovative 'fintech' companies are capitalising on these developments, along with incumbent firms in the sector (appendix F).

Big data underpins greater knowledge and enhanced decision making

Big data is essentially data that has 'high volume, high velocity and high variety' characteristics. The rapid growth in the generation, availability and connectivity of big data and the ever-expanding power of data analytics have the potential to deliver greater insights and knowledge about the world, and to help analysts and policymakers to engage in more effective, better-informed decision making, leading to efficiency gains and productivity growth (OECD 2015a) (chapter 2).

This is not to suggest that data analytics programs, or otherwise automated big data techniques and processes, must (or should) *replace* human decision-making and service

delivery (see, for example, Wong 2017). Rather, big data, when analysed with care and oversight, can provide a more detailed and accurate level of analysis than can manual inquiry alone. Such analysis can then be used to support research, policy and service delivery decisions — in other words, improved services and outcomes for consumers.

Data remains underutilised

Despite rapid technical developments and large potential benefits, much data being generated remains underutilised (even allowing for the fact that a significant portion may prove to have no value). 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). Usefulness is dynamic — some data that was previously of limited value may become valuable as analytical capabilities improve, or with investments made to improve its quality.

There is broad scope to increase the availability and productive use of data. Greater access would empower individuals to make more-informed decisions about the products and services they choose, 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 improve its efficiency and accountability.

This Report offers guidance on where the benefits of greater data use may be most evident, and ways that governments might engage with the public to generate community acceptance of the costs, risks, and benefits associated with data sharing and use.

The process of reforming Australia's data landscape, and creating a new framework to fit the thousands of different types of data generated, is not without its risks or costs. But, given the enormous range of potential benefits such reform can bring (chapter 2), such reform is a worthwhile task. Risk identification and management is, of course, a crucial topic, but the policy debate is not assisted by risk *aversion* (itself a central topic of this Report) and so it is important to draw a distinction between the two.

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 most goods and services. However, it has several remarkable features that distinguish it from other forms of capital:

- 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.

- Relatedly, 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.
- Data in a digital format is virtually costless to reproduce.
- Additionally, 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, some datasets will have significant scarcity value. And the creation of new datasets will sometimes alter the value of others.

These properties, combined with lower costs of using data and of combining various datasets, have increased the benefits of reusing data. They have also increased the importance of, and need for, clarity around the rules regulating data access and the capacity to repurpose data.

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 the decision-making of individuals, businesses and governments. The information derived from data analysis can alleviate information asymmetry and reduce inefficiencies in market operation (box 1.2). It can stimulate competitive responses from suppliers in a way no other asset can. But the extent to which data can be used to improve all these market and non-market operations, including individual decision making, will be constrained by restrictions on data access and use.

Box 1.2 How data can alleviate information asymmetries

- Comprehensive credit reporting seeks to address the information asymmetry between lenders and borrowers — that is, the situation of borrowers typically having more information on their creditworthiness than lenders (appendix F), leading to instances where relatively creditworthy applicants are denied credit or 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.
- There is often an information asymmetry between providers of insurance and those purchasing insurance. 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.

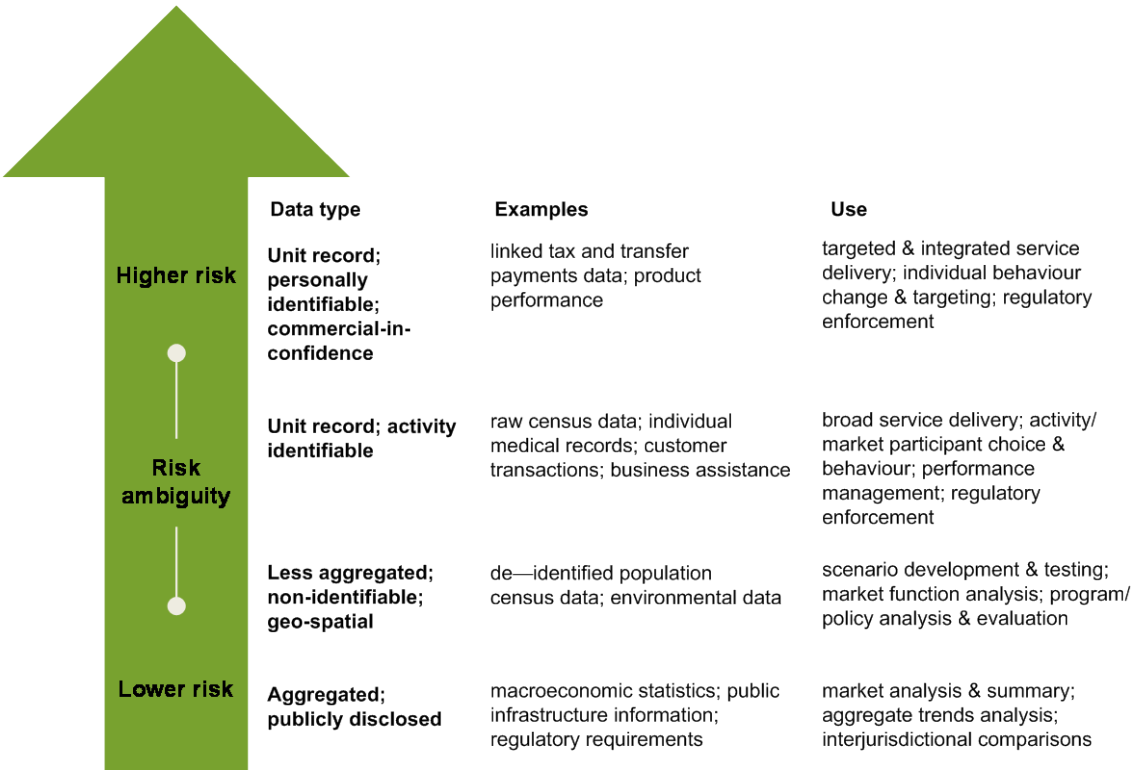
Source: Lane et al (2014)

Different data, different uses and different risks

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

Near real-time data that identifies individuals 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 the improved delivery of many products or services, de-identified, non-real-time data can be perfectly adequate.

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



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 users' medical data and use it to 'intelligently' remind them when to see their doctor, based upon their risks (Health& 2016, p. 1)).

There are undoubtedly areas of ambiguity where it is debatable who should have access to particular 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.

1.3 Stakeholders in data management and access

Assessing the rights and interests over data is complex, as there are potentially many competing interests in, and claims to, a particular dataset (box 1.3). For example, transactions data — detailing the supplier and purchaser of the good or service, the date, what was exchanged, and the price — may be viewed by some 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 1.3 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 or conducts 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 — 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, adding value by tailoring it for specific markets or their own use (such as data brokers or data aggregators)
- 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 custodian — a party that stores data; normally responsible for maintaining its security and deciding which other parties may access the data. The data custodian may very well be the data collector or data user, but may equally be a separate party entirely.
- 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 have different powers and interests in data, depending on their role. For

example, in the context of smart meters for electricity consumption, a Privacy Impact Assessment carried out for 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 … 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. (Lockstep Consulting nd, p. 25)

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 it may be possible to claim ownership over a processed dataset. Privacy law regulates how personal information is collected, used and disclosed.

How copyright law applies to data

Copyright law can give particular rights over data's form or expression if it is sufficiently 'original', and expressed in 'material form'.

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 their skill in, for instance, interpreting test results (*Breen v Williams* (1996) 186 CLR 71). Databases are subject to copyright if they:

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

For instance, in *Telstra Corporation Limited v Phone Directories Company* [2010] FCAFC 149, the Full Federal Court held that phone directories compiled automatically by a computer did not have sufficient human effort of a literary nature to be covered by copyright.

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 licence. In recent years, some public sector agencies and some large search entities (such as Google) have supported a move away from releasing data under restrictive licences, towards using more permissive Creative Commons licences that allow the data to be reused (appendix D). The terms of the use are set out in the conditions of the licence. Creative Commons can, however, potentially impede some re-use 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 (information that could identify, or reasonably identify, an individual; box 1.1). It applies to private sector and not-for-profit entities with an annual turnover of at least $3 million, all health service providers (and certain other specialised businesses) and *most* Commonwealth Government agencies. Similar legislation exists in most States and Territories (applicable to that jurisdiction's government data and entities not regulated by Commonwealth legislation), the exceptions being Western Australia and South Australia (appendix D).

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. These obligations are set out in the Australian Privacy Principles (APPs) and ss. 16A and 16B of the Privacy Act, and include to:

- ensure personal information is collected fairly and lawfully
- inform the individual of their information being collected and how it will be handled
- 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.

These obligations are open to interpretation by organisations that collect and manage data.

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

Protecting the identity of individuals and organisations

Given the sophistication of some data analytics, even data that does not directly relate to individual people or organisations — for example, traffic patterns — can often be used to identify individuals. The ill-considered sharing and release of such data may compromise the privacy, security or wellbeing of individuals. Businesses too may suffer detriment as a result of commercially sensitive information being released — although claiming information to be commercial in confidence can be a default position used for a variety of purposes (such as hindering competition), and businesses are often better placed, legally and financially, to react to the harmful release of sensitive information than are individuals.

Personal data and privacy

Claims to privacy differ with context: that is, social norms around the treatment of personal information vary depending on who the subjects and the recipients of data are. For

example, in a health care context, 'informational norms' — some of which are embedded in legislation — 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). Privacy may be achieved by having control over information about oneself, and the power to decide who one shares personal information with and the terms under which one agrees to share it.

Defining when data is private can be complicated by the circumstances presented by modern communication. For instance, an individual may benefit from disclosing certain personal information to a company in return for services (for example, access to social media). On the other hand, the same individual may wish to keep this same personal information private in other circumstances (for example, from a telemarketer). In effect, a price is being set for access to the information.

There may be an expectation that such risks would be managed by a public policy response, and case-specific policy has attempted to do that. But no general policy or regulatory construct recognises this trade. And to the extent that it occurs (as it is, increasingly) at least some of the responsibility to keep data private and secure must then lie with the individual. Many social media providers, for example, have also blurred the line between personal and non-personal information (appendix G provides examples).

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 take place. A considerable amount of data is collected about individuals without their explicit consent and, sometimes, without their knowledge.

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 an online service (appendix G). In contrast, *implied consent* to share some personal details is given in situations where the details are necessary to perform a service or transaction desired by the individual, such as a credit card transaction.

The APPs Guidelines state that *informed consent* — which can be either express or implied — involves several key elements, including requirements that the individual is adequately informed before giving their consent, the individual gives consent voluntarily, and the consent is current and specific (OAIC 2015).

Meaningful consent involves the consent-giver having a substantive understanding of exactly what they have consented to. In practice, this means that the organisation seeking consent must achieve a balance between providing the necessary information and avoiding 'notice fatigue' on the part of the consent-giver.

Recent years have seen criticisms of privacy legislation's reliance on consent. Evidence shows that many individuals are unable to understand how their information is being used and/or do not read use disclosures (Davis 2016; KPMG International 2016; OAIC 2013; see also chapter 3 for more detail). The Office of the Australian Information Commissioner (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 3).

Confidentiality

Confidentiality refers to the common law duty that collectors of identifiable information about people or businesses have to keep the data 'in confidence' — to restrict it from disclosure to unauthorised individuals, organisations or processes.

One method of providing confidentiality is to de-identify datasets by removing names, addresses and other variables that could be linked to individuals or organisations. The risk of re-identification can then be managed through 'confidentialising' the data by taking additional steps to help protect sensitive information from parties viewing a dataset (appendix C). Common confidentialisation techniques include 'perturbation' (adding statistical noise to the dataset) and 'small cell suppression' (removing or perturbing variables that only apply to a very small proportion of people in the dataset). Yet as data use opportunity grows, so also grows the analytical power behind re-identification.

Data sharing versus data release

Data sharing

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

Data is sometimes shared with 'trusted users' — approved individuals or organisations. Access tends to be restricted to trusted users when the data being shared is sensitive or personal information. 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). Access levels may be differential, based on the type of data and the characteristics of trusted users.

In some countries, governments have initiated programs to encourage private sector data sharing (such as the UK's 'midata' program, which encourages businesses in selected sectors to allow individual customers to download data about their personal use of specific products and services).

Private sector-initiated environments for data sharing also exist to share data about customers, products or markets either between related firms (such as with businesses that share a parent company or between a business and their key suppliers) or between otherwise unrelated firms. Local examples include Quantium and Data Republic, both of which offer secure digital platforms for firms to exchange anonymised customer data with each other and conduct analysis on that data in order to enhance customer offerings (Data Republic, sub. 176; The Quantium Group Pty Ltd, sub. 187; Eyers 2016); and My Mob Tracker, which allows farmers to share real-time sheep data with farm staff and other livestock producers (Cleeland 2016).

Data release (open data)

By definition, open data is available to all potential users without restrictions, notwithstanding the possibility of charging for access. The corollary is that 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 — that is, if its publication is legally mandated, 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 licensing arrangements (appendix D).

1.4 What is the current landscape for data sharing and release in the *public* sector?

Public sector agencies at all levels collect, store and use a wide array of data, ranging from individual health records to Australia-wide maps. Data is collected mostly for administrative purposes (for example, as proof of agencies and individuals meeting the terms of programs), and occasionally for research.

Much of this data remains 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, if it can be done in a manner that maintains the trust and confidence of the general public (chapter 4, chapter 10).

What information is collected?

Data from individuals

Australians provide a wide variety of information to a multitude of agencies and service providers at all three levels of government, including government line departments such as the Department of Human Services, publicly funded institutions (including research bodies) 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 is provided many times over, as some legislation restricts agencies' ability to share information with each other (chapter 3). For example, 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. However, some attempts have been made to create 'tell us once' initiatives to combat this. In 2011 the Australian Government introduced a pilot to allow changes in contact details to flow across agencies (Kennedy 2011) and more recently, the Digital Transformation Office began implementing a similar 'tell us once' feature for MyGov (Carrasco 2015).

At the State and Territory level, New South Wales, South Australia and the Australian Capital Territory have all begun to deploy similar functionalities, with NSW's MyServiceNSW and ACT's Access Canberra both currently enabling users to transact with at least four different government agencies (ACT Government 2016; Bajkowski 2016; DPC (SA) 2016).

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, and sometimes (such as for importers or exporters) detailed information on products and suppliers.
- 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, such as financial institutions reporting particular transactions to AUSTRAC (the Australian Transactions Reports and Analysis Centre) or providing financial performance information to the Australian Securities and Investment Commission.

Much of the information that businesses provide to governments is provided many times over. The Commission (2009, 2013b) has previously noted widespread burdensome, duplicative and redundant reporting requirements. As with efforts related to individuals (discussed above), 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 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 agencies. Some State and Territory Government online service websites, such as service.nsw.gov.au, business.vic.gov.au and accesscanberra.act.gov.au also provide entry points for aspects of Commonwealth business compliance such as registering for tax file numbers, Australian Business Numbers, or business names.

Data from other sources

Numerous 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. Much of Australia's weather data is collected through the Bureau of Meteorology (BoM), while datasets on water resources, Australia's flora and fauna, fisheries, mineral and energy resources, and forestry and agriculture industries are also collected for varying purposes and at varying levels of detail by Commonwealth agricultural and environmental agencies, State and Territory agencies and, in some cases, local governments.

Information about the state and use of infrastructure, such as road usage and aviation movement of cargo and freight, is also monitored and reported on. In all of these areas, Internet of Things-type connected devices can be used to supplement existing datasets and/or provide new sources of insight. For example, data is currently collected in real time from traffic cameras as well as sensor buoys located several kilometres from the coastline (Gardiner 2015).

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). International agreements and treaties can also require Australia to provide particular information — for instance, the World Trade Organisation requires Australia to report information on 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.

How do governments use data?

Governments use data for a variety of purposes, including (but not limited to):

- *Administering payments and other government services* — Australian governments administer billions of dollars in payments each year. Given Australia's heavy reliance on means-testing for transfer payments, administering these payments accurately requires (and generates) significant volumes of data.
- *Informing service provision* — governments rely on data to enable targeting and effective service delivery. Some examples include:
  - delivering welfare services using data on names, addresses and relationships
  - tracking student outcomes to inform teaching methods (VCAA 2016)
  - monitoring patient diagnoses and treatment history in hospitals to inform funding decisions and further investment.
- *Performance monitoring* — governments use data to assess the effectiveness of the services it provides or funds, such as public transport, courts or 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 broad categories of government services: childcare care, education and training; justice; emergency management; health; community services; and housing and homelessness (PC 2017).
- *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 rely on data when preparing for, and responding to, natural disasters and other emergencies. Examples include:
  - using river flow, terrain and weather data for flood impact assessment (Geoscience Australia 2014)
  - using spatial data on vegetation and weather patterns to predict fire spread (Bushfire CRC 2014)
  - data on people who are immobile or otherwise at risk (such as the locations of aged care facilities) can be used to prioritise evacuations.
- *Enforcing the law and regulatory schemes* — governments use data 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).

- *Research* — across a broad range of fields, researchers within government and publicly funded institutions use, for their investigations, both data collected specifically for research and repurposed administrative data.

Open data policies currently in place

Open access to public sector data is receiving increasing attention from governments in Australia and internationally. A key driver of the open data movement has been the Open Government Partnership (OGP), which has been endorsed by more than 70 countries at the time of writing (OGP 2017). Australia and a number of countries within the Open Government Partnership have set a goal of making public sector data 'open by default' (Turnbull 2015). In Australia, this has been the culmination of a series of initiatives over the last seven years (box 1.4).

The Australian Government Public Data Policy Statement (Turnbull 2015) applies across the Commonwealth Government, as does the Guidance on Data Sharing for Australian Government Entities (DPMC 2016b). These aim to promote open data and improve the use of data across the Australian Government. All State and Territory jurisdictions, except for the Northern Territory, also have an official open data policy and a particular department with ownership and lead implementation responsibilities for that policy.[2]

How well are our open data policies working in practice?

Limited variety: most of Australia's open data is scientific information

Participation in data.gov.au, the Australian Government's open data portal, is dominated by a limited number of major contributors. While the counts of over 111 000 resources from 345 contributing agencies (as at 7 March 2017) are substantial, about 80% of datasets on the site are contributed by just nine agencies (figure 1.3). Most of Australia's open data is scientific — particularly environmental or spatial — information.

How does Australia compare to other countries?

Compared to many similar countries, Australia lags in opening up public sector data. The types of data available in the United Kingdom and the United States (table 1.1) highlight many potentially high value datasets that other countries release while Australia does not.

[2] See Appendix C for more detail on the institutions responsible for public sector open data, and their policies and instruments, at both the Commonwealth and State/Territory levels of government.

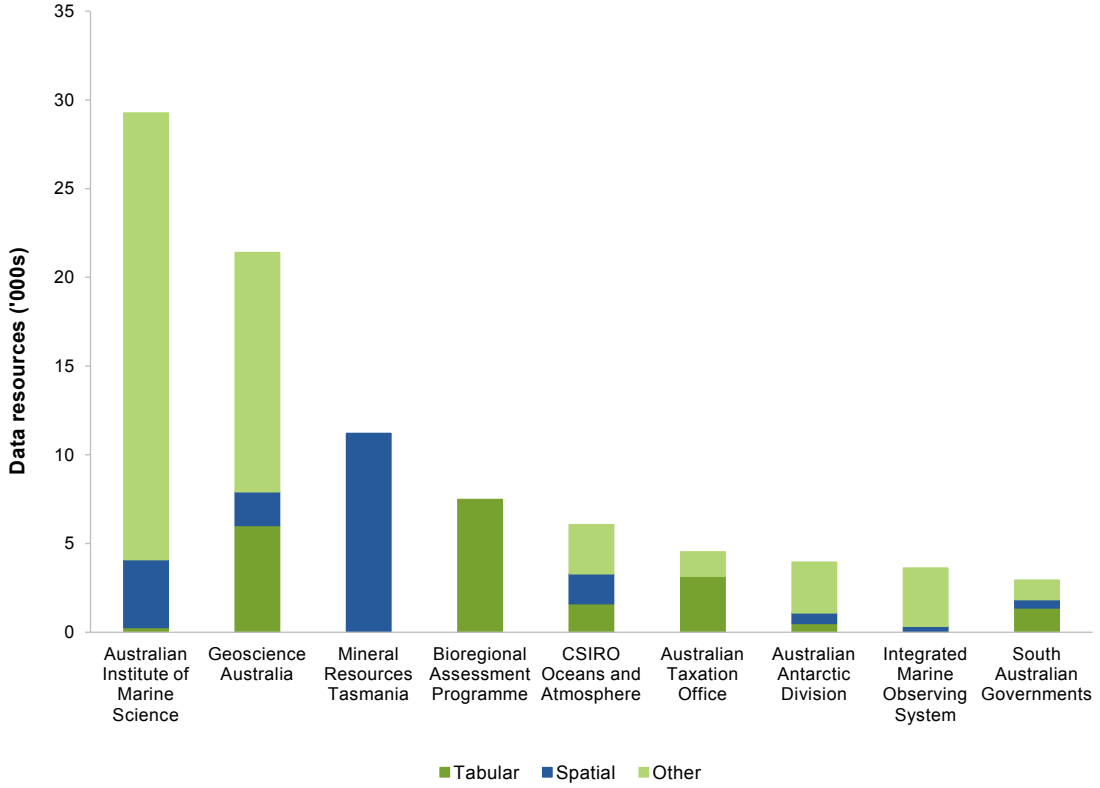
Box 1.4 Key Australian Government open data initiatives

Since 2009, the Australian Government has released a series of policies intended to promote open data:

- In December 2009, the Australian Government released *Engage: Getting on with Government 2.0*, which recommended a 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 afterwards.
- In July 2013, Australia created data.gov.au, a web portal for the publication of open government data (Waugh 2013).
- In November 2015, the Australian Government announced plans to finalise Australia's application for active membership of the Open Government Partnership, after progress had stalled following the announcement in July 2010 (DPMC 2015; Palmer 2015).
- 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 (APIs) with descriptive metadata; using open data standards; and publishing under a Creative Commons Attribution licence (Turnbull 2015).
- In March 2016, the Geocoded National Address File (G–NAF) and the associated Administrative Boundaries dataset were released onto data.gov.au, becoming free for anyone to access.
- Throughout 2016, the Department of the Prime Minister and Cabinet carried out several rounds of consultations on the content of Australia's first Open Government National Action Plan. The draft Action Plan was published for comment in October 2016, with the final version released in December 2016. The finalisation of Australia's first National Action Plan was a major step towards Australia obtaining active membership for the 2016–2018 Open Government Partnership cycle (DPMC 2016a; Easton 2016).
- In October 2016, the Department of Prime Minister and Cabinet commenced a series of consultations to inform the development of a High-Value Dataset Framework, designed to help Governments identify and prioritise the release of high-value datasets (DPMC 2016c).
- In December 2016, the Treasury launched the consultation stage of a review into tax and corporate whistleblower protections (The Treasury 2016), following the commencement of an inquiry into whistleblower protections (across the corporate, public and not-for-profit sectors) by the Joint Parliamentary Committee on Corporations and Financial Services. The Committee is due to report by the end of June 2017 (Parliament of Australia 2016).
- In February 2017, the Treasury released a consultation paper seeking views on the possible implementation of a beneficial company ownership register (The Treasury 2017).

Figure 1.3 **Most open data is scientific information**

Resources published on data.gov.au by agency — biggest contributors



Source: Australian Government (2017)

Australia registers particularly low scores on measures of spending, legislation and health data (table 1.1). While many of these datasets exist and are available in Australia, low scores result from poor update frequency and poor formatting. (Out of the total 33 755 datasets available on data.gov.au as of 7 March 2017, for example, only 6626 were uploaded in a machine-readable format or as APIs (data.gov.au 2017)). This means that while Australia ranks relatively highly on implementation of open data policies, it receives a particularly low score on the impact of its datasets — although data is published, the poor update frequency and formatting mean that it is underutilised (figure 1.4). According to the Australian Data Archive (2016, p. 3):

> 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 [Humanities, Arts and Social Sciences] discipline.

Along with the shortfalls in Australia's current implementation of open data policies, there appears to be little systematic sharing of data between public sector agencies, or between agencies and researchers. Chapter 3 discusses this lack of intra-governmental data sharing in more detail and examines factors that can stymie data sharing and release efforts.

Table 1.1 **Open data availability in selected countries**[a]
2015

Datasets	*Australia*	*United Kingdom*	*Canada*	*United States*	*New Zealand*
Spending[b]	**5**	95	5	80	5
Legislation[b]	**15**	100	80	85	80
Health[b]	**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 out of 100, with points awarded according to whether the data exists, how it is made available and whether it is up to date (WWWF 2016b). Unlike the scores presented in figure 1.4, these are not scaled with respect to the best performing country. Bolded numbers highlight datasets for which Australia has a comparatively low score. **b** Spending data refers to government spending recorded at a transactional level on specific items. Legislation data refers to information on laws at the federal level — while Australian legislation, bills and regulations are available online, they are not published in machine-readable formats, as is done in many other jurisdictions. Health data refers to statistics generated from administrative data that indicates performance of specific service, or the health system as a whole.

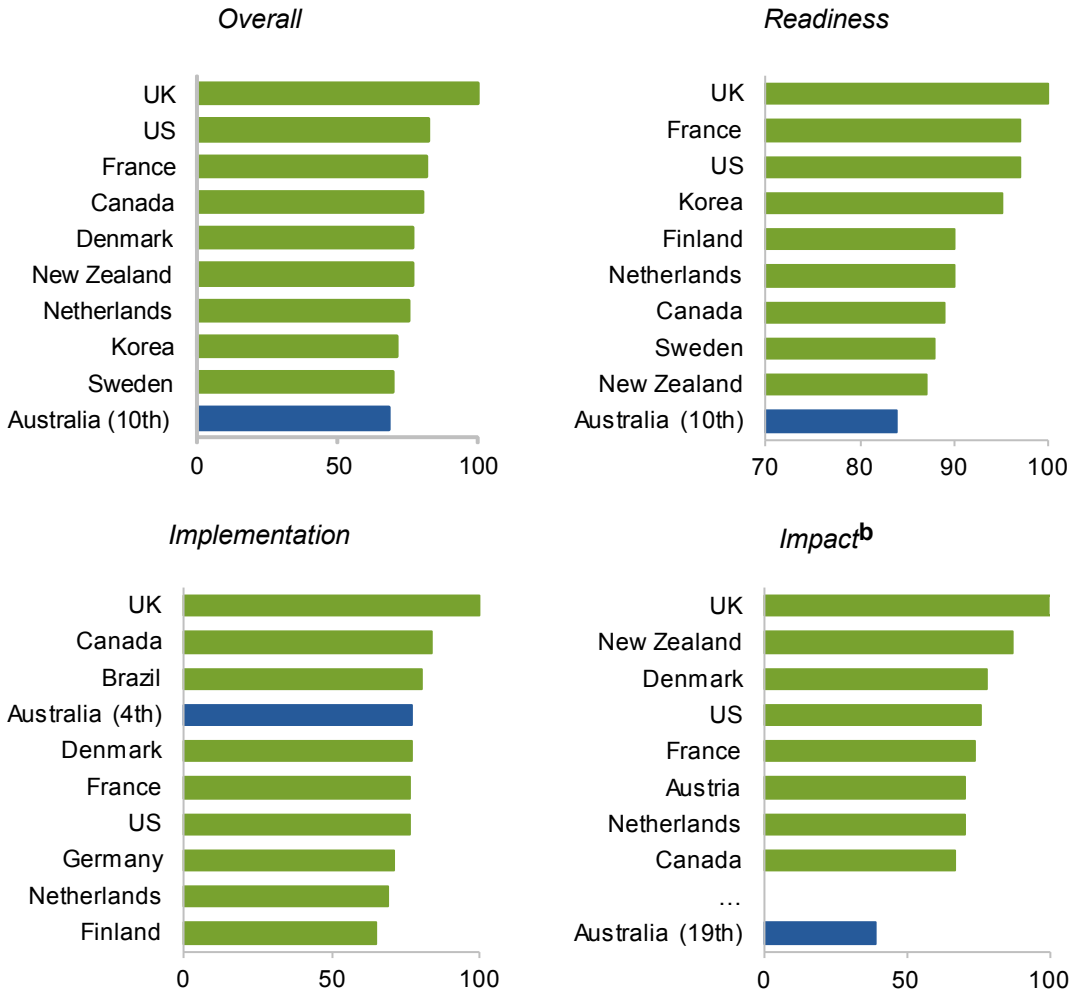
Source: World Wide Web Foundation (2016a)

FINDING 1.1

Australia's provision of open access to public sector data is below comparable countries with similar governance structures, including the United States, the United Kingdom and New Zealand.

While there remains considerable scope to improve the range of datasets published (and, correspondingly, the diversity of agencies and research bodies publicly releasing data), poor formatting and the lack of frequency with which data is publicly updated are reducing data usability.

Figure 1.4 **Measures of Australia's open data performance[a]**
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 impacts. This measure (though it should be interpreted with caution) suggests a lower level of open data use in Australia relative to other countries, including those in which less data is available.

Source: World Wide Web Foundation (2016a)

Publicly funded research data

The public interest in re-use 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)

Often, however, research data is neither kept permanently, nor archived, nor available for reuse by other academics (chapter 3).

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 to other equivalent researchers.

There have been several supporting initiatives aimed at providing access to, and increasing discoverability of, research datasets in Australia (appendix C). A key development in enabling researchers to discover research data has been the establishment of the Australian National Data Service (ANDS) in 2008, funded by the National Collaborative Research Infrastructure Strategy (NCRIS).

The central service of ANDS is its Research Data Australia discovery portal, which at the end of 2016 included entries for 130 000 data collections (ANDS, sub. DR245, p. 6). ANDS does not have a curation role, and data is not directly accessible through the portal — rather, it is a national federated catalogue, with dataset descriptions 'harvested' from over 100 Australian research organisations. 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. In Europe, for example, 15 countries are members of the Consortium of European Social Science Data Archives (CESSDA), which is a network of national social science data archives with a common core of technical infrastructure, policy and data and metadata standards (CESSDA AS 2016). CESSDA and similar organisations are designed to facilitate the re-use of research data adopted by other countries, 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. In contrast, 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 currently under review, with NCRIS having recently completed a consultation on Australia's 2016 National Research Infrastructure Capability Roadmap. 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.

State and Territory data sharing and use initiatives

In recent years, a number of State and Territory governments have implemented legislative reform of data use, access and sharing processes. Others are currently investigating and consulting upon possible reforms, or have committed to undertaking them in the near future. Examples include:

- New South Wales — the *Data Sharing (Government Sector) Act 2015* (NSW) established a Data Analytics Centre (DAC) and contains a number of mechanisms to facilitate data sharing between the DAC and other government entities.
  - For instance, public sector agencies are authorised to share data with the DAC or other agencies in order to identify issues and solutions in policy making, program management and service delivery. Furthermore, the relevant Minister may direct an agency to provide specified government sector data that it controls to the DAC if the Premier advises the Minister it is required for advancing Government policy.
- South Australia — the *Public Sector (Data Sharing) Act 2016* (SA) authorises agencies to provide public sector data that they control (with exceptions) to other public sector agencies for a wide range of purposes relating to policy making, program management and service delivery.
  - The Act specifies that trusted access principles must be applied in respect of the sharing and use of public sector data; this makes South Australia the only jurisdiction to have expressly codified a risk-based approach to data sharing.
- Victoria — in May 2016, the Victorian Government released the *Information Technology Strategy 2016–2020*, which included a commitment to establishing a State data agency to facilitate information sharing between agencies and open up government data to businesses, universities, and the community (Andrews 2016). At the time of writing, the Victorian Centre for Data Insights was set to operate within the Department of Premier and Cabinet, and was close to launching (Cowan 2017b).
- Queensland — in December 2016, the Queensland Government commenced consultation on a review of the *Right to Information Act 2009* (Qld) and the *Information Privacy Act 2009* (Qld).
  - The review's Consultation Paper expressly contemplates intra-governmental data sharing, raising the possibility of treating the sharing of personal information between government agencies (or departments) as a 'use' of information rather than a 'disclosure' (DJAG (Qld) 2016).
  - In February 2017 it was reported that a Data Sharing and Analysis Office would be created within the Department of Science, Information Technology and Innovation (Cowan 2017a).

1.5 What *private* sector data is collected, and to what extent is it made available?

Private entities have always collected data on their customers, including personal details and data relating to transactions. Recent years, however, have seen the privatisation of many government services, a major expansion in Internet–based businesses, and rapid growth in intelligent products — such that much of the overall growth in data collection and analysis now occurs within the private sector.

For example, ANZLIC – the Spatial Information Council noted that:

> … 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 … (sub. 164, p. 4)

A distinction may be drawn between:

- commercial entities subject to a high degree of regulation or in receipt of public funds, much of which expressly requires (or permits) the collection of data; and
- entities that do not acquire their 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 or high entry costs. Entities in these sectors typically provide complex products and services that are of high importance to the community at large — for example, utilities such as power, water, and telecommunication services — which underlies the public interest rationale for a higher degree of regulation and/or funding.

Public interest purposes for data collected under such regulation and/or funding include use for forecasting infrastructure investment, maintaining community protection, ensuring the stability of the financial system, and managing risk over generations. A potential unintended consequence of these regulatory-necessitated data collections, though, is that the rich data holdings built up by existing firms, and the benefits associated with holding those datasets, may create barriers to competition and thereby allow particular firms to dominate large proportions of the markets for their respective goods or services. The datasets may also allow firms to move into other markets where firms have not had similar requirements or opportunities to collect data.

The second group of entities, in contrast, tend to face greater consumer choice (even in sectors where there are dominant firms, such as social media, consumers can choose whether or not to use the class of product or service at all, without adversely affecting their quality of life). As such, they depend to some extent on the continuing quality of their relationships with customers in order to collect and exploit data. They consequently may

have stronger incentives to respond to customer data needs and interests, and to assist customers in using their data.

Commercial and not-for-profit entities in regulated sectors

Banks, health insurance funds, and energy providers are some of the major sectors in which regulation permits or mandates the collection of customers' personal information.

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 (resulting in food labelling requirements); real estate markets (standard contracts and clear title); or employment in particular fields (Working With Children checks)
- where regulatory compliance requires creating or lodging records
- where governments are major or primary purchasers, such as health services
- where governments act on behalf of the community as a steward — for example, requirements for resource companies to lodge mineral and energy exploration data.

Many regulated businesses are already required to make some of their data available to regulators and allow customers to access data collected about them. But digitisation of processes and transactions means that considerably more information is now collected from customers than when current regulatory frameworks and reporting obligations were established. And customers are becoming aware of how important their data is, if they are to exercise choice or otherwise act to enhance their own welfare.

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 would 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 recommends a Framework under which consumers would have access to their information, which would, if implemented, address

this question across all industries regardless of regulatory status — which in principle is the preferred course of action given the potential for significant benefits discussed below in this Report.

Financial institutions

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 email address) and data on the customer's financial position (such as their income, expenditure, savings, and credit history). While much of this data is collected directly from the individual, some entities (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 to share information for marketing purposes; marketing businesses; and brokers and other parties who may introduce customers to the institution (ANZ 2016). Similarly, personal information may be collected from public registers, customers' referees, employers, social media, or other information made publicly available by third parties (PCU nd; CBA 2014). Westpac (2014) indicated that it uses cookies to collect data on the Internet browsing habits of those who visit its website, and of course it is not alone in this practice.

How financial institutions collect, hold, and share personal information is governed by the *Privacy Act 1988* (Cth) obligations for consumer credit reporting, and other specific obligations (appendix F). Under common law, financial institutions generally 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 is used to fulfil legislative requirements (appendix F). For instance, financial institutions 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.

As part of their business operations, many financial institutions share data (including customers' personal information) with third parties, such as superannuation and managed funds organisations and their advisers, insurance companies and fraud reporting agencies (NAB nd).

Some of these organisations 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). However, the bank also notes 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. The *National Consumer Credit Reporting Protection Act 2009* (Cth) requires credit providers to enquire into the financial situation of consumer credit applicants. This requirement not only applies to financial institutions, but also to other entities (such as telecommunication and utility companies) that similarly offer customer credit. These entities also face information use and disclosure responsibilities under Part IIIA of the Privacy Act and under the Privacy (Credit Reporting) Code 2014. The ability to produce (and obtain) credit reports on customers hinges on these various entities (particularly financial institutions) having previously shared data on individual consumers with credit reporting bodies (discussed in more detail in chapter 4 of PC 2016a).

As is the case in many sectors, there is likely to be benefit in permitting consumers to have greater access to their financial data. For example, access to such data could help facilitate data transfer when a person wishes to switch banks or insurers, or enable them to negotiate a tailored contract with their new service provider. Later sections of this Report (chapter 5, appendix F) deal with this in detail. It may also assist individuals in determining the likelihood of whether applications for credit would be accepted or rejected — since it is presently not customary for financial institutions to explain the reasons for the rejection of a credit card application.

Telecommunications

Telecommunications businesses routinely collect personal information including identifying and financial information, some of which may be used, for instance, when providing support to customers. In the course of their operations, telecommunications

providers can also collect data on individuals' locations, the phone calls they make and the websites they look at.[3] 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.

Telecommunications businesses also collect information indirectly. 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). Realistically, it should be expected by subscribers that almost any service provider may take such steps.

Like many other businesses, telecommunications businesses make use of cookies and other digital identifiers to collect information from individuals' online activity. Vodafone states that it uses cookies and web beacons to measure website traffic patterns and may log Internet Protocol (IP) addresses to track user movement and gather 'broad' demographic information (VHA 2016, p. 1). Singtel Optus (2016) reports that it may collect biometric information for use with new technology such as voice and fingerprint recognition software.

Telecommunications businesses share data with a variety of third parties, including technicians, market research, telemarketing and marketing businesses, and debt-collection agencies.

As noted in appendix D, the APPs provide guidance and limits on the extent of information disclosure to other entities, as does industry-specific legislation such as the *Telecommunications Act 1997* (Cth). 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)

[3] Although outside the scope of our approach to this Inquiry, there are legislative requirements that compel telecommunications businesses to collect and retain certain types of data. The *Telecommunications (Interception and Access) Amendment (Data Retention) Act 2015* (Cth) 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 (AGD 2015). That Act deems (in s. 187LA) that all information retained under the Act is automatically personal information about an individual; as such, individuals are entitled to access all of their retained metadata pursuant to APP 12.

A customer's telecommunications data, in their own hands, has the potential to be quite valuable. It could be used, for example, to solicit competitive offers for service delivery from other providers based on actual consumption patterns and needs.

Factors such as complex tariff structures and discount schemes, technological advances, and changing consumer experiences have contributed to low levels of consumer switching between providers (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)

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 (in the latter circumstance, consumers' access to their own usage data would obviously have no effect on competition, but may enable behavioural changes to their utility use, as discussed below). 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 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 use to reduce utility costs. In Victoria, where smart meters were made compulsory, many energy companies have launched smart meter-compatible web portals. Potential 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).

However, there is some question as to whether these intended outcomes have 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 (VAGO 2015). Other factors — for example, the absence of incentive-based pricing structures — have compounded this issue, according to utility suppliers (Australian Energy Council, sub. DR281).

Insurance

Insurers collect basic personal data, 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).

The range of data that insurance companies collect on their customers varies with the type of insurance offered. 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).

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 1.5). 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.

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:

> Advances in scientific research and other digital modelling [have] been particularly important in understanding natural hazards and other catastrophic risks. … 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 insurers Bupa Australia (2015) and Australian Unity (sub. 95) use their customer and claims data 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.

Box 1.5 Using customer data to estimate risk and set insurance prices

One of the reasons insurance companies endeavour to obtain 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 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 (Insurance Council of Australia, sub. 66). Geocoding and provision of real time vehicle data over the Internet opens up the possibility for insurance that varies with factors such as location, timing and amount of vehicle use.

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 Australian Dental Association noted that that there are some private health insurers that are vertically integrated and operate dental 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 centralised electronic health claims system). 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.

Entities in less regulated sectors

The prevalence of private 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 — social media organisations, and others considered below, undertake their data collection activity without either regulatory fiat or public funding. Many, however, are able to combine hundreds of detailed observations about consumer characteristics to build accurate profiles of individuals. Indeed, in many fields, it is probable that private entities know much more about market trends and current economic activity than governments.

Regulation has not typically been a factor in the growth of data collected by these entities; indeed, it is highly probable that the absence of regulation has helped the rapidity of innovation in data collection and use.

Supermarkets and other large retailers

Supermarkets collect large quantities of data that relate to the shopping habits of customers. Aggregate insights can also be gleaned from electronic payments technology, even if the identity of the purchaser is not known. In addition, data collected via reward and loyalty cards — which are held by eight in ten Australians (Directivity, Citrus and First Point Research 2015) — gives retailers access to very detailed demographic and spending information on individual customers who may then have online and/or print advertising tailored to their purchasing habits. 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 smartphone or tablet applications
- rewards and redemption details applicable to membership and loyalty programs — that is, what, if anything, a customer has redeemed their reward points for
- whether a customer has connections with other people whose personal information the business 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 collect similar information, including through in-store surveillance cameras and in-store wireless Internet. Some of this data offers a significant range of secondary uses (box 1.6), not only allowing supermarkets to achieve a deeper understanding of consumer attributes but, given the diversification of supermarket business structures and changes in technology, offering the potential to change the way market participants interact.

Consumer group CHOICE has submitted that allowing customers to access the data held on their consumption habits is equally important in retail markets as in the more regulated markets for services:

> 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)

The Australian Food and Grocery Council (AFGC) echoed many of these arguments, and stated that consumer benefit is highly restricted when individuals' datasets are held exclusively by a single retailer.

The AFGC further contended that, compared to the benefits accruing to the supermarkets operating loyalty programs, the direct rewards available to the individuals who provide the data are of marginal value — essentially, that consumers are *donating* their data to the retailer, and that a mechanism to enable consumers to gain a greater benefit is needed (AFGC, sub. DR284, p. 5).

Box 1.6 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' 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.

Social media organisations

Social media allows its users to create, publish and share information with each other in web-based environments. There are numerous social media platforms in operation — prominent examples include Facebook, Twitter, Google+, LinkedIn, Snapchat, Instagram, Pinterest and WhatsApp. In December 2016, for example, some 15 million Australians

were using Facebook, over 14.6 million were using YouTube, and 5 million were on Instagram (Cowling 2016).

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 (see appendix G for more detail on the types of data collected and retained by social media organisations, and the way in which that data is monetised).

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 G).

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 postings that they make on the platform, from linked apps, and from non-Facebook users who visit websites in the facebook.com domain. Facebook (2015, p. 1) 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.

As briefly mentioned above, some of the datasets held by social media organisations exist at such a large scale, and simultaneously such a granular level of detail, that they may exceed anything held by governments in respect of the same topic — particularly when high velocity is also a factor. Examples include:

- LinkedIn is a repository of detailed data on individuals' employment characteristics, including professional qualifications and employment experience, which included details of 3.6 million Australians as at December 2016 (Cowling 2016). Other (non-social media) online services such as Seek.com.au 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), 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 outperforming the original efforts with updated analytics.

- Facebook also operates 'Safety Check', a tool with which users can mark themselves as 'safe' after an emergency such as a natural disaster or violent attack, which not only reassures their family and friends but can assist public authorities in emergency management (see appendix G for more detail on Safety Check).

For large social media providers, a social licence to operate is essential to the business model — without community acceptance there will not be sufficient users or sufficient data to monetise. Therefore, a firm's reputation among consumers — in terms of data collection, use, and on-selling — matters a great deal, and large firms will tend to self-regulate, or even regulate their third-party contracted firms, according to prevailing public attitudes (see, for example, Facebook (sub. 172)). For 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 — for example, 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 loath to admit this. 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: not only is the value of a social network *to its users* partially contingent upon its extensive use — especially upon their *own* social circles — but the capabilities of technology and software are not always transparent in any meaningful way to an individual once initial access has been granted. Therefore, even if users of the larger social media companies quit and move their business to a different provider, their browsing data, purchasing history or contacts may still be tracked (see appendix G for more detail). For these circumstances, greater clarity in the law concerning consumers' ability to cease being tracked or monitored may be needed.

Wearable technology 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. Fitness trackers (the most popular category of wearables) are capable of collecting an array of data related to an individual's fitness and health outcomes (box 1.7). On an international scale, 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.

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 the use of the data generated.

Smartphone and tablet applications (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 G).

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. The data is typically collected by app owners, and may be passed on to third parties, such as marketing businesses.

Box 1.7 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 identification number. Jawbone uses the information it collects to refine its 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).

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. While the basic technology underpinning the IoT has existed for some time, three key factors have facilitated its expansion (Heydon and Zeichner 2015):

- 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.

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 2016b). 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 to an iPad in the cab of the vehicle.

The Industrial IoT — which combines the IoT in manufacturing with big data analytics — has also expanded rapidly (Accenture and General Electric 2014). General Electric, traditionally a provider of industrial hardware and repair services, now includes digital sensors and microprocessors 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.

1.6 The challenges for governments and society

Realising who benefits from data use

In many cases, the benefits individuals derive from sharing their data are clear and simple to see; many products and services, from health care to online travel booking, cannot be accessed without providing personal information. There are also cases where people realise benefits from trading their data (usually once only), for example by receiving personalised discounts in exchange for providing data through a loyalty card.

The challenge for many data custodians is that the benefits from increasing the sharing and linkage of data can be far more difficult for individuals to see. Similarly, consumers may well benefit from the personalised services offered by banks and other organisations that use the personal information they hold (see, for example, Westpac, sub. 197). However, the benefits to the individual consumers are more difficult to identify when such data is shared on a larger scale between private organisations — and realistically, in such cases, the majority of the benefits obtained are most likely to remain with the organisations that collect and share the data.

In the public sector, the benefits of data use often accrue to the community as a whole, and as such can be even more difficult for individuals to see, let alone to quantify.

Further, the benefits from increasing the sharing and linkage of data can be obvious within an agency but difficult to explain to a public unfamiliar with program development and implementation processes; often such changes may not even appear to present clear benefits to the agencies themselves. Rather, benefits may accrue to the community as a whole (a desirable outcome) but may take some time to emerge or be difficult to quantify.

For example, the Multi-Agency Data Integration Project (Department of Prime Minister and Cabinet, sub. 20) links together data collected about individuals by five different Australian Government agencies and uses it to investigate large-scale social issues. It is likely that a substantial period of time will pass before any tangible benefits to individuals emerge from this initiative (such as better-targeted service delivery by governments) and some individuals may not receive a direct benefit at all.

When the expected benefits are clear, individuals have a stronger incentive to share their data, or to actively consent to their data being shared by those who already have it. Health research is a case in point. A recent survey found 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).

However, health policymakers as a group tend to be very conservative about encouraging such research linkages, even where the resulting improvements would clearly benefit the entire community. For governments wanting to influence the availability and usefulness of data, and for private sector organisations wanting to make more of use the data they collect, such conservatism poses several challenges.

First, research has found that individuals are motivated to share their data either by personal benefit or social gain (Moore and Niemi 2016). Therefore data custodians and users must either create a direct personal gain from use of individuals' data — by saving consumers either time or money — or they must become adept at explaining to the public how the community will benefit from increased data sharing.

Second, data custodians (and the relevant policymakers) themselves may need to take a wider perspective than just their agency when considering the benefits of sharing data, as illustrated by Richie and Welpton (2014, p. 4):

> A problem for government bodies is that they are often not the beneficiaries of the secondary analysis of their data. For example, if the Social Security department makes its benefits data available to academic researchers, this may lead to many research articles; but it is difficult to quantify the 'value' of such research to the social security department.

In Australia's current data policy environment, limited benefits accrue to public sector data custodians as a result of sharing their data, while risk to the custodians *increases*. For example, there is the potential for embarrassment at the quality of data, or the facts it contains; these risks create disincentives to share the data. Legal restrictions and a

risk-averse culture need to be addressed before such benefits can be realised. These issues are also evident with releasing data generated by the research community. The release of research data opens up opportunities for others to verify the results and expand research in new directions, but the individual researcher who generated the data does not necessarily benefit directly from this — and they may even perceive a threat.

Third, along with benefits, there are substantial costs involved in making data more available and usable. For example, in making public transport data open to app developers, governments typically incur processing and standardisation costs, but the benefits accrue to the developers, such as Google, Apple, and other companies producing mapping apps, and to individuals who use public transport. In some ways governments will eventually receive spillover benefits from, for example, more people using existing public transport routes. But this disconnect between the costs and benefits of greater data availability can affect the incentives of the parties involved, particularly when — as for governments — there can be little or no profit motive.

Finally, data held by the private sector can also offer community-wide benefits. Some of this data is collected by entities that are regulated or funded by governments (such as private hospitals), because they provide services or products that are vital to the community. The data held by these entities is equally important in enabling better decisions — by the businesses themselves, consumers and regulators. The potential benefit to society from increasing access to such data must be weighed against the cost to the business that has collected it.

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. Today, a significant proportion of data being created can identify individual people or businesses, or activities undertaken by them. For example, seemingly innocuous digital photos of landscapes or objects will often contain metadata that reveals the device used to take the photos and the time and location each photo was taken. The identity of the person uploading the photos may also be apparent or deducible. The identifiable nature of some data therefore poses challenges for individuals, governments and businesses in finding ways to increase access to particular data and insights while not revealing sensitive information about the individual people or businesses to which the data relates.

In *some* contexts, individual identification is desirable. For example, government agencies dealing in targeted public services or compliance need to be certain they are using data on the correct individuals, or, if they are monitoring business performance, the correct businesses. Similarly, individuals dealing with government agencies through 'tell us once'-style portals or multi-agency identity verification systems need to know that the agencies on the other end of the interaction are in fact using the correct details. Consumers using their purchasing or usage histories to obtain a better deal from competing businesses

need to be able to show that those datasets specifically apply to them. And healthcare providers of all stripes, if using shared electronic health records (appendix E), need to be confident that they have received the correct medical histories for their patients.

In other circumstances, individual identification is not only unnecessary but undesirable. Where the data records details about individuals' behaviour or characteristics, 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). Medical research, for example, can involve the collation of highly sensitive information that is important to a study (and therefore should be maintained as a unit record) but could equally put individual subjects at risk of humiliation or discrimination upon disclosure. Personal security concerns can also arise if information such as contact details or addresses are disclosed. Businesses can equally suffer reputational harm or financial loss (or both of these) if commercial-in-confidence information is released. A Framework for data availability and use must therefore be able to deal with all the situations described, and any other contexts along the spectrum from total openness of data through to complete restrictions on data use.

Dealing with the unknown

The regulatory arrangements for collecting, analysing and linking data (as well as for passing data from one party to another), and the obligations, rights and opportunities between parties — in both the public and private sectors — have developed in reaction to, rather than in anticipation of (or even in parallel with) digital technology improvements. This is the case both in Australia and overseas.

This reactive approach is somewhat inevitable. Designing for the advent of the World Wide Web and mobile computing could hardly have occurred in the 1970s; nor could Facebook and Google's near-ubiquitous nature have been planned for, even as recently as 2000. But a more strategic and holistic approach is now feasible and appears necessary.

Fears about how data will be used and what it will reveal are unsurprising in an era of rapid social and technological change, particularly when the outcomes of this exponential increase in data generation and use are largely unknown. This is true for both the potential benefits — few people would have imagined ten years ago that we would all have access to live traffic streams on our smartphones via Google Maps — and for the potential costs, such as the sheer volume of database hacking every year.[4] Both technology and social norms change in such unpredictable ways that it is impossible to know what 'the next Facebook' or 'the next Twitter' might be and what impacts might result. Regulation, and the development of trust, need flexible forms in order to cope. This Report proposes some.

4 See, for example, McCandless (2017).

Specifically, it designs a legal and policy framework that would give people a level of active control over the data collected about them, and the right to know what is being done with that data, regardless of the platform. Such a framework — principles-based for the most part rather than prescriptive — would give policymakers and citizens alike the knowledge that data regulations are flexible enough to adapt to new platforms and services, rather than needing to be reactive well after the fact.

For policymakers, who are a core focus of this Report, preventative behaviour is often the order of the day when dealing with the unknowable. But where benefits are large and change is rapid, to act in a purely restrictive fashion in respect of data is certain to be costly.

Broad engagement and trust are required to best manage risks

For individuals, control over one's personal data is a cornerstone of trust in the systems that collect and handle that data (NZDFF 2014). This is true irrespective of whether the data is collected and held by government or by private entities:

> 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).

The challenge for policymakers is to provide structures that aim to create a shared sense of opportunity to participate — that is, to foster both the individual and the collective benefit from the economic, environmental and social uses of data while simultaneously managing the potential costs. This requires the recognition of change as an opportunity and not just a threat and, through that, lifting the appetite for considered risk across governments, 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, mitigation strategies tailored to them, and a recognition of the enormous potential for benefits, the Commission's view is that we, as a society, should lean more towards data openness. Appetite for risk, as with many societal values and priorities, can be context-specific and change over time. This is already observable in today's data-rich areas, as people's preparedness to trade data for services increases significantly.

Key to achieving our optimal outcome will be building and retaining community trust in how data is managed and used. If individuals can see benefits arising from the use of data sourced from them; have control over how such data is being used; and use it (and reuse it)

themselves so that its value becomes more obvious, a virtuous cycle of greater trust and larger benefits can be established (chapter 4).

Yet conversely, incidents that damage that trust (such as poor data storage resulting in security breaches, or secondary uses of data that individuals were not told about beforehand) can have longer-term, further-reaching effects than the immediate impacts on the people involved, causing public opinion to move like a pendulum. Today's poor reporting practices — such as where data breaches that are a result of human error or poor security practices are nevertheless publicly attributed to data sharing or use practices — can ensure that if approval swings away from a particular use of data, there is every chance that it will move even further in that direction before it can be repaired (see, for example, Pillar 2013).

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. The Commission proposes a policy framework 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.

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.
- Numerous 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 to the extent that opportunities being taken overseas exemplify, and lack a comprehensive plan to do so in most cases. That said, there are some notable examples of governments seeking to use data in innovative, value-adding ways, most of which rely on the personal leadership of a handful of individuals rather than systemic commitment.
- Opportunities to make more productive use of datasets — such as by applying analytical algorithms or linking them to other datasets — and the benefits that could be achieved, are largely unknown until the data sources themselves are made known and a wide range of users have had opportunities to investigate the data. This underscores the substantive argument in favour of greater data access.
- The distribution of benefits from greater access to data would be widespread, but the Commission is of the view that it is individual consumers who have the most to gain, collectively, from action on Australia's data sharing and release arrangements.

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 are often large and wide-ranging.

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. These are many and various. The examples presented highlight what Australia will forfeit, conversely, through not making data available for wider use. Some of the innovations presented are already happening on a small scale in parts of Australia, whereas other opportunities have not yet been realised at all.

Supplementing this chapter, appendixes E and F illustrate the scope for innovative uses of data in the health and finance sectors respectively.

Potential risks associated with data release and sharing need to be weighed against the opportunities foregone. The potential for breaches of privacy law is only one, and perhaps not always the biggest risk — there is also the potential for financial or reputational damage — for example, to businesses or governments if they mishandle data or if their systems for securing sensitive information are breached.

For some datasets, greater access and sharing will involve greater risk, and occasionally, this will not be able to be fully mitigated or managed, given the rapid pace of change and the inherent uncertainties in how data could be used in the future. For other datasets, risks are largely predictable and manageable, even with widespread release of the data concerned. As the nature of risks evolve over time, data management techniques will also need to evolve.

In deciding which datasets to make more available, the risks and costs of wider release need to be carefully considered, along with policy frameworks and precautions that might be adopted to mitigate these (examined in detail in other sections of this Report). But the potential benefits of wider data use should not be dismissed in favour of undue risk aversion. The benefits already being achieved from innovative uses of data — and the open-ended potential for uses that have yet to be conceived — are simply too large to ignore. Perhaps the biggest risk is that Australia will be left behind in a world that is increasingly embracing and harnessing the opportunities data presents.

2.1 Opportunities for individuals

Australians stand to benefit individually if data is able to be used in new, innovative ways — both by using data directly themselves, and from the innovative use of data by businesses, not-for-profits, governments and researchers that ultimately benefits them.

Finding the right product, getting a better deal

Increased access to data can enable consumers to make better choices about the products and services they purchase. There are examples of this already happening — indeed, the success of firms providing comparisons of products and services demonstrates that consumers benefit from such services. If individuals were better able to share their historical transaction and consumption data with other parties, the potential for benefits would clearly be enhanced. It would be easier for consumers to know how their current products and providers compare, and to switch in cases where alternatives offer better value (witness the reams of ‘paperwork’ still required to change banks while preserving payment arrangements).

Improve personal finance decisions

Streamlining access to consumers' data — transaction data for instance — could improve personal financial decisions (Centre for International Finance and Regulation (CIFR), sub. 9; ANZ, sub. DR231). The potential cost savings to consumers that could arise through better product choices have the potential to be very material, as noted by iSelect (sub. DR266). Further, a research report published by the Queensland University of Technology (Silva-Goncalves 2015) estimated that Australians could save up to $11.6 billion annually by switching essential service providers.

These types of consumer data are already being leveraged by budgeting services and apps — such as Mint and Spending Tracker — to help individuals make better personal spending and budgeting decisions (Manyika et al. 2013). In the banking sector, many consumers are currently seeking to realise these benefits and share data by providing their online banking log-in details to third party data aggregators, which creates risks of fraudulent activity (appendix F) — far safer and more practical options are available in other developed economies. Properly designed measures to facilitate data sharing could also ensure that consumers have more *secure*, and easier, means for sharing their data.

Compare complex product offerings

In sectors where comparisons between products are difficult for consumers, third-party services are emerging and putting pressure on businesses to make data more readily available. 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. As an energy provider, AGL Energy (sub. DR251) has made electricity usage data available to its customers through a web portal and smartphone apps, which allows them to better manage their energy use and thus lower their electricity bills. Customers are also able to share data with other parties in order to seek advice on other energy plans that would better suit their usage patterns. AGL Energy also offers a service that monitors the performance of rooftop solar systems, allowing customers and non-customers alike to determine the most advantageous times to use home appliances.

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). 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. Such comparison services can be particularly beneficial for products with complex pricing structures, such as those in banking and telecommunications.

Yet these are all, at best, partial solutions to a broad failure on the part of data holders to provide access to data.

Personalised products and services

To the extent that businesses can use data to 'segment' customers, there would also be scope for providing bespoke products and services that are more reflective of the needs and wants of individual consumers. Of course, such tailoring would provide commercial benefits as well. As noted by the Australian Food and Grocery Council (sub. DR284, p. 6):

> At the heart of the disruptive power of online shopping are the individual insights about the purchaser provided to the seller … Far more precise matching of products to an individual's wants are possible as the seller builds a detailed profile of the consumer based on purchase decisions. As technology continues to develop and data analytics continues to improve the potential for mass personalisation and the consumer benefits that flow from it become realisable.

The music service provider Pandora operates in a similar manner, using 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. This includes genres and artists that they may like, based on the data previously reported, but have never previously heard of. 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 for consumers — there are nearly 80 million active subscribers to the free service (Morey, Forbath and Schoop 2015).

Big data techniques can also assist governments to better understand the needs of different groups of citizens, and thereby provide personalised services (Department of Finance 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) and to design better services to meet public interest objectives (Department of Finance 2013).

Drive competition in prices

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, in turn inducing them to price more competitively and service customer segments that may currently be underserviced. Increased competition would also drive increased innovation, in turn leading to lower prices for consumers. This sentiment was echoed by the Australian Food and Grocery Council (sub. DR284) in the context of data collected by supermarket loyalty programs — customers being able to transfer this data (which would include data on past purchases) could help new market entrants and enhance competition.

Another 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 5; appendix F).

Enable decision making that better reflects social and ethical preferences

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). A variety of other firms have sought to similarly increase transparency (in the absence of legislative obligation).

Bringing employers and employees together

Seek, an online employment advertiser, publishes a range of information about Australian employers, including employee reviews, to help job seekers identify companies that would be a suitable match for them (SEEK nd). This would be expected to improve workplace productivity and job satisfaction, and in turn reduce staff turnover costs.

Matching buyers and sellers

Greater access to data can also help to reduce search costs for consumers (as well as businesses) in a range of markets.

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 and minimise the likelihood of buying from unscrupulous sellers (Frontier Economics 2008).

Similar services exist for travellers. Tripadvisor collates information about businesses (as well as other attractions such as public parks and beaches) to help travellers make more informed decisions. Booking.com and Airbnb provide platforms for consumers to find and compare accommodation options, with user ratings providing a basis for comparison. Rome2rio offers detailed trip planning — for instance, it suggests different routes to a destination with supplemental information about travel times and costs — and helps travellers to find suitable service providers.

Uber uses a rating system for drivers, which helps it to maintain a high quality of service, but also allows drivers to rate passengers and to not accept passengers with poor ratings (Betters 2014).

Improved service delivery for individuals

In addition to helping consumers find products and services that better suit their needs, access to data can also enable service providers to improve the quality of services for individuals.

Healthcare

Improved access to healthcare data could enable more effective and timely healthcare services for Australians, including the development of new products and services (Research Australia, sub. DR282), earlier identification of population health issues, and more rigorous and better targeted health research (discussed in section 2.3).

PricewaterhouseCoopers (2014) noted that data sharing through greater adoption of electronic health records could improve healthcare provision 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 multiple different healthcare providers.

Already, some businesses use data to improve the quality of healthcare provided to individuals. DoseMe, an Australian-developed software application, uses patient-specific data — such as height, weight, gender and previous test results — to accurately calculate the appropriate dosage for that patient. Demonstrated benefits of DoseMe include:

- increased survival rates for childhood leukaemia patients
- fewer adverse outcomes from drug administration (such as the risk of toxic drug levels)
- fewer side effects for chemotherapy patients
- reduced costs, including those flowing from reduced length of hospital stays (DoseMe nd).

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.

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 and early interventions, which could improve long term health outcomes for their customers.

eTax pre-filling

Over recent years, the Australian Tax Office (ATO) has used data from a range of sources to 'pre-fill' information relevant to an individual's tax return (such as income). Pre-filled data can include:

- employment income (from employers)
- dividend and interest income (from businesses and financial institutions)
- welfare income (from Centrelink)
- private health insurance cover (from private health insurers)
- other relevant information provided by government departments (such as the Department of Education and Training).

As well as helping the ATO to identify fraudulent tax returns and lower auditing costs, pre-filling streamlines the process of lodging tax returns, saving taxpayers time and minimising the risk of mistaken manual data entry. The ATO estimated the compliance costs savings for personal taxpayers as being about $535 million per annum (ATO, pers. comm., 7 March 2017).

Insurance

Information asymmetry between insurers and the parties they insure can inhibit the efficiency of insurance markets, increasing premiums and favouring higher risk customers at the expense of lower risk customers. It could also result in particular population segments being underserviced by insurers.

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). These two innovative uses of data would tend to lower insurance premiums, particularly for those consumers assessed to be low risk.

As noted by MLC Life Insurance (sub. DR298), the use of big data techniques in life insurance markets could generate a range of benefits for customers. In addition to improving pricing and the types of products and services available, big data also has the potential to improve the quality of services delivered to customers by life insurers, such as by:

- streamlining the underwriting process and shortening waiting times for receiving and accessing information
- reducing the length of time taken to assess life insurance claims.

Reduced travel times

Over recent years, growth in the number of cars on Australian roads has contributed to congestion and made it harder to find available parking spaces. As noted by Deloitte (2013), real-time availability of traffic data would allow third parties to deliver apps that could reduce congestion, therefore cutting fuel consumption and carbon emissions as well as saving drivers and passengers time.

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 (CMTEDD (ACT) 2016). The South Australian Government recently released a smartphone app, BlueBays, which uses crowdsourced data to help disability permit holders find accessible disabled parking spaces (South Australian Department for Communities and Social Inclusion 2016).

Google Maps is yet another example of an app that can reduce search costs, by using geographic data to guide drivers and pedestrians to their destinations. By collecting data in real time from many users, Google Maps is also able to provide guidance on the quickest and/or shortest route to a destination, alongside estimates on how long it will take to reach the destination, thereby saving users time and money. Given the effect of traffic volumes on travel times, the utility of Google Maps would be limited if it were not able to collect information from users.

Providers of alternative modes of transport, such as via buses and trains, are also increasingly using real-time data in an attempt to provide consumers with more accurate information about service times.

The use of traffic data to improve road management in Singapore provides a more holistic example of how data can be used to improve road management (box 2.4).

2.2 Opportunities for business

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 pre-eminent examples of businesses that have monetised the collection and use of data.

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.

Improve the efficiency of processes and products

Data on the quality or performance of inputs in production can be used to improve the efficiency of production processes, while data on consumer demand can enable closer alignment of 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).

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).

UPS, a major package delivery service, uses data from telematics sensors in its delivery vehicles, along with mapping 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).

Improved basis for decision making

Data can also be used internally to drive operational 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). 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).

Wal-Mart established a data sharing system in the early 1990s, in which sales data — by item, store and day — was provided to 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). In Australia, the Australian Food and Grocery Council (sub. DR284) noted that access to data generated through retail loyalty programs (such as flybuys) could enable the development of new retail business models, including 'direct to consumers' sales models. Retailers currently charge suppliers for access to the insights derived from such data.

Large amounts of data are also a requisite input for businesses deploying algorithmic decision making tools, which can lead to improved decision making and risk management. As noted by Azcende (sub. DR274), the release of data held by Australian governments

could help businesses to better understand local market conditions, and thus make better decisions (rather than relying on global benchmarking as a basis for decision making). The Financial System Inquiry (Murray et al. 2014) highlighted the potential for small businesses, in particular, to benefit from increased availability of data for benchmarking and decision making purposes.

A basis for product innovation

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)

The benefits flowing from data use are not realised just by large business — indeed, even very small businesses can benefit. For example, Ebay subsidiary, PayPal, has begun using data it collects from Ebay businesses — some of which are very small — as a basis for providing 'working capital' credit to those businesses (appendix F). In other words, small vendors who sell via Ebay are able to access credit, which is underwritten on the basis of their Ebay sales history, and secured against future cash flows. Alipay (an offshoot of Alibaba) does likewise in China, and is reportedly planning to extend this to India.

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

In Tasmania, the Sense-T project, which involves the deployment of sensors across the Tasmanian landmass, has provided insights into the value that can be generated for and by businesses through the use of spatially-enabled data (box 2.2).

2.3 Benefits for society

In addition to benefits realised by either individuals or businesses, there is significant scope for increased data access to generate broader social benefits, including those that are not traded through markets (and thus may not be easy to value).

Better monitoring and use of resources

In 2011, New York City began releasing detailed energy and water use data for commercial buildings, with the datasets being used by building owners to benchmark the energy efficiency of their buildings and to prioritise energy-saving investments (Chui,

Farrell and Jackson 2014). More recently, the Mayor's Office of Data Analytics in New York have used data techniques to underpin a range of innovative initiatives, such as the establishment of a risk-based system for building inspections by the Fire Department (City of New York 2013).

One of the outcomes of Tasmania's Sense-T project (box 2.2) is a smartphone app that provides near real-time information and forecasts on air quality by integrating data from a range of sources, including air pollution, meteorological and allergen data, as well as user-reported health data, such as data related to asthma attacks. The app also uses data provided by users to form a 'profile' of each user, allowing it to provide individualised reports on potential environmental health triggers. The data collected '… also supports community-wide air pollution health advisories, heatwave forecasting and alerts, and fire weather mapping to assist firefighters, landowners and government' (Sense-T 2015).

Box 2.1 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; environmental 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 and itinerary planning; transport planning; vehicle tracking; traffic and congestion management; transport operations in air and rail; intelligent transport systems
- utilities: asset management; supply and 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)

Box 2.2 **Sense-T**

The Sense-T project demonstrates the potential benefits from collaboration between the private, public and research sectors. The project is predicated on making data collected by sensors (deployed by both the public and private sectors) available to researchers and industry. As well as bringing together existing sensor networks, the project also involved the deployment of new cutting-edge sensors, and integration of sensor data with historical and spatial data.

This project has already created a range of benefits for industry groups.

- Data collected from sensors deployed on 'sentinel' salmon (held in commercial fish pens), and from environmental sensors deployed on and around fish pens and in upstream areas, was used to evaluate fish behaviour and environmental conditions. This led to improved feeding practices (leading to lower food costs), improved decision making, and a better understanding of the effect of salmon farming on environmental variables (such as water oxygen levels) and the ways in which the industry can improve environmental practices.
- To improve the management and operation of vineyards, an innovative smartphone app — which provides information such as real-time weather and forecasts, vine growth conditions and warnings of high risk events (such as frost) — has been developed and will be released for pre-industry testing. The data collected will also allow for benchmarking across seasons and individual vineyards.
- Oyster and environmental biosensors have been deployed to collect a range of data, including data on oyster heartbeat and metabolism, and changes in water temperature, salinity, pH, dissolved oxygen, turbidity and algal abundance. The insights drawn from this project have enabled oyster farmers to better manage their stocks and improve oyster growth rates.

Source: Sense-T (nd).

Risk management

In response to the dying out of bee colonies worldwide, the CSIRO is leading the Global Initiative for Honeybee Health to better understand the factors contributing to the decline of bee populations. The initiative is centred around the collection of large datasets on bee hives all over the world, including data collected from micro-sensors fitted to bees to log their movements. This data allows researchers to evaluate which factors, such as pesticides and other contaminants, are having an effect on honeybee behaviour. The intent of the project is to protect honeybee populations and secure crop pollination, which is estimated to contribute up to $6 billion annually to Australia's agricultural sector (ANDS nd).

As another example, the Commission, in a report on natural disaster management (PC 2014), highlighted the importance of data for mitigating the impacts of natural disasters. In particular, it was noted that detailed data on natural hazards is an essential element of improving land use planning and better managing the risks of natural disasters, and that such data could also help households to make better decisions about where to build. The Insurance Council of Australia (sub. DR318) highlighted that natural hazard data could:

- help communities identify and understand hazards
- reduce duplication of mitigation efforts between jurisdictions
- help governments to plan activities to improve resilience against natural disasters.

Improved governance structures

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 potential for data openness 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; Cabinet Office (UK) 2013).

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 the Sunlight Foundation in the United States, and OpenAustralia in Australia (Chui, Farrell and Jackson 2014; OpenAustralia nd).

Improved government service delivery

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, including by directing resources to where they are most needed (Research Australia, sub. DR282; Australian Child Rights Taskforce, sub. DR291). 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)

Data can facilitate comparative performance monitoring of agencies and agency employees, and — knowing that performance and quality of services would 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.

The Productivity Commission's own Report on Government Services (SCRGSP 2016) presents a range of information (across thousands of data points) related to the delivery of government services and provides a basis for cross-jurisdictional comparisons. Such transparency keeps pressure on governments to improve their performance and service delivery.

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.

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. New Zealand has used its integrated data holdings to improve outcomes in the delivery of human services (box 2.3).

Box 2.3 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, in order 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 Australia, the NSW and Australian Governments have collaborated with Indigenous groups to establish the 'Maranguka Justice Reinvestment Project' to reduce youth reoffending rates in Bourke (Forsyth, Armytage and Lawrence 2016). A range of data (including Australian Bureau of Statistics demographic data, and child welfare, education, employment, justice and health data from the NSW and Australian Governments) was collected to '… understand and develop an assessment of the current state, progress to date, and next steps' (Forsyth, Armytage and Lawrence 2016, p. 7).

Analysis of this data allowed the project to identify four key areas of focus: early childhood and parenting; children and young people aged 8 to 18 years; the role of men; and reform in the delivery of government services. While the project is ongoing, and reform pathways have not yet been finalised and implemented, there will be a clear need for further collection and analysis of data to undertake planned evaluations of the project's effectiveness on an ongoing basis.

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

- reducing data management costs where agencies could re-use data from other agencies rather than individually collecting and maintaining data themselves (Department of Industry, Innovation and Science, sub. DR235; Department of Finance 2013)
- reducing the burden on individuals providing information to governments (Department of Social Services, sub. 10; Department of Industry, Innovation and Science, sub. DR235)
- 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.

Better government 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 2.4). In Australia, Uber (sub. DR311) partnered with Infrastructure Partnerships

Australia to develop a 'transport metric' to provide insights into travel times by location in several Australian cities. Underlying this initiative is a belief that it '… can help governments better target infrastructure investment and build a greater understanding of how Australia is tracking in developing better cities with increased mobility for residents' (Uber 2016).

Box 2.4 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 charge records, 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% and above 90% in peak periods (ITS International 2010; Singapore LTA 2008). This was complemented by an algorithm designed to 'fill in' traffic data on unmonitored sections of the road network, 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.

Expose government waste or corruption

Greater scrutiny of government spending can lead to significant social benefits, such as prevention of wasteful expenditure (Australian Taxpayers' Alliance, sub. DR279). As noted by Rosie Williams (sub. DR239), specific datasets that could contribute to greater

financial and political transparency include those related to government budgets, tenders and grants, donations to political parties and lobbyist registers.

Data on public sector contract procurement can also 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 research can improve social outcomes

Inquiry participants highlighted that making data more widely available could significantly improve the productivity of, and the *social 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; Complementary Medicines Australia, sub. DR223; ANU, sub. DR226).

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 could:

- create opportunities for repurposing and re-using 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.

The Telethon Kids Institute (sub. 5) highlighted the importance of data for health outcomes, including a range of linked data on educational, economic, geographic and racial factors. Western Australia has been touted as a world leader in this area and has had a data linking unit for many years (appendix E). The ability to link various datasets allowed Western Australian researchers to undertake:

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

Moreover, health data can help policy makers and researchers to identify factors that contribute to illnesses and to assess the safety of pharmaceuticals (appendix E; Jones et al. 2014; PHRN 2016). Researchers in the United Kingdom are using administrative health

records to improve the diagnosis of cancer (box 2.5), which has directly led to an increase in cancer survival rates.

The Australian Institute of Health Innovation, Macquarie University (sub. DR229) pointed to several research projects made possible by the linking of police-reported road crash data with mortality data and emergency department presentation and hospital admissions data in New South Wales. One such project identified that rear vehicle passengers face relatively higher risks of severe injuries, highlighting the need for protective mechanisms for rear occupants to be reviewed. A similar analysis was subsequently undertaken in the United States.

Box 2.5 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 program
- data from the National Breast Screening program.

This analysis determined that, of the cancer diagnoses in the linked datasets, on average almost 25% were diagnosed following presentation at an emergency department (rather than through other routes, such as 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%.

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

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 E). Given these challenges, it is perhaps not surprising that similar research has not yet been undertaken in Australia.

Source: Elliss-Brookes et al. (2012)

2.4 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 (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.

The wideness of this range reflects that:

- different measures of benefit or value are being estimated
- estimates are based on different types of data (some use spatial data only, while others use a broader range of public sector data)
- there is 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 estimate for Australia).

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, while 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 public sector data alone — could easily be much larger 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 value of data, by some estimates, is immense; second, it is impossible to 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).[5] 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.

[5] Although, as noted elsewhere in this report, standard cost recovery or charging principles generally do not reflect the true value of data held in government hands. The Australian Government announced in December 2016 that it was not proceeding with a proposed sale of ASIC Registry because final bids received did not deliver a net financial benefit for the Commonwealth.

Table 2.1 **Estimates of the value of public sector dataa**

	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 (2 010) ($7.6-$14.8 billion) Vickery (2 010) ($25 billion) Deloitte (2 013) ($7 billion) Lateral Economics (2 014) ($34 billion)
Direct net economic value	Consumer and producer surplus from the collection and sale of public sector data[c]	DotEcon (2 006) ($625 million - $1.2 billion)
Direct use value	Value added plus market value	Deloitte (2 013) ($1.9 billion)
Value added	Value added by entities that use public sector data to generate other goods and services[d]	Pira (2 000) ($22 billion) ACIL Tasman (2 010) ($682 million Vickery (2 010) ($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 (2 006) ($3.9 billion)
Investment value	Costs of making public sector data available to the public Costs of collecting public sector data	Pira (2 000) ($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. [d] 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 underestimate actual market value.

Source: Lateral Economics (2 014)

2.5 The sum of opportunities

As this chapter, and large sections of this Report in general show, there are many and varied contexts where Australia could be doing much more with its data. In many

instances, the full extent of the unexploited benefits from *not* utilising data, or what Jones et al. (2014) call 'the harm from non-use', remains unknown, as is the case with many other areas of economic reform. And indeed, mere observation of opportunities taken up in practice overseas but not here, or common in Australia in one or two jurisdictions but not in others, demonstrates a serious loss of benefit.

The distribution of benefits from comprehensive reform would be widespread, but we are of the view that the balance of the overall gains clearly favours, most of all, individuals in their role as consumers. Further argument to support this is made in later chapters. Prospective impacts for government, business, researchers and nonprofit entities are also promising, at least where data sharing and release can be made more effective and widespread. But for individual Australians there is likely to be a significant upside to reforms in this area, as has been shown by the examples above from the health sector, from banking and finance, from transport, from social service delivery and police and emergency services, and from insurance markets and the markets for individual products and services.

Opportunities from data release would also, by and large, beget further opportunities and, subject to consideration of the risks and costs involved, the balance should generally lie in favour of facilitating greater data access where the benefits are uncertain.

FINDING 2.1

The benefits from greater access to data would be widespread, but consumers, in particular, have much to gain, collectively, from action on Australia's data sharing and release arrangements.

3 What holds us back?

Key points

- A key to achieving the many potential benefits of data use will be building and retaining community trust in how data is managed and used and building a shared understanding of the benefits that flow from better data access and use, including by consumers themselves.
- Community surveys indicate some concerns about the privacy and security of personal information. Such concerns are compounded by individuals' lack of knowledge of what data about them is collected, how it is used and what rights they have over such data, and the often impenetrable nature of data-related terms and conditions.
- Legislation restricting access to data was formulated up to a century ago, and much is no longer fit for purpose. The primary legal impediment to more effective use of data is typically *not* the Privacy Act, but regulations and guidelines specific to the field in which the data is collected.
- A culture of risk aversion among public servants has led to overly cautious interpretation of relevant legislation, lack of willingness to make it known that some data exists (lest the agency be asked for it), and complex and lengthy approval processes for data access (including duplicative ethics committee requirements).
- A lack of national leadership has contributed to piecemeal bureaucratic processes for data sharing and release. Sharing of restricted access data between government agencies and certain trusted parties has been limited at best. There are numerous examples of data siloing, despite the obvious benefits of data sharing.
- The extent of productive linking and integrating of datasets varies substantially across jurisdictions, but is generally inadequate when viewed against the potential opportunities or practices in some other countries.
- For the vast majority of publicly funded research, neither the public nor most of the research community has access to datasets generated by projects, despite there being a clear public interest in this occurring. Data releases are infrequent and very dated. Destruction of datasets — a particular policy of the Commonwealth — increases the costs of data re-use.
- Technical challenges are used to justify risk averse practices: the risk of data breaches and re-identification of de-identified personal data; fragmented data collection and release; lack of common standards; and a shortage of skills and dedicated resources. None of these are driven by increased data sharing, but are nevertheless cited as reasons not to do so.
- The collection and use of data allows businesses to improve their competitiveness, but on their terms. While individuals can currently see their own data, there is no obligation on any data collector to provide this in a useful form.
- Better access to, and the opportunity for individuals to use data held about them, would likely spark additional competition and innovation to the benefit of consumers.

3.1 Fragile community understanding and trust

Maintaining and building community understanding and trust about the ways in which data is collected, managed and used must be at the forefront of any consideration of reform to data and information policies and practices.

The damage created by a loss of trust is particularly evident in examples in both the public and private sectors of poor data handling or practice. The fact that none of the oft-quoted examples — Yahoo addresses and passwords, Home Depot credit card details, Red Cross Blood Bank, and the 2016 Census — was related to data sharing or planned public release is an important fact for reducing community fears or misperceptions. All development of data practice — whether in the private sector or public sector — must take the creation and preservation of understanding and trust as its first consideration.

The community has concerns about privacy and data access

In chapter 1, the Commission noted that despite an apparent willingness to share information about themselves in a wide range of contexts, individuals have voiced concerns about the privacy and security of their personal information. The latter is borne out by community surveys that indicate that the level of concern about the privacy and security of personal information has been increasing. 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 (OAIC 2013a).

Common types of concerns include:

- invasion of privacy through misuse or mishandling of sensitive personal details
- invasive use of data, where individuals are targeted, which can be merely annoying or actually harmful
- discrimination and exclusion from services based on correct or incorrect information
- malicious use for criminal purposes, including stalking, identity theft and fraud
- 'big brother', where the line between state power and individual liberty is crossed
- commercial detriment through use of information to gain commercial advantage or inflict commercial harm.

Cybersecurity remains a serious issue. In 2015, Australian retailer Kmart's online product ordering system was hacked, with the theft of customers' names, email addresses, delivery and billing addresses, phone numbers and details of past product purchases, while denial of service attacks afflicted Australia's 2016 online census (Coyne 2015; MacGibbon 2016). In the United States, health insurance providers Anthem and Premera were both hacked in 2015, with the theft of the personal records of around 79 million and 11 million people respectively (Kirk 2015).

Where community concern exists, it mostly relates to online services, particularly 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 (figures 3.1 and 3.2). Nearly all users of online services take some measures to ensure their information is secure (OAIC 2013a). But they also expect government to play a role in keeping data safe — there is an expectation that the responsibility for the protection of personal information is shared between users, service providers and government (ACMA 2013).

Particularly relevant given the global reach of many entities that collect data is the matter of data being sent offshore. The Office of the Australian Information Commissioner (OAIC) found only 10% of survey respondents were not concerned about their personal information being sent overseas, while 62% were 'very concerned' with the prospect of their personal information being sent overseas (OAIC 2013a). 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 sizable minority, 28%, indicated they would give personal information in exchange for a discount, 14% for a prize, and 34% would give information for better service (OAIC 2013a).

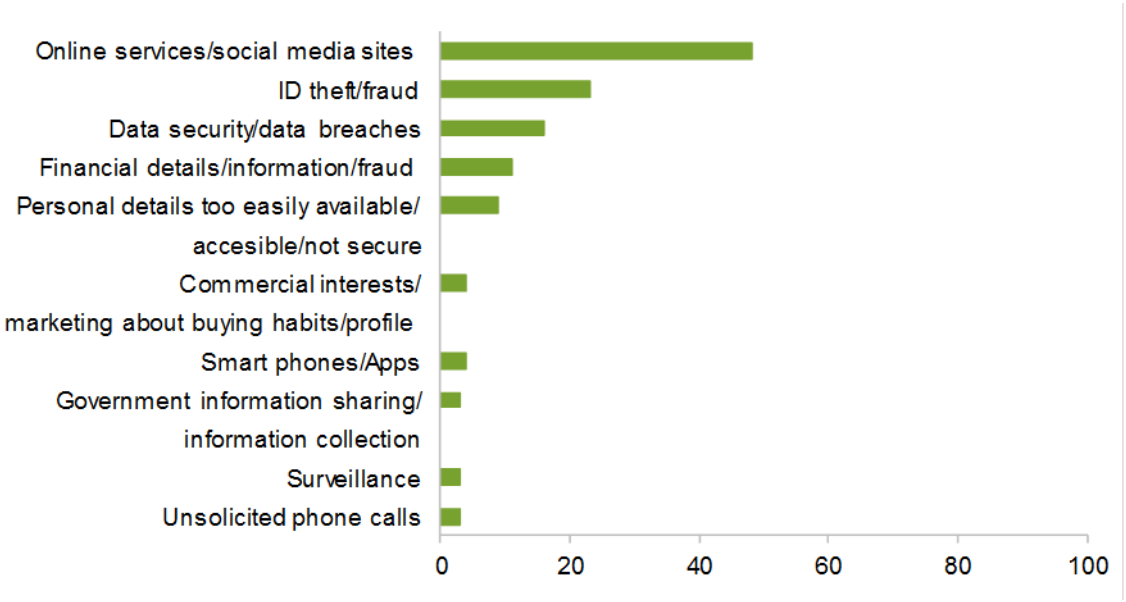
Similarly, despite the highly sensitive nature of medical information, individuals are willing to share this data in order to promote health research, particularly if they are well informed about the research and the organisation conducting it, and if they give their permission for the data to be used (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).

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, the community generally does not view information sharing between departments as a major threat to privacy (OAIC 2013a). In fact, according to the Queensland Government (sub. 207), most people expect that different parts of government 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 see benefits in sharing information between government departments, as they believe it will enable more efficient and accessible government services (Bruce and Bruce 2015). But at the same time, individuals expect governments to share individuals' information with their consent, only when strictly necessary, and to 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.

Figure 3.1 **Biggest privacy risks perceived by people**

% respondents[a]

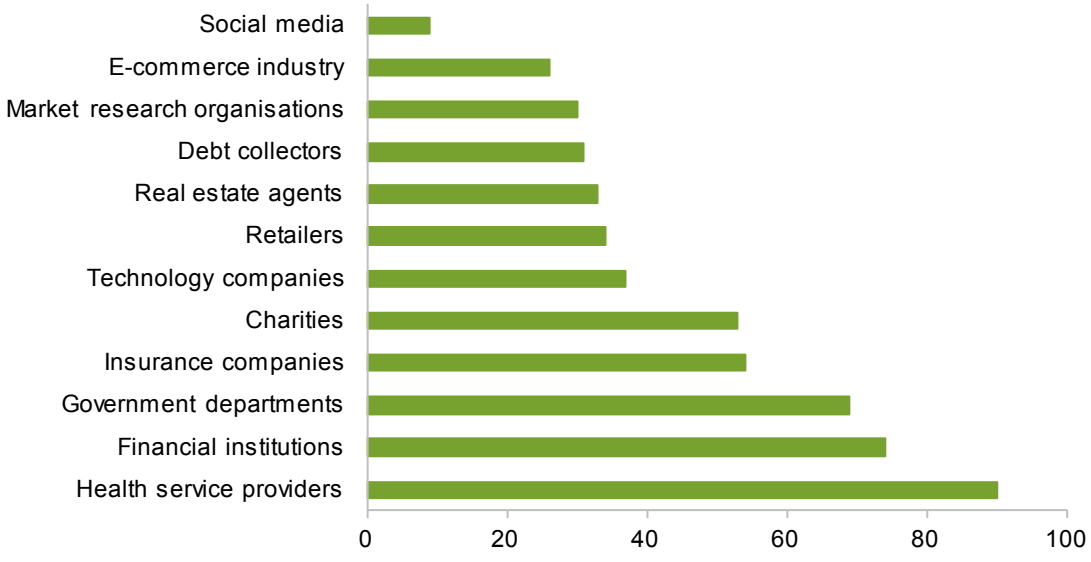


[a] Sample size = 1000

Source: OAIC (2013a)

Figure 3.2 **Trust in organisations to handle personal information**

% respondents trusting organisation types listed[a]



[a] Sample size = 1000

Source: OAIC (2013a)

FINDING 3.1

Individuals are likely to be more willing to allow data about themselves to be used by private and public organisations, provided they understand why and how the data is being used, can see tangible benefits, and have control over who the data is shared with.

The privacy paradox

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 such as wearables, raising what Acquisti, Taylor and Wagman (2016) refer to as the 'privacy paradox'. Explanations for this apparent paradox include that:

- individuals may place a high value on the immediate benefit they receive from making certain data available (such as friends 'liking' their social media updates, or a business providing a free service) and discount the future effects (such as a prospective future employer reading the same social media updates);
- individuals may be unable to assess reductions in privacy as a result of not reading or understanding privacy policies;
- knowledge of possible solutions, such as privacy-enhancing software, may be limited (Acquisti, Taylor and Wagman 2016); and
- degree of separation is a consideration — an individual may be less concerned about what Amazon knows about their personal preferences than what their colleagues/acquaintances know (Wittes and Kohse 2017).

Privacy matters — it matters a lot to some people. But so too do the benefits of the many services that create data and rely on it for their viability. The challenge is to mitigate risk where possible and to balance risks against benefits derived.

Individuals often lack understanding of data collection, data use and their rights

With a rapidly expanding array of data collection approaches being used by businesses and governments (chapter 1), consumers are often 'given' certain benefits or services that are 'free' of charge in notional terms but come with a requirement to provide data to the entity supplying the benefit or service. Even some government websites collect information on the browsing choices of those who use their sites.

Most people are aware that their data is being collected, but few are aware of just how extensive that collection process can be:

> 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 2013, p. 12)

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 their data footprint included information on their location, and just 14% realised that they were sharing their web surfing history (Morey, Forbath and Schoop 2015).

Of course, there is an onus on individuals to make responsible choices about those to whom they provide personal information and for what purposes. But this is difficult with limited knowledge of what is being collected and why.

Consent is often given without genuine understanding

Entities covered by the Privacy Act are required to have a privacy policy that sets out how the entity collects and manages personal information. However, even when 'consent' has been given, there can be a clear lack of understanding of terms and conditions of consent, particularly as their 'take it or leave it' nature discourages consumer engagement (Law Institute of Victoria, sub. 184). Genuine consent requires that people are able to understand their options and make a meaningful choice.

Many individuals do not read disclosure statements because of their complexity. 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). In other words, providing a disclosure statement does not guarantee either understanding or agreement.

A lack of control

Public and private data custodians typically 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).

Individuals can ask to see their own data, but there is *no obligation on any data collector to provide this in a form useful to them* or to their preferred third party. Consequently, access to privately held data (and the desirability of having more access) varies considerably between industries.

The Financial System 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.

Individuals can also request access to government records about themselves under the provisions of the *Freedom of Information Act 1982* (Cth). These provisions are used often in some areas; for example, the Department of Immigration and Border Protection (sub. 168) handled over 20 000 requests in 2014-15.

Trust is integral to data quality and access

A lack of trust in the way data is collected, stored and shared can result in valuable data assets being underutilised and the erosion of potential benefits. In the public sector, the issues that arose with the 2016 census had a detrimental impact on trust and may yet have carryover effects to subsequent censuses, even though the eventual level of participation in the 2016 census was similar to the previous census in 2011 (box 3.1).

In the private sector, loss of community trust has been detrimental for businesses perceived as being invasive of their customers' privacy or having poor information security practices. The failure of Phorm (box 3.1) shows that a lack of social acceptance can prove fatal to new business models — legal authority does not necessarily command social legitimacy (Carter, Laurie and Dixon-Woods 2015).

Individuals concerned about the use (and potential misuse) of their data may have an incentive to provide false or misleading information in order to protect their privacy — thus compromising the integrity of the data collected. By contrast, people appear more willing to share information when they understand and trust how it is being used, and feel like they have control over its use (Acquisti 2010). Facebook and Google — exemplifying data-dependent businesses — have sought to respond to community concern over particular publicised uses of data. For instance, when Facebook introduced a default 'Instant Personalization' feature that shared users' personal information with other Facebook users and third party sites without users' consent, thousands of users expressed their frustration at having to manually disable the feature and Facebook responded by changing its privacy settings.

Box 3.1 How inadequate collection, storage, sharing or use of data can erode community trust

Insufficient care in the way data is collected, stored, shared or used can erode community trust and reduce the potential benefits of data.

2016 Census — security and continuity of data collection

On the evening of the 2016 Census, the Australian Bureau of Statistics (ABS) closed down its online Census system to ensure the integrity of personal information lodged by the public after several 'denial-of-service' attacks. The Census remained offline for over 40 hours. Aside from the inconvenience to the public, some groups expressed concern that the data collected through the Census may be less reliable than in the past because of households deliberately withholding or incorrectly supplying information in response to data security concerns. Ultimately, response rates were similar to the previous Census and the ABS has reported low levels of non-response to sensitive questions.

Phorm — lack of social acceptance for data collection

Phorm was a digital technology company that developed an advertisement business based on monitoring Internet users' browsing habits and providing them with targeted advertisements. The widespread reaction to Phorm's proposed service highlighted concerns over its effect on the privacy of users' browsing habits. Despite changing to an 'opt-in' approach, under which the company would not collect any data from users who had not explicitly opted in to its services, Phorm was unable to recover from the setback and ceased trading in 2016.

Ashley Madison — insecure data storage

The customer data of the Canadian online dating service Ashley Madison, which had the slogan '*Life is short. Have an affair*' was hacked in July 2015. The data — including emails, names, home addresses, sexual fantasies and credit card information — was subsequently made public. A $576 million class action lawsuit was filed against the company.

Medicare Benefits Schedule data — re-identification of shared data

In September 2016, academics at the University of Melbourne used re-identification techniques to recover the ID numbers of some service providers from de-identified Medicare Benefits Schedule and Pharmaceutical Benefits Schedule datasets the Australian Government Department of Health had published the previous month. The department removed the data from public access after being notified by the academics.

Source: Department of Computing and Information Systems at The University of Melbourne (sub. DR303); Basu (2015); Wikipedia (2017a, 2017b); MacGibbon (2016); Williams (2016).

In an era in which big data is ubiquitous and data flows are essential to major consumer and business service products, it is imperative that data frameworks and policies are able to address and respond to a wide range of potential concerns, such as:

- the changing role of anonymity and de-identification — situations where private facts (diabetes) can be inferred from data that may not be sensitive (purchase history of sugar-free food), and matched to other data to re-identify an individual
- information asymmetry — when data collectors know more about an individual's data and its implications than the individual

- the complexity of the data landscape — for instance, the University of Western Australia submitted that many people would be unable to:

  … assess the potential privacy impact of that data even if they had expressed interest in the details in the first place. There is also the question of whether they would have any understanding of the possible privacy leakage of data through secondary interpretation of the data. For example, a temperature/humidity sensor in a room measuring the room's temperature could quite easily be used to work out what a person was doing in that room and at what time from the potentially minute changes in environment that they themselves could bring about. (sub. 296, p. 2)

Accountability

Accountability of data collectors and custodians is also important to building trust and confidence. Recent cases such as the 2016 census (box 3.1) aroused considerable public interest. But it is unclear whether the public has been reassured that the parties actually responsible have taken responsibility and, just as importantly, that someone has taken responsibility for ensuring the same mistakes do not happen again. Without high levels of visible accountability, public trust and confidence is undermined.

In the absence of support for the most active group today in providing data unreservedly — individuals, consuming services — to take a role in the data-driven economy other than a defensive one via protection of their privacy are the seeds for a substantial loss of trust.

Most consumers know implicitly that they are exposed, but as submitters often noted (for example, Pirate Party Australia, sub. DR242; Consumer Action Law Centre, sub. DR308) and as others point out (for example, Tim Berners-Lee (2017)), perhaps not to the degree that applies already today. In the future, that feeling of exposure — and the limited right to participate on reasonable terms — is a serious risk to data collection and use activities in the public and private sector alike.

3.2 Legislative complexity

In chapter 1, the Commission noted that Australia's provision of open access to public sector data lags countries with similar governance structures — including the United States, the United Kingdom and New Zealand. There are many factors that contribute to Australia's comparatively poor release and sharing of public sector data, including:

- a dense web of legislative requirements
- a culture of risk aversion, leading to overly cautious interpretation of the legislations, and approval process complexity
- lack of a whole of government approach (including failures to adequately address machinery of government changes)
- jurisdictional barriers — within and between jurisdictions
- intellectual property and licensing issues.

These factors are not unique to Australia, but have already been overcome to varying degrees in other countries, highlighting the need for clear and strong leadership in this area of reform.

A tangled web of legislative requirements

Legislation — at the Commonwealth and State/Territory level — *restricts access* to identifiable data. This legislation centres on the protection of privacy and government secrecy. While such restrictions are ostensibly designed to protect the privacy of individuals, they may sometimes be used to cloud the transparency of government activities and performance. Individual agencies are often also required to consider their enabling legislation, which may contain specific data handling requirements and a range of other legislative instruments, including privacy legislation and regulations specific to their portfolio or program area.

The authority to *release* information is far more limited. At the Commonwealth level, one such authority is included in the *Census and Statistics Act 1905* (Cth), which requires the Australian Bureau of Statistics (ABS) to publish and disseminate compilations and analyses of statistical information (appendix D). 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 would benefit the community.[6]

Overall, however, even where legislation includes provisions that allow for data use, access is often impeded. As an illustrative example, sharing of information for child protection purposes often fails to occur even when it is possible (section 3.3).

Commonwealth privacy legislation

The *Privacy Act 1988* (Cth) (chapter 1 and appendix D) 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'.

There are several features of the Privacy Act that cause uncertainty and may characterise its limited application in a highly data-driven future. But the key factor in impeding greater access to and sharing of data, is that the focus of regulatory effort under the Privacy Act is *not* on opening up opportunity but on limitation of threat. In the 30 years since the Act first applied, what is 'data' has been altered by digital technology and legislation has not kept

[6] 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*, s. 57).

pace, despite proliferating over this period. The Australian Law Reform Commission made the following observation in its 2008 report on Australian privacy law and practice:

> It became clear during the course of the current Inquiry that these rapid advances in information, communication and surveillance technologies have created a range of previously unforeseen privacy issues. …
>
> [T]he *Privacy Act* has undergone significant amendment since its enactment in 1988, resulting in an unwieldy and overly complex piece of legislation. (ALRC 2008, p. 105)

In the face of ongoing developments in the collection and use of data, a number of further amendments to the Privacy Act have been made since 2008. The Office of the Australian Information Commissioner (OAIC, sub. 200) has submitted that the principle-based approach of the Privacy Act allows it to remain relevant despite ongoing technological advances in the collection and use of data; however, recent court proceedings (*Privacy Commissioner v Telstra Corporation Limited [2017] FCAFC 4*) show that there is still much room for ambiguity.

State/Territory privacy legislation

All States and Territories, except for Western Australia and South Australia, have privacy legislation. This legislation applies to the State and Territory public sector,[7] but not the private sector in most cases (unlike the Commonwealth Act) (see appendix D).

State legislation is based on similar principles to the Commonwealth Privacy Act, but the existence of multiple regulatory schemes severely impedes data sharing by the public sector across jurisdictions (WA Data Linkage Branch, sub. 13; NSW Government, sub. 80). Data sharing involves multiple approval processes by agencies that adhere to different legislation and adopt different policies on data access (OAIC 2016a). 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.

Secrecy protections in specific Acts and Codes

Beyond the general protections in the Privacy Act, a wide range of other Acts contain provisions preventing disclosure of information about people and businesses (appendix D). At the Australian Government level, the Australian Law Reform Commission

[7] 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 privacy of health information (appendix D).

(ALRC 2010b) identified 506 secrecy provisions in 176 different pieces of legislation. In contrast to most provisions of the Privacy Act, more than 350 of these provisions create criminal offences and are punishable by imprisonment (box 3.2).

Box 3.2 Secrecy protections

Secrecy provisions have been restricting access to Australian Government data for over a century (ALRC 2010b). Secrecy provisions can impose substantial limitations on access to and use of identifiable data.

- Section 135A of the *National Health Act 1953* (Cth) generally prohibits divulging any information about individuals collected through the operation of the Act, unless authorised by the Minister. The penalty for divulging such information without authorisation is either a $5000 fine, two years imprisonment, or both.
- 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 [and impede 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 Australian Taxation Office 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 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 [National Assessment Program — Literacy and Numeracy] 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. (Grattan Institute, sub. 12, p. 4)

Given the potential penalties that these secrecy provisions carry, they 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 2010b, p. 537).

Data custodians and research institutions have created complex and very inefficient approval processes for data access to meet legislative requirements (discussed later in this section).

FINDING 3.2

A wide range of more than 500 secrecy 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 would no longer be fit for purpose.

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

Legal limitations on data linking

Data linkages, where different datasets containing information about the same individuals are brought together, add substantial value, enabling more insights to be derived from information already collected. 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).

Data linkage is currently carried out by accredited bodies — six state-based linkage 'nodes' (part of the Population Health Research network) that can link state data, and three accredited integrating authorities (the Australian Bureau of Statistics (ABS), Australian Institute of Health and Welfare (AIHW), and Australian Institute of Family Studies (AIFS)) that can link Commonwealth data (AIHW, sub. 162). Australian Government agencies are able to link data without an accredited authority in some circumstances (such as where a third party is not involved).

Data linkages in the health space are particularly restricted. For example, section 135AA of the *National Health Act 1953* (Cth) 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 only in a very 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 valuable insights into clinical outcomes, access to services and cost-effectiveness of health policies that are currently not available (appendix E).

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 2016a).[8]

Destruction of datasets — not just linked datasets, but shared data more generally — after projects are completed can lead to duplication of effort and waste of public funds. In the case of government-held MBS and PBS data, the Information Commissioner has acknowledged the need for a review of the guidelines that regulate its handling, in consultation with the Department of Health (OAIC, sub. 200). However, incremental change such as this may not be sufficient: *permission* to take a different path across the whole of the Commonwealth — that is, transformative behaviour — is desirable, lest individual custodians remain in their risk averse modus operandi.

The Commonwealth's policy is anachronistic. Linkage keys are successfully reused in projects that are based on data at the State level. For example, the Centre for Health Record Linkage, a data linkage unit managed by the New South Wales Ministry of Health, maintains a master linkage key system that can create links between the health records of over 11 million people in NSW and the ACT (CHeReL 2016b). These linkages meet the requirements of privacy legislation (see appendix C for further information on data linkage systems).

There has been some ad hoc progress in overcoming the restrictions imposed on users of Commonwealth data, and specifically data linkages. For example:

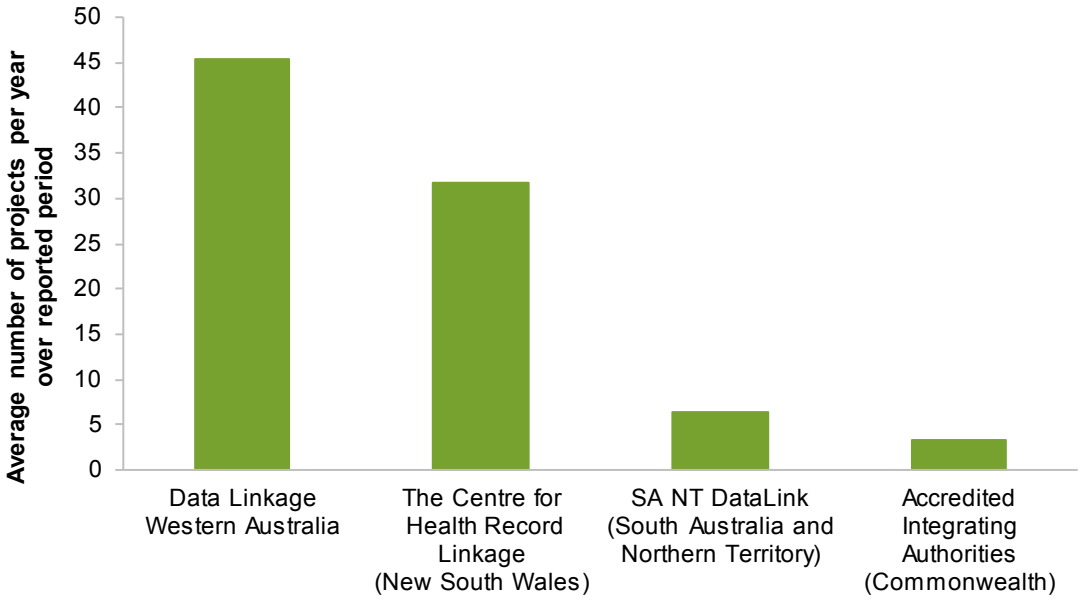
- The Multi Agency Data Integration Project (MADIP) undertaken by the ABS along with four major Commonwealth agencies (the Department of Health, the Department of Social Services, the Department of Human Services and the Australian Taxation Office) and the Business Longitudinal Analytical Data Environment (BLADE), using data from the Australian Taxation Office and ABS surveys, in partnership with the Department of Industry, Innovation and Science (ABS, sub. 94) are both enduring linked datasets.
- A trial was conducted by the AIHW (sub. 162) to establish arrangements for ongoing linkage keys involving the Commonwealth, New South Wales and Victorian health departments.

[8] 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 2016a).

Despite these developments, the potential for data linkage and integration at the Commonwealth level is clearly not fully realised. Each involves complex effort to work around conflicting legislative or 'usual practice' requirements with 'permission to try' tenuously anchored in the efforts of a Committee of Deputy Secretaries or individual agency leadership, both of which must inevitably be prone to shifting with bureaucratic whim or staffing change. In a data-driven world, this approach will ultimately prove inadequate.

Western Australia and New South Wales are the leading jurisdictions for data linkage. Data Linkage Western Australia has been integrating unit record administrative datasets since 1995, averaging 45 integration projects per year. The Centre for Health Record Linkage (New South Wales) has averaged 32 data integrations projects per year (figure 3.3). But these worthy examples illustrate how little is being done in other jurisdictions.

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



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

Overall, data linkage and integration in Australia is limited. There are numerous examples of successful enduring data linkages undertaken overseas, such as New Zealand's Integrated Data Infrastructure managed by Statistics New Zealand. More generally, compared to the policy and institutional reforms in other jurisdictions designed to facilitate data integration, Australia's institutional arrangements (appendix C) also appear to lag international best practice. But the efforts of New South Wales and Western Australia demonstrate that progress can be made with the right leadership and resourcing.

FINDING 3.3

Data integration in some jurisdictions (particularly Western Australia and New South Wales) has progressed in some fields, but highlights a lack of action in equivalent fields at both Commonwealth and State level, and reveals the large unmet potential in data integration opportunities.

Privacy legislation is designed to restrict, not allow, use of data

Instead of establishing clear rights for individuals to actively utilise their data, privacy legislation and secrecy provisions seek to establish protective boundaries around personal information that apply across the public and private sectors (chapter 1 and appendix D). These boundaries are under pressure from burgeoning demand (public and private alike) for data on the one hand and lack of clarity about interpretation on the other.

While a principles-based approach to information access 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), there are drawbacks. By being open to interpretation, access outcomes reflect the culture of the entity holding data — the 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. This is not a matter of attributing fault, but rather a description of an unfortunate reality.

If demand was perhaps not so sweeping and pervasive in its pressure for individuals' data, this lack of clarity may be of little consequence — the regular updating of guidelines could be a successful and low cost regulatory strategy. Indeed, it has been seen as such by the Productivity Commission, among others, in other circumstances. But these are not today's circumstances for data and surely will not be tomorrow's.

The ongoing debate whether IP addresses constitute personal information is illustrative. 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. In January 2017, the Full Court of the Federal Court (*Privacy Commissioner v Telstra Corporation Limited*) ruled that mobile phone location metadata is not 'personal information' to be accessible to individuals under the Privacy Act.

Elements of the legal definition of personal information have long given rise to ambiguity, without legislators reconsidering the matter. The Australian Law Reform Commission concluded in 2008 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).

Data Republic (sub. 176) indicates that their use of data is a matter of significant judgement — 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.

This issue will become increasingly important as public and private sector entities collect more and more information about people and analytical techniques continue to improve. Legislative guidelines may struggle to cope with a global shift in the collection and movement of data across borders. Larger entities concerned with corporate reputation will most likely nevertheless seek to comply despite uncertainty. However, many small businesses are outside the scope of the Privacy Act — notably those that have an annual turnover of $3 million or less and are not in one of the categories of business (such as health service providers and entities that trade in personal information) specified in section 6D of the Privacy Act.

FINDING 3.4

The boundaries of personal information are constantly shifting in response to technological advances and new digital products, along with community expectations.

The legal definition of personal information, contained in the *Privacy Act 1988* (Cth), has always had an element of uncertainty, and is managed by guidelines. In the face of rapid changes in sources and types of data, outcome-focused data definitions remain essential. But practical guidance (that data custodians and users can rely on) is required on what sorts of data are covered by the definitions.

Is privacy legislation the culprit?

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 of the Privacy Act and its accompanying Privacy Principles by data collectors and custodians, that, 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 – Queensland, sub. 42; and the Independent Schools Council of Australia, sub. DR257).

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 what is permitted and how likely are the associated risks.

Agencies' views on the complex web of legislative requirements

Participants have noted the difficulty in complying with law that varies across jurisdictions, managing inconsistencies in data collection practices and coordinating permissions across multiple and diverse data custodians (Australian Indigenous Governance Institute, sub. 60, Department of Social Services, sub. 10; Department of Environment and Energy, sub. 120; Department of the Prime Minister and Cabinet, sub. 20). Indeed, WA Data Linkage (sub. 13, attachment 6) 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:

> 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. The Queensland Government (sub. 207) similarly noted the detrimental effect that multiple and inconsistent privacy laws have on sharing and the need for simplification:

> 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, by creating a whole-of-government system for inter-agency data sharing (the *Data Sharing (Government Sector) Act* was passed in late 2015). More recently, South Australia has introduced the *Public Sector (Data Sharing) Act 2016* (SA) that legislates the 'five safes' model (discussed in appendix C) to enable data sharing across trusted entities.

Intellectual property issues

Intellectual property rights can limit the sharing of data. In some cases, businesses and governments cannot share or release certain data because a third party holds intellectual property rights over it. In other instances, businesses and governments hold intellectual property rights over data and use licences that limit the extent to which it can be redistributed and reused.

Australian intellectual property law provides for copyright in datasets where compiling information involves sufficient 'authorship' (chapter 1 and appendix D). It applies to some but not all datasets. 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 — discussed in appendix D) 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.

The Australia Government Public Data Policy Statement required that Australian Government entities publish 'appropriately anonymised data by default … under a Creative Commons By Attribution licence unless a clear case is made to the [Department of Prime Minister and Cabinet] for another open licence' (Turnbull 2015). The Bureau of Meteorology, in turn, noted the importance of retaining flexibility in licence tools available to use (sub. DR322). Appendix D details policies for licensing arrangements in Australian jurisdictions.

Insufficient data access under some government contacts with private sector

Where public sector agencies 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. In many cases, this is unavoidable.

Particular issues arise where governments contract private sector organisations to provide services — such as public transport — and are unable to access data relating to the operation of those services. In cases where there is a public interest in the data generated — for example, where such data is required to monitor performance — adverse outcomes could be avoided through the establishment of contractual obligations to provide data. At present there appears to be no consistent approach in the public sector for achieving this.

The Australian Government's default contract template (the Commonwealth Contract Suite) does provide for a wide range of uses by the Australian Government of any IP created under contracts between Australian Government agencies and the private sector. However, we have been advised by agencies that not all data generated through contracts will always be covered by intellectual property rights. If it is not covered, the Australian Government's access to potentially valuable data may be constrained.

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 reused.

3.3 Risk aversion and lack of leadership

An entrenched culture of risk aversion in the public sector

Governments have proven to be poorly equipped in understanding the emerging benefits and value of data and hence consumer and business demand for data. They tend to focus on identifying potential risks, but are less practiced in understanding and evaluating the benefits that flow to them and others from access to data. As a result, government agencies have preferred to not allow wider access to data rather than take steps to clearly assess, mitigate and manage potential risks.

Inquiry participants (such as the Office of the Information Commissioner – Queensland (sub. 42), the NSW Government (sub. 80) and the Cancer Council Australia (sub. 141)) and OAIC (2013b) suggested that a culture of risk aversion within government agencies, as well as concerns about agency reputation and data being misinterpreted and misrepresented, make agencies reluctant to allow access to government data holdings. 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 !

While in some cases genuine legislative barriers exist, a risk averse culture has been described as perhaps posing an even greater barrier to data sharing and release than the actual legislation.[9] The Government 2.0 Taskforce (2009) described 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)

However, not all instances of a reluctance to provide data are evidence of risk aversion. We have also been made aware during the course of this Inquiry of a number of instances where a government agency has shared data with another agency and the recipient has misused the data (for example, provided it to third parties without the consent of the original data custodian). Understandably, such occurrences leave data custodians somewhat hesitant to engage further in data sharing, particularly where scope for effective recourse action is limited.

Reluctance to release data — be it because of inherent risk aversion or prior bad experiences — is compounded by a paucity of positive incentives for agencies to share and release their data holdings. Government agencies are generally not entrepreneurial. Compared to private sector data holders, public sector data custodians have few price signals to illuminate the value of data holdings and any value generated from data tend to accrue to others rather to the data custodian. Aside from dampening agency incentives to release data, this reality also clouds agency decisions on how much additional processing to perform on any data that they do intend to share or release. It also has implications for funding decisions. These issues are discussed in chapter 9.

Data custodians: between a rock and a hard place

Individual 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 (and this can only occur if the data is used), but performance against this

9 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'.

expectation is very difficult to measure.[10] At the same time, custodians must also ensure that data is only released in accordance with all relevant legislation. In some cases, breach of legislative requirements can result in imprisonment or substantial fines for the data custodian personally. In light of this, it is unsurprising that data custodians would be risk averse. Internal procedures, which in some cases have evolved to take on a more substantial role than that originally intended in the legislation, can also have a 'chilling effect' on data release.

Summarising this situation, the Attorney-General's Department (sub. 209) stated:

> Agencies regularly point to concerns such as privacy or security, but the real issue can be … 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 focused on the actions and performance of their own agency, whereas sharing is likely to be for the purposes of other 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)

If this culture is to change, senior leadership within the public sector — including agency heads and at the political level — must clearly and consistently communicate the aspirations, expectations and intent behind data policies. The public servants who implement these policies need to be empowered to act, rather than toeing a risk-averse line.

Lengthy approval processes limit the value of data

In an attempt to satisfy the web of legal requirements, and address the concerns around increasing data use, custodians have created complex and inefficient approval processes for data access. In particular, 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 (appendix D).

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 (HREC) 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

[10] 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).

custodian separately and often sequentially (see, for example, Archerfish Consulting, sub. 30; Centre for Big Data Research in Health, sub. 21). In some cases, the situation may be further complicated if custodians do not accept the ethics committees' approval. A noteworthy example involves the Australian Electoral Commission (box 3.3).

The overall result is governance processes, particularly in regard to health data, that are extremely lengthy and inefficient. For example, the publicly funded Vaccine Assessment Using Linked Data Safety Study requires data from both the Australian and State Governments. According to Research Australia (sub. 117), obtaining data from the Australian Government took six and a half years and state data is still yet to be linked, more than eight years after the project commenced.

Box 3.3 The Australian Electoral Commission: a matter of more than ethics

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 research ethics committee (HREC). Data custodians give their final approval only after the HREC signs-off on the application (see, for example, PHRN 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 may refuse to allow access to data even after HREC approvals. Further, there is limited transparency in the process. For example, data custodians are not required to provide potential users with an explanation as to why their application was refused (Adams and Allen 2013).

The Australian Electoral Commission (AEC) recently 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. 33).

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. 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).

An associate professor at the Australian Institute of Health Innovation, Macquarie University, commented:

> Currently, I spend more time filling in paperwork and getting approvals to access linked data extracts and conduct research that involves record linkage than I do in conducting the actual research. (sub. DR229, p. 2)

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. Researchers have reported that data custodians are often under-resourced in data provision, which only serves to compound delays in processing applications (SSCH 2016)

Inefficient approval processes substantially reduce the benefits of research, and in the case of some health and medical research, unduly delay work that could materially improve the quality of people's lives (chapter 2).

Human research ethics committees have an important role, but create further duplication of effort

HRECs[11] 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 2016a).

While the intent of the privacy legislation exceptions for use of identifiable information is to increase access to identifiable information for use in health and medical research, the complexities of the HREC application and approval processes act as additional barriers 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).

The recent case of the National Hand Hygiene Evaluation project provides an example of such burdensome internal procedures. The overall cost incurred by the project in getting approvals to access data was estimated to be $348 000, some 38% of the total study budget. An ethical review body required 30 hard copies of the application paperwork, including the 200-page National Ethics Application Form — over 7000 pages in total. This ethical review body was also aware that the application had already been approved by another ethical review body. An estimated 25% of application time was spent on the original ethics application, with 75% on repeated applications. The planned timeline for ethical and site-specific approval underestimated the actual time by 17 months (Barnett et al. 2016).

Apart from using a national application form, there is little similarity in the way the HREC committees operate. Not only do they require vastly different documentation to be submitted, but there is limited mutual recognition between committees. Furthermore,

[11] In addition to human research ethics committees, universities have ethics committees that oversee research that is conducted on, or using, animals (University of Melbourne, sub. 148).

justifying multiple ethics committee processes by saying they are each dealing with different ethical aspects is not helpful in streamlining data access approval processes. Researchers still must submit applications to each separate committee and negotiate with each 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) 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, 2016a). In addition, a number of jurisdictions have developed a National Mutual Acceptance Scheme, intended to streamline the approval of research conducted in public hospitals (NHMRC 2016b).

While these issues are particularly prevalent in the health sector governance arrangements present a challenge in all cases where access to identifiable information is required.

Streamlining governance arrangements — 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 (chapter 6).

Wanted: national leadership on data sharing and release

The key issue underlying the present malaise regarding data sharing across the public sector, and between the public and private sectors, is an absence of clear enabling frameworks — *permission*, putting it bluntly — to deliver a more cohesive, whole of government approach.

Sharing data in the public sector: unsustainable disarray

The evidence examined in this Inquiry points to data being systematically siloed in the public sector with little sharing between agencies or beyond. As the Centre for International Finance and Regulation (sub. 9) stated:

> 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)

There has been a lack of a coordinated approach to data policy development and a reluctance to tackle head on the morass of legislation, policy and practice. At the Commonwealth level, the Department of Social Services (sub. 10, p. 6) noted that, in effect, 'Commonwealth data sharing initiatives are currently undertaken on a largely ad hoc basis between individual agencies.'

Equivalently for data sharing between jurisdictions, the Australian Institute of Health and Welfare (AIHW) (sub. 162) reported that there are few ongoing arrangements for information sharing:

> 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)

Poor data sharing in the field of child protection is a case in point. 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 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).

Even programs set up with the explicit intention of sharing data across jurisdictions run into difficulties. A Tri-Borders Project was intended to track and provide continuity of learning for remote Indigenous students across all school sectors in WA, SA and NT. However, issues arose with jurisdictional privacy legislation including the types of data that could and could not be shared. These issues undermined the overall effectiveness of the project and, as a result, it is unlikely that it will be extended nationally (Independent Schools Council of Australia, sub. DR257, p. 3).

Other cases where improved data sharing across jurisdictions could lead to substantial community benefits are detailed in box 3.4.

To overcome risk aversion and cultural barriers to data sharing and release, data custodians need to see a clear indication of jurisdiction-wide permission. In Australia, the NSW and SA Acts come closest to offering this (appendix D).

A public sector patchwork of enabling processes

Data sharing between government agencies generally proceeds via a patchwork of mechanisms. In the Commonwealth and most of Australia's States and Territories, it is often a piecemeal process with a mix of different types of data sharing agreements including memoranda of understanding (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.

Box 3.4 Opportunities for improved data sharing across jurisdictions

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)

The Australian Child Rights Taskforce (sub. DR291) noted that there were issues sharing data within and between jurisdictions:

> Information is often not shared which should, subject to appropriate safeguards, be shared across and between public authorities and other authorised agencies to respond to children with holistic, interdisciplinary, timely and child-centred responses across services. This is particularly important for children at risk of neglect or abuse. These problems exist at the following levels:
>
> a. An inter-governmental level (for example, between the federal family law jurisdiction and the state/territory care and protection systems, and between states and territories); and
>
> b. An intra-governmental level (for example, between law enforcement, education, health and child protection authorities). (p. 4)

Information exchange *between service providers* in different jurisdictions can remain severely limited even where there is a clear public interest in doing so. For instance, the Australian Law Reform Commission (2010a) 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)

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 D 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 (and therefore

are within the same legal entity) or even under the same portfolio within the same government.[12]

A number of participants have noted the complexity of MOUs for sharing data between agencies. 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)

In March 2016, the Department of Prime Minister and Cabinet (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.

Machinery of government changes: a spanner in the works

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.

Although policy to govern these transitions is in place (under which the National Archives is responsible for providing guidance on policy, mechanisms and standards for the transfer of information), MOG changes are nonetheless disruptive. As the Department of Employment (sub. 18) states:

[12] For situations where government entities *do* reach an agreement to share data within or between jurisdictions, the Australian Government operates a physical infrastructure can be used for secure, encrypted information sharing, called FedLink. See appendix C for more detail.

> 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)

DPMC (2015) reported that it can take several years and multiple memoranda of understanding to establish data sharing arrangements between government agencies.

At the Commonwealth level, the frequency of MOG changes in recent years has proven significantly challenging for the management of some datasets, and has compounded the difficulties of data access across jurisdictions.

Lack of guidance for data exchange between the public and private sectors

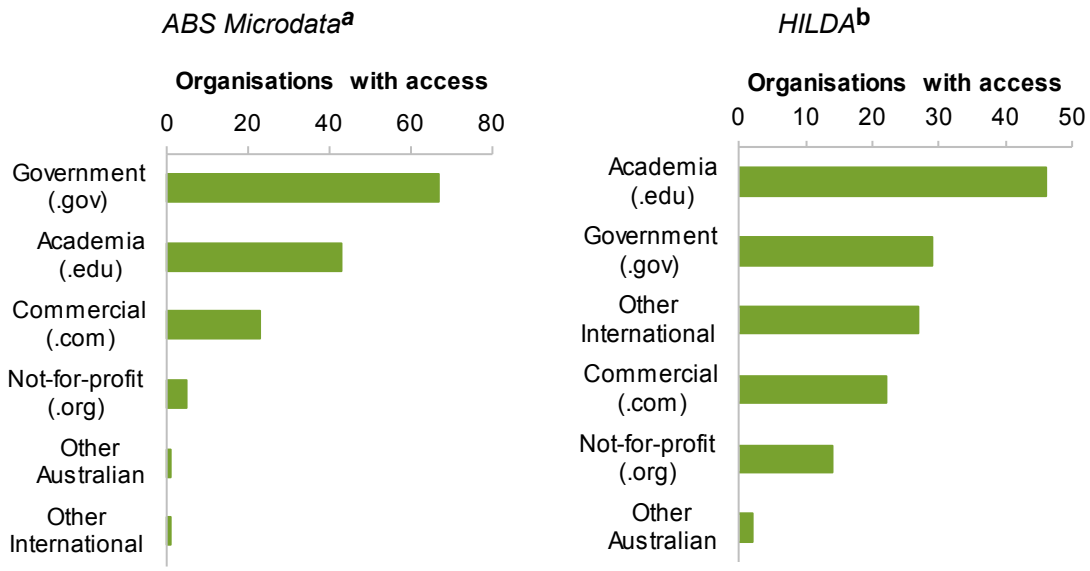
For all the challenges faced by government agencies in sharing data with each other, they typically fare better than private organisation and not-for-profits in accessing public sector data. Access to public sector data typically favours known, trusted parties such as other government agencies and selected academics.

Figure 3.4 presents a measure of access for two groups of well-known restricted access public sector datasets — ABS Microdata and HILDA (Household Income and Labour Dynamics in Australia). These datasets are leading examples of Australian practice, offering sound standards for developing trust, and clear specifications for users. Yet, 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.

There is no reason to suspect that skills needed to use these datasets would be less abundant in the private sector than in government and academia. This, in turn, suggests the comparatively low private sector access rates reflect obstacles to private sector entities gaining access to the datasets. One obstacle can be that of discoverability (it may be easier for those in government agencies to know where to find data in other government agencies). However, a means of access that favours some types of users over others or access eligibility criteria that explicitly discriminates between types of users are also likely to limit use by some.

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 **Types of organisations accessing restricted data**
Number of organisations



[a] Access to detailed microdata is available either on-site at ABS offices or (as part of a current trial) via a secure web login. [b] Access to HILDA by individuals not in a tertiary institution, government agency or community organisation, is limited to those within statutory bodies and is granted only on a 'case-by-case' basis.

Source: ABS (2016a); Melbourne Institute (nd)

A number of submissions outlined difficulties that businesses and not-for-profit organisations have experienced in accessing public sector data. For instance, Archerfish Consulting noted:

> 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 of 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. This has a chilling effect on the range and quality of independent research and policy analysis available in Australia. (sub. 30, p. 1)

Difficulties with data sharing between the public and private sectors also commonly arise where services are provided by a combination of public bodies (such as public hospitals), and by publicly funded private or not-for-profit providers (such as private hospitals) (appendix E). Such information is often of potential value to both the funder (to ensure efficacy of funding) and providers (to improve service delivery).

To the extent that these biases in data access do not represent application of robust risk management practices, consistent guidance is required for public sector data custodians in dealing with potential private sector data users.

Data exchanges with service providers are convoluted and inefficient

The public sector funds many non-government organisation to provide a range of services, particularly community and welfare services. However, the exchange of data between government agencies and service providers seems convoluted and inefficient.

On the one hand, service providers report difficulties accessing information that is necessary for them to do their work. 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)

The Joint Council of Social Service Network (sub. 170) submitted that negotiations to access data can be long and time consuming and that where data is provided, it is often in formats that are difficult to work with or require specialised software.

On the other hand, information provision to government from service providers is piecemeal. Some information generated by service providers is shared with governments for reporting requirements, and may be required under legislation — health, education, aged care and community services are sectors with numerous such reporting requirements.

Sometimes government service providers are required under contract to provide information, yet contracting practices vary widely. Many examples provided to the Commission 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.

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.

Data sharing is even more complicated when public-private partnerships are involved. The realities and complications created by multiple jurisdictions and the boundaries between the public and private sector are nevertheless manageable with a more effective approach in place (chapters 7 to 9). Data sharing is even more complicated when public-private partnerships are involved. The realities and complications created by multiple jurisdictions are nevertheless manageable with a more effective approach in place (chapters 7 to 9).

There are pockets of innovation, but no central vision or leadership

Despite the ad hoc state of data sharing, some instances of good practice have emerged in a number of jurisdictions (box 3.5) — and other agencies and their leaders should be looking to emulate these.

In the course of this Inquiry, we have come across numerous examples of initiatives to encourage government agencies to improve the use of their data. Two of these 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.

Box 3.5 Isolated examples of good practice

- New South Wales 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.
- 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 policy support and, as noted earlier, South Australia has recently introduced legislation that aims to facilitate public sector data sharing.
- The Commonwealth Department of Social Services (DSS), within the limits of fairly anachronistic legislation, operates the National Centre for Longitudinal Data to support the management of critical national data assets (Department of Social Services, sub. 10), and is developing its own trusted access arrangements.
- 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.

DPMC is working towards greater coordination between the various organisations working in this space (DPMC 2015). In recent times there have also been a number of constructive declarations and initiatives from DPMC, encouraging all government agencies to improve the use of their data, but these are yet to be fully implemented (see, for example, Department of Industry, Innovation and Science, sub. 69) and some may be contingent on the Government's response to this Report's recommendations.

In order to achieve large scale reform, ad hoc progress is not enough. Governments must implement broad, systemic changes, coupled with strong leadership and dedicated funding. *Permission* to be proactive in developing data opportunities and to address the gulf between the Commonwealth and the States and Territories is necessary.

The form of this permission needs to:

- be substantive enough to survive changes in key senior personnel (who to date have been a substantial force for change)
- extend more broadly than one or two agencies engaged in single project
- offset the legalism that constantly creeps in via MOUs and guidance documents that now pervade the policy landscape and cannot easily be ignored by middle level public servants.

Some of these changes are already occurring in some States — examples include the NSW Data Analytics Centre and Victoria's forthcoming data agency (Andrews 2016; DFSI (NSW) 2016a). National leadership is also necessary.

FINDING 3.5

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, and often perverse incentives, greatly inhibits data discovery, analysis and use.

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.

3.4 Data breaches and re-identification

Data breaches

As identifiable information is collected and generated in increasing volumes across disparate entities, the risk of data breaches increases. A breach of personal information occurs when 'information held by an agency or organisation is lost or subjected to unauthorised access, modification, disclosure, or other misuse or interference' (OAIC 2014). Inadvertent disclosure of commercially sensitive information that identifies individual businesses can also be considered a breach. Businesses are typically better able to deal with the consequences of data breaches than are individuals, and so much of the focus of data breach risks is on sensitive information about individuals rather than businesses.

Data breaches can be a result of human error or systems failure; however, in most cases, they occur as a result of malicious or criminal activity, such as data theft or hacking (Ponemon Institute 2016). Only rarely do they occur as a result of data sharing.

Commercial entities typically 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, and the data itself is of value to the entity. Researchers similarly value access to data and are strongly motivated to maintain necessary safeguards to ensure their ongoing access. For government agencies, transparency and the resulting public accountability can be strong motivators for safeguards, and there is often little to gain for an individual from allowing misuse of sensitive data. But even for the most well intentioned data custodians, human error and malicious efforts mean risks cannot entirely be eliminated and considerable diligence is required to keep protections up to date.

For businesses, researchers and government agencies that have been involved in prominent data breaches, sustained damage to their reputation and lost trust is a significant risk. Not surprisingly, studies highlight the benefits of transparency around data collection and protection in building and retaining consumer trust (Morey, Forbath and Schoop 2015).

How real are the risks from data misuse?

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).[13]

Data breaches can result in identity theft, which affected about 126 000 Australians in 2014-15 (ABS 2016b).[14] 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).

For the organisations affected by a data breach, there may be 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, there can be substantial implications for their operations and the level of community trust in the public sector's 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 this 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 [Her Majesty's Revenue and Customs] on the understanding that [the Government] will keep it secure. (Poynter 2008, p. 54)

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

---

13 On the other hand, for some individuals, increased access to data may result in criminal liabilities, such as the discovery of welfare fraud.

14 Overall, 1.6 million Australians experienced a form of personal fraud (including card fraud, identity theft or scams) in 2014-15 (ABS 2016b).

proportion of the overall number of incidents involving misuse of personal information.[15] 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). 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). These proportions are broadly in line with international trends (Huq 2015; McCandless 2017).

While neither human error nor systems issues can be completely eliminated as sources of breaches, organisations can use risk management processes and staff education and training to mitigate risks while increasing access to and use of data.

FINDING 3.6

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

Breaches of personal data, often compounded by individuals' unwary approach to offering data, are largely dominated by malicious database hacking or criminal activity. By comparison, breaches due to sharing or release are rare.

The risks of re-identification

Risks of re-identification change as more datasets become available and analytical techniques advance

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 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 identify individuals is rapidly evolving, and the risk this poses to privacy may increase as more datasets become available (WA Data Linkage

15 Notification of data breaches is mostly voluntary. However, in February 2017 the Australian Government amended the Privacy Act with the *Privacy Amendment (Notifiable Data Breaches) Act 2017* (Cth). Prior to this amendment, only eHealth records required mandatory notification 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).

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. In the recent past, there have been a number of re-identification incidents overseas, caused by linking apparently de-identified datasets. But the number appears to be small compared to the range and volume of data breaches caused by hackers and human error.

Nonetheless, while it is the case that if data is not thoroughly de-identified the risk of re-identification increases significantly (El Emam et al. 2011), several Inquiry participants submitted that de-identification cannot permanently eliminate all such risk:

> De-identification is fallible due to the dynamic nature of the data release environment … What data will be released, and when, is neither predictable, nor controllable. A hypothetical calculation of re-identification risk can only be valid on the day it is performed, because it makes specific assumptions about the auxiliary information available to the attacker. The data release environment will be changing every day and the risk of re-identification will change with it.
>
> Given that the risk of future re-identification cannot be accurately predicted, and is dependent on uncontrollable events, it is unadvisable to publicly release data whose privacy is dependent on the accuracy of that predicted risk. What is more, publicly released data cannot be recalled. Its release is a onetime event, once public it is public forever. (Department of Computing and Information Systems at the University of Melbourne, sub. DR303, p. 5)

The OAIC expressed a similar view:

> … it is unlikely that any high-value datasets containing personal information will be able to be sufficiently de-identified to enable general, open publication (in a manner that also preserves the integrity of that data). (sub. DR236, p. 4)

The level of re-identification risk that the community is prepared to accept is an important consideration. Once data exists, zero risk is an impossibility, so the issue is one of managing risk in a way that is acceptable to the community (and balanced against the benefits brought by access to data). In light of this, there remains a role for de-identification:

> De-identification can still serve a purpose — it is a useful technique for protecting against accidental or observational privacy breaches. For example, there will be datasets for which identity is not desired, nor sought, and for which de-identification is useful to allow easy access and analysis while mitigating fear of accidental identification. (Department of Computing and Information Systems at the University of Melbourne, sub. DR303, p. 6)

Addressing the risks of re-identification does not have to rely solely on de-identifying data. Providing a secure environment for data access can also play a crucial role. Chapter 4 of this Report outlines such an approach.

Current guidance on de-identification is inadequate

De-identified information — information that is no longer about an identifiable individual or an individual who is reasonably identifiable — can be shared and published freely. And some government agencies, such as the Australian Taxation Office (ATO, sub. 204), are doing this successfully. 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). But these examples are rare, at least in Australia.

More commonly, focus is drawn to the worst case scenarios, and data custodians avoid releasing data altogether (Ritchie and Welpton 2014). In some cases where data is confidentialised, Inquiry participants have argued that agencies adopt standards exceeding the requirements of the legislation. The AIHW (sub. 162, p. 18) stated 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'.

Consequently, there is need for better guidance on robust de-identification. This need has become increasingly pressing following the reported re-identification of personal information from data released by the Department of Health (The University of Melbourne, sub. DR303). Researchers at the University of Melbourne found that it was possible to use a trial release of de-identified MBS and PBS data that was published on data.gov.au to uncover provider numbers issued to doctors by Medicare. However, in this instance no individual's privacy was compromised.

Rather than celebrate the close relationship between researcher and agency that ensured no breach actually occurred — public commentary instead emphasised the failure. Data custodians all over the Commonwealth and States could hardly fail to notice and fear trying again. While this is perhaps inevitable, it is deeply undesirable, because such activity — an online Census or a de-identified MBS/PBS — or its near equivalent will be repeated in the future. Denial is not a solution. Thus the real questions are what lessons can be learned and what safeguards are needed in order to do this effectively in future.

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). The OAIC (2016b) reported in its submission to the Commission's Education Evidence Base Inquiry that it is consulting and producing further guidance on de-identification; this guidance has not yet been made available.

Deterrence can also play a role. To deter malicious re-identification of data, the Australian Government has introduced to parliament amendments to the Privacy Act to create new

criminal offences of re-identifying government data that has been de-identified and disclosing such data — the *Privacy Amendment (Re-identification Offence) Bill 2016* (Cth). However, queries have already been raised about the potential for the amendments to have unintended effects on those legitimately testing the robustness of data de-identification.

3.5 Poor usability of data

The usability of a dataset — the ease and extent to which data can be transformed and employed to achieve specified goals — is contingent on dataset quality. Key aspects of dataset quality are its machine readability, and the extent to which it makes use of open standards and commonly accepted definitions and methodologies (DFSI (NSW) 2016b). The ABS's Data Quality Framework lists a further seven different characteristics of a high-quality dataset: institutional environment, relevance, timeliness, accuracy, coherence, interpretability and accessibility.

In most sectors, Australia's datasets are characterised by a lack of adoption of standards, which has severely limited the extent to which data can be moved around, linked and made use of. For the most part, the standards exist (see, for example, AIHW 2007) — they are just not used.

You can't use data if you don't know it exists

Lack of discoverability in the public sector

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 (2013b) 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 it became obvious that some departments were uncertain about what datasets they held' (SSCH 2016, p. 18).

The DPMC (2015) recommended that all Australian Government agencies publish details of all their non-sensitive major datasets (and metadata) on data.gov.au. Encouragingly, several datasets previously listed as unavailable have since been made available on data.gov.au — for example, data on visa grants and lodgements, temporary entrants to Australia and permanent migration outcomes (Department of Immigration and Border Protection, pers. comm., 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, the OAIC (2013b) found that over 40% of agencies do not regularly include accompanying metadata for their information holdings; progress in making metadata registries available remains very poor (appendix C). Additionally, lack of standardisation of metadata means the registries that are available are not reciprocally searchable. As noted by the AIHW (sub. DR299):

> To realise the vision for an integrated data system that enhances data availability and use, end users need to know what data exists and how to access it. The first step in making data available for use is comprehensive discoverability. The second step is accessibility which includes easy to use access points and streamlined approval processes. Underpinning this is strong infrastructure and clear, comprehensive and consistent metadata. (p. 1)

There have been steps taken recently to improve the use of data registries and metadata in the Australian public sector. The Australian Government committed Australian Government entities to publish (anonymised) government data, with descriptive metadata**,** through data.gov.au, with the aim of aiding discoverability (Turnbull 2015). However, publication of datasets on data.gov.au has been concentrated among a small number of government agencies (chapter 1).

Lack of discoverability and access in the research sector

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.

There is a plethora of out-of-date, disjointed and difficult-to-find registers of data in the research sector — as well as significant gaps. It is impossible to reuse research data if people do not know what data exists. Failing to reuse publicly funded data means significant duplication is likely to occur as funds are wasted on repeatedly collecting the same data. There is also the opportunity cost of underutilised data and information.

Data standards exist but are not widely used

Coherence and interpretability are necessary for quality data use

Dataset coherence (internal consistency and comparability) and interpretability (information that provides insights into the dataset) are key features of datasets that support quality data analytics. Consistent use of standard definitions and units of measurement are necessary to achieve coherence. Information that could assist interpretation include the variables used, the availability of metadata, concepts, classifications, and measures of

accuracy (ABS 2009). Apart from supporting data usability and the potential for dataset linkage (and thereby increasing the value of data), adherence to standards and data management can foster trust in data (National Computational Infrastructure, sub. 189).

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 DPMC (sub. 20).

The work of the 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. Having a central institution that develops and maintains standards makes the health sector unique in Australia's data landscape. 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.

Difficulties that have arisen in sharing data across jurisdictions in the education sector illustrate the importance of data standards for enabling use of a common platform for data analysis (box 3.6).

Box 3.6 The lack of standards impedes sharing of education data

The education sector is one where substantial volumes of data are collected, but the ability to use this data to conduct meaningful analysis is limited:

> Increasingly there is a focus on providing more data on an increasing number of characteristics. The difficulty is that for some of these characteristics there is no agreed definition, single measure or way of collecting information that allows reporting at the national level.
>
> Reporting at the national level requires a robust and rigorous agreed definition and collection process. Data which is not rigorous or nationally consistent should not be reported publically at this level as this is a potential for data to be presented in misleading or inaccurate formats and interpretations. (Independent Schools Council of Australia, sub. DR257, p. 2)

This issue arises in part due to the different education policies implemented across Australia. For example, the difference in school starting ages means that comparing results across school years is problematic as there are significant age differences between the students (Tasmanian Government 2016).

However, beyond these issues, there is limited implementation of national data standards even where these are applicable. For example, socioeconomic status in education data collections can use two different definitions, and there are numerous definitions for disability (National Catholic Education Commission 2016). While there have been attempts to improve the use of data standards across the education sector, the lack of standardisation continues to impede the sharing and linking of education data across jurisdictions (PC 2016).

Implementing standards can be costly …

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.

Enforcing uniform standards without consultation 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).

Stakeholders have suggested that standardisation can have the unintended effect of delaying 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 recommends that:

> … a single format to 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)

Guidance on the extent to which government agencies should invest in datasets prior to their release is provided in chapter 9.

… but not implementing standards can be more costly

The absence of, or inconsistencies in, data standards and management can significantly 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 2016).

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).

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) because measurement practices can change over time, even for the same data collection, in the absence of standardisation. Given the strong demand for analysis using longitudinal datasets (DPMC 2015), greater adherence to standards has the potential to significantly increase the value of data use.

Inconsistent use of standards is also an issue with private sector data. Whether voluntarily agreed-upon or mandatory, varied use of standards (such as Standard Business Reporting) contributes to difficulties 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.

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.

The Commission considers that common standards across data collections are highly desirable in some cases, such as where the public interest is clear and data merging or sharing offers real long-term benefit. But a sweeping edict to adopt common standards is unlikely to be justifiable and would have a number offsetting negative effects, such as delaying the availability of data or even reducing its ultimate usefulness.

The development of Accredited Release Authorities (chapter 6) — resourced to develop datasets for sharing, integration and release — would provide a structure within which to determine where standards must be rigorously adopted and where this is not a priority, due to net cost or other evident disbenefit.

Lack of skills and dedicated resources also affects usability

Maintaining and maximising a dataset's usefulness require specialised skills and an understanding of its potential uses. However, these skills are in relatively short supply, particularly in the public sector, which can have substantial consequences for both the quality and the beneficial use of data (Dun & Bradstreet, sub. 135; DPMC, sub. 20; Australian Statistics Advisory Council, sub. 25; NSW Government, sub. 80; CSIRO, sub. 161).

Skills development has been identified as a priority by governments seeking to increase the use of their data (DFSI (NSW) 2015; Victorian Government 2016), with several having introduced initiatives to address skills shortages. 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 (APS) that proposes both foundational data literacy training for all APS employees and specialised 'data fellowships' with Data61 (DPMC 2016). DPMC 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 re-use 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.

Legacy IT systems hinder automation of data provision

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 delayed 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. (ANAO 2015, p. 20)

There are many other examples of continued use of decades-old software and hardware systems, and of system incompatibility, across the entire public sector. The Department of Industry, Innovation and Science (sub. 69) noted that data contained in legacy systems poses challenges such as: inability to link data; standardisation issues between data systems; gaps in metadata availability; and inconsistent storage formats — all of which can reduce data quality and cause difficulty in automating data provision.

Inevitably, IT legacy issues also affect the private sector, even though businesses have the benefit of stronger incentives and price signals to guide their IT investments. For instance, the surge of interest in agricultural data access and analysis has seen issues of program interoperability and proprietary data formats arise. Data generated by one type of software can sometimes not be viewed or aggregated in another type of software, making data pooling or sharing between individual farmers difficult (Beef Central 2016a, 2016b; Pawsey 2015). Appendix E contains details of how IT interoperability can limit information transfer and the quality of services provided in both public and private healthcare.

Fragmented collection and limited and sporadic releases detract from data usability

Collection

Data collection in Australia is highly fragmented. What data already exists is often not widely known and so the same (or similar) data gets collected over again by a different sector or different entity. This is the case 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)).

Fragmented data collection leads to the wasteful repetition of surveys and compliance-based reporting, thus placing an unnecessarily large regulatory burden on individuals, services, suppliers and businesses. In some cases, it may result in those parties submitting poor-quality survey data to save time and effort (ABS 2016c).

Several Inquiry participants — including the Australian Statistics Advisory Council (ASAC) (sub. 25; sub. DR237), Leading Age Services Australia (sub. 47), and the Property Council of Australia (sub. DR293) — have indicated that significant duplicated data collection takes place in Australia. 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).

Minimising the duplication of data collection is an important step in improving data quality. Much greater collaboration between governments at all levels is needed.

Release

Despite the public sector open data policies of the Australian Government and all State and Territory governments, other than the Northern Territory (chapter 1), the accessibility of Australia's public sector data remains limited. Reasons for this could include:

- *The focus on releasing datasets as a metric:* in other words, a focus on quantity over quality or impact, comes at the expense of releasing comprehensive datasets. The absolute number of datasets released may not provide a good measure of Australia's open data progress.
- *Australia's federal system*: portals have arisen that contain data that reflects jurisdictional responsibilities rather than a sector (for example, health or education). This contributes to fragmentation of data collection and release.

- *Difficulties in standardising the data or metadata*: poor metadata reduces the discoverability of datasets which, in turn, contributes to their fragmentation and duplication.

3.6 There is scope for more effective use of business and consumer data

The private sector is awake to the value of data and today much of the services sector (which comprises around three quarters of most developed economies' production) is organised around access to, or insights from, digitally collected data (for examples of use of digital data in the services sector see OECD (2015)). Entrepreneurial forces, supported by price signals and the potential for profits, are driving innovative collection and analytics. The private sector typically identifies a purpose for data and determines how best to use it for strategic or financial advantage (or both).

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 to offer discounts and rewards to attract customer data. Much of this activity has developed rapidly with minimal government involvement — although substantial government investment in research (such as wifi in Australia) has greatly contributed to both the technology hardware and the analytics that underpin this transformation (Dobbs, Manyika and Woetzel 2015).

The case for government intervention to ensure one business has the same access to data as another business, a suggestion that has featured in some submissions (notably, in regard to fintech) is weak. A clear link to some public interest benefit (such as with credit reporting) is required.

Industry-wide data sharing arrangements can skew 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; ANZ, sub. 64).

In business to business exchanges, there is, in general, limited cause for policy intervention.

Wider sharing of private sector data could deliver public benefits

Yet some of the data collected by the private sector may also have the potential to deliver significant public benefits if shared more widely than when guided only 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 has previously 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).

Data release that would allow accountability of public spending is another area where a case could be made. The commercial-in-confidence classification is invoked by public sector agencies and private tenderers alike to prevent the release or sharing of information obtained in the course of tender processes. This has been a perpetual and often misleading defence used by governments in the course of making commitments to infrastructure projects (PC 2014). Much the same might be said of data collected on the operation of hospitals (both public and private). The absence of such data prevents public analysis of the efficacy of investments made by taxpayers.

While business data use can, by the evidence we have seen, be generally left to market development (particularly business to business data sharing), the case for improving the lot of consumers — most often the original data source for a new product or service — to participate more effectively in a data-driven world of their own making is nevertheless strong — not least, through building community confidence and trust. Chapter 5 examines this in detail.

4 A way forward: what we must aim at

Key points

- Data is a strategic asset with great potential and should be treated and managed as such.
- Increased data collection and use is an economy-wide tidal wave already upon us. While it moves inexorably forward, we must move policy and practice forward in a similarly comprehensive fashion.
  - Piecemeal reforms would not achieve the same benefits, nor would they enable effective management of data risks.
- The Commission's recommended reforms are designed to create a scalable, risk-based framework for data sharing and release that will support Australia well into the future.
- Key elements of the reforms include:
  - a legislative framework designed to provide a clear and modernised approach to data access and use
  - new rights for consumers to enable them to share in the benefits of data — these demand-focused reforms would drive better choices and more competitive markets
  - establishing a scalable, risk-based institutional framework to allow integration and sharing or release of Australia's data
  - recognising that some datasets are of such significance they should be treated as national assets.
- The reforms are designed to work together to provide mutual assurance and benefit. Splitting off one principal element from another risks a lack of community trust and support if opportunities for individuals are abandoned; and a serious loss to community welfare if public dataset integration, sharing and release is delayed.
- Building and maintaining a social licence is front and centre of our reform package. Social licence will develop if people:
  - have a sound basis for believing in the integrity and accountability of institutions (public and private) managing data
  - have an inalienable ability to *choose* to participate in extracting benefit from data sharing
  - have some control over how their data is used
  - better understand the potential community-wide benefits of data sharing and use.
- Robust institutional and governance arrangements regarding data use build public trust and maintain incentives for better practice over time. Well-resourced and capable institutions are critical.

4.1 The world is changing, and Australia must adapt

Where are we now?

Increased data collection and use is a tidal wave that is already upon us. It will continue to create sweeping changes in the way Australians work and live.

This new paradigm has the potential to enable a range of new benefits and opportunities, including:

- more 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 built and natural environments, 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.

A key aspect of this ubiquitous collecting of all sorts of data is that the value of new data uses, and the ability to link separate datasets together, are increasingly recognised.

Some people and entities have already embraced this new environment. Many of us are sharing data about ourselves on the Internet through social networking sites. And many businesses are surging ahead with innovative products, services and uses of data (chapter 2). Data handling is not just a technical issue; it is a social phenomenon that affects the lives of every Australian, and is changing the way society fundamentally operates.

At the same time, increased data collection and use give rise to additional risks, including data security concerns (chapter 3) — actual and perceived — which need to be carefully managed in order to ensure that the benefits of increased data use are realised, and that public trust is maintained. There is a lack of understanding in parts of the community about how individuals' data is being used, and there are serious concerns about the lack of control that people have over the use of their data. This lack of understanding and these concerns can breed mistrust, regardless of opportunity. These issues are exacerbated by the complexity of the data environment and lack of transparency surrounding data collection and use. Some Inquiry participants have suggested that the data environment is too complex for the average individual to ever understand (The University of Western Australia, sub. DR296).

Data policies in Australia, and elsewhere, have developed in a reactive fashion to the tremendous development of digital technology. They are outdated, and do not adequately manage emerging risks associated with mass digitisation of data and Internet connectivity, nor do they enable benefits from new uses of data to be realised.

The detailed shape of future developments surrounding data is necessarily still unclear, but the timeliness and necessity of the reforms recommended in this Report should not be in doubt. Other countries have demonstrated better reflexes in reacting to these trends and developments. The direction and strength of change to a data-dependent society is now sufficiently apparent that it is necessary to assess whether existing arrangements are fit for purpose.

Failing to adapt means Australia risks being left behind as other nations surf the tidal wave of data. We are too small a country to stem or redirect the global tide of data generation, sharing and use, nor can we exert much control over what large overseas companies that collect Australians' data do with it. As examples discussed in chapter 2 demonstrate, the future (and its prospective benefits and risks) is already upon us — the gate is open and the horses have already bolted. Australia must adapt or be left behind.

Previous chapters have shown that fundamental reform of Australia's data infrastructure is needed. Retaining the status quo is not a viable option. Nor will piecemeal reforms of existing structures go far toward attaining the benefits achievable through a fundamental overhaul of Australia's data sharing framework. The foundations of current regulatory practice in data — the *Privacy Act 1988* (Cth) on the one hand, and literally hundreds of different data collection rules and regulations on the other — are unsuitable for comprehensive reform of the kind that could match the economy-wide sweep of data-driven products and services. Tacking new institutions or investments onto a framework from a pre-digital world will fail.

The Commission has recommended a new Framework in this and subsequent chapters that fundamentally reforms Australia's data infrastructure to allow Australia to benefit from its data holdings. This is not to say that completely open data is always realistic or desirable. Confidence in data use will not be fostered by ill-advised release practice or poor quality de-identification and access arrangements. Moreover, at the margin, perhaps, but substantively for those affected, it could put Australians at real risk of harm. Further, it is evident from an examination of data use opportunities, both public and private, that he who bears the risks of data use is not always he who benefits.

We believe that Australia can 'have its cake and eat it too' when it comes to data availability and use — that the benefits of data can be achieved, *and* the risks associated with data use can be managed, by adopting a risk-based approach. The rest of this chapter discusses how we have designed this new Framework to ensure data access provides benefits for the entire Australian community.

Data today is neither a threat nor a novelty but a strategic asset for individuals and for the economy, and we can and should use it to drive economic and social value and to create a

comparative advantage for Australia. If not, others will (to the extent of their personal advantage) do this anyway, but without the safeguards recommended by the Commission, explicitly designed to manage risk well and allow Australians to benefit from their own data.

4.2 A Framework

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.

Experience in other jurisdictions, such as New Zealand and the United Kingdom (table C.1), 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.

FINDING 4.1

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.

Without a framework in which policy and practice is mutually reinforced in favour of both individual and societal benefit from burgeoning data supply and use, the necessary permission and expectation to dislodge risk aversion in the public sector and desirable confidence in the private sector will remain aspirational (chapter 3).

What we need is a data sharing framework that can scale up, evolve and improve in line with technology. Principles and objectives can be set and institutional arrangements funded judiciously to allow for the modern realities of ubiquitous data collection and use, and to manage risk *directly* rather than by avoidance or transfer of liability to the less informed (generally an individual, acting either as consumer or respondent to governments). The status quo is not addressing opportunity or risk particularly well.

And we need a mechanism in which States and Territories, as holders of vital nationally-relevant data, can feel confident to participate.

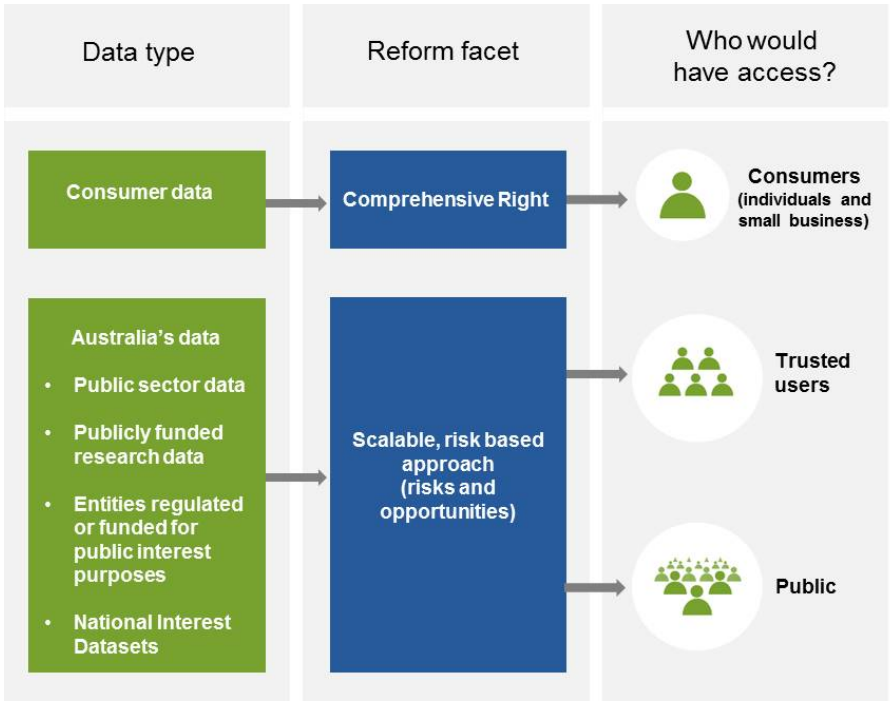
This Report recommends a Framework that is designed to manage these tensions in a way that enables the benefits of data use to be achieved while managing the risks of harm that could occur.

The Commission's recommended Framework has two facets that are designed to work together to create a comprehensive approach to Australia's data management for the future (figure 4.1):

1. a new right that enables both opportunities for active data use by consumers and fundamental reform in Australia's competition policy
2. a structure for data sharing and release that would allow access arrangements to be dialled up or down according to the different risks associated with different types of data, uses and use environments.

Each of these facets recognises the spectrum of risk associated with different types and uses of data and the corresponding need for different controls and approaches to manage that particular risk. These facets broadly remain unchanged from the Draft Report, and were predominantly supported by Inquiry participants, although questions about the detail and implementation of the reforms were raised and are addressed in this final report.

Figure 4.1 **Framework of the recommended approach**



A Comprehensive Right for consumers

The scope to provide consumers with benefits from data use through a Comprehensive Right to digital data held about them is considerable. Increased access and transferability would give individuals more control over the valuable asset they so substantially contribute to, afford individuals more choice about the products and services they consume, and be an avenue to improve market competitiveness. In an era where data collection and use is becoming ubiquitous, it seems counterintuitive that consumers cannot easily benefit from data about themselves. Demand-focused reforms, such as giving consumers the right to transfer data from one provider to another, and making it simple for them to exercise that right, enables better consumer choice and more competitive markets.

Chapter 5 outlines the Commission's recommended consumer data reforms.

A scalable, risk-based institutional Framework

The Commission has recommended a scalable, risk-based institutional Framework be established to share and release Australia's data. The Commission's approach recognises that different types of data have different risks and can generate different benefits — ranging from highly sensitive, identifiable real-time data shared between agencies to improve service delivery, to non-sensitive open data that provides greater transparency on matters of interest to the general public, such as bin collection times and food safety ratings. Our approach allows access arrangements to be dialled up or down according to the nature of risks associated with different types of data, uses and use environments. Where higher risk, disaggregated data is required, the Commission has recommended that it be provided in a secure environment to trusted users.

We also recognise that we should not lose sight of the benefits of data use while managing risks, and that it is difficult to assess the potential benefits of data while it is locked in a vault and able to be used by no one.

Our Framework is designed to facilitate all of these uses in a harmonised way by ensuring entities responsible for integrating, sharing and releasing data have the resourcing, capability and public trust to carry out their functions. While data custodians would still have responsibility for business-as-usual sharing and releasing data, the Commission has recommended that Accredited Release Authorities (ARAs) be established to bear the primary responsibility for curating, linking, sharing and releasing data within a particular sector.

Importantly, ARAs would be *nationally focused,* and would operate free of the mistrust and, indeed, the potentially conflicting objectives that can characterise some of the more fraught relationships in Commonwealth-State data exchanges (or lack thereof). We also recognise that some cross-ARA collaboration including joint standards development may be necessary to facilitate cross-sectoral sharing and integration (for example, health data

may be relevant to educational performance). A National Data Custodian would be established to guide the operation and integrity of the reformed data Framework.

Chapter 6 discusses this scalable institutional model in more detail.

National Interest Datasets should be treated as national assets

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. In particular, the Commission considers — applying the principle of subsidiarity — that there may be some datasets that are of such national significance or national benefit that decisions about them should be led by national effort.

Designating datasets as national interest collections would signify their value as resources collected in the national interest. But National Interest Datasets (NIDs) are not about labelling a collection as important in principle; the purpose of designation would be the scope for additional national benefits in practice.

For instance, significant additional national benefits could come from aggregating data across the States and Territories in health, education, social welfare, child support, aged care, and better linking them with elements of datasets from other fields — the population census, taxation, employment, business ownership, telecommunications, private health insurance or housing. State and Territory datasets in justice, infrastructure, land use and property ownership could be similar candidates for permanent linkages and curation. Subsets of such NIDs could be made available to trusted researchers.

In chapter 7 of this Report the Commission has recommended a framework whereby wider access would be enabled to select NIDs. The intention is to promote the development of a valuable suite of datasets that are only subject to contemporary safeguards and restrictions and are to be treated — and funded — as the valuable strategic assets that they are.

A simplified legislative Framework

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 approaches and practice for data management does not support this intent and reinforces in practice a risk avoidance approach to data release.

The sharing and release of data is currently stymied by a complex, restrictive and conflicting legislative framework that may have been suitable for when it was drafted but is not fit for purpose in a modern data-driven environment. The culture of risk aversion in the public sector is likely amplified by this legislative complexity. The opaque nature of legal obligations — when often combined with what can be quite disparate legal advice — undoubtedly contributes to excessive risk aversion through the focus on lore, as opposed to

law (Australian Institute of Health and Welfare, sub. 162; Queensland Government, sub. 207; Research Australia, sub. 117). Additionally, the existing legislative framework does not explicitly recognise that a risk-based approach — that is, an active approach to risk management — would result in better outcomes than the current risk-avoidant approach.

There is no doubt that a well-designed move beyond current legislation will be needed if serious improvements in data sharing and release are to be delivered and the Australian Government's (and most State and Territory Governments' — appendix C) policies on openness are to be delivered, along with the cultural change to bring about genuine shifts in approach.

The Commission has recommended a new Data Sharing and Release Act be established to simplify the existing legislative framework for data access. For instance, there are certain nationally significant datasets which the Commission intends will, once declared, be subject to the modern requirements contained in the Data Sharing and Release Act rather than the plethora of anachronistic restrictions on sharing and release in the data custodian's legislation. As such, our risk-based approach marks a departure from existing arrangements which do not allow re-use of many datasets — for instance, policy development and improved service delivery can hugely benefit the community, but these uses are infrequently recognised or allowed in legislation compared with sharing for compliance purposes (Privacy Act) or statistical purposes (*Census and Statistics Act 1905* (Cth)). The Commission has recommended this inconsistency be addressed. Chapter 8 outlines the Commission's recommended legislative framework.

4.3 What outcomes is this Framework designed to achieve?

We have designed the Framework to benefit all Australians by:

- putting at the front and centre of any data use the requirement to maintain a social licence
- embedding good risk management practices
- creating good incentives for all participants in the data Framework, including through:
  - transparent and accountable data uses
  - effective and capable institutions
  - instituting robust safeguards
  - promoting coordination and cooperation between data custodians, ARAs, sectors and jurisdictions.

This section outlines elements of our reforms that achieve these objectives.

Maintaining a social licence should be front and centre

Trust is the 'oil that can make the data-use machinery really work' (NZDFF 2014, p. 61).

It is a critical component of Australia's data Framework for both principled and pragmatic reasons:

- *Principled reasons* — Public institutions operate under a social licence that obliges them to discharge their responsibilities in a manner consistent with the public trust placed in them, which includes ensuring that identifying information is managed fairly and respectfully. This is most acutely the case when individuals are *required* to provide information to government to obtain a service or as otherwise required by law.
  - For the private sector, community expectations of how firms will operate can be both express and implied, and many private sector organisations have suffered the consequences of failing to meet community standards (chapter 3).
- *Pragmatic reasons* — Data initiatives and uses will fail if the community is not accepting of the need for data or does not understand (or believe in) the benefits to them from its use. Individuals have shown themselves more willing to share their data when they trust how it is being used and feel that they have some level of control over their data (chapter 3).

For entities seeking to increase access and use of data, the level of trust within the community and the effective communication of benefits derived from data use are important considerations.

To enable this, it is crucial that Australia builds and maintains a genuine social licence for data sharing and use. How we propose to do this is discussed below.

Securing a social licence is important — but it also must be maintained

Social licence builds on trust and credibility between an individual and a certain organisation — be it different levels of government, a company or a research institute — and is usually defined as a 'community's acceptance or approval of … an organisation's operations in their area' (Moore and Niemi 2016; Yates and Horvath 2013). Social licence requires:

- trust in institutions, processes and people
- community expectations about how data use is conducted and regulated being met
- a general belief in the public value of data use (Carter, Laurie and Dixon-Woods 2015).

Trust is difficult to gain, takes time to build, and can be lost in an instant (Conroy et al. 2014). When community expectations are not met and trust is breached, this can result in a withdrawal of explicit or implicit permission to use data. Issues that arise when considering social licence include: data ownership and control (including protecting any cultural value of data), finding suitable methods of benefit sharing, and communicating

societal and individual benefits that might be gained from sharing personal health data (Gluckman 2015).

It is difficult to be prescriptive about how to build social licence — 'like chicken soup, [trust] appears to work somewhat mysteriously' (Shneiderman 2000, p. 57). Nevertheless, there are some obvious, fundamental values that Australia's data Framework should embed in order to maintain a social licence for data sharing and use: a sense of *shared* control; an inalienable right to choose to participate, where possible, in the benefits of data collection and use (such as better service delivery, or personal benefits in areas such as health and finance); and a sound basis for a belief in the accountability and integrity of data collectors and users (box 4.1).

Embedding these fundamental values as design features into Australia's data framework, rather than merely asserting their importance, is central to realising the full value of Australia's data.

Box 4.1 Maintaining a social licence requires establishing shared value, control, trust 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 be educated on who holds their data and how it is used, and be able to exercise control over this, subject to the context in which the data is being used.
- *Trust*: Embedding genuine safeguards into Australia's data framework to assure people their data is being used safely.
- *Genuine accountability*: Data management in Australia should build trust and confidence in the system by being transparent, promoting responsible data stewardship, and safeguarding privacy and data security.

Source: Adapted from the NZ Data Futures Forum (2014)

The Commission's recommended Framework is designed to embed these factors in all elements of its reforms, in support of a social licence.

In chapter 5, the Commission recommends *empowering consumers* to make better use of data. This forms an important part of the social contract — if people can themselves benefit from Australia's new data economy, they are also more likely to support other people using data. There are genuine benefits to be gained from empowering consumers to make better choices. The Murray Financial System Inquiry (2014) and the Harper Competition Policy Review (2015) found that giving consumers more access to their data would give them the power to make better purchasing choices. This in turn would benefit the economy as a whole.

The reforms set out in chapter 5 include significant changes to consumer rights that give individuals an inalienable right to transfer their data to another provider. These

arrangements are intended to apply to both public sector and private sector data — we can find no legitimate basis for distinguishing between the two sectors in affording consumers more power to access and control data for which, in one way or another, they were the source.

In chapter 6, the Commission recognises that *shared value and control* can be provided to the community through greater analytical access to data, recognising that there are significant community benefits from data use that are currently not supported by present-day regulation. The Commission proposes special access arrangements to enable the value of National Interest Datasets to be shared with the community (chapter 7). The Commission also advises a democratically supported approach to development and implementation of the reform package (chapter 10) as a primary way of continuously engaging the community via its representatives.

And embedding *integrity and accountability* in this Framework is fundamental, with a number of recommended reforms to achieve this in chapters 6 and 7, including:

- accrediting institutions that have the capability, resourcing, and primary purpose of implementing a risk-based approach to Australia's data management
- improving awareness and understanding of data security and information management practices
- maximising transparency of purpose, security and access arrangements, particularly where data could identify individuals or decisions on data use are made collectively
- ensuring data holders and users are held genuinely accountable
- requiring well-designed information security practices
- establishing robust safeguards around arrangements for trusted users of data.

And across all of these, we expect and anticipate strong political and agency leadership, with good communication of the overall Framework and its intent. This is vital.

Far from being trade-offs in achieving desired benefits, high levels of trust, inclusion and greater control are critical enablers for value creation through collaborative data sharing. Trust in how institutions manage and use data; confidence that all parts of society will benefit from data-sharing; and greater control by individuals of their data and its use, will all contribute to willing support for data sharing initiatives and active participation by individuals in the data use ecosystem. And if more data is shared and used in trusted, protected and inclusive ways, this then will drive even more value that can, in turn, create more trust, inclusion and control (NZDFF 2014, p. 11). The latent value of data will be unlocked.

FINDING 4.2

Community trust and acceptance will be vital for the implementation of any reforms to Australia's data infrastructure. These can be built through enhancement of consumer rights, genuine safeguards, transparency, and effective management of risk.

Context of use is important

Community expectations and behaviours can change over time. Today's almost ubiquitous sharing of previously 'private' information on social media is a major social shift unrecognised by statute or practice. As governments, we know about it but are far removed from it. It is trite now to say that approaches to information will continue to shift. What was exceptional in 2005 — sexting is an clear example — is not today. The technology predates the social trend.

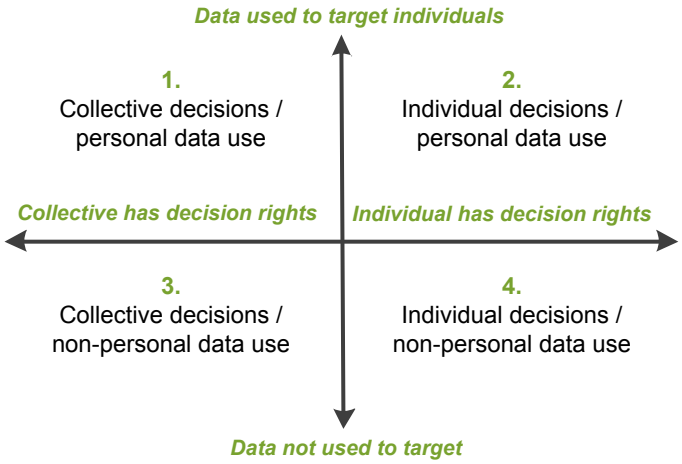
And evidence of the context-specific nature of attitudes to data use is often ignored as simple positions are taken for or against change to public policy. While responses to an Office of the Australian Information Commissioner (OAIC) (2013) survey on community attitudes to privacy showed that 97% of people are concerned about secondary uses of personal information, at the same time over 90% of Australians are apparently willing to share their de-identified health data to advance medical research and improve patient care (Research Australia 2016).

Throughout this report, a number of data use scenarios provided context to our thinking, and to the Framework's conscious effort to move access forward in a way that is cognisant of assessed risks. There are important differences in the way we recommend treating data that identifies an individual versus data that does not.

There are also differences for data management if there is a public interest decision being made about the proposed use versus if the interests affected are solely private. Where there is a clear and compelling public interest, the decisions about personal data may need to sit with the government, as the administrator of public services and the entity democratically positioned to determine and balance public interests (figure 4.2). But to make such judgements, governments need to have fit-for-purpose institutions and processes. Furthermore, where data may be used to understand and influence collective decisions and to deliver public benefits — for instance, to cure a debilitating medical condition — it may be socially detrimental to have the quality of the dataset reduced by individuals opting out.

The need to favour the public interest is to a degree already reflected in Australia's existing legislative approaches which recognise that there are circumstances in which individuals should not be able to limit how their information is used — such as where the information needs to be disclosed to prevent a serious threat to life, health or safety, or where the information is needed for public interest medical research. The Commission's recommended Framework recognises and supports these public interest uses with safeguards (chapters 6, 7, 8).

Figure 4.2 **A framework for choice over data use**



Source: New Zealand Data Futures Forum (2014, p. 18)

Embedding good risk management practices

Australia's existing data use approaches do not manage risk well. Too often data is locked up in the public sector, with risks managed through inaction and avoidance; while the private sector shifts the liability to individuals via Terms and Conditions that are too dense to understand (and therefore could hardly be said to be receiving informed consent).

Neither approach will support community trust, particularly when lax approaches to data security see breaches of credit card or other information on a regular basis. In 2016 alone, it was estimated that, globally, 3.1 billion records were subject to some form of data breach (Morgan 2016).

Dealing with risk through avoidance — 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, security and handling procedures rather than deliberate misuse of data — that is, a failure to properly manage risk (chapter 3). Adopting an explicitly risk-based approach, that requires effective identification, assessment, mitigation and management of risks (box 4.2), is a response that must surely reduce the number of breaches compared with the status quo.

The Commission's Framework is designed to support this. In particular, the Framework requires rigorous assessment of *genuine* risk to inform the development of effective risk management strategies and controls via accredited, capable institutions who are primarily focused on data integration and sharing, and have the capability and resourcing to carry out this role effectively.

Box 4.2 Risk management in data handling and release

For data custodians faced with deciding whether a dataset should be shared or released, a risk management approach provides a useful framework within which to consider risks (such as re-identification of individuals) and opportunities (positive risks) of data dissemination options.

Risk management assesses the activities and actions taken to ensure that an entity is conscious of the risks it faces, makes coordinated and informed decisions in managing those risks, and identifies potential opportunities (Department of Finance 2016).

While there are numerous ways of thinking about risk management, most approaches come down to implementing steps that include:

- identifying risks and opportunities, including the sources of these and the likelihood and significance of particular outcomes;
- assessing which of these can and should be addressed (given community appetite for data opportunities and tolerance of risks);
- determining who is best placed to take responsibility for action (should responsibility rest primarily with the data custodian or the data user, for example);
- taking action to reduce those risks which it is necessary to reduce (recognising that not all risks can or should be eliminated) and secure opportunities where possible;
- reviewing risk management outcomes for improvement over time.

Of particular note is that risk management is not just about the negative risks of data access, but also about positive risks — the opportunities.

There are undoubtedly areas of ambiguity where it is debatable what the 'right' approach is to data access, given trade-offs between the usability of data and confidentialisation. 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 (i.e. context, once again).

The only workable solution in a deeply data-driven environment is to improve the quality of data management practices (and the focus on handling data) in institutions — the practices of researchers, of data collectors and of parties that seek commercially-driven opportunities in accessing data — and adopt an approach that embeds robust risk management practices and guidelines in Australia's data Framework.

Different types of data have different risks

General risk aversion embodied in a reluctance to share or release data can overlook that different types of data have widely different risks. Release of some data poses very low risks, and often it is not in the legitimate sharing or release of data where the main risks arise, but in human error or poor data security practices.

How data is collected (voluntarily or otherwise), the characteristics of data (personal, de-identified or non-sensitive), the environment within which it is used and the purpose of

use, 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.

By applying an explicit risk-based approach to data access, government agencies would be required to clarify, understand and manage the nature of data risks. Risk should be assessed based on both the likelihood, and probable consequence of, data breaches.

Data management frameworks must reflect and address the attendant risks associated with different datasets — different data uses require different echnical solutions and different rules (NZDFF 2014). Such a tailored approach has significant merit, subject to the proviso that adequate safeguards are in place, and that the approach operates with clear guidelines.

The Commission's recommended 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 in submissions to this Inquiry that current frameworks are not working well.

The Commission has also recognised that the proposed *use* of data affects the risk of the proposed disclosure. Sharing data in a secure environment with trusted users carries less attendant risk (but also fewer benefits) than making data public. This is in line with current practice — for instance, guidelines issued by the OAIC (2014). It is this thinking that underpins the distinction we make between *sharing* and *release*:

- *sharing* is providing data to a number of accredited trusted users in a secure environment with safeguards and controls placed on how that data can be used
- *release* is making data public and usable for any purpose (not necessarily at zero charge).

It is important not to lose sight of the benefits of data use when managing the risks.

Where the risks associated with release are significant, the Commission's approach would reflect a sliding scale of release strategies and controls commensurate with the potential risks and benefits of potential release, as outlined in figure 4.1.

We have recommended a Framework that recognises more sensitive data should have more restrictive uses (be *shared* only with trusted users, or not released at all), and less sensitive data should be made more freely available.

- Non-confidential, non-sensitive data, or data that is aggregated and/or heavily de-identified, is suitable for open access or public release.
- More lightly de-identified data may still be at the unit record level. It should not be publicly released as there is still the potential for re-identification, but instead should be provided to trusted users (discussed below) in a secure computing environment.

- There are some limited cases where identifiable data may need to be shared, and indeed, is already shared (such as for service delivery purposes). Recognising the sensitivity of this information, this information should be provided to a more limited set of trusted individuals for defined purposes in a highly secure environment.
- Some data is too confidential or sensitive to be dealt with in a Framework of this kind (for example, that relating to national security). The Commission's reforms are not directed to this data.

Compared with existing practices that often fail to manage risk effectively, we anticipate that introduction of the recommended Framework (with explicit risk management) would both increase the usability of data, and reduce the risk associated with its sharing and use.

Managing the risk of re-identification

De-identification (which we use to include confidentialisation where needed given the context of release) is a technique used to manage disclosure risk, both to protect people's privacy or commercially-sensitive business information, and because privacy and secrecy legislation restrict disclosure of identifiable information. For instance:

- the Privacy Act limits disclosure of information or an opinion about an *individual* that identifies, or could reasonably identify that person
- the *Taxation Administration Act 1953* (Cth) prohibits disclosure of protected information, that is information that was disclosed or obtained under or for the purposes of a law that was a taxation law when the information was disclosed or obtained; and relates to the affairs of an entity (that is, an individual or a business); and identifies, or is reasonably capable of being used to identify, the entity
- the Census and Statistics Act prohibits disclosure of information that could identify entities, and prevents use of the data for other than statistical purposes.

Effectively de-identified data (appendix C covers some advances in this area, and some residual exposures) by contrast is not subject to these legislative restrictions on disclosure as it cannot, by definition, be used to identify an individual or a business.

However, there remains a tension between de-identifying or confidentialising data and providing data at a sufficient level of disaggregation to be useful for answering complex policy and research questions.

Increasingly agencies are recognising that de-identifying data is only one way to manage re-identification risk and the environment the data is provided in can also be relevant to assessing the risk of re-identification (OAIC 2014). Put another way, data that would otherwise be considered 're-identifiable' may be effectively de-identified when placed within a secure environment and made available only to trusted users. The Commission's Framework supports greater use of trusted user models that apply a 'five safes' approach to managing risk (box 4.3) — chapters 6 and 8 contain more details on our recommended approach.

Box 4.3 The five safes model of managing risk

The five safes model 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?

More recent approaches using this model 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 virtual laboratories or trusted access models which provide more risky data in a more safe environment.

- The Secure Unified Research Environment, provided through the Sax Institute (sub. 56), provides a high security environment that facilitates the use of data from different custodians and also the collaboration of researchers working across multiple institutions, including overseas-based researchers.
- A remote access environment is under development at the Australian Bureau of Statistics (ABS). It will allow researchers to 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).
- The Australian Taxation Office (ATO) is developing a longitudinal individuals tax file for policy research. The Australian Longitudinal Individuals File (aLife) will be made available to researchers in the Sax Institute's Secure Unified Research Environment (the SURE) from mid-2017. This approach represents a shift in the ATO's practice from confidentialising and perturbing data inputs to (more lightly) de-identifying data inputs and then checking confidentialised aggregate outputs leaving the secure environment (ATO, sub. DR314).
- The Department of Social Services (DSS) has been working with the Australian Institute of Health and Welfare on a proof-of-concept project that improves data access by researchers. The project has achieved its goal of successfully enabling researchers to access selected social services data remotely, subject to relevant legislative requirements. Access to the data is via a 'curated gateway.' Behind the gateway is de-identified information about individuals to which queries are applied and aggregated answers extracted. However, individual records cannot be extracted. Now successful, the project will be scaled up and developed under a Trusted User Model (DSS, sub. 10). However, there is currently no single clearly defined Trusted User Model across the Commonwealth. Several organisations are considering this broad concept and developing options with the intent of creating a whole-of-government model (DSS, sub. 10).

Source: Desai, Ritchie and Welpton (2016).

Finally, the Commission observes that systems-based approaches to compliance and risk management are gaining recognition as a more effective way of managing risk and meeting legislative obligations (appendix C). Building on this, the Commission has supported a

compliance-based approach to recognising the benefits and managing the risk of data sharing and release (chapter 8).

Creating good incentives

In designing these reforms, our overarching goal is to design a Framework that provides net benefits to the entire community. Likewise, ensuring the recommended approach does not genuinely undermine incentives for businesses, academics and others to continue to collect data and engage in value adding processes, is critical.

Transparent and accountable data uses, and robust safeguards

As noted earlier (box 4.1), transparency and accountability are key components of trust.

In the public sector, genuine accountability requires robust institutional and governance arrangements; transparency and accountability over how the data is used; and clear consequences for breaches (loss of funding, loss of access, loss of accreditation). Transparency and accountability build public trust and help maintain good incentives.

In the private sector, better disclosure obligations would also allow consumers to make an informed choice about which businesses are going to be beneficiaries of their individual data. This can also create incentives for businesses to strengthen their information management practices and transparency as a way of retaining consumer confidence. This is a compelling rationale for introduction of mandatory data breach notification laws (box 4.4), and simplified disclosures to consumers that enable meaningful, informed consent.

Transparency and accountability in Australia's data Framework can be enhanced by:

- making more data available to the public to provide greater transparency over the effective working of government and the performance of publicly funded institutions and services — chapters 6, 7, and 8 outline the Commission's recommended institutional and legal approach to achieve this, including through the creation of National Interest Datasets
- ensuring institutions have the resourcing and capability to undertake their functions — chapter 6 recommends that the National Data Custodian accredit Release Authorities on the basis of their capability and expertise to carry out data curation, integration, sharing and release, and manage risk effectively
- improving public understanding of how data is collected, managed and used through better disclosure that is simple and easy to understand
- clarifying the rules of the game — ensuring institutions know what their obligations are and what decision-making framework they should apply — chapter 8 recommends that the Data Sharing and Release Act establish a clear legal framework for enhanced data

use, with the National Data Custodian issuing system-wide guidance on effective risk management

- establishing a robust data-use ecosystem with agile, responsive institutions and effective and clear rules of the game that provide real safeguards over how data is managed and used — chapter 6 recommends the National Data Custodian be established as the guardian of system integrity
- clearly setting out the consequences for individuals or institutions for breaches, and creating good incentives for compliance — the Data Sharing and Release Act would harmonise penalties and establish robust safeguards.

Box 4.4 Australia's mandatory data breach notification laws

As a result of the recommendations made by the Australian Law Reform Commission in its 2008 Review of Australia's privacy laws and the Parliamentary Joint Committee on Intelligence and Security's *Advisory Report on the Telecommunications (Interception and Access) Amendment (Data Retention) Bill 2014* the Australian Government recently amended the Privacy Act 1988 via the *Privacy Amendment (Notifiable Data Breaches) Act 2017* (Cth).

The amendment requires notice to be provided to affected individuals and the Office of the Australian Information Commissioner when:

- there has been unauthorised access to or disclosure of personal information, or where that information has been lost, and
- there is a likely risk of serious harm to any of the affected individuals as a result of that breach.

However, exceptions exist:

- where remedial action has been taken so the data breach is unlikely to cause serious harm
- where providing notice would prejudice enforcement related activities or be inconsistent with secrecy provisions
- the Office of the Australian Information Commissioner has granted an exemption.

Effective and capable institutions

It is difficult to overstate the importance of having capable institutions with adequate technical skills and resourcing to perform their functions effectively. As discussed in chapter 3, the majority of data breaches can be attributed to poor data security practices, incompetence or human error. Having capable people and institutions and effective data security is vital to building community trust and ensuring that Australia's data framework actually works. It would be naïve to ask an institution to do something for which it simply lacks the technical capability and resourcing.

The Commission's Framework is designed to build and maintain institutional capability, and ensure that institutions are adequately resourced to carry out their functions.

We have recommended that release authorities be accredited to provide sectoral expertise for data sharing and release in a particular sector. However, it would also be short-sighted to assume that institutional capability is a fixed quantity and does not evolve over time. Australia needs a scalable institutional framework that is flexible enough to adapt to changing circumstances and evolves as Australia's data needs and uses evolve and grow. Chapter 6 recommends such an institutional model based on an approach to accrediting release authorities that is outcome focused and flexible enough to adapt to changing circumstances over time.

We intend that the National Data Custodian be the guardian of the integrity of Australia's data Framework.

Wise selection of ways to assess performance is also a key consideration — for example, rewarding institutions for the *number* of datasets they release can lead to a multiplicity of fragmented, low-value datasets being released, allowing agencies to claim they have met their open data target — by the letter at least. By contrast, meaningful and transparent performance reporting with indicators that provide a good metric for the quality of the data released can aid in incentivising good performance. The National Data Custodian would be responsible for reporting on the performance of Australia's new data Framework.

Finally, the Commission's recommended right for consumers to transfer their data between entities is designed to create demand-driven incentives for performance.

Commonwealth, State and Territory cooperation and coordination

Allowing individual data custodians to themselves decide what data should be shared and released can lead to highly fragmented datasets that are of reduced value (chapter 3), particularly when compared, for instance, to the value that can be derived from an integrated, longitudinal dataset. Further, Australia is a federation and different jurisdictions collect different data. 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 (that is, to enable datasets that achieve Commonwealth-State integration). We are recommending a model that allows participation from all the Australian jurisdictions, *not* one that is focused simply on accrediting Australian Government bodies. Our Framework features the volunteering of datasets and the creation of joint governance structures to allow the development of trust between different levels of government.

For the avoidance of doubt, it is *not* proposed that ARAs must be Commonwealth in nature, even if Commonwealth-funded.

Mechanisms for jurisdictional cooperation that we have recommended in subsequent chapters include:

- establishing the National Data Custodian, charged with taking a national approach to liberating Australia's data

- a requirement that Accredited Release Authorities take a national approach to sharing and releasing data, which can include entering into cooperative agreements with data custodians from different jurisdictions so as not to privilege one jurisdiction over another
- establishing a general principle that entities contributing data to integration projects or National Interest Datasets should be able to get the integrated dataset back, if suitable safeguards are applied to prevent re-identification of that data
- a new legislative framework — the Data Sharing and Release Act — designed to create consistent rules of the game across sectors and jurisdictions
- extensive consultation processes and a staged approach to implementation.

We have also recognised that, in practice, issues that the community is strongly interested in, such as health service provision, will not only require data from multiple jurisdictions to assess them effectively, but also from multiple sectors. For example, to meaningfully assess hospital performance, it is necessary to have data from both the private and public sectors. Similarly, the Multi-Agency Data Integration Project (MADIP) is a project coordinated by the Australian Bureau of Statistics that integrates data from the Department of Social Services, the Department of Health the Australian Taxation Office and the Australian Bureau of Statistics itself to create an integrated, longitudinal dataset that provides a valuable ability to research cross-sectoral policy issues (ABS 2016).

The reforms recommended in chapters 5 to 9 are designed to offer an opportunity (and resourcing, via the concept of Accredited Release Authorities) to states and territories to develop datasets that are coordinated within sectors and across sectors to allow complex policy problems to be analysed in a holistic way.

Mechanisms for promoting cross-sectoral cooperation recommended in subsequent chapters include:

- The creation of National Interest Datasets designed to cut across jurisdictions and sectors — these would include private sector data where there is a compelling public interest to do so (for instance, in cross-sectoral policy areas such as schools and hospitals).
- Applying a more consistent approach in government to public benefit datasets held by the private sector. This might include a greater use of contracts and licence conditions to access data, an examination of increased regulatory disclosure requirements, and perhaps the purchasing of data held by the private sector for inclusion in National Interest Datasets.
- Encouraging ARAs to cooperate with each other on cross-sectoral integration projects (for instance, between health and education), and enabling the National Data Custodian to promote this cooperation.

5 New competition policy — a right to use your data

Key points

- Maintaining a social licence for wider data use, both public and private, can be actively supported by offering individuals the opportunity to participate, as firms and governments do, in accessing and re-using their own data.
- Rights to use data will give better outcomes for consumers than ownership: the concept of *your* data always being *your* data suggests a more inalienable right than one of ownership (which can be contracted away or sold). And in any event, consumers do not own their data in Australia.
- A new Comprehensive Right for consumers (including small/medium-sized businesses) would provide greater insight and control for individuals over how data that is collected on them is used. Consumer data for this Right should be defined broadly and with a focus on desired outcomes, but with an opportunity for inclusion of data that is merely imputed to be about the consumer.
- The Comprehensive Right is significant in a policy sense beyond its ability to support a social licence for better data use economy-wide. It may offer the capacity to underpin a new wave of competition policy, similar in its catalytic effect to the Hilmer reforms of the 1990s.
- Under the new Right, all consumers would have a right to obtain a machine-readable copy of their own digital data, provided to them and/or to a nominated third party, such as a potential new service provider.
  - Existing privacy provisions to view and request edits or corrections to personal information would remain, with the new Right also applying these to consumer data.
  - Consumer data would be a joint asset between the individual consumer and the entity holding the data. Exercise of the Right by a consumer would not alter the ability of the initial data holder to retain and keep using the data.
- Determining what data the new Right should apply to will be controversial at times. The coverage of the Right and transfer method should be agreed within each industry through a standard-setting and data-specification process, the outcome of which would require the Australian Competition and Consumer Commission's approval.
  - Data available for transfer must, at a minimum, be sufficient to enable consumers to meaningfully transfer their custom and obtain service from another supplier.
  - Absence of industry agreement would mean the consumer data defaults to a broad definition.
- Participation in comprehensive credit reporting has been low to date and the associated benefits are far from fully realised. A target for participation of 40% of all active credit accounts provided by Australian Securities and Investments Commission-licensed credit providers should be set for 30 June 2017. Legislation to mandate participation should be circulated for consultation by the end of 2017, if the target is not met.

There is a tremendous amount of data provided by and collected on consumers that is used in the delivery and development of products and services. The breadth of data now available, coupled with advancing data analytics, is enabling data holders to apply data-derived insights to deliver better and new products for consumers, and to improve their own competitiveness.

Giving consumers more control over their data would enable them to have more influence in how value is created and extracted from their data. We see significant benefit in this, for consumers clearly, but also more broadly as this will bolster the social licence for data use (by broadening the distribution of benefits) and by encouraging further data use and innovation by business.

While Australians may be generally unaccustomed to having much control at all over how their data is used, it would be unwise in a social licence sense, and costly in a competitive market context, to take consumers for granted as opportunities for data use increase rapidly.

The recommended reforms explicitly intend to allow consumers to benefit from their data and, via this, build trust and confidence — vital prerequisites for maintaining community support, above and beyond simply the service provided today or the public benefit asserted by a government (chapter 4). Individuals providing data to governments and firms must themselves be allowed to actively participate in a data-driven world, 'trading' access to their own data just as data holders in the public and private sector do. Filling the 'data bank' every day then means being able to draw on the account tomorrow.

A useful aspect of data as an asset is that it can be reused over and over again without diminishing its value. A mechanism that draws on this re-use to match personal interest with public interest would be (a rarely available) powerful policy tool. This chapter presents a fundamental reform in competition policy through enabling opportunities for active re-use by consumers of their data.

While the recommended new Comprehensive Right for consumers has some features — an ability to view information and request edits — that mirror those in privacy provisions, it is not a subset or a cousin of privacy law. It is about consumers, and their continuing willingness to supply data — a crucial input to business, research and public policy.

Other countries — such as the United Kingdom and European Union members — have already recognised and responded to the important role of consumers in the cycle of data collection and use, albeit in limited forms.

The reforms set out below include significant changes to access arrangements for consumers. These access arrangements are intended to apply to both public sector and private sector data. As noted in chapter 4, we see no legitimate basis for distinguishing between these in affording consumers more power to access and control data (or information more generally) on themselves.

5.1 Why give consumers new rights?

Opportunities arise when consumers have more control over their data

The opportunities for individuals to benefit from greater use of data are detailed in chapter 2. Discussed more specifically here are the opportunities available to individuals from accessing their own data in their roles as consumers. We are also recommending that these opportunities be extended to small and medium businesses (SMEs), when acting as consumers.

The benefits of giving consumers more control over their own data are — as with much novel public policy — much harder to demonstrate than the costs. The beneficiaries themselves are potentially many but are scattered. As such, they do not present the ready picture that businesses and agencies, in presenting cost estimates and compliance burdens, do. Such analysis would need to be performed on the final scope of the Comprehensive Right, once the precise structure is defined (before which substantiation of claims to both cost and benefit would not be possible).

Giving consumers more rights over their data could:

- benefit them directly, where being able to share data with third parties enables:
  - products and services to be better tailored to their needs or circumstances, or where it allows them to identify which existing products and services are most suitable — for instance, Energy Tailors already uses energy consumption data to advise consumers on which energy plans would be cheapest based on their usage patterns
  - access to new products and services — a range of smartphone apps already use financial transactions data to help consumers invest and better manage their money
  - reductions in everyday 'costs' — Google Maps is a good example of a service that incorporates real-time data to lower costs for its consumers (in this case, the time and fuel costs of map users not taking the most direct or quickest route to a destination)
- lead to beneficial impacts on competition and innovation by:
  - lowering barriers to entry (from uneven data availability) for new market entrants — although market power afforded by data holdings often proves transitory unless entrenched by licensing regulation or intellectual property protections
  - promoting linked services and interoperability of technology, and providing a knowledge basis for innovation
- address the market failure of asymmetric information — the role of credit reporting in improving pricing of risk and the allocation of credit is an illustrative example (assessed in more detail in section 5.5 and appendix F)
- lead to greater trust and acceptance among the broader community of data availability and use — that is, enhance social licence.

Existing consumer rights to data are limited

In Australia, individuals can currently access information about themselves under privacy and freedom of information legislation (appendix D), such as view and request a copy of information about themselves. A range of exceptions apply to this (as set out in Australian Privacy Principle 12), including where access to the information would be prohibited by another 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. There are a number of specific barriers to consumers being able to access and share data, which in turn materially limits their ability to benefit from their data (as noted in chapter 3).

- Consumers are often *unaware of what information is held on them* by businesses or governments, or how to go about accessing it. The complexity of the privacy policies and the terms and conditions of many businesses — which are dense, multi-layered and poorly designed — compounds this, and makes it near on impossible for consumers to opt out of a behind-the-screen data collection process, or to exercise control over their data in any meaningful sense.
- In many cases consumer consent is not meaningfully given — consumers having 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.
- To share their data today, 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. There is generally no obligation on any data collector to provide data in a form useful to consumers or to a business that a consumer may wish to share the data with (for example, a competitor — in order to seek a better offer of service — or a third party that offers advisory services). This is a serious potential impediment to the ready transfer of information.
- Control of data by consumers is limited by the fact that, in most industries, consumers are unable to authorise trusted third parties to access their personal information directly from their service provider. In all cases where data sharing is currently available in Australia, the recipient of the data is the consumer.

Data access for consumers in some sectors has improved

The Australian Energy Market Commission changed the National Electricity Rules in 2014 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 access for third parties. 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).

Also in 2014, amendments to the Privacy Act expanded the scope of consumer credit reporting (section 5.5).

The My Health Record system has had some recent success, but implementation has been difficult because of poor incentives to participate and apparent reluctance within the medical profession (appendix E). There appears to be significant unmet potential for more consumer-oriented data management in the health sector.

Notwithstanding these developments, there is no general regulatory or policy presumption in favour of information-sharing with individuals or for an individual to generate direct benefits from their data. The current regulatory structure adopts an essentially defensive outlook on data. Consequently, while consumers can see some of their data, the degree of difficulty in translating this to any useful action is high.

Recent developments overseas

There have been some moves overseas to boost consumer control over information about themselves.

The *Enterprise and Regulatory Reform Act 2013* (UK) enables the UK Government to compel businesses to release consumer information if they do not do so voluntarily, or to require that they transfer it to an authorised party — in the electricity, financial services and telecommunications sectors — on the consumer's behalf. Prior to introduction of this Act, progress on midata — the UK's program to encourage businesses in selected sectors to allow customers to download data about their use of specific products — was slow.

The UK Government has stated it will use this power if it considers insufficient progress has been made under the voluntary approach to support the aims of enhancing empowerment, innovation and competitive markets. A review of the midata program (DBIS (UK) 2014) found that progress had been sufficient to not require the legislative power of compulsion to be exercised; although even today the system is not fully effective. As part of a review into financial services, the Competition and Markets Authority, however, found that larger banks did not have to work hard enough to win new customers, and has announced that it will compel selected banks to facilitate (limited) data transfer (via Application Programming Interfaces (APIs)) (CMA (UK) 2016).

The EU has introduced a right to data portability under the EU General Data Protection Regulation 2016/679 (EU GDPR), effective from 2018. The regulation provides that the consumer has the right to transmit the EU's version of 'personal data' to another body without hindrance from the data holder. Transfer may occur as long as it is agreed to be technically feasible and the data is available. 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 (article 68, GDPR). This right is 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.

Both are thus limited in their scope.

Access rights — 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 (Sam Toady, 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 case overseas too, although copyright and various other laws can ascribe various rights to parties — databases and medical records can be covered by copyright, for example.

In the view of some very large users of consumer data (such as the Digital Industry Group Inc (DIGI), sub. DR326), individuals already own their data, but the concept of ownership in this context is quite nebulous. If a consumer cannot trade with their data, then it is hardly accurate to contend there is data ownership. Alternatively, if the trade that occurs is only with the incumbent provider — that is, the data cannot be transferred effectively to obtain a better or a competing service — this too is nebulous. Such narrow control over data may well over time only serve to decrease the willingness of consumers to provide data.

The Commission does not consider that the legal position on ownership should be changed — people should not be given ownership over 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 consultation notes because they have created them by virtue of their skill or expertise. 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 of data 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 in practice.

By contrast, thinking about individual's information in the context of consumer 'rights', as many other countries do (United States, United Kingdom, members of the European Union, New Zealand, and Canada), avoids many of these problems. 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 or emerging new opportunities for beneficial data use (Xamax, sub. 3; ANZ, sub. 64; Australian Bankers' Association, sub. 93; Dun & Bradstreet, sub. 135). This

conclusion is consistent with that reached by the Australian Law Reform Commission (ALRC) (2003) in their consideration of ownership of genetic information — that data rights give a more enduring and workable outcome for individuals.

When talking about 'rights', the Commission considers that it is preferable to implement new consumer rights in a way that is consistent with Australia's existing legal frameworks — for instance, in line with the Australian Consumer Law and the Privacy Act 1988 (Cth), which rely on regulators to drive compliance with these consumer protections. This is in contrast to the US approach, which grants individuals a right to sue, and relies on individuals to enforce it. And since the concept here is about consumers, the appropriate regulator would be the Australian Competition and Consumer Commission (ACCC).

5.2 A Comprehensive Right for consumers

The Commission identified five broad areas where consumers could directly benefit from greater control of their data (whether held by a private or public sector data holder), by giving them the right to:

- access a copy of data
  - provided directly by the consumer
  - collected in the course of other actions (and including administrative datasets) and identifiable to that consumer (whether aggregated or not)
  - held by the data holder even though created by others — for example through screen-scraping or tracking, purchase of data about a consumer, or re-identification
- request edits or corrections for reasons of accuracy
- direct holders of such data to copy the data in machine-readable form, either to the consumer directly or to a nominated third party (the 'transfer right')
- be informed about the trade of any element of this data to third parties
- be advised of disclosures of data to third parties.

These five new rights to information defined as 'consumer data' make up the Comprehensive Right (figure 5.1). It is comprehensive because it is intended to apply across the economy, to all data holding entities — whether private or public sector.

This Right is distinct from the existing and important rights of individuals to access personal information, as set out in privacy legislation. It does not replace them.

While privacy is one aspect of how to manage data use *by others*, the proposed Comprehensive Right focuses squarely on enabling consumers *themselves* to better control and leverage the use and benefits of their data. The opportunity cost of not enabling this consumer control and benefit will surely increase if data continues its inexorable rise as a driver of services.

Figure 5.1 **The Comprehensive Right**

Individuals who qualify as consumers for the purposes of the Comprehensive Right would include single persons, family groups or other groups resident at a single address in the data holder's dataset, and any entity with an Australian Business Number (ABN) and turnover of $3 million per annum or less.[16]

The scope of businesses able to exercise rights as consumers under the Comprehensive Right would be considerably narrower than the scope of 'consumers' under Australian consumer law. This is intentional. We do not see the Comprehensive Right as a vehicle by which large businesses would improve their access to data (other recommended reforms better facilitate this); nor do we see significant additional benefits in improved competition or innovation with data from allowing large businesses a Comprehensive Right to data.

16 This threshold is consistent with that used for provision of personal information under the Privacy Act. However, under the Comprehensive Right, some SMEs will be providers of data to consumers, as well as consumers seeking greater access to their own data.

All businesses, not-for-profits and government agencies should be required to comply with the provisions of the new Comprehensive Right, although as noted in chapter 1, security data (such as that pertaining to military organisations, much police investigatory work, or secret intelligence services) would be excluded from scope.

RECOMMENDATION 5.1

Consumer data must be provided on request to consumers or directly to a designated third party in order to exercise a number of rights, summarised as the Comprehensive Right to access and use digital data. This Comprehensive Right would enable consumers to:

- share in perpetuity joint access to and use of their consumer data with the data holder
- receive a copy of their consumer data
- request edits or corrections to it for reasons of accuracy
- be informed of the trade or other disclosure of consumer data to third parties
- direct data holders to transfer data in machine-readable form, either to the individual or to a nominated third party.

Where a transfer is requested outside of an industry (such as from a medical service provider to an insurance provider) and the agreed scope of consumer data is different in the source industry and the destination industry, the scope that applies would be that of the data sender.

Support for a Comprehensive Right for consumers is widespread

A range of Inquiry participants — across a wide range of sectors — expressed support for a Comprehensive Right, noting it could:

- promote greater transparency (Research Industry Council of Australia, sub. DR263) and provide much needed legal clarity (AgriDigital, sub. DR249)
- provide consumers with an enhanced ability to choose between different products and/or service providers (or utilise services that do so on their behalf) (Australian Automotive Association, sub. DR262; Consumer Action Law Centre, sub. DR308; Choice, sub. DR328)
- increase competition in a range of markets (Choice, sub. DR328)
- foster innovation and the development of new products and services (Envestnet Yodlee, sub. DR280; Australian Food and Grocery Council, sub. DR284)
- contribute to greater public trust in how data is collected and used (Insurance Council of Australia, sub. DR318).

In light of this widespread support, we have focused in the Final Report on areas of concern raised by Inquiry participants and on developing further guidance about how a Comprehensive Right could be implemented effectively. Much of the analysis and evidence in the Draft Report is not repeated here; but nevertheless remains relevant.

What data should the Comprehensive Right apply to?

The establishment of a Comprehensive Right naturally raises questions about the scope of data that it would apply to — in other words, how should consumer data be defined?

We envisage a consumer data right defined by outcome — it should be that which is sufficient to afford consumers greater choice in services and spur competitive pressures to this end.

In the Draft Report, the Commission put forward a broad outcome-focused definition of consumer data that would include digital personal information, information provided by a consumer either directly (such as through transactions) or indirectly (such as through online or Internet-connected activity) and all other data necessary for the transfer of a consumer's activities to a third party.

Energy Consumers Australia (sub. DR316) suggested that this definition did not go far enough and argued that there was a need for 'near real time' data. Other participants, however, were concerned about the broad scope of new rights that would be afforded to consumers, and in particular, the breadth of data covered by the Right.

Participant concerns about the implications for compliance costs of broad data coverage

The *additional* compliance costs to business and government agencies of the Comprehensive Right compared to the status quo (individuals' access rights under privacy legislation) would most likely be primarily associated with the new right for consumers to transfer their data to a third party, but in practice this would depend on the precise definition of consumer data for a given industry.

Some Inquiry participants argued that a broad definition of consumer data would impose high compliance costs on businesses with little incremental benefit for consumers. It was argued that limiting the definition of consumer data to 'transaction data' would still allow consumers to benefit from the sharing of this data but would in practice mean less costly upgrades to IT systems, which may be necessary for some businesses to transfer data (AGL Energy, sub. DR251). Telstra (sub. DR312) suggested that basing the definition of consumer data on clearly defined 'transaction data' would: be clearer, simpler and more practical than the Commission's proposed approach; and entail lower compliance costs for entities with large customer bases (such as Telstra).

Participant concerns about impacts on business incentives

Another source of concern for some participants was that a broad definition of consumer data that included data subject to value adding (such as analytical processes) might reduce business incentives to invest in such activities and use data in innovative ways (ADMA, sub. DR275). Telstra (sub. DR312), for example, argued for a definition of consumer data that avoided encroachment on proprietary rights to data that has had value added to it. The Business Council of Australia (sub. DR317) had similar concerns.

Even if desirable, it is hard to imagine how a concept like value added is able to be translated into legally effective language other than through a process of authorised negotiation.

In a similar vein, some participants suggested that holders of consumer data may be motivated to de-identify data in order to avoid the need to make it available to consumers (RIM — Professionals Australasia, sub. DR227; Australian Information Industry Association, sub. DR244; and the Australian Automotive Association, sub. DR262).

Broadly speaking, the intent of the Comprehensive Right is to facilitate enhanced competition and innovation. Taking account of the substance, if not the literality, of these participant comments, it is our intention that the new Right would operate within the bounds of Australia's intellectual property right arrangements, as described in chapter 1. For clarity, however, we do not consider that data that has been cleansed of errors, made better through simple statistical means such as aggregated or averaged for each consumer but otherwise unaltered, or made machine-readable could singly or collectively be construed to be value added (as some might argue).

Nor do we consider it likely that many businesses or agencies would be motivated to desecrate the internal value of their data assets by destroying identifiable information of consumers while it remained necessary for other purposes. And if not necessary, we note that privacy rules seek to see it destroyed anyway, thus rendering any claim of adverse action moot.

Nevertheless, to the extent such incentives do exist, they could be handled via standards established in conjunction with the data-specification process (discussed below).

Participant concerns about duplication of data access rights

Other participants argued that having a definition of consumer data might replicate the definition of personal data for individuals under the Privacy Act, and lead to added complexity with little consumer benefit (Office of the Australian Information Commissioner, sub. DR236; and the Office of NSW Privacy Commissioner, sub. DR268).

This Report has sought to clarify that, while there may be overlap for some consumers (for individuals but not for SMEs, which are not currently covered by the Privacy Act) and for some data, there is unlikely to be confusion about what is covered under each right.

The Australian Privacy Principles would continue to apply to personal information of individuals. Separately, industry-specific definitions of the data that consumers would need to switch their custom from one provider to another would apply to the exercise of the Comprehensive Right. If a firm or agency receives a *transfer* request, it would be clear that this is not through the channel of privacy legislation — this is evident because privacy legislation has no such right to request the transfer of data to third parties. Duplication of rights is thus not a relevant claim with regards to the right to transfer data.

The incentives for industry to participate in the process for determining the definition of consumer data (discussed below) are clear and deliberate.

The process that allows industry groups to define consumer data for their sector may well reduce, if not eliminate, confusion about overlap between consumer data and personal information (as defined in the Privacy Act). A firm, or industry, can do that by deciding to offer all personal information, plus data specified through the data-specification process, as consumer data. In this way, taking personal information as a 'base' ensures that if a request for information is of an uncertain nature (that is, it is not a transfer request) a firm can easily satisfy it by providing the 'personal plus' response (once satisfied of the individual's bona fides, which is a common requirement of both rights). In other words, businesses and agencies can simplify the process by establishing a single universal 'batch' of data to give to consumers, regardless of whether the request originated under privacy law or the Comprehensive Right.

We remain of the view that including both access and edit powers as part of the Comprehensive Right would enhance outcomes for consumers and help build the level of community trust.

Rights to opt out, however, are no longer being included in the new Right. Across the spectrum, submissions argued that there would be a need for various exceptions and qualifications to such a right, to the point that we can no longer in good faith suggest that this is applicable *comprehensively* — that is, to all sectors and data holders (as the other new rights are).

There are clear benefits to consumers in a broader definition

There is a range of data, beyond transactional data, that could be beneficial to consumers if they were able to easily share it with other parties. It would take pages to be definitive about all of the specific data types this might include, but there are some notable examples.

- Health records contain a range of information, potentially including test results, scans and x-rays, prescribed medications and notes transcribed by doctors. Providing individuals with the ability to transfer their health record (between GPs, specialists and hospitals) could improve the information available to doctors, lead to better diagnoses and result in the individual being subjected to fewer tests.
- Financial services providers are required, under various legislation, to verify the identity of consumers prior to the provision of financial services (so-called 'know your

customer'). The ability for consumers to access innovative new financial services from Fintech firms, many of which operate online only, may be enhanced if they were able to securely share data that could be used to satisfy know your customer requirements — this data would, in essence, allow the Fintech firm to rely on identity verification performed by other firms.

- Data that provides insights about an individual's lifestyle, but which is not captured in transactional data, could also be useful to consumers. For example, an individual might choose to share their data from wearable fitness trackers with a life or health insurer if it would demonstrate that they lead an active lifestyle and are a relatively low risk from an insurance underwriting standpoint.
- The use of telematics devices in vehicles — which can record data related to location, acceleration, deceleration and travel speeds — is a valuable resource for insurers seeking to price risk for individual drivers. Indeed, some insurance companies have begun deploying telematics devices into vehicles driven by policyholders (chapter 2; Allianz Australia nd). An ability for those policyholders to transfer data captured by such devices to a different insurance company could spur competitive pressures, and ultimately lead to lower premiums (for safe drivers at least).

Start with a broad definition …

The overarching outcome of the consumer data definition — which should be legislated — is the scope of consumer data that is sufficient to enable the provision of a competing or complementary service or product. Broadly, consumer data would be digital data, provided in machine-readable format, that is:

- held by a product or service provider and
- identified with a consumer and
- associated with a product or service provided to that consumer.

In other words, consumer data is the type of data held on a consumer or SME that a competing or complementary business would themselves need, and reasonably expect to obtain, in order to make a reasonable offer for a consumer's patronage. The benefits to competition and innovation would flow from enabling switching by consumers.

At its broadest level, the definition of consumer data would include, in machine-readable form, all:

- personal information (as defined in the *Privacy Act 1988* (Cth)) that is in digital form
- files posted online by the consumer
- data created from consumers' online transactions, Internet-connected activity or digital devices
- data purchased or obtained from a third party that is about the identified consumer
- other data associated with transactions or activity that is held in digital form and relevant to the transfer of data to a nominated third party.

As noted above, all digital data that is, or has become, reasonably identifiable with a consumer would be subject to the Comprehensive Right. The Right would not require entities with paper records to digitise these in order to supply consumer data. However, entities are increasingly holding data in digital formats and it is inevitable that the value of the Comprehensive Right for consumers would grow over time. And where entities, for their own purposes, do convert legacy records to digital form, such information would potentially become consumer data.

'Default' protocols around security should also be established by Government to ensure that all parties transferring and accessing data under the Comprehensive Right do so in a way that manages the risks associated with the transfer of data under the Comprehensive Right.

… that can be further refined by industry

Information provided by Inquiry participants highlights that what constitutes a reasonable definition of consumer data, to *achieve the intended outcome*, would necessarily vary between industries or sectors. For instance, the Communications Alliance and the Australian Mobile Telecommunications Association noted that the different costs and benefits of the Comprehensive Right would have across sectors warranted a sectoral approach:

> [W]e also note that the costs and benefits of the Comprehensive Right may vary considerably across different industries and, therefore, ought to be assessed accordingly rather than using a rather 'crude' economy-wide approach. These costs, and the effect of such costs, will also vary depending on the size of the organisation that holds the data. In particular, small businesses and start-ups might find the costs of compliance prohibitive. (sub. DR250, p. 4)

The key question is how to develop a definition for different industries.

A process for industry to determine what 'consumer data' should comprise

The Commission recommends that an industry data-specification process should be established to review and reach agreement on the exact definition of consumer data for that industry (based on an understanding of the relevance of specific types of data for provision of a consumer product or service in that industry). This would allow the scope of the Right to move with the march of technology/data, as is highly desirable for any regulation. And, in the absence of industry agreement, the broadest level definition of consumer data would apply as the default, for that industry.

Some submissions argued that transaction data, generated in the direct course of the relationship between the consumer and the data custodian, should be sufficient alone to satisfy consumer objectives under these reforms. Often this would be data that a consumer could compile themselves if they were fastidious record keepers and insisted on a copy of every interaction between themselves and an organisation (including records such as receipts, charges, usage and medical test results).

The Commission has serious doubts, but recognises that in a few industries transactions data may be sufficient to achieve a competitive outcome for a consumer — if so, this would be the demonstrated outcome (to the ACCC's satisfaction) of negotiations between businesses (who would likely be both incumbent suppliers and enthusiastic recipients of consumer data) and consumers or consumer representatives.

But with the exhaustive depth and breadth of data collection today and the heightened relevance of knowing your customer to achieve the outcome specified in our approach, the likelihood of this already seems limited. As noted above, there are clear benefits to consumers and to market competition from a broader definition of consumer data.

Nevertheless, the flexibility inherent in an outcome-based definition and data-specification process allows for industry or sectoral variations to the way the definition of data is satisfied, rather than seeking black letter precision where none is likely in a legal sense (particularly since the data necessary to achieve the intended outcomes will change over time, at least for some industries). It also avoids mandating provision of data that is of no use to consumers, for transfer of their custom, in a particular industry.

Importantly, such an industry-based approach would not preclude any businesses within an industry from offering a 'gold standard' in terms of the breadth or quality of data it provides to a consumer — for instance, greater detail in the data or enhanced formatting and presentation — and so differentiate themselves from their competitors. In many industries, competitive pressures are likely to drive such actions.

Where requested by industry or the ACCC, an officer from an agency of the Treasury portfolio could be a member of the industry data-specification group (this might allow, for example, the Australian Securities and Investment Commission (ASIC) or the Reserve Bank of Australia (RBA) to participate in the banking sector data-specification process). In industries such as education, health and law enforcement (that include a mix of public and private sector service provision), relevant government agencies should be active participants in the industry data-specification.

In addition to defining the scope of consumer data, data-specification agreements should also articulate: transfer mechanisms, including security protocols, to ensure that data handling is practical and robust to technology updates; and the requirements necessary to authenticate a consumer request prior to any transfer. This means that industries would be able to tailor (with agreement) data security requirements for their particular sector — such an approach recognises that data sharable in each sector is likely to have different risks, and therefore different approaches to managing these risks. In practice, this means that parties would have responsibility only for the data they hold, not for data they have transferred to a third party at the consumer's direction — that responsibility should lie with the third party (section 5.3).

Data-specification agreements should be registered with the ACCC

The industry-agreed definition of consumer data, including agreed transfer mechanisms and security protocols, should be registered with the ACCC, who would assess whether the definition would be sufficient to achieve the intended outcomes (in terms of what consumers should be able to obtain from the new Right).

In the absence of industry agreement on the definition of consumer data, *all* data included in the broad level default definition above would be deemed relevant to a consumer's request for their data from an entity in that industry. The ACCC would determine, through the presence or absence of registered industry data-specification, what level of access a consumer was entitled to should a dispute arise. And in the absence of an industry agreement, security arrangements to be applied for that industry would be the default security protocols established by Government.

Given that the right would apply across the economy, sectors would need to be prioritised in registering their definition of consumer data with the ACCC, who should be allowed to offer interim approval where a data-specification is ready but there are higher priority sectors drawing the regulator's attention.

To illustrate the general approach, the banking sector (including fintech) is an obvious high priority sector. By comparison, the video games industry is perhaps an obvious candidate for interim registration, should it choose to develop its definition of consumer data at an early stage after passage of the legislation.

The make-up of industry groups is an aspect that might be contestable, should incumbents wish to limit which firms are able to participate in the data-specification process. Such issues could be resolved through the active involvement, in the data-specification process, of relevant Australian Government agencies. And, the ACCC would have the ultimate say given their role in approving industry-developed definitions of consumer data.

Incentives for industry agreement

It is evident that one size will not fit all — attempts to develop a single solution would see the Comprehensive Right reduced to an ineffective lowest common denominator. And we see it as a necessary feature of the new Right (at least initially) to get industry-wide buy-in and to allow the definition of consumer data to be broader in those sectors where it makes sense.

The Commission has in mind a period of 12 months after passage of legislation before the Comprehensive Right comes into effect, to give data holders the ability to adapt systems, and for the data-specification and standards-setting process to work. Government may, in developing an Implementation Plan for this Report as a whole, allow for variation perhaps by prioritising its own presence on those data-specification working groups where greatest consumer interest is evident.

Starting with a broad default definition — that would apply in the absence of agreement on an alternative — would create a serious incentive for businesses to engage constructively in the data-specification process. There is a risk, however, that parties wanting the broadest possible definition of consumer data to apply (irrespective of the value of such an definition) may simply insist on adoption of the default definition, and choose to wait out the 12 month period before the Right comes into effect.

To counter this possibility, we have recommended specific exclusions in the default definition of consumer data for some types of data (discussed further below). One of these data types — imputed data — could, however, be included in the industry-developed definition via negotiations in the data-specification process.

Imputed data

It is relatively common for a firm (less likely a government agency) to have some data that is sourced from or otherwise identifiable with a specific consumer, and some data that is imputed — that is, recorded as a characteristic of a consumer by that entity, but which is not collected directly from the consumer nor considered to be identifiable data. For example, this may include the likelihood of a person having difficulty repaying their debt based on characteristics that could be ascribed to them — such as whether they reside in a particular location, their age or marital status, the type of vehicle they own, and whether they have an insurance policy (these are illustrative examples only; in reality, the Commission understands that data imputation is often much more sophisticated).

The basic principle would be that when multiple data sources are transformed to an extent that it is merely probable (but not certain) that a characteristic is associated with an individual consumer, then this data would most likely be proprietary information of the data holder. Determining whether such transformed data was sourced at some time in the past from an individual would be an unreasonable burden to place on a data holder.

While such data may be very attractive to a competitor or complementary service provider, it is not and was not at any point an individual's data and so would not be considered consumer data, as such. Another party seeking access to it should invest accordingly, as others have done before them.

Whether data held by a company is imputed or not will vary considerably between industries and while such data would not be included in the default definition of consumer data, it is reasonable to expect that it would be a point of strong negotiation between industry participants in developing their data-specification agreement. In some instances, firms may be willing to include imputed data in the definition of consumer data for their industry in return for access to different data held by other firms in the same industry.

The source of such data may also not always be clear to the consumer. We expect that where the Right is being exercised, other industry members to whom a consumer is trying to transfer data would have sufficient knowledge to determine if data is being unreasonably

withheld by a data holder who is claiming it to be imputed when it was not. Thus again, incentives in the data-specification process should work in consumers' favour.

Purchased data — included if it identifies an individual and is legally transferable

The broad default definition of consumer data includes data that a business has indirectly acquired about a consumer, such as from a data broker. Such data may sometimes be the intellectual property of another entity and be contractually non-transferable by the data holder.

It might be within the bounds of legal possibility to consider requiring that data be provided, at the consumer's request, regardless of intellectual property. However, this is an extreme step and not justified by the general intent of the Right — which is centred on data as provided to the entity by the consumer.

Under existing privacy law, data holders might not be required to provide information to an individual consumer where that information was purchased by the entity under contractual arrangements that prohibit the entity from providing it to any other party as long as it remains identified with the individual.

Where the consumer is the initial source of the data and it remains identified with them, however, such data should in principle be supplied as part of consumer data in response to a consumer requesting access to it. It should not matter whether the entity collected the information themselves or sourced it from another party.

The significance of purchased data would vary by industry and service. As would, under the Commission's approach, its inclusion in an industry-specific definition of consumer data. There should thus be no *legislated* exclusion of purchased data from consideration in the data-specification process; but the in-principle position noted above should guide data-specification processes.

To the extent that purchased data qualifies as consumer data, the cost of obtaining the data supplied by the consumer should explicitly allow the recovery of that purchase cost.

Data collected for regulatory enforcement — excluded if specified in legislation

Definition aside, several submissions argued that exclusions would need to be put in place for certain types of data related to consumers, such as data collected and provided to regulators for enforcement activities. For instance, the National Australia Bank submitted:

> NAB believes that exceptions should also be included for data collected and provided to regulators such as AUSTRAC and ASIC, where provision of the data to the consumer could prejudice enforcement related activities. (sub. DR270, p. 9)

We consider that such exceptions are likely to be readily dealt with in the drafting process for the new legislation.

Circumstances where consumers are not capable of exercising the right

How the Comprehensive Right might operate in situations of personal incapacity or disability will need to be clarified. There are also questions about the ability of parents to exercise such a right on behalf of their children, and indeed whether they should be able to and when (for example, the circumstances in which parents should be able to access and share the medical records of their teenage children).

The Commission notes that there is an existing legal framework for governing individual rights, and the rights of parents, guardians and carers, in such circumstances. Power of attorney is one such legal framework.

As a matter of principle, the operation of the Comprehensive Right should recognise the intent of those legislative frameworks. Moreover, the Commission has not been presented with any evidence that the existing legal frameworks would be unsatisfactory for managing consumer data access and transfer rights.

Consumer right to view data and request corrections

In its Draft Report, aside from the right to transfer to third parties, the Commission also proposed that consumers would have a similar ability to that available under privacy legislation to view information a business or government agency holds on them and to request edits or corrections to improve its accuracy.

The reasoning for replicating these two rights in the new Comprehensive Right, rather than creating only the one new right (transfer) and relying on the Privacy Act for the remaining two, is as follows.

At the highest level, consumers in the Commission's recommendations (in all three rights) would include SMEs. SMEs are not currently covered by privacy rules (in terms of being able to access data). It would be perverse and a failure of policy to apparently offer individuals who contracted for a service in a business name rather than personally the ability to seek access to their data, only to see it denied because the mechanism that legally delivers such a right (privacy legislation) does not cover them.

That ensures that we need to legislate, in one manner or another. It may appear to be simpler to expand the Privacy Act to cover SMEs for viewing and editing rights. It is not. This expansion was considered closely, but two factors make it a poor choice. First, the question of opening up the Privacy Act to SMEs brings with it wider considerations than consumer data. The current exclusion appears anomalous to privacy advocates. This is not an Inquiry into privacy legislation and that question should be settled elsewhere. Second, a specific and flexible definition of consumer data would differ from personal information. While the Privacy Commissioner could undoubtedly administer such rights if required to do so, there is a culturally supportive regulator already available in the ACCC (along with some industry ombudsmen). Chapter 8 delves deeper into submissions from a number of privacy entities that suggest the consumer right is not a matter with which they find much sympathy.

RECOMMENDATION 5.2

The Australian Government should introduce an outcome-based definition of consumer data that is, as an overarching objective, data that is sufficient to enable the provision of a competing or complementary service or product for a consumer.

In the relevant service or product context, consumer data is digital data, provided in machine–readable format, that is:

- held by a product or service provider, and
- identified with a consumer, and
- associated with a product or service provided to that consumer.

Participants in an industry should determine the scope of consumer data relevant to their industry (where an industry in this context would be determined by a broad description of the service). This should be in the form of a data-specification agreement.

Data-specification agreements should also articulate: transfer mechanisms, and security of data, to ensure that data use is practical and robust to technology updates; and the requirements necessary to authenticate a consumer request prior to any transfer.

These agreements should be registered with the ACCC, which may offer interim approval where an agreement has been reached but other industry agreements have been prioritised for approval.

In the absence of such agreement, consumer data must be in machine-readable form and include all of:

- personal information, as defined in the *Privacy Act 1988* (Cth), that is in digital form
- information posted online by the consumer
- data created from consumers' online transactions, Internet-connected activity, or digital devices
- data purchased or obtained from a third party that is about the identified consumer
- other data associated with transactions or activity that is relevant to the transfer of data to a nominated third party.

Data that is solely imputed by a data holder to be about a consumer may only be included, with industry-negotiated agreement. Data that is collected for security purposes or is subject to intellectual property rights would be excluded from consumer data.

A consumer for the purposes of consumer data should include a natural person and an ABN holder with a turnover of less than $3m pa in the most recent financial year.

Data that is not able to be re-identified to a consumer in the normal course of business within a data holder should not be considered consumer data.

The definition should be included in a new Act for data sharing and release (Recommendation 8.1). Given the need for consumer data to have broad applicability, the outer boundary definition and reference to ACCC registered industry-specific definitions should also be included within the *Acts Interpretation Act 1901* (Cth). Consequential amendments to other legislation in the future would ensure harmonisation across federal laws.

At a practical level, a consumer may not know when they seek access to their consumer data what they intend to do with it. Having found an error (the existence of which is by definition only discoverable by obtaining it), the consumer may seek to correct it, perhaps because the party to whom it was transferred has drawn it to their attention. The right to request edits should not at that point solely lie with privacy law, lest a new application have to be submitted through the 'channel' of privacy law.

In the end, it was decided that the ability to do all three — that is to view, edit and transfer — is fundamental to enhancing the value of data to consumers and ought to be kept together.

Another alternative was not to have any difference in coverage between personal information and consumer information. But the Commission's intention is firmly that potential scope of consumer data (now or in the future) be more than just personal information. In a digital world, we are determined to see data holdings scrutinised anew, with a consumer in mind, creating less likelihood that prior decisions on privacy would close off access to data that would help consumers to better leverage their data.

Moreover, individuals are generally unaware of their rights to take these kind of actions under privacy laws (chapter 3). This is understandable — the legislation has been on the books for many years, and the ability to keep refreshing public awareness is limited.

Inclusion of this Comprehensive Right in the reform package would refresh awareness that there is in fact an ability to both protect yourself and now to actively apply your data.

Right to transfer data to a third party

A key component of the Comprehensive Right would allow consumers to direct a data holder to transfer to a third party a copy of their data — the 'transfer right'. In practice, this might be applied in the following manner:

- Copying of data would be initiated by the consumer making a request to their existing service provider to release a copy of their consumer data to another identified service provider or advisory service.
- The data would be copied in machine-readable form to a standard agreed and via a secure process set by industry parties as effective and relevant to that industry. If the third party is in another industry (for example, health data to the insurance sector), the in-principle position we adopt is that health data is transferred in a particular form relevant to health and the insurance industry would need to adapt to utilise it in that form.
- The initial service provider would be permitted to retain their holdings of the customer's data after transfer, and continue to use it.

Underlying this transfer right, and maintaining incentives for data holders, is the idea that the data is a *joint asset* shared between the consumer and the businesses or government

agencies that hold the consumer's data. Thus the consumer's decision to transfer data to a third party, and perhaps switch service providers as a result, should not, in the Commission's view, alter the right of the initial service provider to the data that they collected while providing a service to the consumer. Nor would it alter the right of the recipient of the data to retain consumer data transferred to it (though in practice some companies may elect to not retain data if the consumer does not switch their custom, either to lower storage costs or to provide incentives for consumers to transfer data in the first place).

Where a transfer is requested outside of an industry (the example above is from a medical service provider to an insurance provider) and the agreed scope of consumer data is different in the source industry and the destination industry, the recipient industry would need to adapt to the standard of the original data holder.

An example of the types of outcomes the transfer right could deliver in the health sector is illustrated in box 5.1 (appendix B explores the effects of the new Comprehensive Right on community and welfare services provided by governments). Practical issues that would need to be addressed to action the right to transfer data are considered in section 5.3.

Right to be informed if data related to a consumer is traded to a third party

One of the most potentially pernicious practices with data is the onward trade or disclosure of data to third parties, leaving consumers unaware of who knows what about them. The damage is not so much in cost terms but in the feeling of exploitation. This has great capacity to undermine social licence over time, if misused. The OAIC has found that almost half of the surveyed population were concerned about how unknown organisations had obtained their personal information and almost all did not like their personal information being used for a secondary purpose (2013).

Being informed of data disclosures to third parties could thus be considered an important means for consumers to have greater control over data related to them.

Businesses have expressed concerns, however, that being required to inform consumers of *each instance* of data disclosure would be onerous and impose significant costs on them with little consumer benefit. Moreover, it would risk irritating consumers (Insurance Council of Australia, sub. DR318).

Origin Energy (sub. DR269) highlighted that in today's business environment, businesses often contract out a range of functions — such as those related to metering, billing, operation of call centres, marketing and credit collections. It argued that requiring disclosure for each instance of data sharing with third parties would not be operationally cost effective, and could lead to increased levels of customer dissatisfaction.

Box 5.1 Accessing and transferring medical records — the Comprehensive Right in practice

Patients are often 'flying blind' when it comes to their own medical records. While nearly 90% of people believe they should have access to their electronic health records, this access is often curtailed by a complex web of legislation, at the Commonwealth and State level, which leads to a range of access arrangements and fees imposed on individuals (Srinivasan et al. 2016). Further, patients can request that their medical records be transferred to third parties (for example, a new clinic or a specialist) but this process can be fraught with difficulties:

> It is common practice for a patient to have to 'carry' medical history information (usually provided in a printed form by the referring clinician) and physical diagnostic records (e.g. x-rays, pathology results) to each new service provider. The extent to which this information is captured into the receiving provider's system and then passed on in turn to further service providers is not assured.
>
> On the one hand, this represents a considerable waste of time and effort resulting in ongoing data quality issues. At a deeper level, it means that providers do not have access to a fully integrated patient history, with the potential for unnecessary duplication of services, particularly pathology and radiology services. This inability to see the 'big picture' of a patient's health is potentially dangerous, especially in an emergency setting …
>
> Consumers' lack of access to their own healthcare data and intelligence regarding the appropriateness of treatments for their condition, and the quality and cost of service providers means that they cannot be confident about making fully informed choices (Srinivasan et al. 2016, p. 12).

The Comprehensive Consumer Right would address many of these difficulties, by giving patients the right to access digital records about themselves and authorise health care providers to share these records with third parties.

In practice, by being able to view their records, patients would be able to understand more about their own health and make more informed choices. They would be able to authorise their entire file be sent on to specialists or other health care providers, which has the potential to minimise medical errors and cut costs. The potential benefits of improving this transfer of information are substantial — in 2015-16, Australian GPs made at least 21 million referrals to specialists, allied health professional or hospitals (Britt et al. 2016). And most of these referrals involved a printed letter, rather than a complete record of the patient's history, either delivered by the patient themselves or sent via a fax machine.

The roll out of My Health Record, which gives Australians the option to have a eHealth record that can be shared with any health care provider they choose, would make it easier for patients to exercise their Comprehensive Right. However, My Health Record is not designed as a complete reflection of a patient's health status (appendix E); the Comprehensive Right ensures that the patient can access their complete record, and use this information to make better choices about their health care.

Source: Srinivasan et al. (2016, p. 12); Britt et al. (2016)

The Interactive Games and Entertainment Association (sub. DR267) queried if this right would obligate data holders to notify individual consumers, or whether broader disclosure requirements (in terms and conditions accepted by consumers) would suffice.

The Commission does not propose that consumers be advised *on each occasion* data is traded or otherwise disclosed to a third party. But we recognise that being advised on the trade or disclosure of data is a potentially reasonable expectation of what is, after all, a

joint right to data. In our view, consumers should have a clear understanding about who their data is being shared with.

Firms and agencies should therefore be required to include in their privacy policies, terms and conditions or on their websites a list of parties to whom consumer data has been traded or otherwise disclosed to within the past 12 months. These lists would be updated at regular intervals. Such a requirement should be enacted through the Data Sharing and Release Act (DSR Act). Basing this obligation on the trade and disclosure of data, as opposed to just the *sale* of data, recognises that there are many ways for data custodians to exchange data for commercial gain without it being classified as a sale.

One other obvious change in the holder of data is that which may occur on the wind up of a firm. In such circumstances, consumers should always be advised of who now holds their data. For the small number of windups that occur through insolvency processes, insolvency practitioners (regulated by ASIC) should be required to ensure consumers are informed of the trade of data to which they hold a joint right. For the vast majority of businesses that windup without entering formal insolvency processes, the dataset owner should advise consumers that their joint asset has been separately traded. It is in the interests of purchasers of such datasets to ensure that this has occurred.

RECOMMENDATION 5.3

All holders of consumer data should include in their privacy policies, terms and conditions, or on their websites a list of parties to whom consumer data has been traded or otherwise disclosed over the past 12 months.

On the windup of an entity that holds consumer data, consumers should be informed if data to which they hold a joint right has been traded or transferred to another entity. For businesses entering formal insolvency processes, insolvency practitioners should ensure consumers have been informed. For businesses closing but not in insolvency proceedings, the entity acquiring consumer data should inform consumers of this fact and give them the opportunity for data collection to cease.

More effective disclosure

A number of Inquiry participants advocated greater rights and clarity around data disclosure.

InFact Decisions and Verifier (sub. DR232) suggested that consumers should also be given a right to be notified when their information is used to link datasets, while MLC Life Insurance (sub. DR298) suggested that consumers should be able to request such linkages. iSelect Limited (sub. DR266) wanted consumers to have additional rights to authorise third parties to access consumer data; T'Mir Julius suggested the right should extend to consumers requesting that third parties destroy sensitive information (sub. DR264). The Commissioner for Privacy and Data Protection (sub. DR320) suggested the Right should include a right to be informed on how consumer data is used.

We consider that the extent to which these additional forms of disclosure are afforded to consumers should, in the first instance, be determined on a business by business basis, with what is best practice for an industry allowed to emerge over time. In other words, it is not proposed that they be incorporated in the DSR Act.

The Draft Report requested information on specific methods of disclosure that could assist consumers to make meaningful decisions about how their data is managed in practice. Envestnet Yodlee (sub. DR280) suggested template consent forms and the Australian Dental Association (sub. DR230) suggested that 'Critical Information Summaries' currently provided to consumers in the telecommunications industry provide a template for improved disclosure to consumers and that they should be mandated.

As part of broader data-specification processes, consultation with consumer advocacy groups to develop more simplified and understandable disclosure arrangements is desirable. We recognise that the ACCC, State and Territory offices of fair trading, the OAIC and various industry regulators all have ongoing roles in improving disclosure of information to consumers. There may be scope for these regulators to better highlight best practice examples of disclosure for particular industries or product/service offerings.

Facets of the draft Comprehensive Right no longer recommended

Right to stop collection

In the Draft Report, the Commission recommended that consumers be able to request that a data holder stop collecting information on them (that is, they can choose to 'opt out' of a data collection process). This capacity to opt out was to have been subject to a number of exceptions — specifically, consumers 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.

Some Inquiry participants noted practical difficulties that the right to opt out of a data collection process could entail, for example:

> … I would be concerned to ensure that the new opt-out right does not undermine APP 3, by encouraging entities to collect more information than they need for their business purposes, with the intention of relying on the opt-out provision if individuals object.
>
> On the other hand, I consider that the proposed opt-out right may be beneficial for consumers in some situations, for example where their information is collected by third parties, and there is no direct relationship (i.e. no service delivery) between the data custodian and the individual. In these situations, the opt-out right may provide individuals with greater choice and control over how their personal information is to be used. (Office of the Australian Information Commissioner, sub. DR236, p. 10)

The Insurance Council of Australia (sub. DR318) noted that ongoing data collection would remain vital to delivery of insurance products to consumers, but opt out as specified in the Draft Report recognised that service providers might respond to opt out requests by discontinuing the service. The Australian Automotive Association strongly supported the opt out provision but noted that vehicle manufacturers that collect data prevent consumers from doing this:

> As telematics technology becomes more common in new cars, the ability for consumers to opt out of data collection will be important. …
>
> Currently, consumers do not have a choice when it comes to the transmission of vehicle data, when telematics technology is embedded in the car. As the AAA stated in the submission to the Issues Paper, where opt-out provisions do exist, consumers are sometimes faced with significant operability and safety implications which limit their ability to opt-out. For example, Tesla's website currently states that if a customer opts out of data collection this may result in the vehicle suffering from 'serious damage, or inoperability' (sub. DR262, p. 4)

One difficulty that is common to both the regulatory approach and market-based approaches is the difficulty for a consumer to know if data collection has actually ceased.

Notwithstanding the potential benefits to consumers of an opt-out right (as illustrated by the AAA), we are persuaded by the views of Inquiry participants that this aspect of the Comprehensive Right would have a number of inadvertent consequences and be difficult to implement.

For these reasons, we are no longer recommending that the Comprehensive Right include the capacity for consumers to opt out of a data collection. We note, however, the potential for market forces to respond to consumer preferences for data collection (and non-collection) — for instance, Internet providers in the US have started offering premium services that do not involve the collection of data (Ben-Shahar 2016).

Right to appeal automated decisions

In the Draft Report, the Commission recommended giving consumers the right to appeal automated decisions. Subsequent analysis by the Commission suggests a more cautious approach is necessary.

ANZ (sub. DR231, p. 14) queried if the right would allow consumers ' … to challenge the substantive outcome of such decisions, rather than just the accuracy of the data inputs to those decisions.' They noted that automated decisions are based on clearly defined policies and suggested that outcomes would be the same if the decision were made by non-automated means.

In the context of the EU General Data Protection Regulation, Wallace (2017) argued:

> In short, policymakers should create technology-neutral rules to avoid unnecessarily distorting the market by favoring human decisions over algorithmic ones. If a decision needs an explanation, this should be so regardless of whether technology is used to arrive at that decision.

The issue of incorrect or incomplete data leading to a 'flawed' view of a consumer is a possibility of both automated and non-automated decision making. Statistical techniques used to 'fill in' gaps in data collected about a consumer can create a misleading picture of that consumer, regardless of the final decision-making process (Australian Dental Association, sub. DR230). Consumers should be able to query the accuracy of the data *on which the decision was based* (automated or not) and request that it be updated if incorrect. This would, for instance, mean that a credit applicant could query data related to their application, such as details related to their income and expenses, but not the credit assessment methods used to approve or deny their application. Since, under the Comprehensive Right, consumers would have a separate right to request edits to consumer data where it is incorrect, there appears to be no need for a separate right to appeal automated decisions.

The economic impacts of the Comprehensive Right

As noted in chapter 2, the benefits that could flow from greater data availability are potentially very large. It is practically impossible to estimate, with any certainty, the magnitude of these benefits. But opportunities from data release will, by and large, beget further opportunities, and the balance should lie in favour of facilitating greater data access where the benefits are uncertain.

Similarly, it is not possible to precisely assess the relative economic impacts of the Comprehensive Right at this stage — especially as the scope of the Right would vary between industries as negotiated, and over time. Nonetheless, it is possible, in a general sense, to identify some of the impacts of a Comprehensive Right.

Impacts on consumers: The Comprehensive Right would lead to an improved ability for consumers to leverage their own data. This could help them to find better products and services, obtain lower prices, more readily and effectively use third party advice, access better or less invasive services such as healthcare, and provide a greater sense of control over their own data.

- There may be instances, however, where an ability to transfer data away from a service provider leads to that provider withdrawing loyalty 'rewards' that are currently provided in return for the provision of data. That said, some data holders may step-up measures to retain existing customers — that is, offer additional incentives to discourage transfer.
- Consumers who exercise the transfer right have the potential to reap substantial benefits. Those who do not exercise the right are less likely to benefit and some may argue that they could in fact be slightly worse off. However, as consumers collectively grow more digitally literate and empowered — and word of mouth would help this process — the number of consumers that benefit is likely to grow over time.

- There are risks with trading your data, as there are with the many free forms of similar data transfer that occur today. Consumers would be induced to share their data, and the principle of 'seller beware' is as relevant in data transfer as it is on eBay.

Impacts on businesses: Impacts on business would include the costs of complying with the Comprehensive Right, and the potential impact of disruption on business models.

Many businesses are already subject to the requirements of the Privacy Act and thus have in place mechanisms for offering correction to data and providing access to it — facilitating this under the Comprehensive Right is not likely to be costly. A requirement to notify customers of with whom consumers' data has been shared may initially be time-consuming, but all firms and agencies of any size would have contract management processes that enable this information to be discovered and provided (or posted online) to consumers regularly. As noted earlier, a wise course of action for entities might involve adopting arrangements that provide the same information whether the request originated under privacy legislation or the new Data Sharing and Release Act (the 'personal plus' approach).

The bulk of the compliance costs would likely stem from the data transfer right. Broadly, there would be two main aspects to this: the upfront investment in upgrading internal data-management processes if necessary (such as IT systems); and the costs associated with responding to each consumer request to transfer data.

- The costs for a particular data holder would largely depend on the nature of their IT systems, and the exact specification of data that is transferrable (established through industry working groups as discussed above). It seems likely that where IT enhancements are called for to support this Right, there would be broader benefits to the business related to their own data use (to the extent that it makes data more accessible internally).
- The compliance costs could be greater for small data-holder businesses relative to large businesses, in part, because larger businesses would tend to have greater opportunities to harness economies of scale (all other things being equal, bigger companies would have more customers and more requests for data transfer).
  - At the margin, this could affect the relative competitiveness of small and large businesses within a sector.
  - That said, many small businesses — such as small retailers and plumbers — would have limited *digital* data holdings and thus negligible compliance costs; and many small businesses, now treated as consumers, may benefit from access to data previously unavailable to them.
- Compliance costs would also include the costs involved in a business familiarising itself with any change to regulations — generally a one-off cost. However, any change (increase or decrease) in the complexity of legislation would tend to cause an ongoing change (increase or decrease) in the compliance burden. Narrow and prescriptive specifications may afford more certainty to data holders, but would not be responsive to

changing technology and growing data collections, and therefore would likely prove increasingly insufficient and costly over time.

Impacts on competition: The introduction of a new consumer right to transfer would be expected to increase competition. Indeed, this is a major motivation for the establishment of the Right.

- ANZ (sub. DR231), in particular, noted the potential for data to enable informed decision making by consumers and to boost competition between service providers. The Australian Food and Grocery Council (sub. DR284) noted scope for the right to hasten the growth of new business models in that sector. MLC Life Insurance (sub. DR298) noted the benefits from transferring health data. Businesses that have collected and exploited large amounts of customer data on the basis of their market dominance will now face renewed competitive pressures and a reinvigorated consumer base. Agencies that have sought active consumer engagement with their data (such as eHealth) may see a far more engaged community.

On the other side of the coin however:

- the costs of having data management systems in place to comply with the Comprehensive Right could form a barrier to market entry in the short term, particularly for smaller firms or not-for-profit entities. As has been the case in most instances where changes of this magnitude are introduced, off-the-shelf products and services rapidly emerge to assist entities with implementing systems for compliance (thus reducing any adverse impacts on competition).

Impacts on innovation: It is a widely accepted wisdom (and stated by a number of submissions) that data is an invaluable input into innovation — measures that increase data availability would thus increase opportunities for innovation. And innovation enabled by data can in turn drive downstream innovation (beyond that attributable to data).

Some Inquiry participants have suggested that should the right to transfer data include some transformed data, it could create disincentives for businesses to apply data analytics to 'raw data' and compete on the basis of the insights derived from such analysis (ANZ, sub. DR231, Business Council of Australia, sub. DR317).

However, as the general intent is not to automatically include genuinely transformed data. Since the extent to which transformed data is included within the scope of consumer data would be subject to an industry-specific data-specification agreement, it would seem unlikely that entities would negotiate a position that substantially reduced the value of their data analytics. To do so would suggest quite a static view of markets, and in the data sense, markets appear anything but static.

5.3 Actioning the right to transfer data

The practical issues that would need to be addressed in implementing the right to transfer data to third parties are detailed below. Approaches adopted overseas are drawn on where applicable.

How many years of data could be transferrable?

In the United Kingdom, the data transfer scheme is limited to one year of data — apparently to limit data storage costs. Several Inquiry participants argued that the Comprehensive Right should not apply to all data generated by the data holder, but rather to data that is readily accessible, such as that maintained to comply with existing requirements to retain data (see for example, ANZ, sub. DR231 and National Australia Bank, sub. DR270). The Business Council of Australia (sub. DR317) echoed this sentiment, arguing that the right should apply to data for a limited time only (in line with data retention requirements that already apply to businesses).

This is not sufficient to meet the intended outcomes that underpin the definition of consumer data. Twelve months of data would not be of much use for services that use long-term data series, such as the repayment history of a mortgage or use of an insurance policy. Similarly, for health-related data, substantial benefits are likely to be foregone if the sharing between providers of patients' medical histories is restricted to 12 months.

Most businesses and governments retain data on their customers and clients well beyond a single year — including but not only as required by other legislation. Moreover, when it comes to financial and insurance system participation, records can go back decades and the whole of a consumer's history may be relevant (for example, in establishing an understanding of the risk of providing credit to the consumer). The offset to this is that the Commission's recommended Right applies to digital data, and neither firms nor agencies would be required to undertake any digitisation that did not occur for other business reasons.

The preferable standard is that the duration of data able to be transferred is what the entity actually holds on the consumer at the point of making the request. That is, the right should, ideally, extend as far back as the data holder's digital files reach. And if the data is held off-site or restricted to an identity key held by a separate party, the test is not whether it is still held by the data holder of record, but whether it is *reasonably* accessible to them. This is not unfamiliar to firms and agencies as it is in line with existing Privacy Act requirements.

Where businesses cease to operate, the term of cover would necessarily change. Consistent with the position taken above on purchased data, should the failed firm's dataset be purchased by a new firm, the consumer would be informed of the sale and have the right to transfer their data to another entity before it may be lost.

Number of times a consumer can request data be transferred

In principle, under the proposed Right a consumer could make a request every day, and the Commission does not recommend any restriction on the number of requests — simply because any particular number would be arbitrary.

But the Commission does accept that a fee determined by the data holder (with the fee model open to scrutiny by the ACCC) may be charged for each request. The ability of entities to charge consumers for transferring data should curb spurious or time-wasting requests on businesses. Data-specification processes could also deal with this matter, if it was deemed necessary by those who are parties to it.

For clarity, we note that a single request for data by a consumer must relate to the data held at the time of the request. A request cannot create a constant stream of updated information from one party to another, as Fintech Australia (sub. DR315) had suggested would be desirable. This would in effect make the data holder a 'data bank' for their competitors.

As the Right is a joint right, the more third parties that a consumer transfers data to (in the process of seeking better service offers), the more widely their data would be dispersed. In other words, consumers should be mindful that the process of them trying out alternative services increases the number of entities who formally hold information on them — just as it does somewhat already today (perhaps to the ignorance of many consumers).

Ensuring data is usable by others

To enable efficient access and transfer of consumer data, and ensure high levels of usability, industry-relevant standards for data are required. This sentiment was a recurring theme in submissions.

Industry relevant data standards are needed

To facilitate efficient use, data transferred under the Comprehensive Right should be transferred in an accessible, machine-readable, standardised, timely and interoperable format (CHOICE 2014; Harper et al. 2015; OIRA (US) 2011).

In the Commission's view, the agencies and industries that would have to deliver the Comprehensive Right are likely to be best placed to determine industry-specific standards for its data sharing as part of data transfer arrangements under the Comprehensive Right. If negotiations for determining standards are not successful in making data usable, governments should reconsider the desirability of the approach at a later time, but initially it is preferable to apply industry capability, assisted by the ACCC.

The notion that data holders should be free to determine standards around data sharing (as distinct from the definition of consumer data) was a common theme in submissions (see for

example, Envestnet Yodlee, sub. DR280 and Westpac, sub. DR324). ASIC (sub. DR325) also noted the role that regulators can play in utilising or modifying existing regulation to create incentives for the development of appropriate standards (to reduce the costs of complying with those modified regulations).

In industries (such as education, health and law enforcement) that are characterised by substantial public sector provision, government agencies should be closely involved in the setting of industry-specific standards.

Some participants supported use of Application Programming Interfaces …

In the Draft Report, the Commission requested further information on the relative merits of Application Programming Interfaces (APIs, see appendix C). While many Inquiry participants noted the advantages of APIs, there were differing views on whether their use should be mandated, over alternative options, for transferring consumer data, and whether there is a role for governments in developing standards for implementing APIs.

ASIC (sub. DR325) noted that APIs would be more likely to offer a seamless experience to customers and reduce the 'friction costs' of sharing data, thus increasing the ease with which consumers could utilise the Comprehensive Right. Westpac (sub. DR324) recommended that the financial services sector develop common API standards to facilitate broad adoption of APIs (and to allocate the costs of developing APIs across industry).

The expense of using APIs — despite their apparent value as a low-cost method supporting the explosion of apps on Android and Apple ecosystems in recent years — particularly concerns some businesses with large legacy data bases. The Communications Alliance and the Australian Mobile Telecommunications Association (sub. DR250) said:

> We note that generally speaking, it is fair to assume that the costs of transfer would be substantial, and are likely to be huge or even prohibitive should API-based availability be mandated.

Others suggested to the contrary. iSelect Limited (sub. DR266) noted that, in their experience, the costs of APIs typically range from $10 000 to $50 000, with most around the $20-25 000 mark. Fintech Australia (sub. 182) rejected the notion that the costs of APIs would be prohibitive, noting that banks in other markets are undertaking 'open API' projects without being compelled by government.

The House of Representatives Standing Committee on Economics, as part of their review of the four major banks, also rejected the notion that APIs would be prohibitively costly. The Committee did not put forward Australian estimates of the costs of APIs, but rather referred to estimates from the UK Open Banking Project that put the costs of establishing an API framework at about £1 million per institution (ODI and Fingleton Associates 2014). The Committee recommended that Deposit Product Providers be compelled to provide open access to customer and small business data by July 2018 and that ASIC develop a

binding overarching framework to facilitate the sharing of data via APIs. The inquiry did not evaluate the merits of APIs in sectors beyond the banking sector.

… others argued against locking in particular technology

The vast majority of Inquiry participants, however, cautioned against locking in a particular technology, such as APIs, for the transfer of data. ANZ (sub. DR231) noted that APIs are more beneficial for frequent, individual and real-time data transfers, whereas one-off transfers — such as those likely with consumer requests for transfers of their own data — might be more efficiently performed using 'batch file' transfers. They argued that a legislative presumption in favour of APIs risks creating disincentives for industry to adopt more effective data sharing mechanisms in the future. The Australian Information Industry Association (sub. DR244) noted that while APIs are a desirable standard now, more suitable alternatives could develop over time. Envestnet Yodlee (sub. DR280) argued that financial institutions should have flexibility to provide data via a standardised API as well as through batch file transfers and the use of screen scraping (so that smaller financial institutions are not unduly burdened with the cost of implementing and maintaining an API).

FinTech Australia (sub. DR315) suggested that government should focus on consumer outcomes (rather than on a particular technology), which could be ensured by establishing a 'minimum expectation' of what constitutes a 'seamless consumer experience'. In other words, industry should be free to choose the technology as long as it provides consumers with, at minimum, a similar experience to what they would get if open APIs were adopted.

It is clear that there is not a 'one size fits all' approach — what works for a particular industry might not be suitable for other industries, or in all situations. Indeed, in some situations, there may be suitable technological solutions that provide acceptable consumer outcomes and are less costly and/or more straightforward to implement, or there might be a strong rationales for flexibility in how consumers are able to transfer data. AGL Energy (sub. DR251) suggested that industry should be permitted to determine the technology solution, and associated standards, for transferring data, and noted that APIs are quite possibly, but not necessarily, the appropriate solution for the transfer of consumer data in the energy industry.

While the Commission has heard a range of views on the costs of building APIs — ranging from relatively small to prohibitively large — the fact that APIs are widely used, and the backbone of many apps, would suggest that some of the larger estimates are overblown. Moreover, the *marginal* costs would be lower in industries where APIs are already widely adopted. Indeed, several Inquiry participants provided detail on their use of APIs — for instance, NAB noted that it had already undertaken significant investment in APIs and it currently allows its customers to transfer transaction data (so-called 'data feeds') to Xero via an API.

In the Commission's view, technology considerations are important but still subsidiary to the intent behind the Comprehensive Right and a strong desire to avoid locking in a particular technology. As long as data is transferred in a machine-readable format, industry should be free to determine *how* data is transferred.

Addressing risk and liability

Providing a right for consumers to direct a data holder to transfer data to a third party naturally raises questions about who has liability and when. Moreover, participants were concerned about the demarcation of risk (for events such as data breaches) across the data ecosystem and sought further clarification on how the framework governing the Comprehensive Right would address risks that consumers might face when sharing data.

Risks and liability issues identified by stakeholders include those that could arise when:

- data transferred by a data holder to a third party is incorrect, leading to a financial loss to the third party
- data breaches occur during the transfer of data
- third-party recipients of data fail to adequately protect that data, leading to data breaches or fraud and identity theft
- unscrupulous third parties exploit consumers who may not understand the risks of sharing data with different parties.

Liability for data breaches, mistaken data transfers and inaccurate data

A number of Inquiry participants expressed concerns about who would be liable for data breaches and mistaken disclosures that occur during the transfer of data, or after the data has been transferred to the third party. For example, Google Australia (sub. DR292) expressed concerns that requirements for data holders to *directly* provide data to third parties could open them up to liability for inadvertent or mistaken disclosures. Googles believes that consumers are best placed to make decisions about to whom they disclose their data — an alternative arrangement, whereby data is made available to consumers who are then able to forward it to third parties, would be preferable.

ANZ (sub. DR231) made a similar point, arguing that ' … data custodians need certainty that, provided they follow certain defined steps in responding to a consumer's request that the custodian transfer the consumer's data to another, they will be absolved from data breaches and other losses that arise from the transfer, use or reliance upon the data.'

Envestnet Yodlee (sub. DR280) recommended that standardised contract terms should be developed to provide a legal framework for enforcement of standards, and to clarify the incidence of liability across the data supply chain. A similar view was put forward by the Business Council of Australia (sub. DR317), who argued that legal liability in instances of data transfer needs to be clarified (and that this could include minimum standards for data

recipients). FinTech Australia (sub. DR315) argued that minimum data security standards should be set by industry working groups, overseen by industry regulators, to ensure that data holders and recipients are consistently maintaining best practice. This would include standards around data transfer.

Issues around the security of technology used to *transfer* data would not appear to be an intractable barrier to the implementation of the Comprehensive Right — as noted elsewhere in this chapter, Australian banks are already routinely using APIs to transfer data to third parties. And the Commission notes that data holders already have obligations to protect data under existing privacy laws.

Assurance of what constitutes a consumer request for data transfer

Some participants, such as Westpac (sub. DR324), highlighted the importance of: consumers being able to give informed consent for the transfer of data to a third party; and data holders being certain that the consumer has actually granted consent for the transfer in a manner that is practical for the data holder. In the context of financial services, online banking platforms, which already require passwords, could be an avenue for consumers granting consent for data transfers. In other industries there might be a need to develop new methods for 'validating' that a consumer has genuinely consented to the data transfer.

As noted by iSelect Limited (sub. DR266, p. 3), these 'issues could be overcome by ensuring the regulatory framework clearly empowers the consumer with the ability to provide informed consent to the transfer of their data.'

Issues around risk and liability will need to be clarified

The Commission broadly agrees that uncertainty around consent, risk and liability will need to be clarified as part of final legislative structure (chapter 8) and, where possible, industry standard-setting processes.

The more worrying situations identified by Inquiry participants (such as the transfer of real-time electricity usage that could be used to identify when an individual is at home — Australian Energy Council, sub. DR281) would not arise under the Comprehensive Right as currently specified, because a request cannot create a constant stream of updated information from one party to another.

Robust security standards are a crucial foundation to realising the benefits of data transfer to the fullest extent. Third-party accreditation (suggested as a solution by the Australian Bankers' Association, sub. DR307, and several of their members as well as FinTech Australia, sub. DR315) could reduce risks for consumers and existing data holders, though it could also create barriers to market entry for new start-ups and lead to lower competition than would otherwise be the case.

In practice, however, we note that data holders would have opportunities to advise consumers, prior to data being transferred, about the risks of data transfer. This would likely include advising the consumer that the data holder has not performed any due diligence about the intended data recipient (and that the consumer should themselves do so), and that they do not accept any liability for the data transfer. In principle, industries should be permitted to develop industry-specific arrangements, to be reviewed by the ACCC as part of its broader responsibilities in approving data sharing under the Comprehensive Right.

Costs of data transfer

Earlier, we noted that data holders may levy a fee for providing access and/or limit the number of free access opportunities per year. The charging for such access, correction and transfer rights would need to take into account the *additional* costs incurred by data holders.

But it is important to acknowledge that charges could be used to limit data transfer in practice, particularly where data brings a significant advantage to incumbents.

The Commission recommends that 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. Guidelines around access charges could also be established as part of broader standards-setting processes within an industry.

5.4 Institutional arrangements to support consumers in using their data

The role of the ACCC and other institutions would be to ensure consumers can either exercise their rights or complain when they cannot.

The recommended consumer data access model would 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, the Australian Communications and Media Authority (ACMA) as regulator for telecommunications, ASIC as corporate regulator, the Australian Prudential Regulation Authority (APRA) as prudential regulator and the Commonwealth Ombudsman as Private Health Insurance Ombudsman). These other regulators have various enforcement powers relating to their respective regulatory responsibilities.

We are recommending the following key regulatory responsibilities:

- The ACCC would play a key regulatory role with regard to consumer data access, since competition and consumer policy lies at the heart of the recommended changes. This could include 'naming and shaming' parties who are seen to be dragging their feet. In performing this role, it should consult closely with other regulators, including the OAIC.
- The OAIC would continue to have overall responsibility for privacy compliance and enforcing privacy complaints. It could also act as a backup where no ombudsman or dispute resolution scheme already exists.

At an operational level, complaints handling and dispute resolution could be by existing external dispute resolution schemes in various industries (for example: the Financial Ombudsman Scheme and the Credit and Investment Ombudsman scheme in the banking sector; and ACMA and the Telecommunication Industry Ombudsman in the telecommunication sector), under arrangement with the ACCC. 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 early on in the operation of the Commission's framework. Some supplementary resourcing would be necessary, particularly to the ACCC. For other agencies, such as Ombudsmen, 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 policy functions would 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.

There should be a 'no wrong door' approach to dealing with consumer data issues and complaints. This means the OAIC and industry ombudsmen should coordinate with the ACCC on the receipt and handling of consumer complaints on data access and use. Consumers should not be left straddling a regulator abyss.

To a large extent the individual access model recommended 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 Commission's recommended changes aim to enable consumers to exercise more control over the collection and use of data on them, the onus remains on them to make responsible choices on who they provide personal information to and for what purposes. This must include decisions to transfer data to third parties (in exercising the Comprehensive Right). We recognise that ensuring consumers are able to make meaningful, informed choices is important. As with other legislative reforms, there would

be a need for education for consumers, firms and agencies on their rights and responsibilities under the Comprehensive Right, and on other aspects related to implementation (such as timing).

RECOMMENDATION 5.4

The Australian Government should provide for broad oversight and complaints handling functions relating to the use of the Comprehensive Right. Accordingly, the Australian Competition and Consumer Commission (ACCC) should be resourced to undertake the following additional responsibilities:

- approving and registering industry data-specification agreements and standards
- handling complaints in relation to a data holder's failure to meet the terms of the Comprehensive Right, including in regard to the scope of consumer data
- educating consumers (in conjunction with State and Territory fair trading offices) on their rights and responsibilities under the Comprehensive Right
- assessing the validity, when requested or at their discretion, of charges levied by data holders for application of the Comprehensive Right.

The Office of the Australian Information Commissioner and industry ombudsmen should, in order to ensure a 'no wrong door' approach to handling consumer engagement, coordinate with the ACCC on the receipt and handling of consumer complaints on data access and use.

5.5 Comprehensive credit reporting

In the Draft Report to this Inquiry, the Commission recommended that:

- the Australian Government adopt a minimum target for voluntary participation of 40% of credit accounts
- if the target is not met by 30 June 2017, draft legislation be circulated by 31 December 2017 to impose mandatory reporting at some point thereafter.

The Draft Report also noted industry concerns around how repayment history information (RHI) should be reported for consumers who seek hardship arrangements and noted that clarification could enable greater voluntary participation in comprehensive credit reporting (CCR).

Participation should only be mandated if voluntary participation is insufficient

In the Draft Report, the Commission concluded that it was too early to be certain that a voluntary approach to CCR would be doomed to fail. (Indeed, the finance sector was a

driving force behind the reform process that led to CCR.) However, it was also recognised that low participation would mean forfeiting the benefits that could flow from CCR.

Many Inquiry participants expressed views about mandating participation in CCR:

- The OAIC (sub. DR236) noted that the industry-developed Principles of Reciprocity and Date Exchange (PRDE) — which establish the reciprocal nature of the new system — were only introduced in December 2015 and thus argued it is too soon to consider mandating participation.
- The Australian Retail Credit Association (ARCA, sub. DR247) argued that the prospect of mandated participation could actually delay its members undertaking the necessary work to participate in CCR (until they had a clearer picture of how the new system would operate).
- Veda (sub. DR259) argued that mandated participation could lead to further delay, complexity and cost.
- The Financial Rights Legal Centre (sub. DR289) presented a number of arguments against mandated CCR, including that: there are no jurisdictions internationally that have mandated CCR; the Australian Law Reform Commission did not recommend it as part of their 2008 review of the Privacy Act; there has yet to be an independent review of the operation of CCR in Australia; progress to date has been steady, with delays stemming from the implementation of regulatory codes and issues around RHI and hardship provisions.
- Financial Institutions & Management Advisory (FIMA) (sub. DR233) offered an alternative approach to realising the benefits of CCR, proposing that mandating the provision of all allowable credit reporting information, with the exception of RHI, would allow for greater participation in CCR without the need to first resolve issues around reporting RHI for borrowers in hardship.
- The Customer Owned Banking Association (sub. DR273) suggested that, should CCR be made mandatory, a materiality threshold — $200 billion in assets — be put in place for mandatory participation, since the costs of participation are likely to outweigh the benefits for smaller credit providers.
- Fintech Australia (sub. DR315), on the other hand, supported the Commission's recommended approach, suggesting that it could go further by imposing penalties on credit providers who failed to provide CCR on 100% of their accounts by June 2018.

The proposed timelines in the Draft Report were also a cause for concern for some participants — for example, NAB suggested that a deadline for the 40% target of 31 December 2017 would be more appropriate given the PRDE was approved only at the end of 2015. Dun & Bradstreet (sub. DR319) recommended that the 40% target could be applied, at June 2017, to revolving credit products (such as credit cards), with the remaining credit portfolios transitioning to CCR over the following 12 months to meet a proposed target (at June 2018) of 80% of accounts. The Attorney General's Department (sub. DR334) also suggested that timelines were too ambitious. Veda (sub. DR259)

recommended that a Treasury-led forum be established to set an indicative timeline for meeting the 40% target, with quarterly updates.

The Attorney General's Department (sub. DR334, p. 5) raised the question of whether the mandating of participation in CCR could 'raise constitutional issues around the acquisition of property and, if so, would require the Australian Government to pay compensation on just terms to credit providers for compelling them to disclose valuable commercial information'. The Department also sought greater clarity around what information and which credit providers mandatory participation would apply to.

A way forward

The legal authorities of the commonwealth itself would have to decide on Constitutional issues. Mandatory participation has been envisaged for some time as a possibility and the impediments do not appear insurmountable from past work by the Commission.

The Commission otherwise has been convinced by arguments that some circuit breaker to the current indifferent performance is necessary. We propose that the Australian Government should establish a target for voluntary participation, and, should that not be achieved by 30 June 2017, set in motion a legislative process for mandatory participation, with draft legislation circulated by 31 December 2017 at the latest. While concerns about timing are noted, the proposed deadline for the circulation of legislation is a full two years after the PRDE was approved by the ACCC.

For clarity, this would include provision of the five new data fields permissible under CCR. In our view, the utility of the credit reporting system would be limited if credit providers were able to provide just RHI, without also providing data related to liabilities (since such data is arguably more important for responsible lending).

Moreover, it is recognised that the target of 40% of accounts might be reached prior to the system being formally mandated (during the legislative process) — that is, the industry could continue efforts to participate in CCR. Were this to happen, the Australian Government should 'suspend' the draft legislation (provided the level of participation remains above the 40% target). The Australian Treasury should work with industry and credit bureaus to monitor and publicly report the level of participation on a regular basis (say, every three months).

The Commission is not convinced by industry arguments that the prospect of mandated participation would automatically lead to credit providers not undertaking the system upgrades necessary for voluntary participation in CCR (while they wait to see how the mandated system would operate). If industry participants are indeed moving towards implementation, there is no apparent reason why the current framework governing the operation of CCR, which was developed with the involvement of industry, should not be adopted for the mandated system. Investment by active participants to date should not be wasted, and is a primary reason to persist with the current structure.

In practice, the Australian Government could mandate participation as a condition of being an ASIC-licensed credit provider, and thus avoid changes to the credit reporting regulations. In the Commission's view, the work necessary to participate in a mandatory CCR system would thus likely be the same as if participation was on a voluntary basis.

A minimum target for voluntary participation should be established

Several submissions noted that the Commission's proposed target of 40% of accounts was vague, and sought clarification on the precise definition of the 40% target, and whether it is to be applied at an industry-wide level, rather than at the institutional level (for example, ANZ, sub. DR231, Customer Owned Banking Association, sub. DR273; Attorney General's Department, sub. DR334). ARCA (sub. DR247) suggested that an appropriate denominator is the number of credit accounts issued by ASIC-licensed credit providers (which does not include credit providers such as utilities providers and telecommunications companies).

The Commission is recommending that the minimum target for voluntary participation in Comprehensive Credit Reporting should be 40% of all active credit accounts (that is, across the industry) that are provided by ASIC-licensed credit providers, for which comprehensive data is supplied to the credit bureaus in public mode.

Hardship arrangements

Several submissions pointed to a recent determination about how RHI should be reported for borrowers in hardship as being a barrier to greater participation. They noted that uncertainty creates a risk for early movers who may need to undertake further system upgrades should RHI reporting requirements change in the future (National Australia Bank, sub. DR270; Customer Owned Banking Association, sub. DR273).

As noted in the Draft Report, and by some Inquiry participants, there has been disagreement about how to report RHI for borrowers who have been granted a hardship variation.

A recent determination by the Financial Ombudsman Service (FOS) (case number 422 745) noted that RHI should be recorded on the basis of whether borrowers are meeting their obligations with respect to payments that are *due and payable* (according to the definition of RHI in the Privacy Act). The key issue then is *what* payments are due and payable for borrowers who have been granted hardship relief.

In the particular case it was adjudicating, the FOS noted that an indulgence was provided by the credit provider, and a new payment plan was established — it determined that this varied the payments that were due and payable, and that RHI should have been reported in line with revised payment arrangements, rather than the original contract. This has been widely interpreted by industry and consumer advocates as meaning that RHI must be

reported as up to date for any and all indulgences extended to borrowers, provided borrowers meet their obligations under any revised agreement (which in practice could mean that borrowers who were not making repayments are still reported as up to date).

Some parties (such as the Australian Bankers' Association, sub. 93) expressed concerns that this ruling makes it more difficult to identify which individuals are currently in hardship, and thus inhibits the ability of credit providers to accurately assess credit risk. In effect, this limits the usefulness of reporting RHI information to credit bureaus, and has been raised as a barrier to greater voluntary participation in CCR (ARCA, sub. 87).

Consumer advocates, such as the Financial Rights Legal Centre (sub. 107), however, have also expressed concerns about how RHI is reported for those in hardship, arguing that any perceived risks to an individual's credit rating as a result of hardship provisions would stop them from seeking those provisions in the first place. They also argued that allowing credit providers to identify borrowers seeking hardship arrangements with one credit provider could lead to them withdrawing existing revolving credit. In other words, a borrower seeking a hardship provision with, say, their mortgage provider might find that their credit card provider lowers their credit limit in response to the hardship application, thereby limiting their access to credit and possibly worsening the borrower's short-term financial position.

ARCA (sub. 87, p. 11) argued that the ' … FOS determination is inconsistent with contract law, and it confuses the two distinct concepts of varied obligations and postponed enforcement.'

The OAIC (pers. comm. 22 March 2017) has since provided clarification on how RHI should be reported for borrowers in financial hardship.

- Where a credit provider responds to hardship notice by varying the credit contract, RHI is to be reported against payment obligations under the revised contract. (This has not been a source of contention.)
- Where there is no such variation, and where a temporary payment arrangement is entered into by the borrower and the credit provider, the reporting of RHI would depend on the nature of the arrangement — specifically, whether the arrangement would in practice limit the right of the credit provider to enforce the contract (which has implications for the amounts considered to be due and payable).
  - If the credit provider has made a representation to the borrower to *not enforce* its rights under the original contract, or otherwise induced the borrower to come to this view, RHI is to be reported against the revised payment arrangement (which could in practice involve a moratorium on payments) rather than the original contract.
  - However, if the temporary payment arrangements involve no such representation, and the credit provider thus retains a right to enforce its contract, RHI can be reported against the original contract — any payments missed by the borrower (even if agreed to by the credit provider) could be reported as late on their credit report.

The Australian Retail Credit Association (pers. comm. 16 March 2017) has since advised the Commission that the OAIC's ruling effectively addresses the industry concerns about RHI that stymied greater participation in CCR. Our expectation is that industry would now move more quickly to participate in the new system.

There remain, however, two potential issues with the current arrangements around reporting RHI.

The first is that the way in which RHI is to be reported could create disincentives for credit providers, discouraging undertakings to consumers to not enforce the credit contract (although at least in some cases this could be balanced by incentives to maintain good relationships with customers). In effect, this could create uncertainty for borrowers, to their detriment, and would be more likely to impact on more vulnerable borrowers — this would not be a positive outcome in our view. This concern was also expressed by consumer advocates (Consumer Action Law Centre, pers. comm. 21 March 2017). The Commission considers that ASIC, in their role as overseeing the hardship provisions, would be well placed to monitor and identify if this becomes a material issue. Trading of experience in this area would be a worthwhile investment for all concerned.

Nonetheless, there would be merit in the already slated review of the credit reporting system (which the ALRC recommended to take place in 2019) examining whether this outcome has arisen in practice, and whether there is a case for additional consumer protections.

The second issue is that the current rules around reporting of RHI could lead to situations where (some) borrowers in significant financial hardship are reported as up to date on their payments, and are still able to access new credit (that they may not be able to afford). This is clearly a poor outcome — for the borrower, industry and for the overall risk of credit markets.

The inclusion of a hardship flag in credit reports has been raised as a way to avoid such situations (ARCA, sub. 87). Specifically, ARCA called for the introduction of two hardship flags — one to be applied when the contract is varied, and another that would be activated upon *receipt* of a hardship notice. Inclusion of a hardship flag (or flags) would undoubtedly improve the ability of credit providers to identify borrowers in hardship, and thus improve their credit underwriting.

At the same time, it is important that borrowers are not discouraged from seeking hardship arrangements from their credit providers in the first place. A hardship flag that is activated upon *receipt* of a hardship notice might discourage some borrowers in practice, particularly since there is no guarantee that a hardship variation or revised payment arrangement would actually be offered. This could lead to situations where the borrower's credit report has a hardship flag activated, without the borrower actually receiving any relief. This would likely discourage at least some borrowers from seeking hardship arrangements.

However, in our view, a temporary hardship flag, activated only when a hardship variation or indulgence was *granted*, might lead to better outcomes for borrowers, since borrowers would have the benefit of relief and it would be preferable to missed or late repayments being listed on the report (as they would be listed for a period of five years). Of course, this implies a hardship flag that is activated only when the credit provider is not able to report RHI against the original contract (that is, report payments as being late). Furthermore, it is also important that borrowers who have entered into hardship arrangements with their credit provider are not 'penalised' through the *unwarranted* loss of existing revolving credit with other credit providers, which could worsen their financial hardship. To protect consumers, credit providers could be limited in their ability to lower credit limits (or otherwise withdraw existing credit) in response to a consumer requesting and being granted a hardship variation or indulgence — this might involve a requirement that credit providers do not automatically withdraw existing credit. In other words, credit providers could be compelled to make such decisions on the basis of whether the borrower could afford the credit product, not because a hardship flag had been activated.

In our view, there is a strong argument for the OAIC, in conjunction with ASIC, considering whether there is a need for a hardship flag in credit reports, and how such a flag would operate in practice. As part of this, regulators, as well as industry and consumer groups, should be consulted, and available evidence from overseas considered. The need for additional safeguards for consumers should be evaluated as part of this process.

Is there a case for further reforms?

A number of Inquiry participants raised the prospect of further reform to the credit reporting system to allow the inclusion of new data fields (such as outstanding balance) and to allow non-ASIC licensed credit providers to share RHI. For example, Veda called for a new recommendation for a review of the reforms that allowed CCR in the first place, and suggested that the review consider the inclusion of additional data fields and users.

Noting that the design of the current system is the result of a long reform process, and that further reviews have already been slated, it would, in our opinion, be premature to recommend further changes at this point in time. Such issues could be assessed as part of systematic reviews.

RECOMMENDATION 5.5

The Australian Government should adopt a minimum target for voluntary participation in Comprehensive Credit Reporting of 40% of all active credit accounts, provided by Australian Securities and Investments Commission (ASIC)-licensed credit providers, for which comprehensive data is supplied to the credit bureaux in public mode.

If this target is not achieved by 30 June 2017, the Government should circulate draft legislation by 31 December 2017, to impose mandatory participation in Comprehensive Credit Reporting (including the reporting of repayment history) by ASIC-licensed credit providers in 2018.

The Office of the Australian Information Commissioner and ASIC should consult with other regulators, industry groups and consumer advocates to collaboratively consider whether there is a need for a hardship flag in credit reporting.

The Department of the Treasury should be given responsibility for monitoring and publicly reporting on a regular basis on participation in Comprehensive Credit Reporting.

6 Sharing and releasing data for community benefits

Key points

- A structure for data sharing and release should be introduced that would allow access arrangements to be 'dialled' up or down according to the nature of risks associated with different types of data, uses and use environments.
- The structure would cover a wide range of data, including identifiable data that can be shared with trusted users, de-identified data, and open access data.
- The recommended changes to the current handling of the public sector's data assets are designed to maintain public confidence and build a social licence for data re-use.
- Implementation and ongoing monitoring would be vested with a new statutory office holder, the National Data Custodian (NDC).
  - The NDC would administer the Data Sharing and Release Act, recommend valuable datasets for designation as National Interest Datasets, provide technical guidance and direction to the data system, accredit and coordinate Accredited Release Authorities (ARAs), report on data use 'good news' and breaches, and manage broader ethical considerations within the data release and sharing system.
  - The NDC would not, however, hold data itself or decide on its release or re-use.
- An integral part of the Commission's recommended data reform Framework is the introduction of a network of ARAs.
  - ARAs would be hubs of sectoral expertise in data curation, de-identification and linkage, would implement a risk-based approach to the broader sharing and release of data through formal, contemporary, NDC-reviewed risk management procedures.
  - ARAs would generally be nationally focused and may be located within the Commonwealth, State or Territory public sectors or be other publicly funded entities that have the necessary expertise, focus and governance structures.
  - ARAs would be funded to undertake data curation, management, linkage, storage and release, with funding determined by the NDC. ARAs could also receive revenue from project work undertaken for clients.
- There is a compelling public interest to see greater use made of public-funded researchers' unique datasets. Progress by individual research institutions that receive government funding in making their research data and metadata widely available to others should be openly published by those institutions, with reference to past performance.
- Existing exceptions in privacy legislation allowing the sharing of personal information for health and medical research purposes without consent should be extended to other areas of public interest research.
- The requirement to destroy linked datasets and statistical linkage keys at the completion of data integration projects should be abolished. It is akin to book burning.

As highlighted in chapters 2 and 4, the Commission has considered a wide spectrum of data (figure 1.2) that could be better used to provide benefits to the Australian community — this includes at one end of the spectrum identifiable information; through to datasets that contain no individually identifiable information.

This chapter recommends an approach for safely enabling the sharing or release of all types of data to occur.

6.1 A structure for future data release and widespread use

As outlined in chapter 4, the Commission's new data Framework is underpinned by a scalable, risk-based approach. Scalability allows for the Framework to adapt to the changing nature and volume of data, as new technologies are developed and others become obsolete, and as new uses for data are realised. Hence, scalability will preserve the Framework's resilience and ability to incorporate innovation in data sharing and release techniques, and ensure that risk management is effective and not overly burdensome.

The Commission recommends institutions to support a risk-based approach that is scalable across all data types, from generic data on public assets through to sensitive data on individuals or businesses. This involves considering the risks that may arise from data management and release activities, and explicitly managing those risks.

The recommended structure outlined in this chapter is primarily directed towards public sector data holdings (including data held by publicly funded research entities) as decisions about the sharing or release of public sector data are typically not based on an assessment that weighs the potential value of that data against a realistic assessment of the likely risks involved. Some reforms though, would also be relevant to private sector data (chapter 7). With the mix of public-private models under which services are delivered to communities, it is desirable to have an approach consistent across all data holdings.

A shift toward actively making data available

At the simplest level, release of data that does not specifically relate to individual people or businesses or approved intellectual property rights should become routinely available for use by governments, consumers, businesses and the research community. Instead of mainly releasing data on request for particular projects, the emphasis *for all agencies* should move towards actively pushing non-sensitive data out in a coordinated way.

While the principle in future should be that all non-sensitive datasets be released, initially, data in those fields where there are burgeoning opportunities and capabilities should be prioritised for access and release as resources and sectoral demand allow.

This may mean that the average quality of publicly held or publicly funded datasets is lower than it would otherwise have been had data custodians delayed release with a view to improving the quality of data or metadata. But the balance of advantage demonstrated in the Draft Report, and not challenged seriously by submitters, is in favour of timeliness of access over additional processing.

Data that should be actively made available 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.

At the next level, data that identifies individual businesses — public, private or not-for-profit — can be considered sensitive. As noted elsewhere in this Report, there is a disincentive for entities (firms or regulators) to release such data where it may create grounds for criticism of their own, or a regulated entity's performance. But a contrary view is that this can be an important mechanism for public transparency and accountability; a case in point could be data on hospital performance, for example. Plus the wider benefits of an open approach may include: reducing information asymmetry, improving consumer choice and outcomes, improving program management and service delivery, and promote competition in markets (chapter 2).

A realistic assessment of the risks associated with, and necessary safeguards for, public release enterprise level identifiable information should be undertaken, with particular attention paid to that information already publicly available in less accessible forms.

Datasets that are primarily public sector administrative data or otherwise collected through public funding (such as may be the case with data generated in much of academic research) should also be readily considered on a risk assessment basis for more rapid release.

For researchers' unique project-level datasets, there are many justifications cited by various groups for why these are not generally released. Few of these objections are genuinely enduring in nature. The National Committee for Data in Science stated:

> … there should be a general expectation that publicly funded research datasets should be made more accessible. This includes grant-funded research at universities, research at publicly funded research agencies, and research conducted by government itself. There may be exceptions to this expectation, such as for sensitive data … (sub. DR265, p. 3)

Governments have for many years failed to press this issue with vigour.

Some data (including some administrative datasets) will be assessed as too sensitive to be publicly released. Typically today, data that identifies individuals falls into this category automatically, but often with little consideration of the genuine risks and the costs of lost opportunities of such categorisation (chapter 1).

The Commission has recommended that trusted user models — already used on an ad hoc basis by many government agencies and researchers — be deployed more broadly and

consistently to enable greater access to sensitive identifiable data, while maintaining necessary safeguards (section 6.4).

Leadership and role-modelling can help. Whereas today the Australian Bureau of Statistics' (ABS) legislation (sub. DR285) has excluded the possibility of it using or sharing identifiable statistical information for 'non-statistical purposes' (such as law enforcement, compliance and service delivery), it is an unusual position for a key data collector to be precluded from having a more integral role in the sharing and release of Australia's data.

Statistics New Zealand manages the Integrated Data Infrastructure, a database containing microdata about people and households from government agencies, Statistics NZ surveys and a population census and non-government organisations (Statistics New Zealand 2016). This database has been used to support research and policy evaluation in areas ranging from crime prevention to the situations of vulnerable children and families (sub. 62).

RECOMMENDATION 6.1

As an immediate objective, all Australian governments should direct the early release of all non-sensitive publicly funded datasets — whether held by a government agency or other body receiving public funding for data collection activities.

A realistic assessment of the risks attached to public release of identifiable information that is already public (in a less accessible form) should be undertaken by all governments, with the intention of releasing low risk data, and mitigating risks where possible to enable far greater public release of data, including that which could be used for program or agency performance management purposes.

Agencies should report annually on the proportions of their datasets made publicly available, shared, and not available for release.

Data linkage

Inquiry participants have raised concerns about legal limitations on linking data and delays in data linkage services, particularly for linkage with Australian Government data (chapter 3) (Centre for Big Data Research in Health, sub. 21). It is likely that both the limited number of accredited integrating authorities and the time required to prepare data for linkage contribute to delays experienced. The Australian Institute of Health and Welfare (AIHW) stated 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).

The capabilities and resources to undertake these tasks exist in institutions beyond those currently allowed to link data. Inquiry participants have suggested, for example, 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 recommended that the Australian Government consider extending accreditation to link Commonwealth data to State linkage units (SSCH 2016). We similarly consider that there is likely to be a wider public interest served by allowing state-based data linkage units to link Australian Government datasets.

RECOMMENDATION 6.2

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 (Recommendation 6.6) to allow them to link Australian Government data.

Government purchases and contracting of private sector data

In its Draft Report, the Commission recommended that governments retain the right to access or purchase data that has public interest value, where that data is the result of a contract, funding or regulatory arrangement with the private sector. Though possessing public interest value, not all such data would necessarily meet the requirements for purchase and designation as National Interest Datasets (NIDs) (chapter 7). Examples were provided in the Draft Report of contracts that governments entered into with the private sector for the provision of public interest services that did not allow for sufficient data access. 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 example, for the quality and efficiency of services delivered to the community, and for contractor performance payments.

Now that it is evident that data is a valuable asset, it is imperative for governments to ensure that the contracts to which they are party, provide a degree of data access rights that is consistent with the interest of the wider public. Much more care, consideration, and where necessary retention of data use and release rights by the public sector is necessary in these situations. Privatisation processes should similarly consider ongoing data access needs.

The Department of Agriculture and Water Resources (sub. DR224) suggested that this could be facilitated by implementing clearly defined and consistent data ownership, access and use arrangements across all government contracts, avoiding the need to negotiate rights for every agreement. One means to achieve this outcome at the Commonwealth level would be to alter existing Department of Finance contract templates for procurement (and equivalent templates at the State and Territory level) to incorporate clauses that would vest

access, or at least purchase rights, to data in governments on a default basis. This would ensure that the interests of the public, with respect to data, are upheld in all private-public contract negotiations. Embedding access and purchase rights in Department of Finance procurement templates would also streamline processes and reduce the costs associated with having to negotiate data rights in all procurement contracts.

RECOMMENDATION 6.3

All Australian governments entering into contracts with the private sector that involve the creation of datasets in the course of delivering public services should assess the strategic significance and public interest value of the data as part of the contracting process.

Where data is assessed to be valuable, governments should retain the right to access or purchase that data in machine-readable form and to subsequently apply any analysis and release strategy that is in the public interest.

The Australian Government Department of Finance should modify template contracts to, by default, vest access and purchase rights in governments, and avoid the need for negotiating separate rights in each contract. State and Territory governments should adopt a similar approach.

A process for assessing datasets for release

The need for a systematic process to determine which government datasets should be released has been recognised by the Department of Prime Minister and Cabinet (2015). Achieving clear actions towards this is a central focus of this Report.

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

The process outlined below would be suitable for determining which data should be *considered* for prioritised release. However, risk management (discussed below) would also need to apply to datasets being made publicly available. Principally, the sensitivity of data should receive closer attention. So while non-sensitive data can readily be made open, the release of data with identifiable elements requires awareness and assessment of risks and the need for transformation prior to any public release.

Discoverability of data

It is impossible to reuse data if no-one knows it exists. Failure to reuse publicly funded data may result in significant duplication and wasted resources. Moreover, scope to

validate research findings or evaluate programs can be limited if underlying data is not known to be available.

Arrangements to ensure that the existence of datasets is known could be substantially improved. Publicly funded research groups are in a unique position to instigate major improvements in data discoverability, and at the same time strengthen the case for others, such as administrative data holders, to also act in a more open fashion. Publication of indicators on the proportion of projects involving datasets not available for release would provide the opportunity to ask what is being lost if knowledge of data is limited.

RECOMMENDATION 6.4

Publicly funded entities, including all Australian Government agencies, should create comprehensive, easy to access registers of data, including metadata and linked datasets, that they fund or hold. These registers should be published on data.gov.au. Where datasets are held or funded but are not available for access or release, the register should indicate this and the reasons why this is so.

States and Territories should create an equivalent model for their agencies where such registers do not exist. These should, in turn, be linked to data.gov.au.

A reasonable timeframe in which to achieve this is within one year (by March 2018).

Crowdsourcing ideas for 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. The Australian Government recently recommended formation of an External Reference Group (comprised of experts from academia, business and non-government organisations), and has created a Deputy Secretaries Data Group. Such bodies may provide clearer channels for some data users to signal which government datasets should be considered for prioritised release.

While these developments are helpful, they lack structural substance. In private sector contexts, the increased significance of data has altered business models and the structure of whole markets. In the public sector, this has not occurred (appointing a Chief Information Officer is not a structural change).

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

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.

A formal process within which prospective data users advocate for release by describing to governments the scope for gains from release could be a useful and early move to develop better advice on where release may create national welfare gains. Data users (groups or individuals) could be asked on data.gov.au to submit detailed proposals — the equivalent of public interest assessments we propose later for the National Data Custodian (NDC) to use when assessing NIDs (chapter 7), but with a focus on private investment opportunity as well as public benefit.

Such an approach would assist some agencies in prioritising datasets for release.[17]

In assessing these proposals, and ultimately determining whether to respond, government agencies should consider whether data users are able to perform as proposed. Central agency co-ordination of responses would be valuable, to ensure responsiveness.

While such processes might be viewed as too slow or clunky by some entrepreneurs, on the whole, seeking views from companies, researchers and other users on what data to release (and why) is less likely to lead to poor data release outcomes.

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

Early adoption of such an initiative 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.

More guidance on periodic data release competitions is in box 6.1.

Under *current* machinery of government arrangements, responsibility for obtaining feedback from data users (on either an ongoing basis or through periodic events) and ensuring responsiveness to a data release competition would reside with a central government agency.

State and Territory governments could consider developing their own structure to be implemented by a central agency at the jurisdictional level.

[17] Quantifying the benefits of data use is difficult (table 2.1). In assessing such proposals, government agencies should carefully assess whether the estimate of 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).

Box 6.1 Data competitions

Formalised data competitions could be useful for drawing attention to the Australian Government's high value dataset process, 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 considered 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.

But successful selection would not be a question of voting or other popularity contests. Boaty McBoatface is not an outcome desirable in data release.

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 any public sector datasets. Second, the timeframes involved would necessarily differ in order to allow participants sufficient time to develop and submit detailed proposals and crucially, for organisers to assess submitted proposals.

Responsiveness and accountability

While there may be a range of legitimate reasons for entities to withhold datasets, inertia and risk aversion are addressable barriers to greater access. A competition-based structure for prioritising datasets for release would need to see agencies properly 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 a failure of the agency, and would trigger remedial efforts to address this.

As with other large cultural and technological shifts in response to digitisation, setbacks are inevitable in some early projects. The new structure should explicitly adopt the principle — and insist on it in crisis-style responses — that agencies primarily seek to learn from these setbacks, in order to transition to a culture of openness.

Agencies (as custodians of data) 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. Where a data request is not met for a legitimate reason (such as privacy), agencies should explain this and consider whether there are other data assets that could be useful in meeting the purpose of the request.

The permission to undertake release is inherent in this Report's recommended new structure, particularly its legislative basis. Modelling for all data custodians a desirable

approach to data re-use and research is a primary reason for why a new Act is necessary. Advice to the Commission suggests that this *permission shift* was inherent in the legislation that supported the NSW Data Analytics Centre.

In the absence of such breadth of change, however, we cannot envisage data contests as being anything other than a minor and ineffective reform in their own right.

RECOMMENDATION 6.5

In determining datasets for public release, a central government agency in each jurisdiction with overarching policy responsibility for data should offer a public process whereby datasets or combinations of datasets can be nominated, with a public interest case made, for release.

A list of requested datasets, and decisions regarding dataset release or otherwise, should be transparent and published online, in the Commonwealth's case on data.gov.au.

6.2 A National Data Custodian

Comprehensive reform to data availability and use will require a change to the culture around the sharing and release of data. While establishing an institutional structure for data sharing and release is necessary, it will also be important to ensure consistent leadership and technical direction for implementation of reforms to Australia's data infrastructure, and in building trust between jurisdictions and across sectors.

Who would do what?

Oversight, broad implementation of Australia's new data Framework, and responsibility for the introduction of a new Data Sharing and Release Act (DSR Act — see chapter 8), would be handled in the first instance by the central government agency with policy responsibility for data (in the Australian Government, this would currently be the Department of Prime Minister and Cabinet).

The Commission has recommended that a new national statutory office holder, the National Data Custodian (NDC), be established that would enable a national (beyond just one Australian government) focus on data access, ensure ongoing leadership and technical direction for data access in Australia, and withstand changes in governments. Once it is up and running, the NDC would assume responsibility for the implementation, reporting on and evaluation of Australia's data Framework.

A new national statutory office holder …

The NDC would be an individual with significant expertise in data and the benefits that can flow from data use, but would also have an understanding of community attitudes to data use and the importance of maintaining a social license for that data use.

… within a small independent authority …

We see the NDC being supported by a small independent authority — the Office of the National Data Custodian. The intention is that the NDC would directly employ no more than a handful of staff, relying on external expertise to supplement its own capabilities as needed. The NDC would report directly to a central portfolio cabinet minister, consistent with the view that data is an asset to be actively managed and invested in but there is a risk of scope and mission creep (or shrinkage) if housed in any other location.

Such a body would have the capability to facilitate a coordinated, whole of government approach to data sharing and release, whilst maintaining public trust. It should have no 'skin in the game' for delivery of any particular dataset, since it is to have a system-wide standards and accreditation role.

… and a technical advisory board

As part of its internal structure, the Office of the NDC should include a small advisory board, with an orientation towards technical skills. The advisory board could comprise appropriately qualified members from the research community, relevant Commonwealth, State and Territory Government agencies, or the private sector. The make-up of the board is important to ensure that the NDC is established as a *national* entity, rather than a federal or state body, and the focus should be squarely on technical expertise needed for performance of the NDC functions, not representation of particular groups or interests.

NDC responsibilities

The NDC would oversight the operation of the national data system and be given a number of specific responsibilities under the DSR Act:

- Guidance — The 'rules of the game' for data sharing and release, established by the DSR Act, would be given shape through guidance issued by the NDC. Or where guidance already exists, the NDC would be able to endorse it as best practice. This guidance would cover de-identification processes, risk management by Accredited Release Authorities (ARAs) and data custodians, curation standards and metadata, secure storage and access procedures, and trusted user approval processes (including a model Data Sharing Agreement). Guidance material would be updated as required to deal with changes in data practice and risk. The NDC 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).

- Responsibilities with regard to ARAs — the NDC would establish, update and implement accreditation processes for release authorities, including certification of the capabilities and standards of ARAs under criteria that may be specified in the DSR Act; decide which ARA is the most appropriate to manage specific datasets; be a party to ARA funding determination and allocation; be responsible for the auditing of ARA functions at its discretion for compliance with standards and terms of Commonwealth funding; remove accreditation for non-compliant ARAs; and promote a cooperative national network of ARAs.
- National Interest Dataset (NID) responsibilities — encourage datasets of significant national value to be nominated for designation as NIDs (chapter 7); accept nominations for NIDs and undertake preliminary assessment of their public interest merits; refer selected nominations to a parliamentary committee for public scrutiny; support the responsible Minister in issuing disallowable instruments for purposes specified under the DSR Act, including designation of NIDs.
- Public reporting and complaints handling — be a focal point for data users (including trusted users) and data custodians with complaints about implementation of the Data Sharing and Release Act; receive reports from data custodians and ARAs of proven or suspected data breaches and misuse; report publicly on proven data breaches and misuse instances; and report publicly on 'good news stories' on beneficial uses of data.
- Ethical oversight and direction — review and advise the responsible Minister on emerging opportunities, risks and ethical issues associated with data use.

In undertaking these responsibilities, it is imperative that the NDC *not* adopt a 'clearance house' approach that would likely lead to the development of bottlenecks, limiting data release (Association for Data-driven Marketing and Advertising, sub. DR275, p. 10). The NDC, as envisaged by the Commission, is primarily a *facilitator* of improved data access and best practice — not an additional step or barrier in access processes.

The issue of ethics received significant attention in submissions to this Inquiry. For example, datanomics (sub. DR300) suggested that a formal structure for ethical considerations be established within the NDC. The Commission considers that there should be a role within the NDC for a *dedicated ethics adviser* with expertise in data, who could give overall guidance on ethical issues, and provide advice and expertise. The ethics adviser would consider strategic and system-wide issues, such as the ethical considerations inherent in using certain types of data, or the use of algorithms. In turn, this would bolster confidence in the system for data release.

Critically, the NDC has to be capable of instigating disallowable instrument processes, allocating funding, and broader stakeholder engagement. Having the right leadership, culture, skills and expertise will be critical to the success of this office.

Early appointment of the NDC (prior to passage of legislation, so using administrative means) would demonstrate a commitment to the introduction of the new Act, and enable early development of guidance material and risk management approaches that will be

critical to enabling organisations to embed best practice and therefore in securing and retaining community support for reforms.

RECOMMENDATION 6.6

The Australian Government should establish an Office of the National Data Custodian (NDC) to take overall responsibility for the implementation of data management policy, in consultation with all levels of Government.

The Office of the NDC should have responsibility for:

- broad oversight and ongoing monitoring of and public reporting on Australia's national data system and the operation of the new Data Sharing and Release Act (recommendation 8.1)
- preliminary assessments for, and recommending designation of, National Interest Datasets (recommendation 7.1)
- accrediting release authorities, be party to determining a funding agreement for Accredited Release Authority (ARA) activities, and promoting cooperation between ARAs
- managing complaints about ARA processes
- providing practical guidance material for ARAs and data custodians on matters such as risk management, data curation and metadata, data security, data de-identification and trusted user models
- advising on ethics and emerging risks and opportunities in data use.

The Office of the NDC should include a small advisory board, comprising members with technical skills related to the NDC's activities; and a dedicated ethics adviser.

The NDC role should be filled administratively by the end of 2017 to be operational by the time that new draft legislation for data access is completed for public consultation (Recommendation 10.2).

RECOMMENDATION 6.7

The National Data Custodian should streamline approval processes for access to data by:

- issuing clear guidance to all Australian Government data custodians on their rights and responsibilities, ensuring that requests for access to data they hold are dealt with in a timely and efficient manner and are consistent with the risk management approach to be adopted by Accredited Release Authorities (ARAs)
- requiring that these data custodians report annually on their handling of requests for data access, including requests from ARAs.

State and Territory governments may opt in to these approaches to enable use of data for jurisdictional comparisons and cross-jurisdictional research.

6.3 Accredited Release Authorities

Although leadership from a central agency, such as is now in evidence in many jurisdictions, could encourage incremental cultural reform, data-holding agencies will, in the absence of institutional change, remain the masters of their own level of commitment to centrally authored change.

Such a model of incremental change will never be broadly effective, given the onslaught of data collection and use pressures now facing data custodians (chapter 1). Complex approval processes, fragmented data releases, distrust within and between jurisdictions, and a general culture of risk aversion will persist.

These issues require a comprehensive solution — namely, institutional reform that delivers a widespread *determination to release*, plus an injection of capability to do so wisely; along with a point of focus and coordination between the Australian, State and Territory governments. Therefore, a key element of the recommended risk-based framework for improving availability of data is the formation of ARAs.

ARAs would play an integral role in facilitating trust and cooperation in the use and release of data. These entities would provide sector-based capability and expertise to increase data sharing and release. ARAs would be responsible for a broad range of technical and implementation functions in dataset curation, sharing and release, as well as interacting with data custodians and data users within their sector, across all relevant jurisdictions. They would have governance processes to make them attractive partners for developing and sharing datasets, once legislative obstacles can be removed (under the DSR Act — see chapter 8).

While discretion and flexibility would be required in operating structures and practices, and not all ARAs would operate identically, minimum safeguards, standards and recommended approaches are needed. For this purpose, the term 'accredited' is given to ARAs with intent: accreditation processes (and the link via them to funding) would be set up and maintained by the NDC to ensure constantly updated, fit-for-purpose process and performance behaviour by ARAs.

What are Accredited Release Authorities?

ARAs should preferably be public sector entities or agencies (Commonwealth, State or Territory), other publicly-funded institutions or not-for-profit entities that have a focus on release of data for public interest uses. Such institutions should, where possible, be built on an existing base, to make use of existing technical and sectoral expertise and social licence. But more than one may need to be purpose-built in sectors where there is great potential for data to be applied but minimal history of doing so. Private entities are not excluded from being ARAs, but it is less likely that a private entity could demonstrate to the NDC that they have the desired characteristics.

ARAs would be accredited by the NDC in their processes, governance and objectives and have responsibility for releasing or sharing data, usually drawn from a particular sector of activity.

Publicly funded bodies such as research organisations may meet the criteria to be an ARA. Indeed, a number of them have indicated that they see themselves as candidates to be an ARA (box 6.2).

Box 6.2 Participants' views on Accredited Release Authorities

A number of participants responding to the Commission's Draft Report identified themselves or other organisations as potential Accredited Release Authorities (ARAs). The Office of the Australian Information Commissioner, for example, pointed to the relative paucity of existing permission and capability:

> … there are only a handful of bodies in Australia who have been assessed as having the requisite expertise and safeguards in place, for example the Australian Bureau of Statistics, the Australian Institute of Health and Welfare, and the Australian Institute of Family Studies. (sub. DR236, p. 17)

The National Archives of Australia submitted:

> The Archives' role in releasing records for public access is largely that envisaged by the Commission for an 'Accredited Release Authority'. The Archives would be happy to share its knowledge and experience in managing and providing access to records of national significance. (sub. DR252, p. 5)

The Australian Urban Research Infrastructure Network (AURIN) stated:

> AURIN's experience providing a complete data release and delivery service would position it well to be an Accredited Release Authority. (sub. DR260, p. 1)

The Australian Government Linked Data Working Group (AGLDWG) argued:

> The AGLDWG as a community of Commonwealth Government experts and champions with members from 10 federal agencies, who have been drafting policy and technical guidance on the implementation of Linked Data for the Australian Government, is keen to become an accredited release authority … (sub. DR278, p. 3)

The Australian National University-based Australian Data Archive submitted:

> … there should be consideration given to providing the capacity for organisations outside government to be approved as ARAs. The Australian Data Archive has in fact fulfilled a role very similar to that described in the Draft Report for numerous government department agencies, including 10 different agencies at the time of writing. (sub. DR288, p. 8)

The Australian Taxation Office (ATO) stated:

> The ATO would be willing to serve as an Accredited Release Authority. (sub. DR314, p. 2)

The Bureau of Meteorology submitted:

> The Bureau … has particular interest in becoming an ARA, given its experience and the established need for it to disseminate critical information and value-added data through a wide range of channels to end-users. (sub. DR322, p. 2)

The Commission considers that a model on which ARAs could be based is the AIHW (box 6.3), which analyses and releases a range of research and other 'people-related' data.

Box 6.3 The Australian Institute of Health and Welfare

The Australian Institute of Health and Welfare (AIHW) is an example of an organisation which could serve as an Accredited Release Authority (ARA) under the Commission's new data structure. The AIHW was established in 1987 as an independent corporate entity under the Australian Government health portfolio.

The AIHW's Data Governance Framework outlines a number of internal data management functions as defined by the Data Management Association International:

- data architecture management — entails defining the blueprint for managing data collections
- data development — comprises analysis, design, implementation, testing, deployment and maintenance
- data operations management — relates to the provision of operational and technical support from data acquisition to purging
- data security management — deals with matters of privacy, confidentiality and appropriate access
- reference and master data management — involves managing information on standards and the business of the organisation
- data warehousing and business intelligence management — enables reporting and analysis
- document and content management — relates to managing data held outside of databases
- metadata management — integrating, controlling, and providing metadata
- data quality management — defining, monitoring, and improving data quality (AIHW 2014b).

Besides these internal data functions, the AIHW also has responsibilities to release data. Its main role is to collect, analyse, and disseminate health and welfare-related information and statistics (AIHW 2014a). In describing its data responsibilities and activities, the AIHW submitted to the Commission:

> We provide timely, reliable and relevant information and statistics on hospitals and other health services, aged care, childcare, services for people with disabilities, housing assistance, youth justice and other community services.
>
> We collect data and manage national data collections in these areas … We also develop, maintain and promote data standards to ensure that data collected are nationally consistent. (sub. 162, p. 1)

A number of committees exist within the AIHW's governance structure, one of which is a Data Governance Committee, tasked with making recommendations in relation to data governance and data-related matters to the Institute's Executive Committee. The Data Governance Committee comprises seven senior staff members of the Institute (AIHW 2014b).

The AIHW data governance structure also incorporates internal 'data custodians'. These custodians are staff members who exercise responsibility for a specific data collection in accordance with the guidelines, policies, legislation, and any specific conditions for use that are applicable to the data collection. Data custodians have the authority to release data to other organisations or persons (AIHW 2014a).

As noted in section 6.2, the NDC would accredit ARAs based on criteria that should be set out in the DSR Act; and take principal responsibility for the continuing improvement of data sharing and release in practice. Actions taken by an ARA to facilitate data use, sharing

and release go beyond what the data custodian can do — for instance, a high risk integration project, or data involving multiple jurisdictions or sectors.

Accreditation criteria should particularly specify that the ARA:

- can demonstrate it has suitable governance arrangements in place to carry out its functions
- formulate, implement and regularly update a risk management plan for data handling and release, that is consistent with guidelines issued by the NDC
- is well placed to take a national, public interest, rather than a jurisdiction-specific approach to data
- has the technical capability to manage the risk of data sharing and release — such as de-identification, administration of trusted user models. The required capability will likely vary depending on the sector and the nature of the data being handled.

These criteria should be applied in a flexible and outcome focused way, recognising that given the variety of data and sectors in which an ARA could have expertise, there would be no one-size fits all approach to ARAs.

A national approach and suitable governance structure

The criteria set out in the DSR Act should not include prescriptive governance arrangements for ARAs. In reality, what an ARA would look like could vary greatly from one body to another, and suitable governance arrangements could vary greatly depending on the body seeking accreditation. Importantly, *not* all ARAs would be Australian Government level bodies.

Jurisdiction-specific bodies such as the NSW Data Analytics Centre might also be a potential ARA if the New South Wales Government chose to expand its remit (underpinned by its experience and resourcing) to take a national, rather than jurisdiction-specific approach to functions, new and old — for instance, by entering into cooperative agreements with other jurisdictions to ensure that data is exchanged and shared on a reciprocal basis. The way a national approach is achieved is not as important as the willingness to allow it to be achieved. The ARA model is designed to *overcome* sectoral distrust and promote cooperation between jurisdictions, not to arm turf wars and deepen entrenched fiefdoms.

The Commission's recommended model for the structure of ARAs is for them to be established as independent and probably incorporated entities, rather than divisions within government departments. An independent structure would have the advantage of giving the ARA autonomy, which is especially important for the purposes of data release, where political considerations may prevent or delay the publication of certain data.

Governance structures will be vital to obtaining better Commonwealth-State cooperation, and for those ARAs whose focus is on areas of split jurisdictional responsibility, this effectively rules out a government department, State or Commonwealth. Similarly, it may affect entities whose authorising legislation prevents such governance.

A further benefit of this structure is that it simplifies the process of holding ARAs accountable for decisions — the ARA alone would be responsible for the actions it undertakes, eschewing the complicated accountability structures that can exist within large government departments. In saying this, the Commission is not sanguine about reputation damage — data collectors would be exposed too if serious issues arise. But the focus of review, learning and responding is clearly that of the ARA.

Where independent corporate entity status is not feasible but the NDC determines that an entity has the required expertise to be an ARA, fit and proper alternative governance structures and procedures and protocols must be in place. This would build trust in the organisation and ensure data is handled in accordance with a risk management approach.

Involvement of States and Territories in enabling wider access

Selected sectoral entities located within the States and Territories could also be established as ARAs where such functions are primarily State-based, but nevertheless of national significance, to promote wider usage and release of data.

Equally, if a particular State agency, through its capabilities and the respect with which its efforts in data are held by counterparts, is the obvious choice for locating an ARA that combines Commonwealth, State and other data, this Report envisages that it should be so appointed; and take on the functions and relationship with the NDC envisaged under the DSR Act.

Based on prior experience across many inquiries since 1998, we could envisage national ARAs perhaps being accredited in areas such as health and welfare, infrastructure, environment, education, geoscience, communications, taxation, immigration, industry and science, and defence. State and Territory ARAs might cover such nationwide cooperative data management for release or sharing in planning, justice, and land transport. Alternatively, State and Territory ARAs could assume data responsibilities for other States and Territories in certain sectors, where an agreement to do so is made.

State or Territory based ARAs should be accredited by the NDC to provide consistency and give the community more confidence in allowing ARAs to handle data from multiple jurisdictions. At a minimum, standards and governance practices must be common. The adoption of the form of an ARA, and its title, should involve a commitment by its authorising government to a long-term investment in capability, with legislated support.

Appointment under the DSR Act would seem to be the simplest approach. Simply re-labelling an existing sectoral data managing entity (a number exist in various forms)

should not qualify for Commonwealth support. New labels do not change cultures or practices. As with designation of datasets as NIDs (chapter 7), investment by the Commonwealth in establishing ARA capability is meant to create additional sharing and release activity.

In an expertise sense, an organisation that could be accredited as an ARA is the WA Data Linkage Branch, established in 1995. The Branch's capabilities and safety record with data handling are well established. It operates however, within the Western Australian Department of Health (also the Branch's major funder), and has responsibility for managing the WA Data Linkage System, as well as offering client services (Data Linkage WA 2016). The Branch securely links data collections from a range of sources to enable such activities as research and planning (appendix C), which are activities that ARAs would perform. The need for governance arrangements that make effective cooperative arrangements amongst jurisdictions would need to be solved; as would the ability to be accredited by the NDC. Earlier, the Commission noted its reservations about this being consistent with the role of a Department of State at Commonwealth level.

Similarly, SA NT Datalink was created in 2009 as a collaboration between South Australia and the Northern Territory, and enables the linkage of administrative and clinical datasets for research and policy use (SA NT DataLink 2016). Again, this organisation performs functions that might provide a foundation for an ARA, including linking administrative and clinical datasets, as well as providing infrastructure for researchers to enable them to gain access to information (appendix C).

RECOMMENDATION 6.8

Selected public sector and public interest entities should be accredited as release authorities. Accreditation should be determined based on sectoral expertise, capability, governance structures, and include consultation throughout the relevant sector.

Accredited Release Authorities (ARAs) would be responsible for:

- deciding (in consultation with original data custodians) whether a dataset is available for public release or limited sharing with trusted users
- collating, curating, linking and ensuring the timely updating of National Interest Datasets and other datasets
- offering advice, services and assistance on matters such as dataset curation, de-identification and linking
- providing risk-based access to trusted users.

ARAs should be fully operational from the beginning of 2019.

The number of ARAs

Given that ARAs are envisaged as sectoral centres of excellence, with an established track record and leadership position within their sector, it is likely that a relatively limited number of such entities would exist (or be created).

Where responsibilities within a sector are split between the Australian Government and State and Territory Government level, such as in health, the Commission envisages that there should be a single *national* ARA. Again, the AIHW serves as a possible model in this regard. National ARAs would have the advantage of reducing the need for a body in each jurisdiction to be accredited as an ARA. This structure would also streamline data processes, such as the ability to combine data from different jurisdictions into a single accessible dataset with uniform standards and definitions.

Furthermore, the establishment of national ARAs could also provide scope for an improvement in public services. Consolidation of data across national and State and Territory levels may enable improved targeting and administration of policies and programs via efficiencies that arise from the collection, release and analysis of data. It will become evident that some data is collected or developed unnecessarily. The Productivity Commission has had its own experience rationalising data collection in the 12 000-odd data points of the annual Review of Government Services.

The Commission considers that it should generally not be necessary to establish multiple ARAs within sectors. While data collected within sub-sectors might be developed to different standards and be of varying quality under present arrangements, consolidating holdings in an ARA has the potential to facilitate an improvement in data quality. Given that ARAs would be responsible for the management and curation of data that has been transferred to them (discussed below), their data expertise could lend itself to improved quality of collections. Where ARAs are unable to improve internally the quality of data holdings due to poor collection methods, for example, ARAs could make users aware of the limitations of particular datasets.

Fear of data being misunderstood by researchers or the public is another principal impediment to data release, and one that ARAs are capable of overcoming by constant reiteration of warnings. The Commission has done this successfully in many areas in the Review of Government Services.

Avoiding appointment of a large number of bodies accredited as ARAs would diminish system-wide complexity and reduce the potential for overlaps. Considering national, and State and Territory ARAs altogether, a limited number, almost certainly with an upper boundary of perhaps 12–15 bodies in Australia, might be accredited as ARAs.

Ethics committees and ARAs

ARAs may need an ethics committee for some of their activities. Ethics committees would be of particular importance for ARAs that handle and share identifiable information with frequency, and where those granted access to identifiable information for such purposes as research do not have similar organisational structures established.

Where an ARA has its own ethics committee, this should contribute to reducing the role of participating data holders' own ethics committees. More broadly, ethics committees should either mutually recognise approvals by other ethics committees or, in the absence of other relevant ethics committee involvement, seek only that information necessary to assess the ethics of a proposal for which identifiable data is requested. In the former case, the ARA may consider having researchers or users sign a deed outlining precisely what use is permitted, and where identifiable data is involved, obtain assurance via the deed that the data would not be re-identified. Data users, and their host institution, should have their right to future access on the line in support of this requirement.

Where those seeking use of data are based overseas, a bond or other form of financial guarantee may be sought where application of Commonwealth law (if the current Bill is approved) and other forms of assurance, or the threat of reputation damage in an Australian context, are insufficient or of uncertain reach.

Where overseas-based researchers and organisations work with, or are sponsored by, Australian institutions, the bulk of responsibility should be borne by the latter. Guarantees and undertakings would be easier to enforce, and domestic institutions would likely have greater incentives to maintain a positive reputation to preserve their ability to continue carrying out research.

Roles of Accredited Release Authorities

ARAs would have a number of roles, and their capabilities could come from a combination of internal proficiency and external contracts. There is no one-size-fits-all approach to this. Some bodies may develop a comparative advantage in one or two technical areas (such as de-identification or coordinate reference systems). The Commission envisages that ARAs would generally limit external contracting to areas in which they did not have a comparative advantage.

Importantly, the Commission does *not* envisage that accreditation of an ARA means that they are necessarily the only entity authorised to hold, link, and/or release that data. Original data custodians may persist with current activities, and/or leave further development and linkage work to ARAs, for example.

But as risk rises or integration and governance of shared datasets requires, so ARAs come into their own. Given the significant benefits from cross-sectoral linkage, there is no reason to entrench one ARA's monopoly position in a particular dataset, or define the appropriate

scope or breadth of an ARA — these things cannot be foreseen. The Commission anticipates the development of data standards (such as on curation practices and trusted user access) by the NDC would go a long way towards facilitating cooperation between entities.

ARAs must encourage a much wider release and use of data than is presently customary in Australia. They would streamline existing processes and act as a sectoral hub that provides a readily identifiable and clear access point for data users (if not, the investment in them would be substantially devalued).

They are not designed to act as an additional approval layer on top of existing data custodians, but rather are to improve datasets, encourage use, minimise the need for multiple ethics committees and release in such a manner that provides societal benefits.

Funding should be withdrawn if it emerges that ARAs found, after evaluation by the NDC, to be impeding access to data. Since it is expected that the Commonwealth would, via the NDC, be the sole funder of ARAs, withdrawal of funding is both a plausible and practical threat.

Data release

As the term 'release authority' implies, the main focus of ARAs would be on releasing data. The word 'release' in the title of an ARA is not intended to be limiting: most ARAs might do as much if not more sharing of data amongst cooperating data holders and trusted researchers. Integration of datasets is another key function, as is curation. The release term signifies the highest (often perhaps most risk prone too) purpose of such an authority. It is release that ultimately should demonstrate how far governments are prepared to go in their quest to be more open with data handling and reuse. Entities seeking to be ARAs should not do so for status reasons, or for the resourcing that must go with it. The purpose is inherent in the name: taking action in support of wider release.

ARAs would be responsible for improving access to the full spectrum of data — it could be data that originally pertained to individuals, NIDs, and/or any other types of data. Depending on risk, data may be entirely open, closely held, shared with trusted users only, and may be de-identified, identifiable, or non-confidential. ARAs would, by their institutional brand's standing, become integral in preserving the social licence to utilise data and disseminate it publicly within a formally established risk management process.

It would fall to ARAs to determine, in applying their risk management framework, what modifications would need to be applied to datasets should they be publicly released in some form, or whether some data should only be shared with trusted users (section 6.4).

Research organisations, government entities, and private sector organisations would still be able to release data that they collect, much as they do under current processes.

An additional benefit of the data release role of ARAs is that it may assist in improving data discoverability, which has been a significant barrier to public sector data use, and the use of research data. By acting as a central sectoral hub for data release, ARAs would be able to make data holdings known to a wide variety of potential users, and potentially be a data 'go to' point for their sector.

Interaction with data custodians

ARAs would act as a hub of expertise for data, and an important element of this status would entail the management of sectoral data assets. Data management would comprise such activities as acquiring, storing and processing datasets belonging to the sector for which the ARA is responsible. ARAs would *not* be primarily responsible for the collection of data.

We envisage that data custodians would also retain the obligation and ability to share and release data on a business as usual basis. For instance, if two government agencies wished to exchange data or collaborate on data projects where they were the original data custodians, they should remain free to do so *without* ARA approval or involvement.

Dataset custodians would retain their legislated responsibility for original data contributed to ARAs and so make the choice regarding its release or sharing in that *original* form; but where possible (given legislative and other requirements), an ARA would have the primary responsibility for *transformed* datasets (for example where original data has been transformed via integration, linkage, de-identification or use of Artificial Intelligence).

The ongoing maintenance and updating of a transformed dataset would necessitate cooperation between ARAs and custodians of component datasets. Dataset custodians that have provided data to ARAs would remain obliged to provide updates to that ARA. Duplication of processes, which may raise doubts about which is the 'true' dataset, must be avoided. One option to minimise the risk of duplication is to set up automatic transfers of data from the custodian to the ARA, as the dataset is updated. After risk assessment, in some cases ARAs could also use remote secure access to transform the data, while the custodian continues to store the dataset.

Building skills

In many agencies that collect and handle datasets, significant (often informal) skills already exist that should not be underestimated and can be built upon. ARAs should be in a position to provide resources to assist with the development of skills in other entities, including data custodian entities. In turn, this would improve data practices in agencies and hence provide support for more informed policymaking, which could be particularly significant given the vast range and types of data that ARAs would manage. There are a number of avenues by which this could occur — ARAs could:

- second staff to entities that make significant use of large, complex datasets, and provide assistance with matters such as managing data and conducting analyses.
- provide training sessions to other entities to share their expertise on data curation and management, and serve as a means of allowing data users to exchange information on analytical techniques and procedures.
- provide (possibly for a charge not greater than cost recovery) formal advice on matters relating to data in their sector.

ARAs may also wish to use bodies such as CSIRO's Data61 as a limited example for skill-building activities, as long as such activities do not divert these bodies from their core purposes (Data61, for example, is not a data releasing entity).

Broader issues around skills requirements and development are discussed in greater detail in chapter 10.

Linkage

The Commission envisages that ARAs would be available to perform linkage for their sector, or be responsible for the coordination of linkage activities. This would have the advantage of drawing on their expertise and extensive knowledge of datasets relevant to their sector, and reduce the fragmentation of responsibility for data across many different bodies. Where linkage is being contemplated by other data custodians or organisations, drawing on the expertise and advice of ARAs remains an option.

Linkage activities could be sub-contracted externally by an ARA where it does not possess sufficient in-house expertise, or where there are competing demands on its resources. ARAs would need to ensure strict control and oversight of sub-contractors, and have formalised risk management plans (discussed below) in place should sub-contractors be used.

Curation

Datasets often require significant ongoing work after collection to ensure usability across time. Active and ongoing data curation would be a primary activity undertaken by ARAs, including curation of NIDs. Some existing organisations already fulfil a similar role — for instance, the Australian Data Archive is a national service for the collection and preservation of digital research data, and makes data available to researchers and other users. While data custodians could curate their freely released data, as sectoral experts, ARAs could advise and assist data custodians on this.

Fit-for-purpose de-identification and encryption

Since ARAs would interact heavily with trusted users (discussed below) and other parties with whom data is shared, ARAs may need to perform de-identification (removing or altering any information that identifies an individual or is reasonably likely to do so) and encryption (transforming information into a scrambled form) on a basis that is suitable for the needs of the trusted user or other organisation in question.

Where an ARA considers that fit-for-purpose de-identification or encryption is necessary prior to sharing or releasing data, or is requested by an organisation seeking access to data, approval should first be sought from the ARA's ethics committee (where one exists).

ARAs should be permitted to charge organisations for fit-for-purpose undertakings on (at most) a cost recovery basis.

Risk management would require careful consideration when an ARA performs fit-for-purpose de-identification or encryption. Procedures to directly address risk of breaches will be necessary and specified as part of accreditation by the NDC. Standard, access-friendly data use agreements (see below) may offer one method, if they do not become a lawyer's picnic.

ARA risk management in action

The release, curation, and management of data entails risks — be they from errors or malicious activities. One of the most contentious risks in this context is the possibility for an individual (or individuals), or individual business to be 're-identified' from a dataset that was 'de-identified' (box 6.4) prior to its public release.

A common sense and proportionate approach is needed. By applying an explicit risk-based approach to data access, ARAs (and data custodians) would be required to ascertain the nature, likelihood and consequences of risks for particular datasets, and establish how they might be mitigated.

ARAs would be required to make a variety of complex decisions relevant to the sensitivity of datasets, as well as how data should be released. While ARAs should release non-confidential, non-sensitive data as a routine process, the question arises as to how judgements would be made about the higher sensitivity datasets — that is, whether such data can be shared, or released to the public, and if so, what transformation or modification is required beforehand.

ARAs would have the responsibility for making an assessment about whether certain data ought to be regarded as sensitive. If the ARA considers that sensitive data should be released, any required transformation of the data to make it non-sensitive and suitable for public release must be outlined to users.

Data release would occur in accordance with accreditation arrangements set by the NDC as well as legal obligations to which the ARA was subject plus any internal policies and guidelines maintained by the ARA — including the organisation's risk management process.

In the case of shared access to data, responsibility for managing the risk of linking and/or sharing would depend on residual legal complexities (NIDs should not suffer from this, but other ARA datasets might). Consistent with any shared legal liability — for example, for breaches — the NDC may be asked to offer advice on better processes and practices necessary to manage these risks properly.

The Commission considers that managing the risks for data held by ARAs should be vested in ARAs. Making ARAs accountable for sharing or release would have the benefit of providing them with strong incentives to maintain effective, sector relevant risk management and mitigation practices and keep these updated in the context of technological, regulatory, and other environmental changes.

To fulfil their obligations, ARAs may wish to establish a risk management committee to review the design, implementation and maintenance of the ARA's formal risk management process. Accreditation by the NDC should address this.

Although the details of risk management approaches would vary among ARAs, it is important to emphasise that processes should be evaluated regularly (both internally and by external experts on a less frequent but still regular basis). The AIHW engages in a regular program of data audits, to ensure that governance arrangements, such as confidentiality requirements, are being observed (AIHW 2016b). The Commission considers this would ensure that processes are kept up to date as new technologies and threats emerge, provide a degree of transparency and raise awareness, and hence, ultimately improve accountability.

Managing relationships with external organisations should be a further element of an ARA's risk management approach, as the transfer of data between organisations can be an area of significant data risk. For this reason, ARAs could make use of Data Sharing Agreements as a means to manage risk.

RECOMMENDATION 6.9

All Accredited Release Authorities must have and publish formal risk management processes to effectively assess and manage the risks associated with sharing and release of data under their control.

Standardised, access-friendly Data Sharing Agreements should be implemented with external data providers and users to formalise the activities that can take place with identifiable and de-identified data.

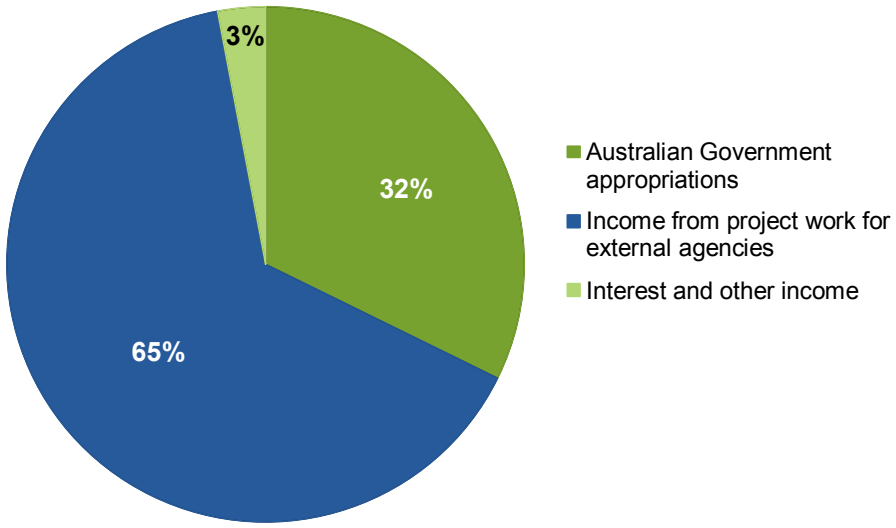
Risk management processes should be regularly reviewed and revised to account for new and emerging risks.

Funding for ARAs

As noted above, the Commission intends that one of the roles of the NDC would be party to the funding agreement allocated to ARAs for data management, curation, storage and access. As the NDC would also accredit ARAs it would have the advantage of understanding the complexity of the tasks required of each ARA. Consequently, the NDC would be able to advise on funding levels based directly on the activity levels of the ARA, and set adjustments as necessary.

ARAs could also source revenue from project work. This is the single most important source of income for the AIHW (figure 6.1), and comprises revenue earned by the AIHW's statistical groups for work undertaken for external agencies. The AIHW uses a pricing template to determine the fees necessary to cover costs — these costs include salaries and on-costs, other direct costs, and a corporate cost-recovery charge to recover infrastructure and corporate support costs (AIHW 2016a). Similar models could be considered by ARAs.

Figure 6.1 **AIHW revenue sources**
2015-16[a]



[a] Between 2010-11 and 2014-15, Australian Government appropriations averaged 33% of AIHW revenue, whilst income from project work for external agencies averaged 66% of revenue.

Source: AIHW (2016a)

Where ARAs undertake work for external agencies, they should have the power to charge and retain fees sufficient to recoup costs incurred.

It is envisaged that, in undertaking work for external parties ARAs would consider the implications for future accreditation and the reason they were resourced in the first place — not to act as paid consultants but to release data.

This policy would not be limited to Commonwealth entities acting as ARAs. Since funding for State/Territory/other appointed ARAs would largely derive funding from appropriations received based on the assessment of the NDC, the judgement as to what is the primary purpose for resourcing ARAs must be based on more than just the prospect of earning revenue from project work. Although the NDC can be expected to develop its own standards, some guidance may initially be obtained from the model of the AIHW and the WA Data Linkage Branch — for example, the latter body stipulates that it:

- charges cost recovery for all linked projects
- reviews charges on at least an annual basis
- maintains a charging structure that considers components including client services, linkage, extraction of linkage keys, geocoding, and cohort selection
- may apply additional charges for cross-jurisdictional projects given the extra complexity and time associated with such requests (Data Linkage WA 2015).

Approaches for pricing data are further discussed in chapter 9.

RECOMMENDATION 6.10

Funding of Accredited Release Authorities (ARAs), for the purposes of data management, curation, storage and access should be set via a funding agreement with the National Data Custodian.

ARAs should have the power to charge fees sufficient to recoup costs where ARAs undertake requested work beyond that envisaged in their funding arrangement with the National Data Custodian.

In assessing the scope to undertake such activities, ARAs must ensure they do not detract from their primary focus on the public benefits of enabling greater access to, and use of, data (which is the basis for their accreditation and funding).

Cross-sectoral issues

The Commission envisages that ARAs would be able to share data amongst themselves. Where a particular dataset requires ARAs to work together (for example, through the exchange or sharing of integrated datasets), they should have the option of deciding via formal agreements precisely how cooperative work would be undertaken.

Collaboration amongst ARAs has the potential to deliver organisational benefits to ARAs, including access to the expertise of staff in other ARAs, which could lead to diffusion of knowledge, new innovations, and resultant improvements in work practices. ARAs could be allocated specific funding for cooperative work, provided there were legitimate grounds and a necessity for collaboration. It is intended that ARAs be funded in order to achieve outcomes as well as quality of effort.

6.4 Trusted user models

Use of a formal risk management approach — widely accepted and endorsed in other spheres of management — is not common for data access. The Commission has endorsed the NDC providing harmonised guidance on a best practice approach to managing the risks of data access that does not lose sight of the opportunities of data use.

The five safes model (outlined in chapter 4) allows different levels of data sharing and release to take place by dialling up or down the various controls. For instance, data held by some ARAs and data custodians will often need to be shared with other agencies for service delivery purposes (think health or education). In such cases, the data shared may be identifiable, but the users are highly trusted, have security clearances and work within a computing environment that has security guided by government-wide criteria. 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 form. This sometimes, but not always, occurs through trusted access models at present. And where data is publicly released (for instance, on data.gov.au) an additional level of confidentialisation needs to be applied because other controls (such as managing access) are not available.

The Commission's model for sharing sensitive data that has not been classified as an NID, 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. Although the process for determining trusted users would be the same for NIDs and other types of data, trusted user access in the case of NIDs would be ongoing (not program-limited). The trusted user model could, and should, equally be used by other Commonwealth and State data custodians in making data sharing and release decisions.

In a similarly transparent fashion to that envisaged for NIDs, the following details should be publicly available on a website, such as on an ARA's website: 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. This would enable potential users of these to know of a dataset's existence and how to be approved for access to it. Further, it may be beneficial for ARAs to disclose those trusted organisations to whom they have granted data access, at the time that access is approved. This would create an additional mechanism for transparency and provide another means through which a social licence could be cultivated.

Safe people: Who are trusted users?

Trusted users would comprise individuals that ARAs regard as capable of using identifiable data in a responsible manner (NIDs and trusted users are discussed in chapter 7). Individual users must be securely and appropriately identified (as opposed to issuing organisational licences to access data, which limit knowledge of individual users).

Trusted users should be based in entities that have a signed legal undertaking that sets out safeguards for data use and recognises relevant privacy requirements (chapter 8). This could cover individuals from both public and private sectors, although most trusted users are likely to be located within government agencies (Commonwealth, State and Territory), universities, and specialist research institutions. Current instances of trusted user models in Australia are largely confined to the public and research sectors — for example, the Department of Social Services has collaborated with the AIHW to develop a trusted user model that allows researchers to access selected social services data remotely (sub. 10).

For the private sector and research institutions outside of governments, having a robust approach to trusted user processes with approvals by ARAs, is expected to improve their access to public sector data in particular. This, in turn, would give these groups a greater incentive to make their own data available for linkage with public sector data, and enable all data holders to benefit from the analytical skills of the broader research community. In the absence of these reforms to trusted user access arrangements, the ability of non-government sector to generate new value from data would be significantly curtailed and much more ad hoc, jeopardising the broader social and economic benefits that may be realised from greater access to data (chapter 2).

The outsourcing of activities to the private and not-for-profit sectors may provide a case for considering the extension of the trusted user model to these areas, in some instances. This should become a core consideration in future outsourcing, the obverse of considering data retention in outsourcing or privatisation.

Access for safe programs, rather than projects

Safe projects requires that the data be used for an 'appropriate' purpose. Once qualified as a trusted researcher, ARAs should not indulge in assessing the merits of using data for a particular project or purpose unless evidently likely to breach any residual restrictions that original custodians have applied to the use of the data. We consider that a *project*-specific accreditation of trusted users is also unnecessarily limiting of opportunities to use data. Many research programs span many years. The Commission considers it preferable for access to be granted on a program-specific basis.

The *sharing* of data where it is not put up for open release should be for a stated *public interest* purpose. A public interest purpose would not preclude there being some commercial benefit — many academics commercialise their research, and many government bodies employ private consulting firms. But it should not predominate.

The selection of trusted users may also carry over to the use of other datasets (such as de-identified data), where risk-assessed by an ARA's processes and capable of avoiding unnecessary duplication of effort and preventing delays in accessing data. This could be revised depending on the risk of the entity — for instance, if a private sector party was seeking access, this may warrant higher scrutiny and/or tighter controls. The NDC should cover this in its guidance.

Safe data

There are inevitable, but varying, risks attaching to data sharing and release. A common sense and proportionate approach, including with regard to processes such as de-identification, is needed.

De-identification can often be a valuable risk management strategy, enabling the release of data that would otherwise be considered too sensitive for dissemination. But de-identification processes are not foolproof (chapter 3) and do not operate in a vacuum. They are instead determined in part by the broader environment and institutions and protocols in operation:

> The decision of how or if to de-identify data should … be made in conjunction with decisions of how the de-identified data will be used, shared or released, since the risk of re-identification can be difficult to estimate. (Garfinkel 2015, p. 1)

As highlighted in section 6.2, we are recommending that de-identification techniques could be one area in which the NDC provides guidance and, where useful, auditing of de-identification processes used and assurance that best practice has been followed. During the course of this inquiry, the Commission has been made aware of examples where a 'boots and all' approach to de-identification has been adopted, even in instances where the broader operating environment has been quite secure. The resulting negative effects, in terms of both cost and data usability, appear to have been out of proportion to the likely risks.

The NDC's guidance should clarify such matters as:

- managing the risk of re-identification from linking a de-identified dataset with another dataset may lead to the re-identification of some records
- pseudonyms derived from identifying information may enable the reversing of a pseudonymisation process
- attributing disclosure (in which a piece of confidential information can be attributed to a data subject (Garfinkel 2015))
- differencing (the process of evaluating the differences between two or more datasets) between aggregated data can highlight features of individuals' data used in calculating aggregates (Australian Taxation Office, sub. DR314).
- distinguishing between the risk of uniqueness itself and uniqueness which may lead to re-identification
- managing the risk of spontaneous recognition, potentially via legal undertakings
- allowing researchers to request draft and final output while managing the risk of cumulative disclosure.

Apart from regulating to deter re-identification, NDC guidance should recognise that there are other strategies that ARAs could adopt to reduce the risk of re-identification attacks

and other misuses of data, such as requiring trusted users to enter into data use agreements. These data use agreements may for instance, be a preferable way of managing the risks of spontaneous recognition rather than perturbing the data. Data Sharing Agreements are discussed further below.

The Commission anticipates that information about firms too would generally be provided within a secure environment in a de-identified manner, where there the costs of being identified are considered significant. Importantly, the Commission recognises that it will sometimes not be possible to provide information about businesses that is not 'likely to re-identify' those businesses, particularly for very concentrated industries where there are two or three players in a market (for instance, airlines). ARAs with such data would need to address this, and any legislative restrictions, as part of their risk management processes.

RECOMMENDATION 6.11

The Office of the National Data Custodian should be afforded the power to require an audit of a data custodian's de-identification processes and issue assurance of de-identification practices used.

Safe settings and transfers

Risks associated with data transfer and storage would be managed through the use of secure computing environments, such as the SURE (Secure Unified Research Environment) system, or other secure environments depending on the risk of the parties. The CSIRO (sub. 161) listed a number of options that can be considered to ensure data is accessed securely, including remote analysis systems (which allow researchers to submit queries and receive results, rather than accessing the data directly) and virtual data centres (where the researchers get full access to the data through a remote link). Experience with these networks could be further built upon to develop a more expansive secure computing environment.

Further, there are measures that organisations can take to reduce the likelihood that data misuse occurs internally. For instance, it may be best practice for organisations to reduce the handling and viewing of sensitive data to the minimum number of staff necessary, to monitor and document all queries made of sensitive data, and take measures such as locating those working with sensitive data in different locations to other staff; or use air gaps (that is, fully disconnected systems).

The particular measures chosen by ARAs to manage risks would depend on the type of data handled by the ARA, the sector in which it operates, and its own internal structure and culture. ARAs would be best placed to tailor measures to their individual circumstances.

Safe output

The extent to which research outputs under the trusted user model should be reviewed by the ARA (prior to publication) should be flexible and reflect the ARA's assessment of risks. Where output is checked, it should be by a process that is not labour-intensive and does not unduly hold up use or publication of the work. The focus of any such check should be confined to assuring confidentiality requirements of a particular dataset have been satisfied.

Importantly, the level of output checking should take into account that trusted users may need to request draft and final output while managing the risk of cumulative disclosure. The NDC should issue guidance on this.

Managing risks through Data Sharing Agreements

Preventing misuse of data by trusted users, and penalising (in a proportional manner) misuse by individuals and their organisations, are further critical determinants of how effective the Commission's recommended approach will be. Penalties are discussed further in chapter 8.

Apart from regulating to deter re-identification, there are several strategies that ARAs could adopt to reduce the risk of re-identification attacks and other misuses of data. In the United States for example, data use agreements are required under the Health Insurance Portability and Accountability Act (US) Privacy Rule. The US's data use agreements typically address issues such as:

- specifying the permitted uses and disclosures of datasets
- stipulating who may use or receive data
- prohibiting data recipients from using or further disclosing data, except as permitted by the agreement and under law
- requiring data recipients to use appropriate safeguards to prevent unauthorised uses and disclosures not permitted under the agreement
- requiring that data recipients bind any subcontractors to the same provisions outlined in the agreement
- prohibiting recipients from re-identification and contacting individuals (Stanford University nd).

Data agreements have also been used in Australia (box 6.4). If utilised more broadly, Data Sharing Agreements should be access-friendly, promoting the use of data, rather than serving as an instrument of restriction. For trusted users in the research or private sector, we consider that both the individual using the data and the organisation should sign a legal undertaking before they access the data as this creates incentives for the organisation to enforce the individuals' compliance with the conditions of access. For public sector agencies sharing with each other or with an ARA, we consider the organisations entering

into a Data Sharing Agreement should be sufficient as the individuals are bound by public sector employment legislation and codes of conduct — any further requirements would be duplicative. This is discussed further in chapter 8.

Box 6.4 Australian Data Sharing Agreements

A number of Australian organisations have employed Data Sharing Agreements to outline the conditions under which data may be shared and used. For example, the City of Stirling and the City of Bayswater, both in Western Australia, formulated a deed of agreement for data sharing in December 2014 (McLeods Barristers & Solicitors 2014).

The Australia and New Zealand Organ Donation Registry (ANZOD) and the Australia and New Zealand Dialysis and Transplant Registry (ANZDATA) have constructed a Data Sharing Agreement for prospective data recipients. The Agreement outlines permitted uses and disclosures of data, terms regarding publications based on data analysis, and termination processes. The permitted uses and disclosures section of the agreement specifies that data recipients agree:

- to use data solely for the purpose of the relevant project, and not conduct additional analyses of the dataset without ANZDATA and ANZOD's written permission
- not to attempt to re-identify, or permit others to re-identify individuals or units/facilities represented in the dataset
- not to attempt to link the dataset to any other datasets from any source without ANZDATA and ANZOD's written permission
- not to release the dataset to a third party, except with the written approval of ANZDATA and ANZOD
- to use appropriate administrative, physical and technical safeguards to prevent the use or disclosure of the dataset, other than as provided for by the agreement
- not to present or publish information contained in the dataset that might make an individual or unit identifiable
- to notify ANZDATA and ANZOD immediately of any known or suspected breach of privacy arising from use of the dataset
- to ensure that any agent, including subcontractors, to whom the recipient provides the dataset, agrees to the same restrictions and conditions that apply throughout the agreement to the recipient (ANZOD and ANZDATA 2013).

We consider that Data Sharing Agreements should be a key part of trusted user models, and should be taken into consideration (where they exist) as part of accreditation by the NDC. A harmonised approach to these could have significant benefits and accordingly, the NDC should develop guidance on this.

The maintenance of the integrity of the system will rely upon the risk management approaches adopted by ARAs, data custodians and trusted users, and the incentives trusted users have to build lasting relationships with ARAs and data custodians. An ARA's risk management plan should seek to prevent data breaches, but also outline the steps to be taken should a breach occur. Breaches or suspected misues of data that have system-wide implications should be reported to the NDC.

Similarly, entities with trusted users should have agreed procedures in place to ensure that only named individuals who require access to identifiable data are given access permission, even taking measures such as monitoring access to the secure data environment, and seating staff in physically separate locations if necessary to help ensure that the integrity of data is maintained.

The necessity of regularly cooperating with ARAs and data custodians provides extra incentives to ensure that data is appropriately handled. Sloppy practices would damage the reputation of a researcher and the trusted user for whom they work; the desire to be granted access by ARAs and data custodians on an ongoing basis would militate against such practices.

The process of linking and/or sharing de-identified data in a secure environment with trusted users would be governed by standards set by the NDC. These standards (as adopted by the ARAs) may cover integration methods, deeds for trusted users to sign, methods of providing access, and other things necessary to ensure risk is managed properly. Audit of compliance with the standards would be done by the NDC — underscoring the need for the NDC be an independent, statutory office.

RECOMMENDATION 6.12

Accredited Release Authorities (ARAs) should be given responsibility to grant, on a continuing program-wide basis, data access to trusted users from a range of potential entities that:

- 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, and
- have a signed legal undertaking that sets out safeguards for data use and recognises relevant privacy requirements.

In assessing trusted user access, the ARAs should accept existing current approvals of the trusted user's work environment.

Trusted user status for use of identifiable data would cease for that user when they leave the approved environment, when a program is completed, or if a data breach or mishandling occurs in that same environment and/or program.

RECOMMENDATION 6.13

Accredited Release Authorities (ARAs) and data custodians should be required to refer suspected and actual violations of data use conditions that have system-wide implications to the National Data Custodian.

Clarification should be issued detailing how this process would interact with the *Privacy Amendment (Notifiable Data Breaches) Act 2017* (Cth).

6.5 Changes to data practices affecting researchers

The reforms outlined in this chapter and chapter 8 are in large part motivated by a desire to increase the ability of researchers to make use of important data that is not currently openly available.

Researchers too need to improve how they offer access whenever generating new datasets in the course of projects or programs of work.

Prioritising the openness of research data

The Commission has recognised a compelling public interest case for greater re-use of researchers' data, but progress on this has been limited to date. Explicit strategies should be adopted to promote greater discoverability and openness of research data, as well as preservation and registration of metadata.

A further important goal should be that individual researchers are not penalised when, in fact, it was their institution that was dragging the chain. Achieving the objective of greater discoverability and openness is important for the Commission's reforms, and would assist ARAs in discovering data relevant to their sector, and provide scope for the ongoing maintenance and curation of research data. As a result, more users would have knowledge of, and access to, a wider range of higher quality data.

In its Draft Report, the Commission recommended that public research funding be prioritised based on progress made by research institutions making data widely available to other trusted users. However, an alternative that would not penalise researchers based on institutional policies — yet still retain incentives promoting greater sharing of data and metadata — could be the construction of comparative tables of performance for release of publicly-funded research data.

The tables would rank the performance of research institutions in making data and metadata available to other researchers, and report changes over time. This would alert users and the broader public to progress made by entities in sharing data. Those entities lagging behind in the tables would have an incentive to improve their sharing practices to maintain their relationships with other entities and other researchers, and to avoid the stigma attached with being regarded as having restrictive policies.

In recent correspondence with the Commission, the NHMRC has indicated it is considering two key changes to facilitate greater access to publicly held and funded health and medical research data. First, the NHMRC is considering amending its Funding Agreement with Administering Institutions to require the latter to have the capability to make the data that arises from publicly funded research searchable at the metadata level, and for the data holdings to be accessible to users approved by data owners. Second, the NHMRC raised the possibility of amending its Research Grants Management System to enable collation and reporting on the metadata of datasets resulting from NHMRC funded

grants, allowing regular compliance reviews (NHMRC, pers. comm., 27 January 2017). The Commission considers that an approach requiring research institutions to specify the degree of data access permitted by others should be a required feature of end of project reports. Regular auditing of reports to determine the progress made in making data accessible to other researchers would enhance the accountability of this system and provide incentives to institutions to make progress.

RECOMMENDATION 6.14

Progress by individual research institutions receiving Australian Government funding in making their unique research data and metadata widely available to others should be openly published by those institutions, with reference to past performance.

All bodies channelling public funds for research, such as the National Health and Medical Research Council and Australian Research Council, should similarly require in future funding agreements with research applicants that data and metadata is to be publicly available, and publish the results of progress on this for their funded projects.

On completion of projects, research institutions should include in their reports details of when and how other researchers can access the project's data and metadata.

Reforming ethics committee processes

Numerous Inquiry participants attested to the substantial delays in the approval processes to access data, requiring reviews by multiple ethics committees and data custodians (chapter 3). Additional guidance to data custodians, as well as increasing the transparency around their operations would contribute to improved data availability and use.

Streamlining ethics approvals

Human research ethics committees (HRECs) have a very valuable integrity preservation role, and bring much needed sectoral and subject expertise to bear. However, under the current system, many research projects (primarily in the health sector) are faced with a daunting task of making a multitude of submissions to separate ethics committees, wasting substantial amounts of time and money (see, for example, Barnett et al. 2016).

The fundamental issue behind these problems is the lack of mutual recognition between the different committees. Although all committees follow the NHMRC's National Statement on Ethical Conduct in Human Research, they can each use different processes and require different types of information from researchers before deciding to approve a project.

Over the years, there have been multiple efforts to implement aspects of mutual recognition in ethics committee approvals. However, as discussed by the Senate Select Committee on Health (2016), progress remains slow. Some jurisdictions have introduced single ethical reviews for their public hospitals (see, for example, DHHS (Vic) 2015); in

other cases, institutions have come together independently to implement mutual recognition (box 6.5). However, there is no existing requirement for HRECs to register with the NHMRC and only a limited number of HRECs have chosen to be certified by the NHMRC (NHMRC 2016b). Further, while some States and Territories have signed up to a National Mutual Acceptance Scheme that provides mutual recognition of ethics approvals for research projects involving public health organisations, the scheme does not cover all types of research or all jurisdictions (ACT Health 2017).

Box 6.5 Simplifying ethical approvals — the Melbourne Academic Centre for Health

Through the Melbourne Academic Centre for Health, 12 institutions, including the University of Melbourne, research institutes and major hospitals, cooperate to promote medical research and improvements to patient care. To overcome the potential delays to research resulting from the need to obtain approvals from a number of separate human research ethics committees (HRECs), these institutions have signed a memorandum of understanding that streamlines both the ethical and governance approvals for research projects (University of Melbourne, sub. 148). This cooperation has a number of unique features.

- It includes among its participants five HRECs that are certified by the National Health and Medical Research council (NHMRC) to approve multi-site projects. Researchers looking to conduct a project in more than one hospital require only one approval, from any of these HRECs. One of the committees convenes each week, so that delays in reviewing projects are minimised. Mutual recognition is retrospective, meaning that new sites can join an approved project without the need for new HREC approval.
- Some of the institutions participating are part of the Victorian Department of Health's single ethical review system. This gives all members access to the National Mutual Acceptance Scheme. Victoria participates in the scheme, which allows multi-site project approvals given by certified HRECs in its public hospitals to be recognised in other jurisdictions (except Western Australia, the Northern Territory and Tasmania. The Commonwealth also does not participate in the scheme).
- Other institutions that have non-certified HRECs are able to continue reviewing projects that only involve one site.
- The institutions have also agreed to streamline and standardise the governance approvals required before a research project commences, including determining agreed timeframes for HREC and governance decisions. The governance approval process remains independent, and each organisation can decide whether to participate in a given research project, regardless of whether it has HREC approval (RMH nd).

The Commission believes that a combined effort will be required to streamline ethical approvals. Unlike the United Kingdom, for example, where a central model is enshrined in legislation and all ethics committees operate within the National Health Research Authority, the fragmented nature of the Australian health system has meant progress is more likely to be achieved by agreement between the various stakeholders (Rahimzadeh and Knoppers 2016).

In effect, this will require a more structured approach to the operation of HRECs, which should reduce the variation in the ways they operate and improve transparency. As part of this, the Australian Health Ethics Committee, which operates within the NHMRC, should develop uniform processes for certified committees to follow when considering multi-site projects. Beyond improving the timeliness and efficiency of reviews, using a uniform process will give confidence to institutions when they accept approvals given by a different HREC that appropriate consideration has been given to the ethical aspects of the research project. At the same time, governments must reach an agreement on mutual recognition of HREC approvals, to facilitate cross-jurisdictional research.

Streamlining ethics reviews will not remove the need for governance approvals, which are given by each institution separately. These approvals allow the institutions to consider whether they have the necessary resources to participate in a project. This is an important approval step and should continue; however, institutions must ensure that this approval process is timely, efficient and transparent (Iedema et al. 2013). To achieve this, the NHMRC (2016a) has developed guidelines that can substantially streamline governance approvals — and these should be implemented across all jurisdictions.

RECOMMENDATION 6.15

Processes for obtaining approval from human research ethics committees (HRECs) should be streamlined.

To achieve this in the health sector:

- All HRECs should be required to register with the National Health and Medical Research Council (NHMRC). The NHMRC should receive funding to expand its current registration process, to include audits of registered HRECs.
- To maintain their registration, HRECs must implement efficient and timely approval processes, which ensure projects are not unduly delayed. The time taken to consider and review projects should be reported to the NHMRC, and included in the annual report on HREC activity.
- As a condition for registration, all HRECs and the institutions they operate in would be required to accept approvals issued by certified HRECs for multi-site projects, without additional reviews. The Australian Health Ethics Committee should develop uniform review processes to be used by certified HRECs.
- The Council of Australian Governments' Health Council should sign an intergovernmental agreement that extends the existing National Mutual Acceptance Scheme to all jurisdictions, including the Commonwealth, and all types of projects. As part of this agreement, all jurisdictions should also implement streamlined governance approvals.

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 the sharing with trusted users 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, as assessed by the NDC. This extends to research with datasets that are not defined as NIDs, but where that research is in the public interest. This would substantially streamline approval processes for access by trusted users to identifiable data, without increasing risks to individuals.

Several submissions following the Draft Report expressed concern about the possibility of personal information being used without individuals' consent for broader public interest research (for example, Rosie Williams, sub. DR239 and the Group of Eight, sub. DR304). The Australian Dental Association argued that 'retention of privacy in respect of records must be retained' (sub. DR230, p. 1) and that any public interest exception would need to be clearly defined.

The Information and Privacy Commission NSW (sub. DR309) suggested that the *Government Information (Public Access) Act 2009* (NSW) provided a framework for defining and assessing the public interest which it considered had matured and operated to good effect. Public interest considerations against disclosure under the Act are grouped under the following headings:

- responsible and effective government
- law enforcement and security
- individual rights, judicial processes and natural justice
- business interests of agencies and other persons
- environment, culture, economy, and general matters
- secrecy provisions
- exempt documents under interstate Freedom of Information legislation.

The Commission acknowledges the reservations of some submissions towards the research exemption recommended, and agrees that the public interest needs to be carefully defined to ensure that data can be used for research whilst minimising the risk of misuse. As part of implementing broader exceptions for research the Commission considers that the separate sets of rules under sections 95 and 95A of the Privacy Act should be combined.

In determining what is in the public interest, the guidelines issued by the NHMRC (under section 95A of the Privacy Act) could serve as part of a template. The guidelines detail the issues that human research ethics committees must consider when determining the public interest in a specific research activity (box 6.6). Research Australia (sub. DR282) considered that research that has been funded by the Australian Research Council, NHMRC, and/or which has received approval from an approved human research ethics

committee should be sufficient to establish that the research is in the public interest. The Commission however, believes that a more comprehensive approach is required, since sources of funding and ethics committee approvals are not in themselves sufficient to establish that research is in the public interest.

Box 6.6 National Health and Medical Research Council guidelines approved under the Privacy Act

The guidelines issued by the National Health and Medical Research Council, approved under section 95A of the *Privacy Act 1988 (Cth)*, in part state that:

> In determining whether the public interest in the proposed activity substantially outweighs, or does not substantially outweigh, the public interest in the protection of privacy, an HREC should consider the following matters:
>
> a) the degree to which the proposed collection, use or disclosure of health information is necessary to the functions or activities of the organisation
> b) the degree to which the research, or compilation or analysis of statistics activity is relevant to public health or public safety
> c) the degree to which the research, or compilation or analysis of statistics or the health service management activity is likely to contribute to:
>   i. the identification, prevention or treatment of illness, injury or disease; or
>   ii. scientific understanding relating to public health or safety; or
>   iii. the protection of the health of individuals and/or communities; or
>   iv. the improved delivery of health services; or
>   v. enhanced scientific understanding or knowledge; or
>   vi. enhanced knowledge of issues within the fields of social science and the humanities relating to public health or public safety
> d) any likely benefits to individuals, to the category of persons to which they belong, or the wider community that will arise from the research, or compilation or analysis of statistics, or management of a health service being undertaken in the manner proposed … (NHMRC 2014, p. 23)

These guidelines could be adapted to enable other types of research, which would enhance community welfare (for example, social science research). For example, part c of the guidelines could read:

> c) the degree to which the research, or compilation or analysis of statistics is likely to contribute to:
>   i. the identification, prevention or interventions to address barriers to improving community or individual welfare; or
>   ii. scientific understanding relating to aspects of community or individual welfare; or
>   iii. the protection of the welfare of individuals and/or communities; or
>   iv. the improved delivery of welfare and other types of social and community services; or
>   v. enhanced scientific understanding or knowledge; or
>   vi. enhanced knowledge of issues within the fields of social science and the humanities.

RECOMMENDATION 6.16

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 in the legislative package that implements these reforms to apply to all research that is determined by the National Data Custodian to be in the public interest.

Retention of linked datasets

One comparatively simple way to substantially increase the value of linked datasets is to retain linked datasets and linkage keys as an ongoing resource for future research work, rather than delete after initial use. Existing policies require that most linked datasets are deleted once the project they were created for is completed. State-based linkage units retain the linkage keys, to create enduring linkage systems that enable datasets to be linked quickly (PHRN, sub. 240). For projects using Commonwealth data, both the datasets and the linkage keys are destroyed (see appendix C for further detail on linkages and related policies).

Deleting linked datasets and linkage keys — which in effect means each linkage project must start again from scratch, rather than build on previous efforts — is a practice that wastes both time and money (PHRN, sub. DR240). Indeed, in some other instances, destruction of data is prohibited by law. For example, the *Archives Act 1983* (Cth) prohibits persons from destroying, damaging, or altering Commonwealth records, with the exception of acts authorised by the National Archives of Australia. In the United States, there is a broad prohibition against destroying, or attempting to destroy, government records (US DoJ nd).

Australia lags other developed countries in both the creation and use of linked data (Health Research Institute – University of Canberra, sub. 115). Some ad hoc progress has already been made in retaining linked datasets (for example, MADIP is an enduring integrated dataset that will not be deleted upon the completion of a specific project (DPMC, sub. 20). In the Commission's view, to ensure such progress is achieved for all data, further actions are required:

- abolishing the requirement to delete linked datasets, and enabling trusted users to access them, using the same risk-based approach that is employed for all other types of data.
- retaining the linkage keys used to create the datasets and working towards the creation of enduring linkage systems.

Submissions to this Inquiry raised potential options for storing linked datasets. The Cancer Council of Australia (sub. DR254), for example, proposed that, where original linked data were placed in the SURE system, the Sax Institute or the AIHW be responsible for storing

data for further use. For data linked and used at a jurisdictional level only, the Cancer Council of Australia recommended that the respective jurisdictions be responsible for storing linked data. Yet another option was put forth by the Australian Data Archive (sub. DR288), who suggested implementation of a metadata registry for linked datasets, as well as the documentation, curation, and management of newly created linked datasets. The Commission considers that these, and similar suggestions, warrant further consideration. These models could also be used by sectoral ARAs, once they are established.

ARAs should play a part in the creation of enduring linkage systems, to make data linkage more efficient. As the central point for data linkage in their sector, ARAs would be well placed to maintain a catalogue of the various links created, and build this into a linkage system. This could also support broader improvements in datasets, as the AIHW explained (sub. DR299, p. 7):

> To ensure this information is retained in such a way as to lead to continual improvement, at a minimum, linkage keys from each linkage project should be retained for re-use. In the case of an appropriately accredited linkage authority, linkage keys could also be retained on a spine file so that updated information could be used to improve the links whenever further information is received.

One submission to this Inquiry suggested the statistical linkage key for data relating to individuals using government services be abandoned, due to the risk of re-identification occurring. Specifically, the Australian Privacy Foundation alleged that:

> The government is increasingly using a Statistical Linkage Key [SLK] for individuals using government services. Both health and social services use this key. The SLK is supposed to be anonymous and de-identified, but is not. We understand it is a combination of part of last name, part of first name, date of birth and gender … The Statistical Linkage Key [should be] abandoned due to the real risk of re-identification … (sub. DR313, p. 5)

Statistical linkage keys have been used for many years by state-based linkage units, while maintaining safeguards around risks such as privacy. Rather than abolishing statistical linkage keys, and suppressing data use which would damage the capacity of researchers to undertake work in the public interest and substantially increase the cost of future research, the possibility of re-identification underlines the need for sound risk management practices (that is, strong linkage keys).

Re-identification risk suggests that data custodians be vigilant with their encryption and de-identification techniques, up to and including paying the equivalent of 'white hat' hackers as occurs elsewhere in digital risk management to attempt to re-identify data. The current re-identification Bill before parliament may make such sensible risk management out of the question. Even so, suppression and destruction of data is not a sensible choice in a world increasingly dependent on it for better service delivery and efficiency in the use of scarce resources.

RECOMMENDATION 6.17

The Australian Government should abolish its requirement to destroy linked datasets and statistical linkage keys at the completion of researchers' data integration projects. Where an Accredited Release Authority is undertaking multiple linkage projects, it should work towards creating enduring linkage systems to increase the efficiency of linkage processes.

Data custodians should be advised as part of early implementation of this reform package to 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 made available to other dataset users.

7 Getting value from Australia's national interest datasets

Key points

- Wider release of data and more effective sharing by governments would likely trigger significant investment (private as well as public) and improvements in national welfare. However, determining which datasets could lead to such improvements remains a serious practical issue for governments.
- Existing government data initiatives, such as data.gov.au, should be leveraged as part of the broadest sharing and release efforts. A framework is needed to formalise how such efforts would be implemented and managed, and is provided in this Report.
- Datasets of high value have a number of distinct characteristics, including that they are unique (or cannot be readily replicated), are of high quality, have a high degree of coverage in the relevant population, and are up-to-date or updated regularly.
- National interests additionally require that coverage is oriented towards nationally significant subject matter. Data with these combined characteristics are likely to generate spillover benefits for the community and should be designated as National Interest Datasets (NIDs).
  - Prioritising selection of NIDs would require flexibility and discretion. A parliamentary committee would be a suitable vehicle to scrutinise this process, rather than try to define national interest in legislation. Engagement with democratic representation would assist in maintaining social licence.
  - The process for selecting datasets for designation as NIDs would be open to the public. As the funder of NIDs, the Australian Government would be the arbiter of their selection.
  - Initial candidates for consideration as NIDs include existing public sector datasets that provide registers of businesses, services or assets, or record activity in key areas of public expenditure, such as health and education, as well as datasets held in private entities that are regulated in the public interest and/or receive public funding.
  - It is expected that other datasets would be nominated as NIDs by a range of parties, including State and Territory governments (preferably acting collectively), private sector entities and not-for-profit organisations. Incentives to do so include provision by the Commonwealth of ongoing funding for integration and maintenance, a desire for national linkage and access, and the removal of current restrictions on sharing and release, replaced by fit-for-purpose safeguards.
  - Access to NIDs would be provided via the processes outlined in chapter 6, by a designated Accredited Release Authority (ARA).
- The National Data Custodian (NDC) would lead the process to identify and develop the case in support of NIDs; and ensure the chosen ARA makes NIDs widely available.
  - The processes used by the NDC should be transparent and risk-based, and a relationship not unlike that of the Auditor General with the Public Accounts Committee is anticipated.
  - A focus on the net benefits likely to accrue from designating a dataset as NID, the possible impacts on intellectual property, impacts on incentives to continue collection, and a detailed consideration of the costs to the collecting party or parties, should be points of primary focus in deliberation.

Over recent decades, increases in computing power, data analytics, and the quantity and availability of data have underpinned an increasing demand for datasets of high value. There is now much more demand to link valuable datasets across Australia, compare related datasets in different jurisdictions, and generally create and use national level datasets. This is driven by impressive revelations possible now through intelligent algorithms and data analytics, a recognition that datasets can be used to discover hidden linkages allowing improved decisions, outcomes and solutions, evidence-based evaluations and targeted service delivery. Chapter 2 and appendixes to this Report illustrate the scope for innovative uses of data in the areas of health and finance alone.

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 (as examined in detail in chapters 3 and 6). 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 would minimise risks from its existence.

In its Draft Report, the Commission proposed a revised approach to data sharing and release, including of high value datasets, and a system for identifying and giving broader access to a subset of these, so-called National Interest Datasets (NIDs). This chapter considers comments from submissions on the Draft proposal and provides greater detail on the recommended reforms, including arrangements that would action them.

7.1 High value datasets

The terms of reference for this Inquiry asked that we consider the determination of high value datasets. The underlying presumption is that the community-wide benefits from improved access and use of these datasets would be significant (and ideally exceed anticipated costs to the community).

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 for the private sector in determining its high value datasets. Submissions to this Inquiry tend to confirm this. 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, readily extending as far 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 on occasion in encouraging private sector data holders to make their data available — commercialisation of basic research is presently topical — 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 7.1) 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.

It is evident from the Terms of Reference that both views can be held simultaneously.

The challenge of determining high value public sector datasets

Literal approaches to valuing public sector datasets can be difficult as public sector data holders do not generally receive market-determined price signals to assist them in identifying value. Most agencies would not regard valuing datasets as 'core' business. One factor, but by no means the only one, is that datasets are not required to be recorded or managed as assets. More often it is simply that data collection proceeds from a legislated set of objectives that has nothing to do with maintaining an asset of value. Agencies can hardly be blamed for sticking to their legislated core business.

Cost recovery for releasing data is practiced at times, but by definition is related to cost not value. 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 are the exception. The price of data sold by the public sector in Australia is thus not broadly indicative of its value or even of its priority for release; other approaches are needed (chapter 9). These would also be more relevant to public interest purposes.

For both private and public data holders, many innovative uses of data cannot be valued now as they are not yet envisioned. Likewise, the way in which a particular dataset is used can influence the value of subsequent uses of that dataset and of other datasets. The value of a particular dataset would 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; Australian Institute of Health Innovation, sub. DR229; National Health Funding Body, sub. DR234; Sax Institute, sub. DR256).

Box 7.1 Why don't governments just release all non-sensitive data?

Non-sensitive data that governments could release includes data of very high value, 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.

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.

While some data cannot be released because it identifies individuals or businesses, 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 excessively cautious interpretation of them — that are responsible.

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

If release becomes the preference for a particular dataset, rather than retention for administration, 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 at the agency level, significant costs may be implied in a move towards open by default.

In other words, the value of data is about more than the revenue streams it might generate. The potential to trigger innovative investments or opportunities for better governance are considerations that would likely outweigh any revenue generation, even if government datasets were priced according to market forces — and a market is in any event rarely present. As recommended in chapter 9, pricing data to researchers may well see fewer discoveries or effective evaluations of inefficient or failing programs.

While assessment of which datasets are of high value and should be prioritised for release clearly relies on input from end users, a necessary precondition is an awareness of what data actually exists.

The starting point must be much better information on what datasets are held by government; that is, comprehensive, complete and transparent information (including metadata) on what datasets are held by government. In the absence of efforts in this area, other agencies, researchers, not-for-profits and commercial entities cannot know enough to effectively participate in the determination of high value datasets (chapter 6).

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. For example, the National Committee for Data in Science (sub. DR265, pp. 2–3) pointed to the following datasets of high priority:

- data on activity and usage of government services and facilities
- datasets produced by National Collaborative Research Infrastructure Strategy (NCRIS) facilities
- health, medical and hospital data; health insurance data, both public and private
- environmental data
- election data
- police, emergency services and courts data.

Many other participants nominated datasets relevant to their sector.

Broadly speaking, data on the administration of government policies and programs (so-called administrative data) was held up as being particularly valuable, since it is comprehensive, directly sourced and usually with an incentive for accuracy (tax or welfare data have clear rules in support of accuracy, not necessarily effective in all cases but much better than inferred or surveyed data). Overall, administrative data offers scope to evaluate and improve decision making that affects broad cross-sections of the community (PC 2013). John Daley from the Grattan Institute noted in particular 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' (DPMC 2015, p. 16).

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; the Australian Institute of Tropical Health and Medicine (AITHM), sub. 52; and Department of Social Services (sub. DR255).
- 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 (MBS) and Pharmaceutical Benefits Scheme (PBS).
- 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, 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 F).
- 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).

In responding to the Draft Report, Inquiry participants also discussed specific high value datasets, in the context of the proposed system of NIDs. These suggestions are discussed in section 7.2 below.

There was a greater focus among Inquiry participants on high value datasets held by the public sector (although private hospitals and private education datasets were raised). The nominated datasets span a broad range of areas and include datasets related to health, natural hazards, education and welfare. AURIN (sub. 116) specifically noted that most of the datasets nominated as being of a high value to the research sector are held by governments. The list of opportunities is rich and long.

Characteristics of high value datasets

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

Characteristics around 'use' that high value datasets *might* possess, include 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 (DTF (Vic) 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 re-use, and a large potential user base (Archer et al. 2014).

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 on the use and updating of 'dataset identifiers' with machinery of government changes), would assist data users to understand the properties and ongoing updating of particular datasets, and thus draw conclusions about their value.

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

Methods to understand demand for public datasets

Whether of interest nationally or to a group of external users, understanding value as a way of prioritising action (and expenditure) is essential.

The level of interest in and 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 (DTF (Vic) 2015; Headd 2016).
- The use of feedback mechanisms on data portals that allow users to suggest and vote on datasets for release (DTF (Vic) 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 user feedback are 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 applications could also provide insights into which datasets should be prioritised for 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 consult 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 (appendix C).

There is room for improvement in determining which datasets are valuable

There is considerable room for improvement in how governments determine datasets that could be valuable to users and the community more broadly (Department of Prime Minister and Cabinet, sub. 20).

First, not all government data holdings are discoverable through open data platforms (and there are numerous data registries, adding to a sense of confusion). Many Australian Government entities do not make their data holdings discoverable on data.gov.au, the Australian Government's primary data portal. And many agencies do not even publish registers of their data holdings on their own websites (chapter 3).

Second, some government datasets are not published in fully structured formats, which enables automated exchange and use. 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 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 all users, it appears 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, across sectors, and in different jurisdictions.

A more comprehensive and coordinated approach is needed. All Australian Government agencies should publish information on their data holdings on a central registry. This would enhance the use of data as it becomes more discoverable, and would reduce duplicated data collection, which burdens government agencies and the public.

There are some circumstances, however, where data is too sensitive to have its metadata published. But, these cases are very limited. 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 (chapter 6).

Although it may seem too hard to contemplate, a start must be made to address these issues comprehensively at Commonwealth level; and as a role model to States and Territories (noting some are in advance of the Commonwealth in limited areas — no jurisdiction is comprehensive in its approach). The Commission is recommending that a start be made now, that is, within six months. Rapid roll-out of a template guide to setting metadata would be very helpful and can be done in this timeframe. Glaring absences should then be quickly filled, the template helping to draw them out. No doubt agencies would learn overtime how to do all of this better, and would engage more fully; either in recognition of the benefits to be had, or because of concern about damage to reputation. 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 find it and use it. A primary purpose of data release or trusted access is to create the opportunity to confirm ideas. This should not be 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, and the Commonwealth has experts in this — from the ABS and Data61, to the Australian Institute of Health and Welfare (AIHW) and the Productivity Commission's own Government Performance Reporting and Analysis team.

Once the data is found, it needs to be usable. 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 on release websites.

7.2 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. Those high value datasets that may be used to generate substantial benefits across a broad swathe of the Australian population are considered to be of national interest (as further discussed below).

The Commission has given considerable thought to establishing a framework whereby wider access would be enabled to select high value NIDs and has received valuable comment from submitters. The intention is to promote the development of a suite of datasets — some of which are released publicly; others that would, at least initially, be shared (rather than released) across all Australian governments and with a group of other trusted users (as defined below).

Through the recommended process detailed in this section, the Commission intends that the designation of a dataset as a NID would take precedence over existing and future restrictions to access. This would include legislative and other 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, where there is a public interest in doing so, while also maintaining appropriate protections around privacy and confidentiality.

Datasets contributed to the suite of NIDs would be shared or released through sectoral Accredited Release Authorities (ARAs) (chapter 6).

Main features of NIDs

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

First, the dataset has spillover benefits, and its use would be likely to generate broader economic and social benefits beyond those accruing to the initial data holders. This could be because the dataset enables wider 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 performance evaluation and comparison between programs and investments by the Commonwealth, States or 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, and data on public hospital performance.

Finally, datasets that might become NIDs are very likely to have an established focus on nationally significant subject matter. Functions that are nationwide in their nature but with a variety of providers (for example, in education) may be good candidates.

While created on the basis of different criteria, the ABS's register of Essential Statistical Assets for Australia may include some examples of NIDs. 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, in part, a practical illustration of how governments could determine some high value datasets of national interest. This point was acknowledged by the Australian Computer Society (sub. DR329, p. 6), who suggested that detailed consideration be given to the register and its development when considering how to conduct the process for identifying NIDs. Governments could also draw on international developments to guide them in determining which datasets are of national interest.

The process for screening proposals for nomination as a NID would involve the National Data Custodian (NDC) (chapter 6) undertaking an assessment of public benefit. This would be based on net gains to be had from taking datasets into a new legislated form supported by Commonwealth powers that:

- remove barriers to integration and re-use
- deliver new, accessible data either for release or sharing
- ensure continuity of data collection by the commitment of Commonwealth funding into the medium term
- and offer clear, high value likelihood of spillover benefits from trusted user access or wider public release than at present.

The last factor is important. The process of establishing NIDs would have costs and must offer additional, nationally-oriented benefits over and above those already to be found from current (presumably somewhat limited) uses of the data. A number of big Commonwealth datasets may appear ready candidates for a status upgrade under this process, but the Productivity Commission is not advocating status upgrades. Additionality is crucial.

The analyses underpinning assessment of nominations would be published. The NDC would submit a recommended nomination first to relevant Commonwealth Ministers (including the NDC's portfolio minister and the minister responsible for the dataset/s) and, if accepted in principle, then to the parliamentary committee for scrutiny. The Committee's objective would be to consider whether this nomination would create an outcome that is a nationally relevant improvement in data use. Researchers and (where relevant) potential commercial users would be expected to provide evidence to, and if necessary before, the Committee to offer their views on what might be done with improved access.

Following such scrutiny, the Committee would record its views and the Commonwealth Ministers would then decide to proceed or not, and a disallowable instrument under the Act would allow the dataset(s) to be taken up by an ARA for integration, preparation and release or sharing. Datasets acquiring NID status would be maintained for a minimum period of ten years.

There is no great science to this number, other than that it should demonstrate sustainability of release or access — as noted earlier in the chapter, value would only be determined subsequent to release. The present 4-year forward estimates process would often be insufficient for such medium term activity and data analytics and proof of outcome. Data custodians (which may in a few cases include very wary private sector custodians in health or education) are likely to be called upon for updates and advice over a medium term period — as the Productivity Commission has done with States and Territories for the Review of Government Services, now in its twenty third year.

Some likely candidates suggested by participants

In the Commission's Draft Report, it was suggested datasets that might potentially be designated as NIDs *could* include land use data, business register data, property and transport data, data on service provider performance, financial systems data and data on public infrastructure projects.

In response to the Draft Report, a number of submissions put forward more detailed examples for consideration:

- The Department of Social Services (sub. DR255, p. 6) listed 15 core longitudinal data assets deemed as of national importance by a recent Review. This included survey datasets such as the Australian Census Longitudinal 5% Dataset; the Household, Income and Labour Dynamics in Australia (HILDA) survey; and the Longitudinal Studies of Australian Children (LSAC) and Indigenous Children (LSIC). Administrative datasets included the Multi-Agency Data Integration Project (MADIP); the MBS/PBS 10% dataset; and the National Assessment Program – Literacy and Numeracy (NAPLAN) dataset.
- The University of Melbourne (sub. DR305, p. 5) suggested that national health, clinical, economic and population data, including more micro-level examples, be included in the list of possible NIDs.
- The Sax Institute (sub. DR256, p. 3) stated that all datasets created as part of the business of governments (Commonwealth, State and Territory, and local) should potentially be designated, with exceptions justified via a transparent process.
- The Insurance Council of Australia (sub. DR318, p. 11) argued that natural hazard data, building standards data and mental health data should be considered in designating NIDs.
- Research Australia (sub. DR282, p. 15) listed a number of State and Territory datasets (including hospital admission records, Maternal and Child Health records, registers of births and deaths, primary care records where delivered by State agencies, educational attainment and school data, child protection, criminal convictions, and prison population data) and both for profit and not for profit private sector datasets (including primary care data, hospital admission records, and data on childcare and aged care).
- CSIRO (sub. DR323, p. 18) suggested: foundational spatial data; integrated health data across MBS, PBS, hospital admissions, treatment, outcomes and e-health records to develop longitudinal analysis of illness, treatment and outcomes; real time economic activity data on housing commencements, to grocery bills and credit card expenditure on a linked, de-identified and aggregated basis; and longitudinal education and employment outcomes for the first few years of employment, also in linked, de-identified and aggregated form.
- The Attorney General's Department (sub. DR334, p. 12) said that family law services data provided by non-government service providers and held by the Department of Social Services was potentially of great value to researchers, in particular when

combined with data held by the Australian Institute of Family Studies and the Department of Human Services.

While these are important initial examples of *possible* NIDs, others may be less immediately obvious but become clear candidates over time. What is crucial is that a process and legislative structure be established to encourage these datasets to emerge for broader use in a simple and adaptable manner. A focus on datasets likely to contain significant spillover benefits is a crucial threshold criteria within the designation process.

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. As has been emphasised throughout this chapter, such access requires the development and ongoing use of robust, risk-based de-identification processes, with related governance and monitoring structures in place.

Initially, the Commission expects the focus would be on known Commonwealth datasets that could readily be designated for broader access, and their State and Territory counterparts if available. Some collations of data collected by local governments may also be found suitable for designation.

At the Commonwealth level, some datasets, such as the Multi-Agency Data Integration Project (MADIP) and the Business Longitudinal Analytic Data Environment (BLADE), would appear at first blush to be ideal candidates for NID status. However, this is not a fait accompli, and would be subject to the same detailed consideration proposed for other potential NIDs, including consideration of additionality of benefits possible through designation. In proposing this process, the Commission is keen to ensure that 'business as usual' data collaborations continue to occur, with the NID process reserved for a select subset of broader high value datasets.

Over time, it is expected that with greater transparency regarding data holdings and engagement with stakeholders, the NDC would more proactively prioritise public datasets for designation.

Resulting benefits

NIDs must generate spillover benefits to the community beyond those derived by just the data holders and data contributors. And they must be greater than is currently being generated. That is, additionality would be delivered from the investment by the Commonwealth.

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 response to the Draft Report, a number of participants emphasised the benefits that would accrue from this proposed system.

> CoreLogic believes that the concept of a National Interest Dataset could enable the release of more state and territory government property data and enable all data to be collated and produced in a standardised and consistent way. That would, in turn, lead to greater benefits for consumers and businesses. (CoreLogic Asia Pacific, sub. DR248, p. 1)

The University of Melbourne (sub. DR305, p. 5) stated:

> These reforms would enable significant improvements in access to data, especially public sector datasets.

The AIHW (sub. DR299, p. 5) pointed to the potential of such an approach to assist in identifying data gaps:

> The development and use of NIDs may highlight information needs in other areas. The framework for establishing NIDs should develop in such a way as to assist with identifying data gaps.

The Department of Industry, Innovation and Science (sub. DR235, p. 1) focused on positive impacts on data quality and maintenance:

> This will improve the availability and accessibility of data that is in the public interest. It would focus attention on significant datasets, encouraging their custodians to ensure that data is well managed, of good quality, and its collection and maintenance appropriately resourced for the good of the whole community.

Telstra (sub. DR312, p. 13) was generally supportive of the Commission's approach but, as with several other parties, argued that greater detail was required around the designation process and consideration of costs and benefits.

In response, the Commission has attempted to provide greater process detail in this Final Report. Several detailed examples of how the introduction of NID status can produce benefits for public and private sector users of existing datasets are provided in appendix B. Further discussion on the recommended approach to private sector datasets is provided below.

How datasets would be designated as NIDs

As above, an important responsibility of the NDC would be to oversight the designation and use of NIDs.

The Data Sharing and Release Act (DSR Act, discussed in greater detail in chapter 8) would establish the NIDs regime, provide that the NDC would have responsibility for designating NIDs, and set out the process by which they can be designated. Designation would begin with a nomination submitted to the NDC. The nomination could be brought by the data custodian, or a third party.

Designation would occur in several stages (figure 7.1). First, the dataset would be designated as an NID by the NDC, following the assessment process noted earlier. The DSR Act (or associated guidance) would set out the criteria for designation, drawing on those noted earlier. This designation would occur by disallowable instrument. By default, being designated as an NID would subject the dataset(s) to the ARA's processes (as outlined in chapter 6 and further discussed below).

Figure 7.1 **Process for getting a dataset designated as an NID**

In some cases, it would be critical to the integrity of the NID that it incorporate information from the private and not-for-profit sectors. This should be done with caution if the information is commercially sensitive. But just as personal privacy is not an absolute right but needs to be balanced against the public interest in using the information, so too does the claim of commercial in confidence need to be balanced against public interest considerations.

Accordingly, the framework should provide that private sector datasets should only be designated when the public interest in designation outweighs the private interest in protecting commercially sensitive information. Parliamentary committee processes would help expose the undoubted contention that may accompany such access proposals. This process does not avoid the issue, but it does allow for exposure to see both sides of a claim against data use examined. Sunlight is the best disinfectant; transparency is useful in exposing any (and all) evils.

Where data from a regulator is recommended to be included, particular care would need to be taken that the designation of the NID does not undermine public trust in the regulatory scheme. However, given the general requirement that NIDs are held and shared within robust risk management frameworks, the Commission does not consider that this issue would frequently arise. The NDC has a core interest in preservation of public confidence in data re-use.

Where data held by the private sector is to be included in an NDC, the designation should specify which method(s) are intended to be used to include these private sector datasets in the NID (box 7.2). Where private sector NIDs are required to be purchased, the process would apply accepted Commonwealth principles for procurement. Further detail on private sector data and NIDs is provided in chapter 8.

The Commission's model requires that NIDs would be shared in a secure environment with trusted users, or released as non-sensitive open data — the same safeguards that apply to other datasets should equally apply for NIDs.

Implementing access arrangements

The approach recommended by the Commission represents a marked expansion in data access in Australia.

In contrast to existing arrangements for access to significant datasets, the approach recommended envisages *open release* to the general public of a range of NIDs. Some of these datasets may already be publicly available, but the process of designating them as NIDs must, by criteria for selection, not only improve their discoverability, but ensure they are more widely accessed, integrated where necessary and adequately maintained into the future, *as a national asset*.

Notification arrangements for future data contributions to a NID would be settled by the relevant ARA with custodians. A specific provision in the new Act should both authorise and require this action by ARAs.

The primary steps for designation would vary depending on whether the datasets concerned are held by multiple custodians or not:

- If the data is held entirely within one Commonwealth Government agency, the primary steps for designation would involve dataset curation, selection by the NDC of an ARA, determination by the ARA of any de-identification and linkage necessary, and whether the dataset would be shared with trusted users (and which ones) or released.
- If the data is held across multiple government agencies or jurisdictions, designation would also require determining which ARA has lead responsibility for ongoing dataset updates and curation, or otherwise how this task would collectively occur in a timely and reliable manner.

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 to expand the range of data users that would be able to be accredited as 'trusted', expand the range of datasets they might access, and the types of uses to which data can be put. *Sharing* of these NIDs with approved trusted users is a first step — a trial of the approach — with *public release* possibly the ultimate objective that increases the use, and therefore the value of the dataset.

Where NIDs that contained identifiable data are intended to be generally made available for release in de-identified form, reducing the risks to individual persons and businesses). Risks associated with data transfer and storage of de-identified data would be managed through use of approved secure computing environments.

Where data forms part of a NID that is *shared* in the first instance, rather than publicly released, the Commission intends that:

- Access be granted on an *ongoing* basis to ARA-approved government personnel (in a Commonwealth, State or Territory government agency) and to approved trusted users under accreditation in a secure computing environment. Access that is ongoing would 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 work for which the data could be used by trusted users.
- The output from research based on the dataset *may* be reviewed by the ARA prior to release, on a risk-assessed basis; and would always be available to the ARA for process and accreditation evaluation, if necessary.
- Responsibility for appropriate use would rest with the accredited trusted user, with clear and significant consequences (see chapter 6) for any breach of this trust.

Data sharing agreements may provide a useful mechanism for setting out guidelines around use by trusted users.

RECOMMENDATION 7.1

The Australian Government, in consultation with State and Territory governments, should establish a process whereby public (and in some exceptional cases, private) datasets are nominated and designated as National Interest Datasets (NIDs).

This process should be public, driven by the National Data Custodian, and involve:

- The National Data Custodian accepting nominations for NIDs, assessing their public interest merits and, after consideration by the Government, referring selected nominations to a public scrutiny process. Designation would occur via a disallowable instrument on the recommendation of the National Data Custodian.
- The establishment of a parliamentary committee, or addition of such a role to the work of an existing parliamentary committee, to conduct public scrutiny of nominations for NIDs.

The process of nomination should be open to the States and Territories in order to cover linked datasets.

This process should be in place by the end of 2018, as part of the legislative package to implement these reforms.

RECOMMENDATION 7.2

In considering nominations for National Interest Datasets (NIDs), the National Data Custodian's public interest test should establish that through sharing or release, the designation of a dataset would be likely to generate significant *additional* community-wide net benefits beyond those obtained by the original data holder.

Once designated, NIDs that contain non-sensitive data should be made available for immediate release.

NIDs that include data on individuals would be available to trusted users only in a manner that reflects the accreditation processes of the relevant Accredited Release Authority, as established and updated by the National Data Custodian, to respect privacy and confidentiality.

Where data from the private and/or not-for-profit sectors is recommended to be included in a NID, the analysis prior to designation should specifically note the ways the designation addresses genuine commercial sensitivity associated with the information and costs (including those related to ongoing dataset maintenance).

Defining and accrediting trusted users for selected NIDs

Robust processes for identifying and accrediting trusted users would be critical to enabling access to NIDs that are not yet available through open access. Trusted users should have:

- governance structures and processes that minimise and address the risk of inappropriate use or release of information, and
- signed a legal undertaking that sets out safeguards for use and recognises relevant privacy requirements, and
- access to appropriate secure technology/technological infrastructure and facilities.

Under the Commission's recommended model, trusted users of NIDs would be vetted and accredited by the relevant ARA. It is possible that, in doing so, the ARA would vary access levels for different types of trusted users.

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 for that user when they leave the approved environment, or if a data breach or mishandling occurs in that same environment and/or program.

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 avoid data breaches. We remain of the view that establishing such a system of accredited trusted users is clearly feasible and desirable, and would provide adequate safeguards and penalties.

The complexities attaching to accreditation processes are acknowledged. In this regard, the Commission concurs with the view of the University of Melbourne's Department of Computing and Information Systems (sub. DR303, p. 8) that:

> … accreditation of trusted users will require an understanding of computer security, including encryption, cybersecurity, risk management, and access control.

While these points are valid, recent experience both here and overseas suggests that such complexities are not insurmountable, and that a flexible system using 'safe' elements is implementable. The current operation of access and use arrangements by the United Kingdom's Data Archive and Statistics New Zealand, and recent provisions within the UK's *Digital Economy Bill 2016-17* (box 7.2), as well as the experience in Australia within the research and government sectors, provide instructive examples here of workable approaches.

The 'five safes' model was discussed in chapter 4. It needs to be applied but also adapted by the NDC in development under the Commission's Framework model. The Australian Data Archive (sub. DR288, p. 9) suggests:

> … the use of de-identified data (Safe Data) and access by approved Trusted Users (Safe People), potentially in secure access environments (Safe Settings) may be sufficient risk management to enable relaxing the requirement for specific project approval (Safe Projects).

Box 7.2 Trusted users in the UK Data System and the Digital Economy Bill

The UK Data Service was established in 2012 and provides unified access to the UK's largest collection of social, economic and population data. It uses a three tiered model of data access, involving:

- open access to many types of data, whereby users can do anything with the data (including sell it), but it is not freely available (still being subject to a licence)
- safeguarded access: to data that has been anonymised/confidentialised — the risk of re-identification is small, so it can be used by anyone for anything, but users have to sign terms and conditions before access. Some of this data is only available to government or academic researchers.
- controlled access to data that has been de-identified but not anonymised, so there is a significant risk of re-identification. The UK Data Archive, which is the lead organisation within the Service, conducts intruder testing (with the permission of the data custodians) to assess this risk.

The Service had more than 22 000 registered users and has had very few breaches to date.

A further relevant development of note from the UK is the introduction of the 2016-17 Digital Economy Bill. This contains detailed arrangements for the accreditation of persons and research for the purposes of information disclosure and use. . The UK Statistics Board is accorded responsibility for oversight of this process and for withdrawal of accreditation in instances where the detailed Code of Practice has not been adhered to.

Source: UK Data Service (2014), UK Parliament (2016)

The Department of Industry, Innovation and Science (sub. DR235, p. 2) also discussed hierarchies in regard to trusted user status:

> … it would be more appropriate to describe trust along a spectrum. A better approach may be to develop trusted user categories, each of which is accorded a level of access commensurate with the level of trust or accountability they have. This would allow for greater flexibility in dealing with the different types of users who would have an interest in accessing National Interest Datasets. Provisions could be made to allow users to move to a more trusted category if they give enforceable undertakings regarding their use of data.

The Commission supports the use of a flexible approach, with an 'escalator' of accreditation, reporting and compliance requirements that vary with the context of use, data type, use environment etc.

RECOMMENDATION 7.3

Trusted users should be accredited by the relevant Accredited Release Authority (ARA) for access to those National Interest Datasets (NIDs) that are not publicly released, under processes accredited and updated as needed by the National Data Custodian.

Trusted users should be personnel from a range of potential entities that:

- 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, and
- have a signed legal undertaking that sets out safeguards for data use and recognises relevant privacy requirements.

The default position should be that after applicants and their institution establish capability to respect the processes and obligations of the ARA's accredited standard, an individual researcher from one of these organisations would be readily approved for access.

For trusted users of NIDs, this status should provide an ongoing access arrangement to specified unreleased datasets that would only cease on completion of a researcher's engagement with their relevant institution, or a loss of trust in the user or their organisation (via processes also established in accreditation of the ARA by the National Data Custodian).

Transparency mechanisms

The parliamentary process noted earlier is an essential factor in building and maintaining social licence in much wider public sector reuse and analysis.

Similarly, the NDC's role in advising the public (by release of its analysis); the government (by submitting this analysis to the funding Minister for approval, first and last); and the parliamentary committee (by appearance before it and in assisting the Committee to a judgement on the national significance or otherwise of designation) is also designed to reinforce public confidence, including through establishing a clear line of accountability.

The Commission understands full well that a dysfunctional application of this otherwise elegant solution to a public confidence risk conundrum may set back data reuse. We remain supportive of it because such structures have worked before and could, in our view, contribute greatly here. Without a public process like this, the opponents of data release and re-use would have open licence to allege any and all malpractice. Transparency of purpose in public policy is the best shield against those who would attempt to drum up fear of data reuse in the community.

The Public Accounts Committee and its relationship with the Auditor General is a useful if imperfect model for the relationship between the NDC and the Committee proposed. Other Committees, for example the Joint Treaties Committee or the Public Works Committee, have often worked in cooperation with a key adviser and with greatest experience in the facts rather than the politics. This may prove to be optimistic but, as with other policy recommendations from time to time by the Commission, optimism about policy-makers' interests is necessary in order to advocate first-best solutions.

Several submissions in response to the Draft Report also flagged the possibility of establishing an additional cross-sectoral forum to play a role in the designation process. For example, the University of Melbourne (sub. DR305, p. 15) recommended that a National Data Consultative and Advisory Forum be established, with one key role being the provision of advice on, and nominating or assisting with, decisions about NIDs. Such forums could play an important role in the community consultation for NID designation.

Broader implications

There are already some high value and enduring public sector datasets and data 'spines' within Australia (as noted earlier). For these datasets, the recommended approach to NIDs would have ongoing and significant implications, regardless of whether or not they met the criteria to be designated as NIDs. But the recommended approach would also have significant broader implications, including for data currently collected by the States and Territories, and for the private sector, and these are discussed in this section.

State and Territory data

The process of NID designation 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 may be integrated.

Consideration should be given to intergovernmental agreements as a potential means through which to facilitate cooperation across jurisdictions. In this regard, the NSW Government (sub. DR327, p. 8) stated:

> The final recognition of NIDs should be made through an agreement by governments (in consultation with the original data custodians). The determination should be made public and reasons provided when a recommendation from the independent assessment is not implemented.

The NDC and parliamentary committee would clearly need to consult closely with States and Territories regarding their data holdings.

Private sector data

In some cases data that would form part of a NID is likely to be held across different sectors. 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 some private sector data in the suite of NIDs would be 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.

There was significant reaction to the Commission's Draft Report concerning the possible inclusion of private sector datasets as NIDs. The Business Council of Australia (sub. DR317, p. 14) expressed concern that there was a lack of criteria to indicate the scope of private sector datasets that might qualify, and that the proposed process did not include a detailed consideration of costs and benefits. Google Australia (sub. DR292, p. 3) said:

> … we remain concerned about the potential that private sector data may be caught up in a National Interest Database, resulting in private businesses being required to disclose proprietary data that has significant value, or for which commercial business models already exist, without an opportunity to outline the reasons why this may not be appropriate or to initiate an administrative review of appeal in cases where the business disagrees with this designation.

Similar concerns were also raised by AGL Energy (sub. DR251, p. 4), Westpac (sub. DR324, p. 3) and DIGI (sub. DR326, p. 8). The Communications Alliance and Australian Mobile Telecommunications Association (sub. DR250, p. 3) called for businesses to be compensated on a cost recovery basis for any data they are obliged or requested to contribute (dealt with earlier in this chapter).

Conversely, CSIRO (sub. DR323, p. 19) argued that significant benefits may accrue, for both business and government, from inclusion of private sector data within NID. A number of Commonwealth Departments also offered supporting comment.

As can be seen by a dispassionate observer of the Commission's recommendations, a process to outline reasons against designation of a particular dataset as an NID is proposed. It is not an administrative appeals tribunal process because the purpose is much wider than whether process was handled correctly. The process focuses clearly on whether there is at stake a *national interest*. And a process via disallowable instrument does in fact allow for business to disagree. But in the open, public sphere of the parliament.

The private data the Commission generally has in mind is data held and most often collected due to services funded or (at times) legislatively authorised by public policy — Commonwealth and State. The root of this thinking is also behind recommendation 6.3, which advocated that before contracting out to the private sector, the public sector should move to proactively consider data retention/access. It is, in practice, a matter of common

sense. Firms that are unable to envisage why the public sector might, for example, need to retain access to (now private) data on transport times for transport planning purposes are in effect questioning the ability of the private sector to deliver public services at all.

Such data has a public interest inherent in it, or the public sector should not be funding the service or investment. It is far from unreasonable to seek to retain access to it; and to use it for better design and evaluation. This does not mean the public interest *must* prevail. But neither does it mean the private interest in limiting such data must be beyond any consideration.

We have not limited our recommendations to just this described data simply because it is impossible to neatly divide it; and because some private data (land transactions was noted earlier, based on a private submitter's view) might be incorporated into a NID, by purchase. Such action occurs today, on an ad hoc basis. It too is far from unreasonable.

Thus consideration should also be given, in select circumstances, to payment for private sector access to NIDs as envisaged above.

Where the NDC is of the view that identified private sector datasets would be likely to deliver significant private spillover benefits and the nature of charging is likely to be consistent with the public interest designation of the NID in the first place, a case for payment may be considered and as noted above, in a manner consistent with existing practice.

Given that candidate datasets would be drawn from across the private and public sectors and, in the latter case, involve datasets collected by the Commonwealth and States and Territories, voluntary provision in the first instance would be ideal. However, as discussed in chapter 6, where key datasets are held by entities that are regulated for a public purpose or publicly funded, a change in data provision requirements may be necessary. The parliamentary scrutiny process would act as a check on any assertion of outrageous behaviour.

RECOMMENDATION 7.4

The Australian Government should make provision, in select circumstances as approved by the funding Minister, for the National Data Custodian to pay for access or linkage to private sector datasets (Recommendation 9.4).

Equally, the National Data Custodian may consider applying charges for access to National Interest Datasets where this would not be inconsistent with the public interest purpose of the National Interest Dataset.

It is expected this would not be a common occurrence, in either case.

8 A modernised regulatory framework

Key points

- Legislation, in the form of the Data Sharing and Release Act, should be introduced to create consistent rules for improved data sharing and release.
- The Data Sharing and Release Act serves two major purposes:
  - First, as a clear and unambiguous signal of the shift in approach of the Commonwealth Government and parliament to the data issues uncovered in this report, a crucial step to achieving cultural change amongst a myriad of data custodians.
  - Second, as a structural framework, principles-based and outcomes-focused as far as practical, to give authority and guidance to effective and responsible use of information in a data-rich future.
- The Data Sharing and Release Act would require a risk-based approach to improved data sharing and release.
- It would, for the purposes of other legislation that impedes consideration of improved sharing and release, authorise the sharing of data within the public sector and with ARAs.
- New institutional arrangements to promote trust and confidence in the data sharing and release system are centred on improved capabilities, accreditation procedures for both users and custodians, and most particularly the National Data Custodian.
- National Interest Datasets would overcome impediments to the effective integration, sharing and use of data of national significance, currently hindered by multiple legislative barriers.
- Privacy is an important human right and existing protections are retained. The Data Sharing and Release Act should, where it deals with release of personal information (as defined in privacy legislation), operate subject to existing Commonwealth privacy legislation.
- To ensure that individuals (including small businesses) can participate in data sharing opportunities, a new definition of data — consumer data — would be created to cover all digital information to deliver the ability of consumers to utilise their data, as outlined in chapter 5. The changes in consumer data rights should be supported proactively by existing consumer-oriented regulators.
- The Privacy Act is not an effective vehicle for the reforms covered in this report, particularly when it comes to consumer data. Privacy regulators seem unconvinced of the need, and emphasised to us their preference for the human right of privacy over consumer interests. Whereas from a national welfare perspective, as required in the Productivity Commission's Act, it is evident that privacy is one aspect of data use, not pre-eminent other than that it is incumbent.
- The Data Sharing and Release Act should, to the maximum extent available under Commonwealth powers, establish a national framework, and offer cooperating jurisdictions (and occasionally private datasets) the ability to become integrated and accessible for research and other authorised purposes.
- The benefits of a consistent framework, that is scalable and adaptive over time, have generally been widely recognised by stakeholders responding to the Draft Report.

8.1 The Data Sharing and Release Act

Previous chapters have recommended reforms to institutional arrangements to empower consumers (chapter 5), build trust and embed a risk-based approach to sharing and releasing Australia's data (chapter 6), and designate certain high value datasets as nationally significant, thereby ensuring their ongoing curation and use (chapter 7).

Implementing this package of reforms has the potential to bring significant economy and community-wide benefits. These reforms are best implemented through:

- legislation, in the form of the Data Sharing and Release Act (DSR Act), designed to create consistent rules for improved data sharing and release; and
- guidelines, issued by the newly created National Data Custodian, to promote best practice and effective risk management (chapter 6).

This chapter explains our recommendations and rationale for creating this new regulatory system.

We have designed our Framework to put trust front and centre of the reform package (chapter 4). Trust is necessary if progress is to be widespread and sustained. The DSR Act contains a number of measures to promote trust and transparency — discussed later.

Why there is a need for a legislative approach

Systemic institutional and cultural change is necessary for Australia to realise the full value of its data. There are too many barriers to better data sharing and use, in the form of existing regulations, incentives and attitudes, to have any confidence that far reaching change (and benefits) could be achieved through piecemeal reform of current policies and legislation, mostly designed for a pre-digital world.

Legislative reform provides the opportunity to enshrine a new approach to data — a permission to use data assets in a way not envisioned when the current rules and practices were established. It signals a different approach to data sharing, to data custodians *and,* perhaps most importantly, their agency and political leaders. Legislation provides clarity around the rules for improved data access, and embeds expectations regarding key issues such as effective risk management and the continuous promulgation of best practice (as it emerges).

These two features — clear signalling of intent and clarity around rules and expectations — are needed to underpin broadly based and sustained cultural change. Unlike individual data projects and policy statements, legislative change endures and becomes embedded in behaviour. Alone, it is not sufficient for cultural change, but without it there would be no such change.

Our recommendations require leadership via central agency responsibility in each jurisdiction, and once established, by the National Data Custodian (NDC).

Establishing the NDC would also help to drive cultural change. As an independent statutory office holder in a central Commonwealth portfolio, the NDC provides a locus of responsibility for the core changes to data risk management and accreditation.

A new DSR Act should serve to grant an authorisation — and indeed create an expectation — to share and release data, subject to modern safeguards; and the NDC should support the delivery of those objectives in practice.

Australia has a unnecessarily restrictive legislative data framework

Too often public sector data custodians and researchers are required to negotiate multiple pieces of legislation, many with criminal offences that apply strict liability for various actions. It forms a complex web, with intimidating consequences for missteps, that reduces the likelihood that data would be put to good use. For governments, and its agents, these circumstances contribute to excessive risk aversion to sharing or releasing data. It is also likely to lead to reluctance to admit breaches or missteps when they occur. And, a complex legislative framework is costly and contributes to a lack of certainty about legal responsibility for data once it has been transferred.

It is now well recognised (internationally, as well as in New South Wales, South Australia and Western Australia to different degrees) that effective risk *management* rather than risk avoidance, produces far better outcomes (chapter 4) — in terms of providing robust safeguards and liberating Australia's data for use. But this more contemporary approach is not reflected in the existing Commonwealth legislative framework — or in that of other States and Territories. This is clearly suboptimal.

The DSR Act should, as far as possible, set overarching and consistent rules of the game and drive best practice approaches to data sharing and release. For this to occur, the DSR Act should have as broad a remit as possible (discussed below).

Participants to this Inquiry (Department of Social Services, sub. 10; Australian Taxation Office, sub. 204) have pointed out the lack of consistently applied best practices across government. With such variance, the reluctance of some data custodians to share data is understandable. Driving the application of consistently applied best practices across entities is therefore a critical function of the DSR Act, and guidance issued by the NDC.

Somewhat self-evidently, piecemeal amendments to existing legislation would not achieve this in a consistent manner. New legislation does more than simply fill a gap in the current system — it creates an overarching Framework to manage risk properly and ensure confidence in the system is maintained. The DSR Act would ensure that the system operates transparently and safeguards are robust. But it should also ensure that good incentives are maintained throughout the system, including incentives to add value to and

use data, for entities in the system to act cooperatively in the national interest, for entities to continue to evolve data management approaches in line with best practice, and for compliance.

Why not just amend the Privacy Act?

Privacy, as its regulators and some supporters told us (OAIC, sub. DR236 NSW Privacy Commissioner, sub. DR268; Xamax, sub. 3) is a fundamental human right. While this was not in dispute as far as this Inquiry was concerned, there are a number of reasons why the Privacy Act is not the appropriate vehicle for reform.

First, continuing to look at consumer rights and data sharing through a lens clearly dedicated to the protection of a human right is unlikely to see cultural change: it encourages data to continue to be viewed as a risk rather than an asset. A more proactive approach is needed. The purpose of the DSR Act is to create and enable cultural change — this requires that we move beyond the lens of privacy.

Second, privacy principles are not the only tool in the data management toolbox, any more than privacy is the only consideration in the use of data. Taking advantage of incentives, such as actively building trust in data use, is equally important.

There is no legislative vehicle other than the Privacy Act at the national level that deals comprehensively with information or data, and even then the Privacy Act does not reflect all relevant considerations. There is no regulator for breach of confidentiality or commercial in confidence (although court action may be taken), no complaints handling body for anything other than privacy, no institution responsible for driving improvements in risk management, de-identification, or trusted user arrangements. Such an absence, in a world now heavily characterised by data analytics and opportunity, is startling enough to justify consideration of a new Act.

Privacy is important, and would be retained in this legislative approach. But the benefits to safer, consistent and more accessible data identified elsewhere in this Report make the case for new, more comprehensive legislation beyond question.

Embedding a broader view of data sharing and release through a new Act allows a unified approach to be taken to the different, competing legislative considerations across dozens of Acts associated with sharing and releasing data — there is a wide range of legislation containing ‘secrecy’ provisions that claim to protect various interests, and in narrow ways authorise release of data for limited purposes (appendix D).

As chapter 3 outlined, this is a complicated space, and it is important the new regulatory regime allows for the full spectrum of complexity to be considered when making decisions about data sharing and release, and to have a legislative instrument that fully reflects the breadth of considerations.

A principles-based, outcomes focused approach

The DSR Act would provide clear and consistent rules for sharing and release of some of Australia's most valuable data — something that is sorely lacking at the moment (chapter 3). It would also implement the scalable, risk-based approach to sharing and releasing Australia's data outlined in chapters 4-7, in conjunction with guidance issued by the NDC.

This new legislation and guidance should be principles-based and outcomes focused. This enables it to adapt to changing circumstances, and evolving practices and technologies over time. A principles-based approach would enable Accredited Release Authorities (ARAs) and data custodians to develop best-practice risk management processes that reflect the characteristics of particular datasets and uses, and to monitor and report on them in the most effective way.

Whereas principles-based legislation is often subject to criticism that it lacks clarity and certainty, in this context such criticism must surely ring hollow — all involved appear to agree flexibility is required to adapt to changing circumstances over time, and guidance can provide more detail where required. The primary locus of current regulation of data — privacy regulation — is *also* principles-based, and this approach has long been familiar to practitioners. Prescriptive regulation is only proposed where other alternatives are likely to prove ineffective for the purposes envisaged. However, the Commission also recognises that principles-based regulation requires more monitoring to occur — discussed later.

Regulatory frameworks governing data should also be as technologically neutral as possible in order to 'future proof' the scheme. Our approach to the debate over APIs is substance in support of this (chapter 5).

Reforms that have broad reach

The DSR Act would have broad reach across all Australian and State and Territory government bodies, and private and not-for-profit sector bodies, including small businesses; although data collected by law enforcement and national security agencies should be specifically exempted (chapter 1).

To maximise the benefits of this legislation, it should draw on the full extent of the Commonwealth's constitutional powers to legislate with respect to the nation's digital data. However, application to States and Territories would be on an *opt-in* basis. New South Wales and South Australia lead the way in better practice and policies in data, and the new Act would also seek to recognise this.

Digital data would preferably be defined along the lines of the *Data Sharing (Government Sector) Act 2015* (NSW) (section 4(1)):

> … any facts, statistics, instructions, concepts or other information in a form that is digitally communicated, analysed or processed (whether by a computer or other automated means).

The motivation for directing the scope of the DSR Act to digital data is to target the information that is: most usable today; transferable at the least cost; with the greatest ability to be analytically plumbed for new insights and opportunities.

RECOMMENDATION 8.1

New Commonwealth legislation — the Data Sharing and Release Act — should be passed drawing on the full range of Commonwealth powers to regulate digital data, in order to authorise the better sharing and release of data.

The new Act should also establish the Comprehensive Right of consumers to access their data from government and private data holders alike, for the purposes of improving the services that are offered to them by alternative providers.

Changing culture

The objectives of the DSR Act should be to:

- facilitate sharing of public sector and other publicly funded data in accordance with the trusted user principles
- promote greater use and curation of nationally significant datasets
- promote greater release of non-sensitive data
- provide safeguards around the sharing and release of data
- advance competition and consumer choice through the Comprehensive Right.

The DSR Act would recognise an explicit distinction between sharing and release (as described in chapter 1).

A key function of the DSR Act is to grant an overriding authorisation to share and release data in order to cut through outdated secrecy provisions and provide a modernised legislative framework for Australia's data.

The DSR Act would provide that data custodians and ARAs *may* share and release data *despite* the terms of other legislation as long as it is done:

- to achieve a public interest purpose; and
- applying the five safes principles, in compliance with one of the sharing and release processes set out by the DSR Act; and
- subject to the safeguards contained in the DSR Act.

To be clear, that Act should signal that since modern risk management approaches would mean that data can be shared and released safely, then this should happen as a matter of course (when there is a public interest in doing so, which is not the case now).

Public interest purposes should include but not be limited to:

- increasing the availability of open data
- promoting better service planning and delivery and program management
- allowing the development of better statistics or analysis
- assessing the performance of public services or programs
- informing policy
- allowing public benefit research to occur
- National Interest Datasets (NIDs)
- exercise of the Comprehensive Right, and
- other purposes as declared by the NDC via disallowable instrument.

The intention in describing the breadth of potential public interest is not to create an expectation of vetting projects or programs against some form of hurdle, but rather to highlight the many and varied ways in which data sharing and release can *and should* contribute to public benefits. Purposes, as specified in the new Act, will need close attention to ensure that, as separately recommended (Recommendation 6.12), trusted users can receive ongoing access on a program rather than project basis.

The aim of specifying purposes in the DSR Act is to build a social licence by assuring participating entities and the community as to what their data would and would not be used for by ARAs and data custodians. We consider that purposes should extend to better service delivery — experience in New Zealand with targeted early interventions shows the lifetime benefits can be significant (chapter 2).

The DSR Act should also establish some data handling principles that should be applied in all sharing and release processes — for instance: that privacy and commercial confidentiality should be managed carefully, but the public benefit of increased data use should also be recognised; and that the five safes principles should be applied in sharing and release decisions (chapter 6).

The DSR Act would enable a culture of data utilisation by establishing the following features:

- *a presumption of open data:* all non-sensitive data should be made open, with guidance issued by the NDC on what constitutes non-sensitive
- confidence in trusted users: authorise risk-based provision of data to trusted users in a secure environment subject to application of the five safes principles.

The processes to deliver these features would be broadly set out in the DSR Act, and subject to more detailed guidance issued by the NDC.

ARAs and public sector data custodians would be required by the DSR Act to adopt a risk-based approach to sharing and release of data. A risk-based approach requires active consideration of both risks and opportunities, and applies a sliding scale of controls commensurate with the sensitivity of the data (as per the five safes principles). More

sensitive data would have more stringent approval processes and controls for sharing. Less sensitive data would be shared more freely or released.

Critically, the DSR Act should be focused on increasing 'access' rather than transfer per se. Providing another entity access to data need not involve a data transfer or 'dump', and the DSR Act should recognise and support this. Modern technological advancements allow real-time access to a live feed of data without any transfer occurring, or integration to occur without holding the original dataset. Data should not be thought of the same as paper files — and providing access in this way is a valuable risk management tool. Our discussion of safeguards later also raises the importance of not simply sending copies of datasets around, but rather, facilitating remote access (with no data downloads) that can be readily terminated in the event of a breach.

Building trust through capable institutions

We have designed our Framework to put trust front and centre of the reform package (chapter 4). Trust is necessary if progress is to be widespread and sustained. Our reforms are designed to build trust by promoting capable institutions and clear accountabilities.

Public sector data custodians: The DSR Act would cover Commonwealth Government agencies, and State and Territory bodies that have opted into the new Framework (as 'public sector data custodians'). As discussed in chapter 6, we envisage that public sector data custodians would not reduce their current activities aimed at sharing, integrating and releasing data on a business as usual basis. For them, the new Act would be most beneficial in its objectives to authorise wider and more proactive data use, and through its new institutions that would provide guidance on better practice.

Accredited Release Authorities: More complex, higher-risk integration projects that cannot be undertaken by an individual data custodian would be undertaken by Accredited Release Authorities (ARAs). These ARAs would be established by the DSR Act and accredited by the NDC (as discussed in Chapter 6). Accreditation would require release authorities to demonstrate: data management and security capability; linkages to, and the ability to assess or accredit, trusted users; and a public interest, national approach to their activities. Given these requirements for accreditation, we envisage that ARAs would be public sector or publicly funded institutions. It would be unlikely today to find a private entity capable of exhibiting all the desired characteristics, but in the future the NDC (in its role as accreditor) should be able to review partnerships that do involve private commitment.

The model recommended by the Commission anticipates that some existing entities (as an example, the Australian Institute of Health and Welfare) would be ARAs. Where existing entities that are well placed to become ARAs are operating under restrictive legislation, and have the data they hold subject to the terms of their existing restrictive legislation, further legislative change may be required to remove these restrictions.

Entities that are candidates to be ARAs and yet already have existing legislative penalty regimes will find participation in this new Framework challenging. Moreover, the ability to

share governance of submitted datasets — which in the case of a few prospective ARAs is also restricted by legislation — is another key ARA characteristic towards making it attractive for parties (private sector, State and Territory) to open up and share their datasets. As noted in the Report, this would either need to addressed in reforms (Part 2 reforms, discussed later) or otherwise mean some existing entities could not be used as ARAs. The ABS has been cited in this context, although its ability to be an ARA is subject to other limitations, specifically on governance.

National Data Custodian: The NDC would be established as an independent statutory authority. In addition to accrediting ARAs, the NDC would be charged with issuing practice guidance that covers ARAs and public sector data custodians, and recommending on the selection of NIDs. This is expected to drive improvements in practices across the system.

Under this model, we intend that the actions of all existing data custodians *and* ARAs would be guided and governed by the NDC under the DSR Act. Yet this NDC role is guidance (and subsequent review of whether advice was taken), not approval. We want to avoid the creation of new bottlenecks. In other words we are not seeking to curtail *in any way* the existing sharing and release by public sector data custodians. For the removal of doubt, the obligations of current data custodians and future ARAs to curate their data on an ongoing basis should also be clear in the terms of the DSR Act.

The DSR Act would further promote trust by establishing risk-based approaches to handling data — which would focus more attention on *all* risks associated with data collection, storage *and* use), providing robust safeguards in the Act, incentives for compliance via accreditation, and other mechanisms designed to ensure the capability of institutions (such as certification and audit of best-practice de-identification processes — chapter 6). It would be complemented by transparency — parliamentary scrutiny and public benefit assessment processes (chapter 7) — and reporting mechanisms.

RECOMMENDATION 8.2

The Data Sharing and Release Act should establish the risk-based approach to data sharing and release and accompanying institutional frameworks.

- All non-sensitive data held by agencies and Accredited Release Authorities (ARAs) should be explicitly presumed to be made public, consistent with the Australian Government's Public Data Policy Statement.
- Data custodians and ARAs would be authorised to provide sensitive data to trusted users in a secure environment, with de-identification where necessary for risk management of the data.
- The National Data Custodian should have the authority to issue guidance on how the risks of *all* sharing of identifiable data between entities should be managed. This guidance should be updated where it judges the risks have shifted.

8.2 Actioning a scalable, risk-based approach

Different users, different risks

The objective of the DSR Act is to facilitate *public interest* sharing and release of data. A way of thinking about the model we have designed is to recognise that there are broadly speaking, three groups that may access data. In increasing order of potential risk, these are:

- Public sector agencies, ARAs or institutions with a data sharing agreement — we consider sharing between these bodies would be mostly low risk, as both are covered by the DSR Act and public sector codes of conduct and more likely to act in the public interest. A signed agreement (such as we envisage would be required of trusted users) would usually be sufficient to describe and manage risks.
- Non-government trusted users — a broader, but still select, group of users that would have varying 'risk profiles' and overall may have a higher risk. This group may need to be subject to more specific obligations or controls.
- Open users — as no control processes could apply after open release, data provided must be non-sensitive or already in the public domain (there is plenty of already-released data which is not easily found or analysable — ARAs and hopefully more original custodians would at times be aggregators and editors).

In chapters 4 and 6 we endorsed embedding the five safes principles within a risk based approach to data access. The DSR Act should explicitly endorse the application of these principles (as 'trusted user principles'), with specific sharing processes and more detailed guidance issued by the NDC. A risk-based approach means that higher risk uses should be subject to more stringent controls, while lower risk uses are likely to have less stringent controls applied. For instance, a public sector employee would be bound by a public sector code of conduct and other conditions of employment. Academic institutions have strong (at times too many, but nevertheless strong) ethics committee approvals.

However, risk is likely to vary between organisations — some public sector organisations for instance might have consistently poor data handling practices, and should be treated accordingly. Risk ratings are discussed further below.

Critically, however, a trusted user model recognises that even for higher risk users and uses, there are ways of managing the risk other than tormenting the data — for instance the environment may be made more secure, and the project vetting may be more stringent. Like an onion, risk management has layers. Each of these five safes can be dialled up or down depending on the risk involved in the project. These should *not be seen as cumulative requirements*, but rather, ways to balance out risk, while still enabling sharing and use.

Access by a public sector agency or an ARA

There is significant scope to improve data flows between Commonwealth, State and Territory agencies, and between these bodies and trusted users (through ARAs, once established). Within the public sector, it can take years to negotiate today's unwise forms of data Memorandums of Understanding (MOUs), where the terms vary considerably, culture opposes any risk no matter how unrealistic, and conditions and safeguards often are simply unenforceable. Providing a proactive form of authorisation would overcome risk aversion, and improve timeliness, while greater accountability would promote trust.

Improving information flows between public sector bodies and ARAs is critical to the operation of our Framework. Getting access to the data is a critical first step to being able to integrate it and share it with trusted users.

In order to put into effect the opportunity for such proactivity, the DSR Act should provide that public sector agencies or ARAs may provide another public sector agency or ARA with access to their data if:

- the sharing entity and the receiving entity are covered by the DSR Act, and
- the original data custodian has agreed to the transfer, and
  - agreement could be obtained via a contract or licence condition requirement, or preferably a modern data sharing agreement of a form issued by the NDC
- the data is being provided to the entity for a DSR Act purpose, and
- the trusted user principles adopted by the NDC have been applied — these principles would cover amongst other matters if and how any identifiable data is used, and
- a Data Sharing Agreement has been entered into (throughout this chapter, we treat Data Sharing Agreements as a preferred method of facilitating clarity in data sharing in future)
  - the Data Sharing Agreement should concentrate on *how to achieve* the transfer, and make sure each party clearly understands the expectations of the other; while observing the DSR Act.

This approach has been used in South Australia under the *Public Sector (Data Sharing) Act 2016* (SA) and it is similar to the Approved Information Sharing Agreements that operate under the *Privacy Act 1993* (NZ). It offers permission to act, by comparison with past incentives to avoid action. It also focuses on proactively embedding safeguards and creating a culture of change, rather than the lack of risk management that has characterised data sharing to date (chapter 3). Importantly, application of the trusted access principles in this case does not require data sharing to occur through a strict trusted user model such as SURE, nor does it necessarily require checking of output for disclosiveness (box 8.1). Secure storage is likely to be sufficient, and sometimes it will be necessary for the output to identify an individual (for instance, to target service delivery). Our model *allows* this to occur subject to safeguards.

Box 8.1 South Australia's approach to applying the trusted user principles

Section 7 of the *Public Sector (Data Sharing Act) 2016* (SA) requires data custodians to *consider* a number of matters in determining whether the trusted user principles have been met, including whether:

- there is a public interest in the proposed program that outweighs the risk of the data use, and what the cost is of not using the data for the proposed purpose
- the receiving agency has the necessary skills and experience
- the providing agency would be able to engage with the receiving agency to support the use of that data
- other persons 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
- data should be of the necessary quality for that purpose, and whether the results of the data analytics work would be audited and whether that process includes the data provider
- the environment in which the data would be stored, accessed and used must be able to keep data secure and not subject to unauthorised access or use
- the receiving agency would limit access to the data to people with the appropriate security clearance
- publication or other disclosure is appropriate having regard to the nature and audience of the proposed publication
- the publication could contribute to the identification of a person.

Source: Public Sector (Data Sharing) Act 2016 (SA)

The DSR Act should also provide that the NDC design a model Data Sharing Agreement which sets out how the data would be provided by each agency, what the data would be used for (including on-sharing) and how these trusted user principles would be applied. Custodians and ARAs covered by the Act would need to apply this standard, once issued.

Agencies sharing within the same jurisdiction, sharing with an ARA, or that have agreed to share data *between jurisdictions* under the DSR Act would all be covered. The model would thus need to consider many circumstances and is unlikely to prove particularly onerous — it is more a way of advising what good practice looks like.

State and Commonwealth data sharing is meant to be facilitated by the DSR Act, and providing a consistent legal framework would go a long way to achieving this. In practice we expect that much new data sharing would generally take place through an ARA, as these entities would be accredited to achieve key dataset integration and curation.

Providing a legislative requirement that data collecting agencies need to consider and embed the five safes principles when sharing information with each other would go a long way to simplifying the current unwise MOU process. Extraneous quasi-legal processes which the Draft Report noted were common in MOUs should be removed. And preferably nothing added to replace them.

With each sharing agency then in effect working to the same objectives, and adopting processes to achieve these, it should be a relatively simple matter to develop a data sharing agreement by comparison with today's multi-year processes.

We envisage that these agreements would allow ongoing access. This means the trusted user principles should be accepted at the agency level — that is, they should not require repetition for each project, or repetition for each individual user, but rather establish trusted user status once per dataset, and review only in the light of performance issues (or an evaluation date, for example, every two or three years). Structure is important, but simplicity and an outcomes focus — thus avoiding repetition of documentary processes for their own sake — are also equally important.

If, over time, Data Sharing Agreements and their negotiation remain a barrier to more effective data sharing, the NDC could issue further guidance, including through the development of model agreements or templates, to better facilitate sharing between agencies. This should help to ensure that agreements are not simply window dressing (DPMC 2015), but actually assist agencies in clarifying the terms of an exchange and the NDC in its task of offering advice on improving risk management.

All agencies sharing information with one another should be required to publicly report (in their annual reports or on their websites) whether they are following NDC guidelines in developing data sharing agreements, and if not, why not and what alternative practices have been adopted to address the underlying intent of the NDC guidance (box 8.2). The purpose of this approach is to recognise that while it is desirable for best practice to be observed, it is of even greater import that there is genuine ownership by the agencies involved for the processes and practices adopted, and that data sharing actually improves.

Legislation should provide that breaches or near breaches of these agreements be reported to the NDC as close to real time as possible. The objective of breach reporting should not be to impose penalties for simple accidental breaches that were accompanied by immediate redress and immediate reporting of them. They should be used, rather, to highlight where data management practices may be failing and to incentivise learning and improvement. The NDC should be required under the Act to advise of breaches in its Annual Report.

Consideration could also be given to how this breach reporting could translate into a 'risk rating' by the NDC so that all data sharing agencies are able to learn from broad experience and adjust their sharing practices to reflect the higher risk record. This is discussed below.

Box 8.2 National Data Custodian guidance on trusted users

Guidance should be provided by the National Data Custodian (NDC) on:

- Standards for de-identification taking into consideration the sensitivity of the data and the security of the trusted user environment, guidance on how to distinguish between uniqueness vs re-identification risk, and on how to manage the risk of re-identification via spontaneous recognition (such as by trusted users signing legal undertakings).
  - Ways of providing lower risk data could include providing a sample rather than the entire population and generating a synthetic dataset. The appropriateness of this would need to be assessed having regard to the purpose for which the data is required.
  - Where data about businesses in highly concentrated industries is being provided, it may not be possible to fully de-identify the data. In such cases, potentially identifiable data should be provided to trusted users. Suitable controls should be developed by the NDC in its guidance.
- Suitable risk-based approval processes for trusted users.
  - Private sector parties are higher risk users and may need to be subject to more stringent safeguards, including assessment by an ARA of whether their research is in the public interest, closer vetting of the research question, and controls on accessing potentially identifiable data about businesses in the same industry.
- Harmonising the legal undertakings signed by trusted users that provide conditions and safeguards on use.
  - The Commission considers that a two-part approach to legal undertakings is best practice. This would require both the individual trusted user and the organisation that the trusted user belongs to sign legal undertakings. This creates incentives for the organisation to enforce individual compliance with the terms and conditions of data use, lest it lose access to data for all its trusted users.
  - Legal undertakings should provide that access would be terminated in the event of a breach and may require trusted users to provide a monetary bond upfront that would be forfeited in the event of a breach where other assurances do not exist or are of limited effect. ARAs must report breaches to the NDC to ensure a consistent approach to trusted user risk.
- Output checking, including ways to manage the risks of cumulative disclosure while recognising that individuals may need to request draft and final output.
- Other risk management considerations — for instance, how to ensure the user environment is secure in a platform such as SURE.

Access to data in exceptional circumstances

There may be exceptional cases where public entities that have not acceded to, and are not otherwise covered by the safeguards in the DSR Act nevertheless have established (such as in a COAG agreement) a need to access data managed under the DSR Act. For instance, a State that has not volunteered its data for inclusion under the DSR Act and does not intend at present to do so.

It would be the antithesis of the reforms in this chapter to prevent data exchanges, even where the parties are yet to agree to place data under modern management via the DSR Act and there is clear public interest in the exchange occurring.

For this unique but predictable circumstance, Commission recommends that the NDC be given the power to issue disallowable instruments authorising the sharing of information between two entities.

This process should still be governed by the same principles applied to Data Sharing Agreements above, including applying the trusted user principles, but might apply more politically-driven or even monetary penalties (such as bonds) for breach. The disallowable instruments themselves may be either enduring or sun-setting, but the DSR Act should require they be publicly reviewed every five years.

The Commission has noted the comments of the Attorney-General's Department (AGD) (sub. DR334, p. 17) that support the general concept of sharing but raise concerns that legislative instruments can be inflexible. The Commission agrees. The use of disallowable instruments should be kept to a minimum — our model here has limited them to a unique but desirable mechanism that enables data exchange in the transition to wide acceptance of the new Act.

Risks should be shared between the agency and the ARA

Even after a public sector data custodian has transferred data under the DSR Act, they still bear reputational (and in some cases, legal) risk for the actions of the receiving data custodian or ARA under their own legislation.

Our Framework recognises that responsibility for managing the risk of the data should be shared between the data custodian providing access and the data custodian or ARA receiving access. This reflects the position at law, as well as good risk management principles that require responsibility for the risk to be allocated to the entity best able to manage it. The Framework should not promote 'set and forget' behaviour.

In practice, requiring the parties to enter into a Data Sharing Agreement outlining how they would implement the trusted user principles would pre-emptively address these risk allocation issues.

The provisions of the DSR Act authorising transfer of data between parties should be drafted to exempt the data provider from liability for the actions of a data receiver where the data provider can demonstrate that they have acted in good faith and for a proper purpose.

What can the receiving entity do with the data?

What the receiving entity can do with the data should be specified in the Data Sharing Agreement and be consistent with the DSR Act — for instance, the receiving entity may integrate the data, provide access to trusted users, or make a non-sensitive version of the dataset open. ARAs and data custodians would also be required to comply with the Commonwealth *Privacy Act 1988* in what they do with the data they have received under the above provisions — for instance, providing data to trusted users or making data open. Application of *Commonwealth* privacy legislation is preferred in order to promote consistency and allow State and Territory ARAs to hold, without question, Commonwealth data under the restrictions imposed by Australian Privacy Principle 8.

A Data Sharing Agreement model issued by the NDC should address directly the question of whether and if so how the receiving entity should be able to on-share untransformed datasets provided by another data custodian without their awareness, particularly where the data is identifiable. We have referred to a Data Sharing Agreement being a vehicle for clarity of activity, recognising that lack of clarity and control by originating data custodians over on-sharing, in particular, have in the past created an unwillingness to share and a general lack of trust.

RECOMMENDATION 8.3

The Data Sharing and Release Act (DSR Act) would, where possible, override secrecy provisions or restrictions on use that prevent original custodians actively providing access to data to other public sector data custodians and Accredited Release Authorities (ARAs).

Access should be governed by Data Sharing Agreements that embed the trusted user principles, actively assist data sharing and create clarity of understanding amongst the parties. The National Data Custodian (NDC) should issue a model Data Sharing Agreement early in its life, and update it from time to time.

The DSR Act should establish modern, clear and supportive standards — the new 'rules of the game' — for data sharing and release. The Commonwealth Privacy Act would continue to apply, as well as any residual obligations emanating from the original data custodian's legislation.

Existing protections would remain on datasets that do not utilise the DSR Act, in order to ensure there is no gap between the accountability obligations on original public sector data custodians and the ARA.

In limited exceptional circumstances as the DSR Act transitions to becoming nationally effective, it may be necessary to provide access to data shared under the new Act to a party that has yet to adopt its provisions. The NDC should be provided with the power to use a disallowable instrument to allow access or sharing for such transitional purposes.

Data integration

Promoting data sharing between data custodians and ARAs would allow more integration to occur within and between sectors and jurisdictions. The above reforms are designed to achieve this. Managing the risk of integration will be one of the matters covered by the NDC in its guidance.

Returning integrated data to linkage participants

The Commission considers that data custodians who have participated in an integration project should have access to the linked, de-identified dataset to which they contributed for use with no restrictions other than necessary to prevent re-identification. This creates a significant incentive for participation in linkage projects. Enabling re-use of linked data assets also prevents overlap and duplication of datasets created under the DSR Act regime.

Access by original data custodians is currently stymied by concerns about re-identifying the entire dataset or use of the data for a different purpose. While these are technically possible, the original data is only likely to be provided where there is a strong interest in the research outcome, and in reality, the incentive for the original data custodian to re-identify data is not obvious. Moreover, in a well-considered risk analysis, only consequences additional to those likely from the original dataset are of relevance.

The Commission considers this to be a risk that should be managed thoughtfully rather than subject to a blanket prohibition. Priority access under trusted user models or other secure storage arrangements, for instance, could be used to ensure the identifiers are kept separate from the de-identified linked dataset. Subject to these controls, parties should be free to transform the linked dataset and reuse it as they see fit.

The DSR Act should provide for this to occur, and the Commission expects that the NDC would issue guidance on managing risks.

Access by a wider range of trusted users

The Commission's Framework has endorsed the use of trusted user models for providing both identifiable and sensitive de-identified data to entities not within the public sector, usually academic researchers, but also potentially private sector parties undertaking public interest research (chapter 6).

Trusted user models provide significant control (commensurate with risks and sensitivities attached to particular data) over who accesses the data, and generally require outputs from any data use to be checked for disclosiveness. This allows sensitive information to be used more so than would be acceptable in a less secure environment. As these users are potentially higher risk, in general, we envisage that this sharing and release will occur through an environment such as SURE.

As the DSR Act operates subject to the Commonwealth Privacy Act for sharing and release to these trusted users, the following framework would be applied when assessing what data should be provided to trusted users:

- We have recommended (chapter 6) that sections 95 and 95A of the Privacy Act, which establish a framework for the provision of identifiable data for health and medical research without consent, be extended to all research that is in the public interest — for example: education and childcare research, and public transport research. As a consequence, in these areas, identifiable data may be provided to trusted researchers under the DSR Act.
- Where these revised exceptions do not apply, trusted user models may still be used to provide de-identified data under the DSR Act (for instance, linked datasets).

We envisage that data analytics would occur fairly freely under our Framework — trusted users would be able to access data from ARAs and data custodians and run code in secure environments, subject to safeguards. Importantly, even though trusted user models allow *sharing* of identifiable or lightly de-identified data, they generally require *release* (output that is published) to go through a number of output checks to ensure it is non-sensitive.

Risk management — best practices

Different data and different users would have different risks, so the calibration of trusted access models by ARAs and data custodians (that is, dialling up or down the controls as discussed earlier) would be governed by a risk assessment process, as discussed above and in chapter 6.

Importantly, we have not ruled out private sector parties being trusted users — indeed this may be necessary as the outsourcing of public functions continues to increase. But private sector entities may be subject to more stringent vetting and controls in line with their potentially higher risk.

A harmonised, best-practice approach to risk assessment in trusted user models is important in order to develop and maintain a social licence. The NDC should be responsible for developing guidance on a best-practice approach to administering trusted user models, with the aim of promoting consistent risk management practices across ARAs and data custodians (chapter 6). Its accreditation processes will provide the foundation for this. Assessment by the NDC of ARA compliance with these best practices should be taken into account in gaining and maintaining accreditation.

The NDC may be asked by data custodians or ARAs to assist with judgements where there is doubt and accredited processes do not provide sufficient guidance. However, the model is predicated on ARAs and data custodians managing the risk of their data — the NDC cannot substitute for such accountability.

Risk assessment — security classifications

The Australian Government's Protective Security Policy Framework (PSPF) sets out the rules for determining the appropriate classification for Australian Government information, and similar policies exist at the State and Territory government level. Classified information is subject to various controls on its handling and use, and these controls increase the higher the level of classification applied to the information. Information is classified depending on the consequences of damage from unauthorised compromise or misuse of the information — for instance, loss of confidence in government. This means a large amount of administrative data would likely be subject to some form of classification.

These policies are vital in protecting government interests. However, the Commission notes with concern two elements of the PSPF that appear to be at odds with data handling law.

First, de-identification of data is widely accepted in law to reduce the risk from release of that dataset, but the PSPF does not explicitly recognise that de-identification of data can reduce its classification, nor does it give guidance on how to assess this reduction in classification. In other words, the PSPF does not explicitly reflect one of the most common approaches to reducing dataset risk (despite purportedly adopting a risk management approach).

Second, the PSPF only assesses classifications on the harm that would result if the data was released. It does not take into account the *probabilit*y of misuse given the security of the environment the data is held in or other controls in place. Again, this seems at odds with the current legal position, and the Commission recommends the PSPF be updated as a matter of urgency to align with the current legal position.

The Commission notes the Australian Government is currently considering its response to the Belcher Red Tape Review which recommended the Australian Government take steps to reduce the over-classification of information.

RECOMMENDATION 8.4

The Australian Government's Protective Security Policy Framework (and equivalent State and Territory policies) should be amended to recognise that the risk and therefore the classification needed for data can be reduced by:

- transforming a dataset, for example through de-identification, such that the risks of misuse on dataset release are reduced
- only making the transformed data available to trusted researchers in a secure computing environment, with usage monitored and output checked for disclosiveness.

This would align the Protective Security Policy Framework with the current legal environment.

The Australian Government should consider doing this as part of its response to the Belcher Review.

Public access: Open data

The DSR Act should contain a clear authorisation that all non-sensitive data held by data custodians and ARAs may be made open. This is intended to strengthen and formalise the existing Australian Government Public Data Policy Statement (and similar State and Territory pronouncements — appendix C).

When data is made open (public), there are no risk management controls other than the characteristics of the data itself. The Commission expects that open data would be non-personal or aggregated and otherwise transformed to meet the descriptor of non-sensitive. ARAs would be required under their conditions of accreditation to deal with this effectively, and the NDC would provide expertise to data custodians to deal with this effectively, and issue standards and guidance on the processes by which potentially sensitive data may be made non-sensitive.

National Interest Datasets should be designed for broad use

We proposed in chapter 7 that the NDC assess and nominate for declaration as National Interest Datasets (NIDs), a particular category of datasets that should be treated and funded as national assets (because of their ability to deliver benefits in the national interest), and available for wider use in an integrated fashion.

NIDs would often cut across jurisdictions and sectors — examples *could* include education, health service provision, family and community services, and social housing. In such cases, access to data from the public, private and not-for-profit sectors in all jurisdictions could be necessary to achieve a compelling public interest purpose.

The DSR Act would establish the NIDs regime, and provide for designation of NIDs where:

- there is a compelling public interest — for instance:
  - the dataset has clear substantial spillover benefits
  - the dataset can be used as a basis for performance evaluation of publicly funded projects or programs; or
  - has an established focus on nationally significant subject matter, such as education or health; and
- access to data from the specified entities is necessary to achieve this compelling public interest.

Designation of NIDs should occur via disallowable instrument issued by the NDC. The DSR Act would set out the process and criteria for declaring NIDs, and specify which ARA should be responsible for curating that dataset. A parliamentary committee (new or existing) would be appointed to scrutinise the process, which has the benefit of transparency and keeping in check any overuse of disallowable instruments.

Providing the NID to an ARA means that the ARA will — as well as its most basic purpose of creating additional benefit through improved data use — be required to comply with the DSR Act and best practice guidance issued by the NDC in curating and managing the dataset.

The benefits of using a disallowable instrument process for NIDs are as follows.

- NIDs are intended to be enduring and exist for at least ten years, making administrative work-arounds (today's approach) unattractive and ineffective.
- NIDs may not be immediately evident in many fields where data use has been greatly constrained, and their declaration by disallowable instrument would offer both certainty of use and clarity of value to the nation.
- Disallowable instruments provide a simple, transparent way of including data that could potentially cut across sectors and jurisdictions.
- NIDs will vary widely, from fairly aggregated data to highly sensitive identifiable data. Disallowable instruments allow flexibility in the risk-based safeguards applied to the data, while providing assurances.
- Disallowable instruments allow transitional arrangements to be put in place for the creation of NIDs (discussed below).

There would be clear processes for accrediting and enabling trusted user access of NIDs (described in earlier chapters). These datasets must be funded — as ARAs themselves would be — to achieve specific additional gains from data use.

The use of NIDs has particular complexities given the compelling public interest in making them more widely available.

We envisage that a subset of more aggregated NIDs will be provided to trusted users in a secure environment to use fairly freely with few if any checks on the output. This could be achieved for example through generation of synthetic data or providing a sample rather than the full population.

Performance reporting data, while sensitive, may need to be shared and released on an identifiable basis to enable effective consumer choice — for instance, particular health service providers or facilities. There is a strong public interest in this data being released, but there are also risks that need to be managed — the disallowable instrument should be used to provide a legal authorisation for this to occur. This is discussed further below.

Finally, specification of the NID should recognise there may need to be fewer controls on on-sharing of these NIDs than on other datasets — for instance, an ARA may need to provide it to a performance reporting body to achieve the compelling public interest the dataset was declared to achieve. The DSR Act should provide permission for this and other compelling public interest uses to occur for NIDs via the NID process.

Data from non-government sectors

While the Commission has canvassed inclusion of non-government (private and not-for-profit) sector data in NIDs, this would only occur under exceptional circumstances where there is a compelling public interest.

In the majority of cases compelling public interest would arise where public services are delivered by both government and non-government organisations (for example, health or education). In such case, access to the non-government data is required to meaningfully examine performance and outcomes for the entire sector. And there can be significant benefits from releasing data on the performance of individual businesses to enable effective consumer choice (Harper et al. 2015).

Any compelling public interest in sharing and release of non-government data must be balanced against considerations of maintaining business incentives to collect and add value to data (including though their own analysis), and claims of intellectual property issues. But in the vast majority of cases, NIDs will come from businesses that are regulated or significantly funded for public purposes. Providing data for public benefit purposes should be business-as-usual for these entities.

Ideally this would occur via contract as part of the terms and conditions of receiving public funding (chapter 6). But there are other levers. It is currently government policy to assert ownership over intellectual property developed as a result of public funding (PC 2016). This should be extended to apparently copyrighted data. Many of these businesses are licensed or regulated — again, it could be part of the terms and conditions of gaining the licence that data generated under that licence is available for public benefit use.

While we do not underestimate the difficulties, it is unarguable that private hospital or private education data should be made available, where it is in the public interest to complete a comprehensive picture of public spending and national welfare, or the performance of publicly funded services. We expect that putting claims to the contrary to public scrutiny via a parliamentary committee will prove to be inherently desirable.

Another option is re-use of existing regulator data holdings for this purpose. This occurs already — for instance, ACMA's statutory performance reporting obligations (Part 27 of the *Telecommunications Act 1997* (Cth)). While possible, there are difficulties associated with this. For instance, there may be common law liability issues associated with re-use of such data holdings for another purpose, such as the potential for causes of action related to breach of confidentiality or economic loss. Loss of trust in regulators if they use their data holdings for another purpose has also been raised with the Commission. But these difficulties can be solved by implementing proper processes. NDC guidance should cover this.

Voluntary provision following negotiation would usually be preferable, and it is envisaged that in many situations there would be mutual interest in seeing datasets linked in support of a deeper understanding of how to improve the effectiveness of service delivery.

Intergovernmental arrangements have been struck at times — States and Territories provide hospital performance data under the Performance and Accountability Framework agreed by COAG in 2009, while private hospitals provide this data on a voluntary basis for publication on the MyHospitals website, currently managed by the AIHW.

But the DSR Act should thus provide for disallowable instruments to be used to extract relevant data holdings in circumstances where such agreement cannot be reached and the benefits are demonstrable. Any steps to force access to data would require acquisition on just terms. The high profile nature of such a step would itself provide an incentive towards negotiation rather than use of regulation.

Other levers may be more effective in the longer term. Sectors where there may be a compelling public interest in accessing non-government sector data are generally also sectors that are heavily government funded and/or regulated. Including in funding contracts, a provision that allows for data to be reused (as discussed in chapter 6) is an option. Licence and associated regulatory provisions might also be reviewed. These appear to be more attractive options than acquisition on just terms, and are likely to be more enduring in the long run.

Safeguards

Robust safeguards are essential to promote trust. There is benefit to the DSR Act providing a consistent approach to safeguards, rather than the piecemeal approach in evidence today. An overarching approach would clarify existing legal confusion, such as who is responsible for data after it has been transferred to another body.

But the Commission strongly counsels against recreating the strict criminal liability culture that has so badly impeded current data sharing (with no evidence of effectiveness). Onerous criminal sanctions — which assume fault for a breach regardless of intent ('strict liability') do not create good incentives for sharing data, or incentives to report breaches. The objective of the DSR Act to provide a flexible and modern framework would not be supported by repeating a failed approach.

Instead, the DSR Act should create incentives for self-management of risks and voluntary improvement of data management practices as follows. We have designed this system to focus on promoting self-regulation, but with subsequent audit of ARAs against accreditation funding agreements and guidance standards, and reporting by all custodians. These are powerful tools and have been structured in this Report to try to keep the NDC to as small as possible an organisation while being a strong force for good practice.

Establishing a principles-based, outcomes focused regime: Data handling principles (such as the trusted user principles) should be contained in the DSR Act and guidance on best practices to achieve these should be issued by the NDC.

Reducing information asymmetry and promoting trust: Entities covered by the DSR Act should be required to publicly report on their compliance with this guidance on an 'if not, why not' basis. This means stating that they either:

- implemented the best practice guidance, or
- if not, explain why not, and how alternative systems and controls address the intent of the guidance.

'If not, why not' reporting provides flexibility in *how* the outcomes are achieved not *whether* they are achieved.

We consider the NDC should be informed of adverse events, including when:

- reporting requirements are not met
- the principles contained in the DSR Act and underpinned by guidance are not achieved — for instance, if a dataset is not properly de-identified before it is made public
- the conditions of a Data Sharing Agreement have been violated — this reporting should be mandatory
- trusted users have breached a legal undertaking
- an ARA has violated its conditions of accreditation.

We do not consider issues of judgement around risk management under the guidance issued by the NDC should be subject to appeal — this judgement is the NDC's core work. What matters is whether the underlying outcomes have been achieved.

The NDC should be required to maintain a public register of adverse events reported to it, such as violation of the conditions of access by public sector agencies. Recognising the extreme sensitivity of some within-government issues, the NDC should have the power to handle them in a confidential manner, but must always publish, at a minimum, the name of the organisation involved.

The purpose of this register is to assist over time in identifying entities or data collections that are higher risk. Over time, as experience and evidence emerge, guidance may be available on request to data custodians or ARAs looking to share data with users new to them. This could take the form of a 'risk rating' issued by the NDC or provided by participants about ARAs or data custodians they are dealing with. Entities with a history of adverse events would have a higher risk rating; entities that have implemented robust data handling processes would have a lower one.

Entities with a higher risk rating might be subject to more stringent restrictions on accessing and using data or lose their access altogether, until such time as improved security and risk management processes can be demonstrated that lower the risk of dealing with that entity. Under our model, a public sector entity that was previously low risk may, as a result of its increased rating, be subject to the same controls as a private sector entity. This is likely to create significant incentives to improve its data handling practices.

Creating incentives for continual improvement: The DSR Act should contain a simple sequence of penalties to enable the NDC to adopt the 'enforcement pyramid' principle — at the top, systemic, deliberate breaches should be subject to the highest penalties. At the bottom, mistakes with no consequences that are immediately reported to the NDC should be used for learning.

The NDC should be allowed to issue a warning and work with the breaching party to improve their data handling processes, in all but the highest categories of breach above. We have also endorsed the NDC being able to accredit best practice de-identification processes and conduct audits to support improvement in risk management processes over time (chapter 6).

Failure to report a breach should fall within the high penalty range. But the DSR Act should not contain criminal penalties. We also consider ARAs should be held to a higher standard because of their trusted position in the Framework.

The NDC is responsible for accrediting ARAs and can manage their compliance with the DSR Act through accreditation conditions. The penalty for breaches can be focused on the ARA's accreditation and funding agreement with the Commonwealth. The NDC should be able to recommend funding for an ARA be withdrawn or continue under revised conditions. Datasets could also be reallocated. Entities that have received access to data may also be required delete their own holdings if they have breached their conditions of access under a Data Sharing Agreement.

The NDC should also be able under the DSR Act to handle complaints of lack of access from trusted users or potential trusted users of datasets covered by the DSR Act. This focus on trusted users is given special preference because it is the basis for the reforms — to encourage greater trusted user access. But these complaints should be limited to non-compliance with the DSR Act, not issues of judgement around application of the NDC's guidance or ARA risk assessments.

8.3 Achieving a consistent approach

There is a wide array of Commonwealth and State restrictions and limitations on data use that are outdated and strongly discourage data custodians from making effective use of data. The Commission is recommending the DSR Act to resolve these complexities and simplify the existing legislative framework.

Australia's legislative framework is complex. In 2010, the ALRC identified 506 secrecy provisions within 176 pieces of Commonwealth legislation, and undoubtedly a similar plethora of secrecy provisions exists at the State and Territory level also (chapter 3). While these secrecy provisions protect legitimate interests, too often they are drafted in overlapping and confusing ways. Significant costs are incurred from legislative requirements that are inconsistent across jurisdictions.

Attempting to amend hundreds of legislated strictures across Commonwealth and State legislation, with the consequent assessment of each that is required to do this job effectively, is impractical and would restrict the opportunity for early benefits to be realised. This section discusses the overarching framework that would be established by the DSR Act, and how it should be implemented.

Consistent rules of the game

The DSR Act is intended to create a consistent legislative framework for Australia's data sharing and release. It should make use of the communications power in section 51 of the Constitution to cover the field with respect to digital data, although other powers could potentially also be relied on such as the corporations and external affairs powers. There are some limitations on this power, but they are not likely to significantly affect the overall efficacy of the scheme (box 8.3). The Commission has undertaken informal due diligence but expects that further legal advice will be sought from the Australian Government Solicitor on this point (as per NSW Government, sub. DR327).

Box 8.3 Limitations on the Commonwealth's power to cover the field

There are some limitations on the scope of legislation under section 51, both express and implied — most relevant is the implied constitutional limitation that a federal law may not discriminate against a State, or prevent a State from continuing to exist and function as an independent unit of the federation.

However, while State powers may be 'effectively restricted or their exercise made more complex or subjected to delaying procedures' by the operation and requirements of the DSR Act, the DSR Act is unlikely to affect the existence and nature of the 'State body politic': *Western Australia v Commonwealth* (1995) 183 CLR 373.

Legislative provisions applying to public sector employees in the higher levels of state government may be one qualification to the Commonwealth's power to cover the field with the DSR Act. The High Court has found that Commonwealth laws that seek to regulate state employees at the 'higher levels of government' (including ministers, ministerial assistants and advisers, heads of departments and judges) may interfere with the existence and nature of a state (*Re Australian Education Union; Ex parte Victoria* (1995) 184 CLR 188, 233; *Austin v Commonwealth* (2003) 215 CLR 185). Another limitation may be if the DSR Act purported to regulate the handling of information that goes to the core of State government functions, such as cabinet-in-confidence documents and other highly sensitive documents (ALRC 2008). Given the DSR Act will exclude data in cabinet-in-confidence documents, serious limitations on its scope are not likely.

Source: ALRC (2008)

Further, it is intended to cover States' and Territories' commitment of datasets on an *opt-in* basis. And, while it is hoped that States and Territories might pay serious heed to the NDC's advice on better practices, it is only when datasets are committed by them to an ARA that NDC guidance becomes a matter that *must* be considered or applied. The reach of the Commonwealth under this new Act is intended primarily to be used in a positive

manner, to effect the removal of restrictions on the commitment of data to the extent constitutionally possible as and when agreement is made on the nation-wide benefits.

Undoubtedly, a range of issues would have to be worked through in developing this national approach, such as the interactions with some State and Territory laws that would need to continue to operate under any national legislative scheme. For this reason, the Commission recommends extensive collaboration and consultation with the States and Territories to implement this national approach. These issues are discussed further in chapter 10.

Two parts of legislative reform would be needed to implement a consistent legislative framework:

- Part 1: Passage of the DSR Act that purports to operate to the exclusion of other legislation.
- Part 2: Amendments to other specific legislation that cannot be overridden in Part 1.

Under Part 1, the DSR Act would authorise the following actions to the exclusion of all other legislation:

- risk-based provision of data to trusted users
- public release of non-sensitive data
- creation of national interest datasets.

Generally, based on legal review, a provision excluding the operation of other legislation appears likely to be effective and sufficient for some but not all possible cases. Its effectiveness would depend on how the legislation is drafted (table 8.1).

For instance, where the data custodian's head legislation provides it can be overridden by another law (such as in the *Telecommunications Act 1997* (Cth)), the DSR Act would be able to override it with certainty (cf. Telstra, sub. DR312; Communications Alliance, sub. DR250). However, where the data custodian's legislation has quite specific provisions, or also contains a provision that purports to exclude all other legislation, it is unlikely the DSR Act would be able to oust this.

Where outmoded restrictions — inconsistent with the purpose of the DSR Act — cannot be assuredly replaced via an explicit authorisation under the DSR Act, specific amendments to the dataset's head legislation should be undertaken to create that assurance — part 2 of the Commission's proposed reforms. These amendments should create an exception to the general secrecy provision to allow data be shared and released in accordance with the DSR Act. Similar exceptions are already common in a range of legislation, but self-evidently not in all the many places where DSR Act datasets might be drawn.

Table 8.1 **Exceptions to some Commonwealth secrecy provisions**

General circumstances where sharing of identifiable information is permitted without consent of individuals

Legislation	*Exceptions*	*Effectiveness*
Telecommunications Act 1997	Authorised by law	Part 1 will be effective
Disability Discrimination Act 1992	Authorised by law	Part 1 will be effective
Australian Securities and Investment Commission Act 2001	Authorised by law	Part 1 will be effective
National Health Act 1953	Public interest certificates	Part 2 likely required
Social Security Administration Act 1999	Public interest certificates	Part 2 likely required
Higher Education Support Act 2003	To assist in a Commonwealth officer's official employment	Part 2 likely required
Taxation Administration Act 1953	For specific uses, to an exhaustive list of public sector entities, under specific legislation only	Part 2 required
Australian Organ and Tissue Donation and Transplantation Authority Act 2008	None	Part 2 required
Census and Statistics Act 1905	None	Part 2 required

Source: Commission analysis

The recommended way of approaching this task is best exemplified by NIDs. In chapter 7, we have suggested a number of datasets that might conceivably be declared as NIDs, including in health and education. These datasets cut across multiple sectors and jurisdictions, and are subject to different legislation with different provisions. For instance, the *National Health Act 1953* (Cth) — just one of many Commonwealth laws that apply to health information — permits general disclosure only via a public interest certificate issued by the relevant Minister, and hence, the Part 1 authorisation is unlikely to be effective in circumventing this requirement. Similar issues (where exceptions to secrecy are too specific to be overridden) are present in much State and Territory legislation (appendix D). Without Part 2 amendments, the dataset would be seriously compromised.

We recommend that an initial set of NIDs be identified by the NDC to accompany the DSR Bill, following processes to establish additionality and public interest, to enable the parliament and various interested parties to see clearly the kind of datasets that the model intends to free up. And to consider cost and benefits in real terms rather than in the abstract. Transparency will be essential in this process to build social licence and community trust.

Although NIDs are the simplest way to conceive of the possible need for Part 2 amendments, it should be noted that they may also be required for non-NID datasets held by ARAs. Such a process would depend on what ARAs are created and what cooperation emerges amongst data holders.

The datasets freed up by the DSR Act would cover both those voluntarily shared with ARAs by the Commonwealth, States and Territories and non-government sector but not specified as NIDs, and those declared by disallowable instrument as NIDs.

RECOMMENDATION 8.5

Legislative reform to implement the Commission's recommendations would need to be undertaken in two parts, moving forward together:

- the **first part** is the passage of the Data Sharing and Release Act (DSR Act) itself, that authorises to the greatest extent practical in a single statute, the sharing and release of data for the purposes of the Act and removes existing Commonwealth and State restrictions on integrating, linking and research uses of datasets by Accredited Release Authorities
- the **second part** is a further legislative amendment process that may be necessary, depending on the particular characteristics of, for example, National Interest Datasets, in order to address residual restrictions on the use of specific datasets that were not able to be effected by the DSR Act itself.

The National Data Custodian should be asked to identify residual legislative restrictions that need removal in its consideration of National Interest Datasets.

Transitional issues

For reasons of transparency, the Commission has recommended that Part 2 amendments occur via passage of bills through parliament in the case of the initial set of NIDs. The necessary amendments if any would be evident from that selected list.

Subsequently, and as the NDC recommends more NIDs, there would be a need to prioritise Part 2 actions.

Unsurprisingly, the Commission's preferred approach is to prioritise Part 2 amendments according to where these restrictions impose greatest cost (and thus opportunity to do much better is high). Part 2 amendments would offer a legislated process under which *for those specified datasets* the relevant restrictions would be lifted and replaced by a common regime of privacy standards (the Commonwealth Privacy Act) and other fit-for-purpose modern standards that achieve integrated, safe, secure, properly-resourced, effective re-use of data with a proactive focus on sharing or release.

In sum, criteria for identifying legislation highly suitable for amendment should be where:

- the data held by the custodian is of high value to the Australian community
- there are minimal alternative mechanisms available to the data custodian under its existing legislation to participate in the DSR Act framework (for instance, the data custodian is not able to use a public interest certificate to release data).

Finally, we have recommended (recommendation 8.3) that in circumstances where it is unclear if Part 2 legislation is actually needed, disallowable instruments to amend other legislation might be a preferable way to allow other data uses to take place and avoid minor doubt (particularly given the cultural shift required here). This should be seen as a transitional mechanism only, and be subject to parliamentary committee oversight along with disallowable instruments used under the broader NID scheme (chapter 7).

Interactions with other legislation

Privacy legislation should remain unaltered

The Commission is not proposing any changes to the *Privacy Act 1988* (Cth), other than extending section 95A to cover all research in the public interest that relates to people (recommendation 6.16). While the DSR Act would authorise identifiable data to be shared with data custodians and ARAs, the Commission intends that data integration, provision to trusted users, and making data open would still be subject to privacy legislation.

Wherever possible, the Commission's preference is for the Commonwealth Privacy Act to apply to all actions occurring under the DSR Act. The DSR Act should contain a provision that applies the Commonwealth privacy legislation to actions taken under the DSR Act, to the exclusion of State and Territory privacy legislation. In practice this means that Western Australia and South Australia would be bound by Commonwealth privacy legislation for all their actions under the DSR Act if a State entity were envisaged as an ARA (chapter 6 noted a couple of high quality examples). This of course may be a desirable step in advancing privacy policy.

Other legislation

The Commission commends the progress made by New South Wales and South Australia with their passage of legislation to facilitate data sharing within their public sectors. The DSR Act is unlikely to be inconsistent with this legislation, but where there is inconsistency, the DSR Act should prevail.

Similarly, other public sector data management schemes, such as those for archives and freedom of information, are also intended to operate in tandem with the DSR Act — there is no inconsistency apparent to the Commission. Specification of the DSR Act should make clear the interactions with these other legislative schemes, in line with the *Public Sector (Data Sharing) Act 2016* (SA) model.

RECOMMENDATION 8.6

The Data Sharing and Release Act (DSR Act) should have national reach — to create a simplified and transparent one-stop location for a national framework for data volunteered, declared or acquired for inclusion under the DSR Act.

The Act should allow for the acquisition of private datasets via disallowable instruments as part of the process of creating National Interest Datasets (NIDs). Acquisition should only occur on just terms after parliamentary scrutiny determines the benefits are demonstrable.

An initial set of NIDs should be identified by the National Data Custodian to accompany the DSR Bill, following processes to establish additionality and public interest.

The DSR Act should apply Commonwealth privacy legislation to datasets managed by Accredited Release Authorities where feasible. It should be drafted with reference to (and with the intention of being consistent with) the *Data Sharing (Government Sector) Act 2015* (NSW) and the *Public Sector (Data Sharing) Act 2016* (SA) to the extent possible.

8.4 Implementing the Comprehensive Right

The Commission has recommended the introduction of a Comprehensive Right for consumers to view and propose edits to correct errors, and be told of actions to trade their digital data, and to direct data holders to transfer their data to a third party provider of a competitive or complementary service (chapter 5).

The Data Sharing and Release Act would house any legislative authorisations needed for the Comprehensive Right. As described in chapter 5, this Right would include a new description of digital data that would form consumer data in the absence of industry-level agreements offering a better option. The aim is to ensure consumers have a right to access a fit-for-purpose dataset generated by their actions, sufficient to generate improved service offers across a variety of services and markets (also non-markets, such as some community services).

A standards-setting process needs to be established under the DSR Act, to allow the Australian Competition and Consumer Commission (ACCC), as the relevant regulator, to register the industry-agreed scope of consumer data and agreed standards for transfer and data security. The registered standards for each industry would give the ACCC a guide against which to judge whether the Comprehensive Right is being delivered or not.

The inclusion of the Comprehensive Right in the DSR Act is crucial to balancing the opportunity to take advantage of data opportunities across all three major data holders — firms, governments and consumers.

It moreover can use the Commonwealth legislative reach to allow consumers to have consistent rights across all sectors and jurisdictions without having to navigate amendments to State and Territory privacy legislation, or grapple with Western Australia and South Australia not having any privacy legislation at all. Given many consumers may want to transfer their data across jurisdictional boundaries (for instance, from a State-regulated public hospital to a Commonwealth-regulated private hospital), a consistent approach is vital.

The purpose of having a suite of rights — view, propose edits, transfer, and be informed — for consumers is simple: if in the course of directing the transfer of their data to third parties (the right that is not accorded to anyone under current Australian law, privacy or otherwise), a consumer discovers an error, they should not have to revert to a different regulatory regime to seek a correction.

Where the right to access and correct information would overlap with the Privacy Act, we propose the drafting of the DSR Act recognise and accept this overlap. It is not an undesirable outcome for overlap to occur, given the separate objectives of each and the net benefits of the reform, so long as regulators are conscious of it and apply a no wrong door approach. Firms and agencies, as we noted in an earlier chapter, can simplify this matter for themselves, if they view the generation of community trust as a positive opportunity to continue data gathering, rather than a compliance burden.

The existing exceptions to access to personal information in the Privacy Act are broadly suitable to be applied to the access and transfer rights in the DSR Act, but should be considered in detail at the final drafting stage. Reasonable exceptions, in principle, may include when giving access would prejudice current legal proceedings, be contrary to another law, or reveal information in conjunction with a commercially sensitive decision making process. This last would for example deal with some submissions' concerns about excluding decision-making algorithms and copyrighted information from transfer.

However, access to energy consumption information should not be excluded from the DSR Act. The current regulated access to consumer information in energy should become consistent with the Comprehensive Right in the DSR Act.

The DSR Act would contain an assurance of transfer that cannot be contracted out of or traded away. Participants raised concerns (chapter 5) about potential liability for mistaken transfers, or where a third party makes a fraudulent request on behalf of a consumer. While initial advice provided to the Commission suggests that this liability would be rare, the Commission considers the DSR Act should protect businesses from liability where they are acting in good faith. Implementing robust security standards (chapter 5) would help to prevent this situation arising in the first place — prevention is better than cure.

Responsibility for administering the Comprehensive Right should rest primarily with the ACCC, other than where an industry ombudsman already serves a purpose. Telecommunications, electricity, gas and water appear to have applicable entities. Banking may also, although the Commission notes that different stakeholders see different parties

(each with some regulatory status) as being most relevant. Choosing a preferred party would be a task for implementation, although some broader issues are discussed below.

8.5 Streamlining regulatory responsibilities

The Commission's recommendations focus on allocation of responsibility between three main regulators:

- the NDC, a new body responsible for oversight of Australia's data sharing and release and administration of the DSR Act
- the ACCC, which would be given additional competition and consumer related responsibilities under the DSR Act — although industry ombudsmen and State and Territory offices of fair trading would also play a role
- the Office of the Australian Information Commissioner (OAIC), which would retain its existing responsibilities under the *Privacy Act 1988* (Cth), but not have any specific new responsibilities under the DSR Act.

We have also recommended a central agency in each jurisdiction bear initial responsibility for implementation of the Commission's Framework until the NDC is established.

Our intent in allocating regulatory responsibilities is to make use of existing expertise, and to fill a major gap in the existing system — there is currently no regulator responsible for ensuring the integrity of Australia's data sharing and release processes. The NDC has been designed to fill this gap.

The responsibilities of the NDC and the additional consumer data responsibilities of the ACCC should be contained in the DSR Act.

Opportunities for cooperation and reform

We have recommended a data-specific regulator on the basis that one is needed to provide whole of system oversight and to drive good practices, applying a lens that is broader than just consumer protection or privacy. Participants have supported this approach (amongst others the Department of Prime Minister and Cabinet, sub. DR286; Australian Taxation Office, sub. DR314; Choice, sub. 167; Energy Consumers Australia, sub. DR316). The Commission recognises that the existing regulatory landscape is a challenging one for consumers to navigate.

We intend that the ACCC would have the primary complaints handling role for consumers under the DSR Act.

However, consumers may be more likely to make first contact with an industry-specific regulator than a more general one, as an industry-specific regulator (such as the Telecommunications Industry Ombudsman) is likely to have greater visibility to the average consumer, who is likely to have very little knowledge of allocations of

responsibilities between regulators. Accordingly, the ACCC should enter into agreements with these industry-specific regulators and the OAIC to ensure there is 'no wrong door' for consumer complaints. Some industry ombudsmen may need additional statutory backing to handle these third party complaints.

Finally, when the Office of the National Data Custodian is established, it may choose to enter into an agreement with the OAIC to the limited extent that the enforcement of the Data Sharing and Release Act interacts with enforcement of privacy legislation.

RECOMMENDATION 8.7

The Australian Competition and Consumer Commission (ACCC) and the Office of the Australian Information Commissioner should enter into working arrangements with each other, industry ombudsmen and other relevant bodies at all levels of government to support a 'no wrong door' approach to how individuals (including small businesses) pursue complaints or queries regarding their rights as consumers to data held on them.

Where an industry data-specification agreement (Recommendation 5.2) seeks to use a recognised industry ombudsman to address consumer complaints, this should be considered by the ACCC as part of its acceptance or rejection of a proposed industry agreement.

9 Transformation and pricing decisions

Key points

- Opportunities for harvesting data and creating new business ventures are important determinants of the extent to which organisations choose to transform data, and/or share it with other parties. The transformation of data to enhance its value, and its sharing with others, has resulted in major structural change in some markets and businesses over the last decade. Structural change in government has been much less substantial.
- Public and research sector datasets also offer attractive opportunities for new and innovative private sector services, but government agencies and research bodies have limited ability to determine the value of data they might release.
- Assessing value could be important for assessing the case for 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 transformation and make the value-added data available; or
  - the agency can perform the transformation more efficiently than either users of the data or 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; and the information technology upgrade risk is assessed and found to be small.
- Beyond this, government agencies should refrain from additional transformation of datasets. The delay incurred by agencies in doing so prior to release (or sharing) can be substantial, and data users generally 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. Which approach is most suitable will vary according to user demand and agency capability to act commercially.
  - For a given level of data quality, making data freely available in a timely way will maximise use and hence deliver the highest level of social benefits. But it will increase the net cost to government of data release.
  - Where agencies undertake substantial transformation 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 separate review.
- Maintaining and increasing the availability of public sector data is not costless. Sharing or release of minimally processed datasets should be funded by agencies from existing budgets. For datasets that have significant and demonstrable public interest and are to be made more available, additional funds should be provided to agencies to ensure the quality of dataset curation.

One of the enduring characteristics of data is that many parties can access it simultaneously and repeatedly, and yet its value remains undiminished. Entrepreneurial forces, supported by price signals and the potential for profits, are driving innovative collection and use of data. The private sector identifies a purpose, values it strategically or financially (or both) and determines how best to utilise data. This valuation guides business decisions on transformation (and pricing if a business decides to market any of its data holdings).

By contrast, the value (and any subsequent consideration of pricing) of data tends to be less clear when it resides in government hands. 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 might be put, particularly private uses.

The context for this chapter is valuation and pricing of data, with a particular focus on the challenges for the public sector, as it seeks to make data more widely available. Costs associated with transformation of data under the Comprehensive Right and National Interest Datasets (NIDs), specifically, are discussed in more detail in chapters 5 and 7 respectively.

9.1 Transformation and sale of private sector data

Market opportunities for harvesting data and the possibilities of competitive advantage are the substantial determinants of the quantity and quality of data that private sector entities (businesses, industry organisations and not-for-profits) collect, add value to, and share with other parties (either by sale or other mutually amenable arrangements).

Where data is shared or released for use in the private sector by other parties, the value placed on it, and how access is charged for, will be determined by a range of factors, including the: strategic competitive advantage created by doing so; level of demand (willingness to pay) for such data; the extent to which data has been or needs to be processed for use and the likely cost of any transformation required to get the data to a marketable standard; and the extent to which the business wants to maintain control over the data by restricting with whom, and on what terms, it shares its data.

The constantly improving capacity of smart technology to extract new insights from data and apply artificial intelligence and learning algorithms, is likely to extend the value of even relatively well-analysed datasets in the future.

These types of activities have induced major structural change in some markets (accommodation, taxis, books, music) and within firms (the application of Watson analytics by IBM in a plethora of fields is an exemplar). But such *structural* change is not evident amongst Australia's governments, although other chapters have cited some examples of practice improvements.

Transformation by businesses

Transformation of data by businesses will be done for either internal purposes — to identify market trends, profile customers, and guide investment and pricing decisions — or for external purposes — to share with or sell data to parties such as:

- suppliers — to enable suppliers to 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 increase sales, build goodwill, for promotional purposes or to comply with regulatory requirements
- industry organisations — for industry development, benchmarking, promotion, or to inform and influence governments.

Separate activity also goes into collecting and framing data in response to various licencing or regulatory obligations, to supply agencies such as the Australian Bureau of Statistics, the Australian Taxation Office, Australian Securities and Investments Commission, State Revenue Offices, or local governments. Specialised regulators (such as the Australian Prudential Regulation Authority and the Australian Transaction Reports and Analysis Centre in the finance sector, the National Health Performance Authority in the health sector and the Australian Curriculum, Assessment and Reporting Authority in education) also induce data collection or transformation in the private sector.

Transformation to enable the *sale* of data is a relatively recent development outside of specialised industries such as advertising or marketing. Today, and much more so in the future, by virtue of more accessible big data analytics, most firms will be able to engage in active management of their data assets, if there is advantage in doing so.

The question of value then becomes central: how much will additional transformation cost and how much is it likely to return? What is the strategic purpose of this activity? Is data suppression in the private interest? While the public sector faces the same conundrums, 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 somewhat opaque public interest objectives.

Data pricing by the private sector

The impact of market power on pricing

Some data intensive businesses — oft-cited exemplars are Amazon, Facebook and Google — have very rapidly gained substantial market share in the sectors in which they operate.

Where market power exists, a concern is that it could result in, or be reinforced by, restricted access to data. Another concern is that the choices of consumers providing data to these entities may become constrained. On the other hand, market power often proves to be transitory (particularly when it is not entrenched by licencing regulation or intellectual property protection), and the interoperability of technology and linked services that may come with a large market share can be to the considerable benefit of consumers.

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), while entities want to use customer data for their own purposes, they also do not want to alienate those parties with policies that may be deemed too invasive.

Governments are already active in requiring data and monitoring of prices where the risk of abuse of market power may be higher, such as in essential services after they have been privatised (for example, energy utilities and telecommunications). Moreover, should adverse circumstances arise, the *Competition and Consumer Act 2010* (Cth) provides the Australian Government with the authority to address certain misuses of market power.

General government intervention to deal with real or perceived market power in a period of rapid change necessitates fine judgements, to be made carefully. 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).

The Commission has recommended in other recent work that governments should not leap to regulate new market entrants (Uber, Airbnb and others). Rather, governments should ensure their existing regulatory arrangements and consumer protections remain effective in delivering policy objectives, but are technology neutral and accommodating of innovative business models (PC 2016).

The actions recommended in this Report are designed very much with such thinking in mind — opportunity to innovate and technological neutrality in particular.

9.2 Transformation and sale of public sector data

In the case of the public sector, broader and wider access to data is equally pertinent. Consumers of public health services should be readily able to transfer their data, both to another provider and also within a health facility. The latter will provide spill-over benefits to the health facility itself. Much investment has been made by various parts of the health system but today, transfer of a consumer's data from ward to ward of a hospital is still an exception rather than a rule (appendix E).

The same cost and benefit queries that apply to private sector data sharing and release are relevant to the public sector. What is the strategic rationale for an agency in preparing data for sharing or release, particularly where the value to be gained from doing so is unlikely

to accrue to that agency? What is the case for transformation? And what charge is relevant, if any?

These issues affect both the ideal level of transformation of data before making it available and the price charged for data shared or released.

Decisions relating to transformation and pricing will be influenced by a range of incentives, disincentives, obligations and constraints faced by government agencies (chapters 1 and 3) 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*, that it is in a usable format and, conversely, that time and resources are not wasted on processing effort that is not valued by data users.

'Basic data' versus 'transformed 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 transformation) to make the data fit for sharing or release
- 'transformed data products' — to which an agency has undertaken more extensive curating or processing effort, often with the purpose of targeting the data at a specific set of uses or users — rather than general release.

Agencies will usually have to undertake some processing of data to make it fit for release (rather than just for internal use), a point noted by the Australian Government Linked Data Working Group:

> [R]eleasing data in its raw form, without for example, adding value by defining the schema of the data, or lacking machine-readability of the data or lacking links to other datasets is, in fact, creating a barrier for private sector entities to add value. (sub. DR278)

A minimal level of data curating and processing for basic data release should aim to make the data:

- machine readable, at least for the relevant sector (but ideally broader)
- readily linkable to other datasets
- understandable — that is, what the data is, how it was collected, and who has provided it (this will usually involve the inclusion of metadata)
- de-identified (to protect the identities of the data subjects in the case of data on individuals or organisations) — unless the data is already publicly accessible in identifiable form.

The current inadequacy of basic data processing in Australia was discussed in chapter 3 and the development of standards around dataset curation and other processing are discussed in chapter 10.

To transform or not — 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 transformation of some of their data holdings in order to meet user interest. Users may have unrealistic expectations of the current capability of agencies and of the funding available to undertake such transformation.

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 well-judged foot in both camps: 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. Volunteers otherwise assist to keep costs low.

In some circumstances, however, agencies may transform data: 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; as a result of historic legacy; or because some level of transformation is often essential prior to using data for internal analysis or for sharing with external parties. Common examples of requirements (usually with a public interest element) for agencies to undertake transformation include:

- 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.

For most agencies, data sharing and release (and the associated processing of the data) are not (yet) a 'core' activity (motor vehicle and land title registries are notable exceptions) and so these activities are accorded low priority and an over-riding wariness (chapter 3). Lack of 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, that moreover creates perverse incentives. And chapter 1 noted that Australia's poor international rankings for data openness is related as much to the poor useability of data (primarily poor update frequency and lack of data formatting) as it is to lack of availability.

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, now a common practice for GBEs, allows for a practical approach to transformation:

- To achieve social objectives, transformation 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 transformation should be funded by the government.
- Transformation 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 transformation is enough?

In determining how much transformation to undertake before sharing or release, the primary focus for all government agencies should be to establish first the public interest objective served by releasing the data in question. This will then guide whether to simply process the data to achieve a basic data product (fit for sharing or release) or undertake additional transformation to produce a transformed data product.

When both the public and private sectors have a role

There are good reasons for agencies to, in general, limit the amount of processing 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 transformation would be sought by potential users (businesses tend to be more 'market oriented' not least because they face more direct market signals)
- undertake such transformation
- distribute the outputs of that transformation (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, that government agencies should exercise caution as they add more value to raw data or provide more specialised services. Advising more specifically, the authors 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)

This view was broadly echoed in submissions. For example, the Medical Software Industry Association (MSIA) stated that it:

> … has seen many examples where [the] private sector could have performed tasks more effectively on behalf of Australian health consumers and tax payers. (sub. DR297, p. 6)

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)

Similarly, the Australian Spatial Information Business Association stated that government agencies should re-focus on the management of good quality and current spatial datasets 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 New South Wales Government stated that it takes the approach that data should be open by default, free where appropriate, and released as close as possible to its primary form and in a timely manner: 'Releasing data in a relatively unprocessed form can increase its reusability and speed of release' (sub. DR327, p. 6).

The Centre for Policy Development (sub. 11) provided an example of misplaced transformation by government agencies (box 9.1).

In markets where there are both public and private sector players, the policy challenge is to determine the efficient boundary between public sector transformation activities and private sector value-adding. 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. The Bureau also provides a number of other services that, in some other countries with larger markets, are provided by the private sector (box 9.2). 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.

Box 9.1 Misplaced investment in transformation of public sector data

The Centre for Policy Development gave an example of well-intentioned — but ultimately misplaced — investment in transformation 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.

Source: Centre for Policy Development (sub. 11, p. 17)

Box 9.2 Transformation 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 transformation is currently undertaken by the Bureau of Meteorology. This could reflect a range of factors, including differences in population spread and 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.

Source: Bureau of Meteorology (sub. 198); Mandel and Noyes (2015); National Weather Service (1991); Stiglitz, Orszag and Orszag (1991); Weiss (2010); Pettifer (2015)

The need for transparency about the quality of released data

Consistent with our recommended level of processing for basic data (above), the Association for Data-driven Marketing & Advertising (ADMA) stated that the release of minimally processed non-sensitive data would put Australia on an equal footing to other international jurisdictions and help drive innovation (sub. DR275, p. 6). It further noted however, that there is a need for transparency around the quality of data released and standards to ensure data is usable.

Timeliness of data release is paramount

A further consideration is the trade-off between timeliness and transformation. 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 government agencies should generally focus on timely release and ensuring basic usability of data, rather than undertaking any additional transformation. Timeliness will, very often, be of more value to users than having the agency address relatively minor issues with data quality. That is, users will generally prefer data 'as is' — albeit machine-readable and accompanied by metadata and a clear statement of the data's quality — and released promptly.

A question of principles

A set of policy principles is needed to guide agencies' decisions on transformation. These 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 transformation 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 transformation; and the IT risk is assessed and found to be small.

There will most likely be relatively few such circumstances, but a policy of open by default or similar sentiment (such as is canvassed in the Commission's recommended risk based approach to release — chapter 6) should allow for circumstances that fully meet these principles.

Datasets should be fit for the purpose of release — meeting the basic requirements for use in the field in which they are offered — but no greater than that unless they meet the criteria set out above. Agencies might usefully consult with potential data users regarding their views on transformation.

In chapter 6, we detailed the concept of Accredited Release Authorities (ARAs) as sectoral hubs of expertise in data curation, sharing and release. These entities could provide advice to agencies with limited skills on the minimum (and maximum) level of data processing required prior to release.

RECOMMENDATION 9.1

The emphasis for government agencies in handling data should be on making data available at a 'fit for release' standard in a timely manner. Beyond this, agencies should only transform data beyond the basic level if there is a clearly identified public interest purpose or legislative requirement for the agency to undertake additional transformation, or:

- the agency can perform the transformation 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.

9.3 Pricing of public sector data

Considerations for the pricing of public sector data differ somewhat from considerations of the private sector in sharing or releasing data. The public sector (ideally) places consideration on the public interest purposes for release, rather than the commercial opportunities.

The logic sequence for public sector data holders in considering the application of a price should commence with assessing in whose hands the data has greatest value and the relative significance of that value versus other calls on an agency's resources. Relatively low value suggests little relevance for pricing, and little transformation beyond the basic level described earlier.

Having established, though, that a dataset has high value in external hands is *not sufficient* to indicate that access must be priced. To illustrate, cancer patient data has high value to cancer research, but elaborate transformation that requires large cost recovery via a price to researchers is unlikely to be in the public interest.

Observation of willingness to pay (this comes in many varieties, noting that it is rational to prefer not to pay at all) and a user with the ability to capture that enhanced value are together a much clearer indicator that pricing (that is, above zero) should be considered.

From that point on, considerations in determining the price of public sector data to be shared or released are the:

- cost incurred by a data holder to get a dataset ready for release
- intent of policy-makers to encourage wider use of the data
- public interest purpose for release
- 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 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 pricing considerations, 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). They also 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 9.3), make it almost impossible to estimate in advance the benefits of increased sharing and release of public sector data (OECD 2015).

Box 9.3 Spillover benefits from data use

Spillover benefits can arise from a variety of uses of data, including by 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. Indeed, generation of community benefits is expected to be a core objective of publicly funded research.

The effect of the 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, transformed data products] would be assessed for cost recovery. (PC 2001, p. XLII)

Source: 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 9.4).

In Australia and elsewhere, given the uncertain — and potential for large — benefits, many open data initiatives encourage provision of data at the lowest possible cost or marginal cost. For instance, 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 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 9.4 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 the Australian Securities and Investments Commission'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).

Source: Australian Government, Department of Finance (2015); michaelwest.com (2016)

Cost recovery

Some governments have endorsed cost recovery, perhaps driven by fiscal concerns. However, upfront costs involved with standardising data and metadata to prepare for re-use can be significant. And, fiscal issues are not a sufficient consideration alone.

Under a cost recovery approach, data is usually 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. Costs apportionable to internal use of the data by the agency are usually excluded, and any charges related to transformation should not attempt to claw back the costs of providing a basic dataset.

In cases where the transformed data product is contestable — that is, where businesses are able to provide products that compete with an agency's transformed data product — competitive neutrality principles would mean that the agency should adopt commercial pricing (discussed below) (PC 2001, p. 25). But under the Commission's principles outlined earlier, such data is unlikely to be transformed by a government agency (Rec 9.1). Private providers would thus have the transformed data field to themselves and competitive neutrality need not arise.

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 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 chance of a 'pay-off' — commercial or otherwise — from the data use is uncertain or relatively low).

And it must also be acknowledged that cost recovery schemes themselves are not costless, with costs of administration (processing requests, payments and the like) and maintaining the effectiveness of the scheme (through licensing and enforcement activities, for example).

Alternatives to cost recovery

Pricing guidance, capable of applying to all data releases, would be desirable. 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 common for government agency products — prices are based on the costs incurred in making the product available to an additional user, while the agency funds all other sunk costs. In the case of data, the marginal cost of data reproduction and release (distribution) is usually so low as to effectively be zero — that is, marginal cost pricing typically equates to data being made available free of charge.

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 9.5).

This approach generates little or no revenue to offset the costs of collecting, producing and curating the data — which means government must fund these costs. An agency 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. In other words, there may be transitional issues to contemplate as well.

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), 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.

Box 9.5 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 small and medium-sized enterprises (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).

Source: Capgemini Consulting (2013); European Commission (2011); OECD (2015); Spatial Information Industry Action Agenda (2001); OpenGov (2016)

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 greatest benefits to users.

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 transformation that is undertaken, including in response to user demand (a 'freemium' model — basic data would be free, but transformed data products would 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
  - if the primary use is non-commercial, 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 the primary use is commercial, then a price discrimination approach would help to achieve any cost recovery objectives (compared with free data), but could also significantly dampen demand, and possibly innovation, amongst some groups of users.
- how practical it is to discriminate between users
  - if expected revenues are small or users are largely indistinguishable, the administrative costs of price discrimination may outweigh any additional revenue possible.

Deloitte (2013) noted that pricing public sector data at a level above marginal cost may help '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 suggested that if prices are paid, this could be called (post fact) a signal of consumers' willingness to pay. The Commission is not convinced this will help establish a price, other than through trial and error, which may affect other objectives.

Conversely, preparedness to pay the demanded price can also create a signal to maintain the dataset over time, and provide an incentive for agencies to be more responsive to the needs and wants of their customers.

Commercial pricing may set off a response from the broader market (prices would be fairly transparent and the annual revenue raised may be discernible in an agency's annual report). This may then:

- provide other data managers with some guidance as to what their data *might* be worth
- encourage private entrants into the market to supply such data.

The shift from public sector provision of crop science data to private sector provision has followed this path in other countries.

There may, in rare cases, be 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, 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 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.

FINDING 9.1

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 would 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

The approach taken to pricing should seek to maximise the opportunities and benefits offered by greater use of data, while retaining the incentive and ability for agencies to collect and release data effectively and 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.

A mix of private and public benefits is also likely to be common. Fish Ranger, a weather forecasting website targeted at fishers and recreational water users, suggested there would be greater use of Bureau of Meteorology data — with associated safety benefits — if the Bureau lowered the price of some of its data:

> The high data cost is a significant financial hurdle to overcome for new internet based businesses … This discourages innovation of new sites that would serve high quality weather data to different user groups around the country if the data was cheaper or free. … Through taxes the Australian public fund development of the best forecast data for Australia but do not realise the benefits of it due to the proliferation of apps and sites that do not use BoM forecast data due to its high cost. … Accurate weather forecasts are a safety concern and many recreational boaters are endangering themselves by trusting inferior weather forecast data when they make decisions on whether or not they should go out on the water. (sub. DR221, p, 1)

The Bureau is an interesting creature in the data world. Substantial transformation is inherently core to many of its activities (a forecast being much more than just basic data) and under the Commission's model (rec 9.1) should persist as a core function. How to price this *transformed* data is, as the range of commentary above might indicate, a tricky conundrum.

While this Report can and does provide clear advice on where *not to* incur cost by transformation, once reasonably incurred the guidance on price is less specific. In sum, it amounts to common sense: not to charge such that it would undermine the public interest or otherwise discourage data access to the detriment of society. Application of such an approach must necessarily occur case by case.

While there is broad consensus that making data freely available is likely to maximise use and hence deliver the highest level of social benefits (box 9.5), this will come at a cost — loss of recurrent revenue for agencies with a dependency on charging, and the potential cost of IT systems upgrades to prepare for a potential large increase in demand in the event that a previously priced dataset is made freely available (as noted by the Bureau of Meteorology, sub. DR322).

Social benefits from the research sector

In view of the exceptional potential for the research sector to generate spillovers, in the form of data-based discoveries and evaluations of inefficient or ineffective programs, the

Commission is of the view that the pricing of data for public interest research purposes should be the subject of a separate review.

Charging one public-funded entity for access to another public-funded entity's holdings is at once observably perverse, but also potentially helpful if investment in data quality is limited. It is a form of effective resource reallocation.

Where transformation is underwritten by other obligations — for example, the Bureau of Meteorology must transform data in order to create a forecast — the question is less clear cut, as noted above. Moreover, an investment by the Australian Government in NIDs (chapter 7) will necessarily result in transformation. Charging research users for access might prove to destroy much of the broad societal benefit; or it might sustain the NID into the long term. The case is much less clear than for commercial users.

Such a review should be undertaken independently of the Department of Finance. It should also specifically consider the position of NIDs, as well as the much larger range of other datasets of interest to researchers.

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.

RECOMMENDATION 9.2

The pricing of public sector datasets for public interest research 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 transformation (because it meets the principles outlined in section 9.2 above, including that there is a willingness and capacity to pay for this), there are generally strong grounds for passing these costs on to data users. However, the potential for spillover benefits should be taken into account. The example of the Bureau of Meteorology discussed above is important in this regard.

In general, where data clearly has high commercial potential (for example, exploration data compiled by Geoscience Australia), there could be equity grounds for supporting a cost recovery approach in line with the user-pays principle (noting that charges should potentially be very low if data has been provided to the agency at no cost). It could be argued in such circumstances that free access or very low (marginal cost) pricing would amount to a subsidy to data recipients, with the potential for efficiency losses. 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 in an Inquiry of this nature over the magnitude of benefits of making public sector data more available in any specific sector, 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. A distinction should be made between basic datasets (minimally processed to a 'fit for release' level) and those datasets that have undergone more substantial transformation by the data holder — but always taking into account the likely public interests and spillover benefits.

Table 9.1 summarises the Commission's assessment of the various pricing approaches for a typical agency selling minimally processed data.

Table 9.1 **Pricing approaches for a typical agency selling minimally processed data**

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	Possibly relevant	Possible
Commercial	Low	Low	Possibly relevant	Likely

[a] Ignores downstream impacts on central government taxation revenue generated by any increase in economic activity resulting from data being made available free of charge or at marginal cost) and takes into account that agency budgets are likely to be adjusted in response to changes in their revenue from sales of data. [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 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.

Where basic datasets are already available and charges are above 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 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 not currently available, the cost of making them available would comprise the cost of curating the dataset to a sufficient minimum standard for release (these costs may be substantial) and the distribution cost (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 transformed data products, the case for cost recovery or commercial pricing is stronger

If there are strong efficiency arguments for an agency undertaking transformation above the 'basic' standard, the case for adopting cost recovery or commercial pricing is strengthened, where users are willing and have capacity to pay, and this is consistent with the 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 access to the value-added data.

On balance, for transformed data products where a public sector dataholder is providing a 'contestable' product, cost recovery pricing is the approach that would most plausibly be in the broad community interest. In this regard, the New South Wales Government said that although it takes the approach that data should be open by default, when formal information access applications are made that require significant additional resources from

an agency, the *Government Information (Public Access) Act 2009* (NSW) allows agencies to impose charges to cover the costs of providing information (sub. DR327, p. 6).

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).

RECOMMENDATION 9.3

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

Where data has been transformed, the transformed dataset may be priced above the marginal cost of release. Data custodians should experiment with low prices initially to gauge the price sensitivity of demand, with a view to sustaining lower prices if demand proves to be reasonably price sensitive.

9.4 Funding support for public sector data release

Pricing will not achieve full recovery of costs for agencies because of normal budget practice (whereby agencies often do not retain proceeds of sales). Further, data release for the purpose of enabling substantial public benefits, is an activity that can be expected to grow — particularly if the Commission's recommendations for National Interest Datasets are adopted — 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 and the Bureau of Meteorology, sub. DR322). 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, organising and preparing data for release and continued investments to keep pace with technological change and the evolving expectations of users. The costs associated with managing sensitive personal data can be particularly significant.

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.

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, where agencies are rewarded for data releases that result in public interest 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 other 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.

This approach was broadly supported by Research Australia:

> The provision of data for research purposes by agencies should be part of their core business, and a reflection of the obligation they have to the individuals whose data has been collected and to the broader public to ensure that the community receives the greatest possible return on the investment it makes in the collection and storage of this data. (sub. DR282, p. 12)

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.

The New South Wales Government, while endorsing that the costs of releasing data should be seen as business-as-usual, noted that supplemental funding may be needed in some instances:

> In general the costs of releasing data should be addressed by agencies as part of business as usual and be factored into commissioning, regulatory and service planning activities. A by-design approach to identifying data for release will enable costs to be incorporated into operating budgets. This approach is being implemented by some NSW agencies but further guidance is needed to widen public sector capability in this area.
>
> Supplemental funding may be appropriate where significant additional effort is required to make data available or where release would impose a significant additional burden on an agency's IT infrastructure or support staff. Supplemental funding may also be required for state datasets identified as having national public interest.
>
> The funding request should be supported by a business case that justifies the additional data effort proposed. (sub. DR327, p. 7)

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 (chapter 7)). 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.

While general budget policy is beyond the scope of this report, 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. If, as argued in chapter 1, data is to be considered as an asset — a form of capital — then arguably its ongoing updating

should be funded as capital maintenance rather than as a recurrent expense (that does not result in the creation of assets). While such an approach would likely considerably improve the quality and useability of Australia's public sector data, given the magnitude and far-reaching consequences of such a change, it would best be considered in the broader context of any future review of public agency accounting procedures.

Lowering data management costs 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 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. An example of outsourcing is the National Cancer Screening Register. It will be managed by Telstra Health with the Australian Government retaining 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 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 deceptive — 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 free access or marginal cost pricing be adopted for the most common form of public sector data holding — minimally processed datasets — 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 undertake data transformation beyond actions required 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 agencies' data investments at no more than marginal cost.

For those agencies with 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 6, 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.

This 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.

And the development of National Interest Datasets and creation of ARAs proposed in other chapters will necessarily involve additional Commonwealth funding obligations into the medium term.

These aside, however, the larger data release activities triggered under an open government approach should not impose costs that larger agencies in particular should fail to absorb, or potentially utilise ARAs to help. The Commission believes that normal government budgetary processes should determine funding for such activity across agencies.

RECOMMENDATION 9.4

Funding should be provided to agencies for the curation and release of those datasets determined through the central data agencies' public request process (Recommendation 6.5) to be of high value with a strong public interest case for their release. This funding should be limited and supplemental in nature, payable only in the event that agencies make the datasets available through public release.

Funding would also be required for the Office of the National Data Custodian, for functions undertaken by Accredited Release Authorities and, in some cases, for the purchase and ongoing maintenance of National Interest Datasets. Additional responsibilities required of the Australian Competition and Consumer Commission in regard to the Comprehensive Right should also be resourced.

Aside from these purposes, no additional supplementary funding appears warranted for agencies' activities related to their data holdings as a consequence of this report.

10 Implementing the new data Framework

Key points

- Implementation processes for Australia's new data Framework must be as open and transparent as possible. All beneficiaries — the community, governments and businesses — should be involved in an exchange of views around data handling and use, in a manner that raises confidence in Australia's data ecosystem. Changes of this order cannot be suddenly launched into the public arena.
- State and Territory Governments will have a major role to play in the new data Framework and must continue to be engaged actively.
  - Some have already made substantial progress, by introducing data sharing legislation and creating agencies that specialise in data. The Commission's recommended reforms build on this progress to promote data sharing within and between jurisdictions.
- The Australian Government should set an ambitious timeline for reform implementation. The Commission envisages that some of its recommendations can — and should — be implemented immediately, including:
  - abolishing the requirements to destroy linked datasets that include Commonwealth data
  - accrediting state-based linkage units to link Commonwealth data
  - creating registers of data held by publicly funded entities.
- Before the end of 2017, the Australian Government should establish the Office of the National Data Custodian (NDC) via administrative means.
  - The NDC can add substance to the debate proposed above, and become a focal point for interaction with parties essential to the passage of the Data Sharing and Release Act.
  - It can also commence developing the processes for accrediting ARAs and trusted users, and enabling stakeholders to debate nominations for National Interest Datasets, to be implemented on passage of the legislation.
- At the same time, the Australian Government should lift the profile of the national conversation on data, using the proposals and evidence in this Report as a base. The emphasis should be on all stakeholders gaining greater certainty from the data reform process.
- The Australian Government should work towards passing the Data Sharing and Release Act before the end of 2018.
- A central agency with data responsibility should be tasked with monitoring the reform implementation process and remain in that role until the legislative steps are completed. It should report publicly on progress at the end of 2017 and again in 2018, to set the tone for maintaining this process as a transparent one. From 2019 onwards, the NDC (presuming passage of legislation) should take on this role.

Throughout this Report the Commission has presented a reform plan for handling Australia's data assets. This chapter focuses on key issues that governments must address as they implement the Commission's recommended reforms, including:

- working with the community, to maintain and enhance public acceptance of the reforms (section 10.1)
- working with each other, to learn from experience and existing data strategies, and to promote a collaborative approach to data (section 10.2)
- setting ambitious, but realistic, timelines to deliver benefits to the community in the short term and in the years to come (section 10.3)
- priorities beyond the implementation of initial reforms (section 10.4)
- and the fundamental importance of leadership, within agencies and at the ministerial level (section 10.5).

Beyond implementing changes to legislation and institutions, these reforms require governments to build — and enhance — community acceptance and trust about the way data is collected, stored, managed and shared.

Such trust is difficult to build and easy to lose.

Meaningful public consultation will be essential to secure citizens' trust in the new data Framework. For the Commission, meaningful consultation involves properly informing people about the reforms proposed, enabling questions to be raised, listening to concerns and issues, and explaining how these are being addressed and why. As it implements the new data Framework, the Australian Government should commence a process of national engagement and education about the proposed reforms, their benefits and costs, and provide mechanisms for public input into the implementation approach for the new policies.

10.1 Working with the community to maintain and enhance social licence

The Commission's reforms aim to address the inefficiencies in Australia's data policies, focussing on the outlook for exponential growth in data use, and ensuring the social licence granted by the community to government and other data users is maintained. Strong social licence can be difficult to build and maintain, but it is vital to successful implementation of these reforms (chapter 4).

The Commission's recommended reforms seek to enhance social licence by strengthening individuals' control over their data, and creating incentives to use data better, underpinned by effective risk management frameworks (figure 10.1).

Figure 10.1 **Building trust in Australia's new data ecosystem**

People and organisations are more willing to share information when they trust how it is being used and can see personal benefits stemming from access to their data that go beyond the immediate service they access (be it Facebook, car insurance, a visit to the GP or a new payments system for a small business). For example, the vast majority of Australians are willing to share their de-identified health data for research purposes, as they see a benefit in the discovery of new drugs and therapies (Research Australia 2016).

As more data is regularly made available, the benefits are likely to grow, which builds public trust in the system and supports continued use of data. For organisations, the benefits can take the form of better decision making or higher profits. For individuals, benefits can include access to a wider range of products and services, more informed decision making and choices, and improved convenience — anything from products that are more suitable to individual preferences to simply finding a parking spot faster.

In already highly competitive markets, the additional gains for consumers might be small. But in heavily concentrated markets, with disrupters at the gate, they could be very large.

Elements of the reforms that support social licence

Three key elements[18] of the Commission's reforms will support data custodians and users in securing and maintaining their social licence.

- *Ensuring that consumers have the right to practical control of their data, and are made aware of how it is shared and used* (recommendation 5.1). Greater control for consumers (*both individuals and small/medium businesses*) over their data should equate to greater trust in firms and government bodies that are collecting data. It is also a very practical way to alert the community to the constant data collection that occurs, often in novel and unexpected ways. The new Comprehensive Right would give consumers the ability and the incentive to find out what information is being held about them and to act on it.
- *Assuring data custodians and users that data can be shared, released or used without adverse consequences to them or the individuals whom the data is about.* Using a risk-based approach, as recommended, to determining how data is shared and used would promote active consideration by data custodians of the ways to protect identifiable data, while enabling its use in ways that would benefit the community (such as better social policy and health care). The recommended trusted user model is a key feature of the Framework, enabling data custodians and Accredited Release Authorities to share more data, confident that users have the expertise and security necessary to do so. Data would only be shared with trusted users in secure environments, and Accredited Release Authorities would have dedicated, resourced expertise on the use of sensitive data and the circumstances under which it could be accessed (recommendation 6.7). These changes cannot completely remove the risks to privacy. However, these risks would be less than those that typically arise through the use of a home computer, a wearable device, or a mobile phone. Sharing data with trusted users in secure environments should ensure that the net trade-off in data use versus risk would substantially favour community benefits from data use.
- *Promoting community trust that data custodians and their authorised users will safeguard their data.* Most Australians trust government agencies, financial institutions and healthcare providers to handle their personal information (OAIC 2013). But at the same time, individuals do not want to be surprised by who has their data and what is being done with it; they expect transparency about data handling practices (Bickers et al. 2015). The Commission's recommended reforms would address these community expectations, by creating a transparent data Framework, including disclosure when consumer data is traded to third parties and acknowledgment by governments of the value of data used in policy and program development (recommendations 5.1 and 8.3). Full implementation of the recommended trusted user model would also convey to individuals that the organisations handling their data — public and private — recognise their obligations to keep data safe and be transparent about the way they use it. The recommended appointment of an ethics adviser to the National Data Custodian should

18 This framework was adapted from Moore and Niemi (2016).

also give the community confidence that consideration will be given to when and how data should not be used — just because it can be used, doesn't always mean it should be.

It's not a 'set and forget' exercise

Implementing the Commission's Framework is only part of the effort required to maintain the social licence for data use. Obtaining this social licence cannot be seen as a one-off transaction, where completing a set of tasks will ensure success forevermore. A tick-the-box approach is unlikely to support ongoing expansion in data use; rather, data custodians and users need to build a relationship with the community that is based on a shared vision, strives to meet both immediate and longer term needs, and delivers real benefit to the public (Quigley and Baines 2014; Yates and Horvath 2013).

All parts of government have a role to play, and must commit to ongoing processes of community engagement, explaining how data use and innovation will benefit all individuals, and how the processes around data sharing and use will be maintained and strengthened over time to ensure data risks are always well managed.

Similarly, community acceptance of a particular data use by one agency does not automatically translate to acceptance of data use for other purposes by that or other agencies. Private sector data custodians are well aware of this; it has yet to become innate to many government agencies that they need to be active in maintaining community goodwill for data collection and use.

The social licence that the public and private sector require to continue to collect and use data can be easily lost (Telstra, sub. 88). Data breaches and negative examples of data use can quickly sway public opinion and erode social licence. The 2016 Census, which endured a number of controversies, will remain for some time in the public memory as an event that weakened community trust in data collection and handling practices. While a few will celebrate this, more should recognise that the Census — and other major data collection exercises — must continue to take on digital form, and will necessarily progress from paper-based systems to more efficient and reliable digital data handling. The Special Adviser to the Prime Minister on Cyber Security (MacGibbon 2016, p. 47)[19] concluded that

> … in light of the Census incident, the ABS [Australian Bureau of Statistics], and the APS [Australian Public Service] more broadly, should enhance its ability to identify and manage privacy risks in a way that maintains and builds public trust for Australia's digital future.

19 The MacGibbon Review made a number of recommendations on improving stakeholder engagement and communications with the community, as well as technical aspects of the Census. The Australian Government accepted all recommendations, and the ABS has commenced implementing them (Treasury 2017).

Increasing data use will continue to create challenges for custodians and users, in their communication with the community:

> Even the most experienced companies may make errors in their engagement with local communities and stakeholders. They may begin too late, fail to share complete information in a timely manner, breach communication protocols and local customs, inadvertently exclude important stakeholders, disregard important issues and concerns, or impose an unfair burden on the community. Successful companies will humbly acknowledge their mistakes; renew their commitment to timely and effective communication, meaningful dialogue, and ethical and responsible behaviour; and thereby begin to rebuild trust and social capital, the currency of social license. (Yates and Horvath 2013, pp. 21–22)

The public sector must think more along these lines.

In all of this, we must not lose sight of the fact that increasing data use will generate significant tangible benefits across the economy (appendix B). Success stories, where demonstrable benefits are realised, play a vital part of building and maintaining social licence. Alongside learning from past mistakes, the future conversation on data, facilitated by the National Data Custodian, must always highlight achievements through better practice and emphasise why we must continue to strive for better use of our valuable data.

The importance of transparency and meaningful consultation

Transparency and clear communication between all participants in the data system are necessary for maintaining social licence (Carter, Laurie and Dixon-Woods 2015; Moore and Niemi 2016).

In response to the Draft Report's emphasis on this objective, the Office of the Australian Information Commissioner (sub. DR236, p. 6, emphasis added) submitted that:

> A social licence for data use will be built on a number of elements. First, governments must be *transparent* about their intentions, so that individuals actually understand what the data reforms may mean for their personal information. Second, there must be *meaningful consultation* with individuals, to find out what uses of data the broader community believes are valuable, and reasonable. Third, governments must respond and *take public opinion into account* when making decisions. This may mean, ultimately, that there is community support for only some proposed uses of data - rather than all those that government and business may desire. However, in a democracy, having broad community support for reforms of this nature is essential.

This Inquiry is such a process of consultation, but it cannot be the final point. Further consultation should be undertaken around the implementation detail of key aspects of our reforms.

Once implemented, the Commission's Framework would enable greater ongoing transparency in Australia's data ecosystem.

- The Commission's recommendation to establish a public process for the designation of National Interest Datasets (recommendation 7.1), and create data registries that will

detail the data holdings of various publicly funded bodies (recommendation 6.4) will increase community awareness about the data that governments hold.

- For researchers and other users, new access processes to be introduced by the National Data Custodian (NDC) and the Accredited Release Authorities will replace the current complex and often opaque system of approvals for data access. Moreover, trusted users will benefit from a transparent system of approval, which should also greatly reduce waiting times to receive data.
- As the body responsible for the overall monitoring of the new data system, the NDC would ensure transparency is maintained in all aspects of data handling, including periodic review and reporting on progress, and ongoing consultation with the community, through its advisory group and other forums.

Governments are already taking steps towards improving transparency. In its recent Open Government National Action Plan, the Australian Government committed to a number of strategies to increase transparency, including providing more information to the community about how it is using and protecting the data it collects. Other measures include consultation to assess barriers to using data, surveys, social media and blog posts (DPMC 2016).

Further avenues for engagement should be considered.

- The Australian Government should convene advisory forums, which could include academics and other community representatives, to advise on data policy implementation. For example, the University of Melbourne (sub. DR305, p. 4) recommends establishing 'a National Data Advisory and Consultative Forum comprised of non-government data experts to advise the NDC, relevant departments and Accredited Release Authorities on public interest data needs, technological aspects, data governance and data release priorities'. The New Zealand experience, setting up the NZ Data Futures Partnership, which coordinates an ongoing public discussion about data use and social licence (Statistics NZ, sub. 62), could serve as a useful blueprint for Australia.
- Using the NDC, Governments should work with data custodians and Accredited Release Authorities (ARAs) to ensure community expectations are addressed. Once ARAs are established, community advocates might complement existing consultation processes. This model is operating successfully in some areas — such advocates work with health researchers, to ensure community and consumer involvement in their projects, through the Consumer & Community Health Research Network at The University of Western Australia (Involving People in Research 2017).
- Online engagement between governments and the community should be expanded, promoting the notion that 'the government should go to the people rather than making people come to it' (McNamara 2010, p. 238, quoted in Holmes 2011). This will require governments to consider more avenues for online interaction, such as forums hosted by third party websites. To be successful, online engagement must be part of an overall culture of transparency (Holmes 2011).

Maintaining a social licence via meaningful community engagement is equally important for private sector entities seeking to expand their use of data (Actuaries Institute, sub. 206; Telstra, sub. DR312). The Australian Communications Consumer Action Network (sub. DR306, p. 3) proposed that organisations be required to 'work with consumers to help them understand what their personal information is being used for and what the benefits are to them'.

Some see a role for the private sector in educating consumers on how their data is used. For example, the Australian Payments Clearing Association (sub. 44, p 11) suggested conducting a 'public awareness campaign to help further [consumers'] understanding of how data collection and sharing occurs today and the associated benefits and risks'. An important role in educating the public will also fall to the institutions recommended by the Commission to ensure that individuals can exercise their Comprehensive Right over their data (section 10.4).

RECOMMENDATION 10.1

The Australian Government should engage actively with the community on matters related to data availability and use.

At a minimum, the National Data Custodian should regularly convene forums for consultation, to ensure community concerns about increased use of data are addressed.

10.2 A collaborative effort: working with States and Territories

State, Territory and Local Governments are key players in Australis's data ecosystem. Indeed, some jurisdictions have made substantial progress in improving data sharing within government and increasing access to data for researchers. Examples include the New South Wales Data Analytics Centre established to facilitate data sharing within government, and the Western Australian data linkage unit, which has been operating within the Department of Health for over two decades (Data Linkage Branch, sub. 13; NSW Government, sub. 80).[20]

The missing piece of the puzzle, however, is a national framework for data, to facilitate coordination between the Australian Government and State and Territory Governments:

> The lack of consistent, comprehensive information release frameworks is a major obstacle. At a national level, one of the most important interfaces affecting the flow and value of public data occurs between jurisdictions. The absence of a consistent and nationally harmonised approach

20 Chapter 1 and appendix C list further policies and initiatives implemented in the States and Territories.

limits the capacity of governments to coordinate services, improve planning, and more generally streamline the proactive release of datasets. This can manifest in a number of ways:

- The additional effort needed by agencies or sectors to create one-off arrangements for the flow of information (e.g. 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 (e.g. 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. (NSW Government, sub. 80, p. 10)

The implementation of the Commission's recommended reforms will address these issues by providing a national framework for data sharing and release. It will create opportunities for collaboration with State and Territory Governments and build on the progress already made in the different jurisdictions in improving access and use of data.

At the same time, the support of State and Territory Governments will be instrumental to the success of the Commission's recommended reforms (Sax Institute, sub. DR256; SA NT DataLink Steering Committee, sub. DR283).

Learning from the experience of States and Territories

In developing the Data Sharing and Release Act (recommendation 8.1), the Australian Government should seek to learn from the experiences of State Governments that have already introduced similar legislation. For example, the Office of the NSW Privacy Commissioner (sub. DR268, p. 7) submitted that the NSW Data Sharing Act 'balances efficiency in data sharing with privacy protections' and could be used as a model to develop legislation at the Commonwealth level, as well as in other jurisdictions.

Within the framework created by the Data Sharing and Release Act, the NDC will be responsible for engaging with data holders and users across all Australian governments, to implement the national data policy. State and Territory Governments must have a substantial role in this interaction, as custodians of some of Australia's most important datasets.

Existing institutions and forums could be used to engage data holders and users across all Australian governments, including:

- the Council of Australian Governments (COAG), which has negotiated several data related matters such as the national minimum datasets and the National Health Information Agreement (COAG 2013);

- advisory forums, such as the Australian Statistics Advisory Council (subs. 25, DR237) which includes representatives from all jurisdictions and has been involved in developing some past data policies.

However, to date, these long-established groups have largely not delivered improvements in sharing and use of data across jurisdictions. This would suggest that the quest to improve inter-jurisdictional data sharing should not be left to these institutions and forums alone.

Some other groups — largely focused on particular types of data — have had some successes in achieving cross-jurisdictional improvements, including:

- cross-jurisdictional bodies that coordinate data collections and standards, such as ANZLIC — the Spatial Information Council (sub. 164), that provide a foundation for jurisdictions to co-ordinate some data collections;
- data governance committees, such as those convened by the Australian Institute of Health and Welfare (AIHW 2017), that bring together data custodians from all jurisdictions and facilitate ongoing development of data standards;
- State and Territory organisations involved in cross-jurisdictional data initiatives, such as the Population Health Research Network, which comprises all the state-based data linkage units (subs. 110, DR240).

As a starting point, the Commission suggests that the NDC should convene a working group that includes representatives from every State and Territory central agency (those tasked with data policy development for their jurisdiction). This working group should focus initially on several key areas in which data sharing between States and Territories and the Commonwealth could generate clear and early benefits.

Improving data availability and access across borders

State and Territory Governments must be actively involved in the process of designating datasets as National Interest Datasets (NIDs). Some probable NIDs will rely heavily on data collected by States and Territories, such as hospital and education data. State and Territory agencies, as accredited trusted users (recommendation 7.3), will be able to access data for analysis and policy development. Currently, the processes to access such data from other jurisdictions are convoluted and time consuming, and often fail (chapter 3).

State and Territory agencies may also be accredited as Release Authorities (ARAs) (recommendation 6.7). There are already numerous examples of State and Territory agencies that manage and release data (for example, the NSW Bureau of Crime Statistics and Research, sub. 23), as well as collaborative data sharing arrangements between jurisdictions (such as those managed by the Australian Criminal Intelligence Commission, sub. 210). The scope of these arrangements tends to be limited, whereas ARAs must be nationwide in their focus and limited in number (chapter 6).

Implementing the Commission's recommendation to accredit State and Territory data linkage units to link Australian Government data may prove to be a relatively easy reform to further improve data availability and use across borders. The Australian Institute of Health and Welfare (AIHW, sub. DR299) indicated that processes to accredit state-based data linkage units are underway.

These reforms, coupled with early implementation of recommendations for the creation of data registries and release of non-sensitive data, will expand the evidence base available for State and Territory agencies to develop better policy, and promote cooperation between jurisdictions.

10.3 Implementation timeline

The Commission has made a number of recommendations that can be implemented in the short term to improve data availability and use — commencing consultations on improving data access in line with community expectations and releasing non-sensitive 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, detailed planning and strong leadership from the highest levels of government (section 10.5).

The Commission recognises that progress has already been made in improving access and use of data through reform initiatives and demonstration projects currently operating in many jurisdictions (chapter 1). On a broader level, the Australian Government has signalled its intent to overhaul access to and use of data in the Open Government National Action Plan, which was submitted to the Open Government Partnership in late 2016. In its action plan, the Government committed to responding to the Commission's Report by September 2017 (DPMC 2016).

A staged approach — with an immediate start

Although building the new data system will require a number of stages, implementation of the Commission's reforms can commence immediately — meaningful change can be achieved without legislation. Key immediate actions include:

- Streamlining data access and eliminating inefficiencies
  - abolishing the requirements to destroy linked datasets and linkage keys at the completion of projects that include Commonwealth data (recommendation 6.17).
  - accrediting state-based linkage units to link Commonwealth data (recommendation 6.2).
  - commencing discussions with the National Health and Medical Research Council and research insitutions on the recommended reforms to the operation of human

research ethics committees, and drafting an intergovernmental agreement on mutual recognition for ethics approvals and streamlining governance approvals (recommendation 6.15).

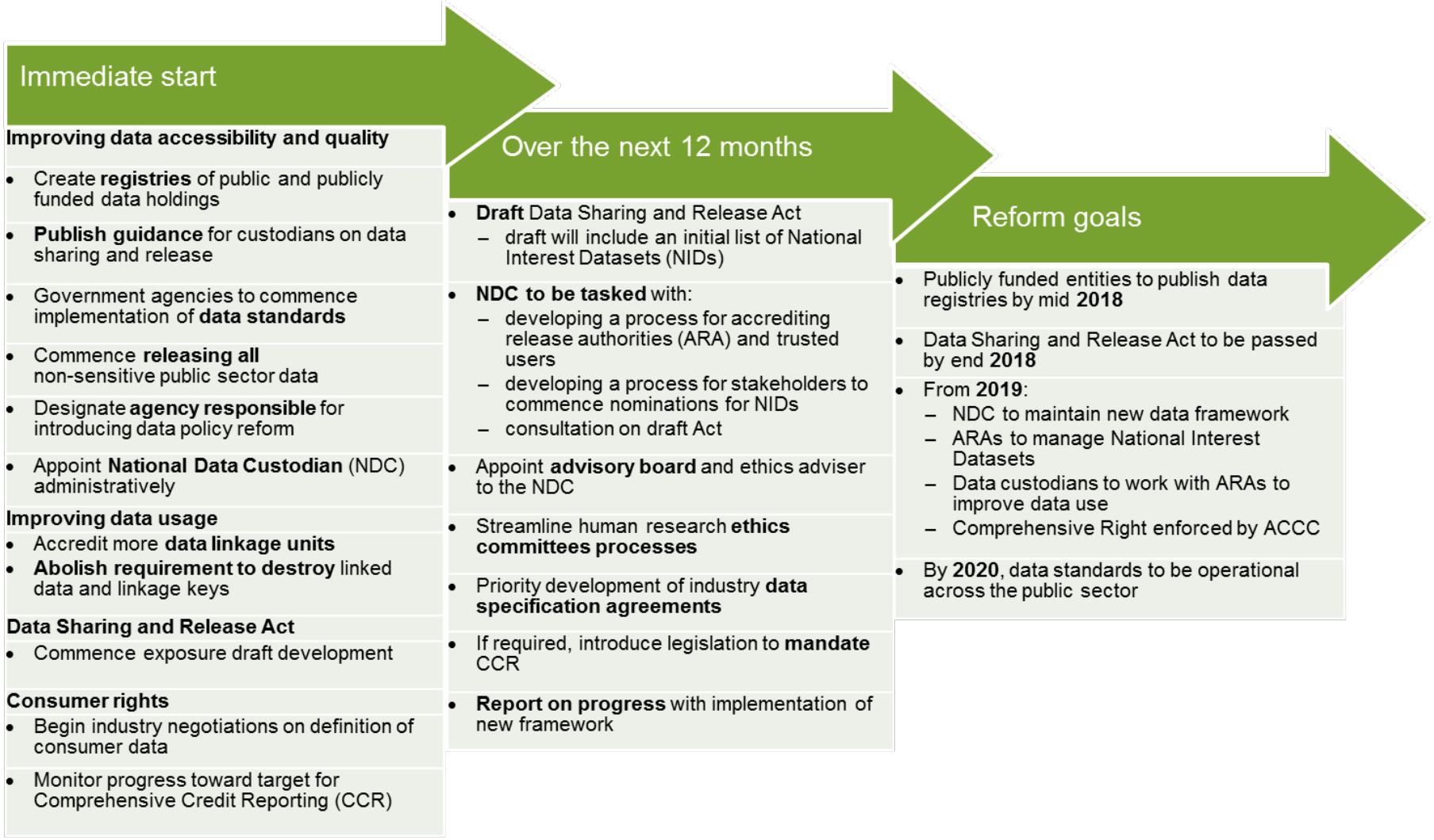
- Improving data availability
  - creating and publishing registers of data (including metadata and linked datasets) held by publicly funded entities. This includes all Australian Government agencies and other bodies that fund research, such as the Australian Research Council and the National Health and Medical Research Council.
  - changes to government contracting will ensure access to publicly-funded data collected by the private sector (recommendation 6.3) can be made without the need to update legislation.
- Setting up the new Data Framework
  - establishing the office of the NDC, and tasking it with developing guidance on de-identification and processes for the accreditation of release authorises and trusted users
  - conducting extensive consultation on the Comprehensive Right, and drafting the Data Sharing and Release Act (chapter 8 discusses the legal aspects of implementing the new data framework).

Further, all Cabinet submissions should increasingly take account of the data aspects relevant to policy issues under consideration. In addition to improving policy decisions, such a move would signal internally the Australian Government's commitment to data use and assist in putting data sharing and release at the centre of government deliberations, at least at the Commonwealth level.

The Commission envisages that — with appropriate resourcing and attention — the full implementation of the Framework can be completed by the end of 2018. From 2019, the NDC and ARAs will manage the ongoing access and release for public sector datasets (figure 10.2); and the NDC will publish an annual report on progress against the Commission's Framework.

The Commission's Framework structurally establishes a central leadership and governance model for all data assets held by the Australian Government. From 2019, the NDC and ARAs must provide such leadership, and a central agency may let go.

Figure 10.2 **Implementation timeline for the Commission's key recommendations**



But creating a Department of Data would be a fundamental error — focusing on the input rather than outputs and benefits that data use can deliver. Better an independent agency with clear accountability to enable collaboration and promulgate best practice, than a Department that by definition is necessarily a subsidiary to another larger Department. The NDC should facilitate more data access and use over time, even if in contest with the objections of service delivery Ministers or central agencies. Only in this manner will the true value of data be exposed.

Accountability and evaluation

The Government's implementation plan must put in place clear ownership and accountability structures. Transparency against milestones in implementing the Framework, reported annually to the public — first by the central agency, later by the NDC — is a vital part of successful reform. Specific milestones will depend on the Government's review of this Report and response to it. However, the key deliverables described in that response should include specific dates by which milestones should be achieved, and the sequencing of events (a critical path).

As part of its annual report, the central agency leading the Framework implementation should provide details of the progress of data custodians in all Commonwealth Departments and key organisations holding publicly funded data, in creating and publishing data registries, up until the NDC is legislatively able to take on this role.

Within different agencies and organisations, clear accountability structures should be put in place to support the implementation of the new Data Framework. Internal leadership (backed by ministerial support — see section 10.5) has a vital role to play in this context. In its submission to this Inquiry, the CSIRO highlighted a number of leadership factors that must be considered in order for data sharing to occur:

- The importance of key individuals, such as trusted leaders, who have built community trust based on strong track records within domains. Strong senior champions are needed to drive the initiative within participating organisations; …
- The critical role of whole-of-government leadership and priority setting to provide top down drivers and coherence for collective activity; and risk of lack of 'buy-in' if some stakeholders are not engaged early. (sub. 161, pp. 36–7)

As with all new policies introduced by government, evaluation will be vital to ensuring the policies are effective. Evaluation should assess:

- progress in the release of all non-sensitive public data
- accreditation of ARAs and state-based data linkage units
- data custodians' timeframes for processing applications for access
- the implementation of human research ethics committees reform, and progress towards mutual recognition

- development of industry data-specification agreements, as part of the implementation of the Comprehensive Right.

Evaluation roles could initially be carried out by the central agency with data responsibility (currently, at the Australian Government level, this is the Department of the Prime Minister and Cabinet). Once the NDC is established, it should be tasked with monitoring progress, supported by the ARAs. Further, the NDC should be tasked with evaluating the impact and efficiency of the data framework overall, within three years of its implementation.

In addition, the Australian Government's progress against the National Action Plan it submitted to the Open Government Partnership will be independently assessed in future years (DPMC 2016). The commitments made in the action plan are closely aligned with improving access and use of data, and the monitoring by the Open Government Partnership can be valuable in ensuring progress is achieved.

RECOMMENDATION 10.2

The Australian Government should set an ambitious — but realistic — timeline for implementation of the Commission's recommended reforms.

A set of actions in this Report can be completed in 2017, to ensure they deliver benefits to the community in the short term.

Passage of the Data Sharing and Release Act and supporting Part 2 amendments for an initial suite of National Interest Datasets should be in place by the end of 2018.

A central agency with data responsibility should actively support the progress made against the implementation plan until the Office of the National Data Custodian is legislatively established.

Once established, the National Data Custodian should assume responsibility for monitoring and evaluating the effects of the new data Framework, reporting annually on progress and with a formal evaluation after three years' experience of the Framework's reforms.

10.4 Ongoing priorities for governments

Improving availability and use of data is a long-term commitment from governments, one that can fundamentally change the way governments operate. Beyond the implementation process detailed by the Commission, a number of areas will require ongoing investment of funding and effort — improving data quality, data skills and public education and consultation.

Improving data quality

If Australia is to improve the usability of its public sector data, the application of consistent data and metadata standards will be a vital component of this reform. Many Inquiry participants recognised the importance of standardisation in order to make data usable — and useful (see, for example, Department of Prime Minister and Cabinet, sub. 20; Department of Industry, Innovation and Science, sub. 69).

Over the past decade, numerous policies have been developed with the aim to introduce standards across government agencies, including the current Digital Continuity policy implemented by the National Archives (sub. 114). In its Public Data Policy Statement, the Australian Government declared that it will make high-value datasets available using high quality standards (Turnbull 2015). However, progress towards this goal has been limited so far (chapter 3). The Commission considers that the implementation of standards should be made a matter of priority, to support increased availability and use of data.

RECOMMENDATION 10.3

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
- deal effectively with 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 would 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 to improve the quality of their data holdings.

The Commission's recommended reforms will support improvements in data standardisation in a number of ways.

First, for datasets designated as National Interest Datasets, the NDC will allocate funding specifically for the purposes of data management and curation (recommendation 9.4).

Second, ARAs will play an important role in the implementation of data standards, as well as improving the quality of data across the board.

- They will provide technical expertise and advice, reflecting best practice within their specific sector. This will be particularly beneficial to those smaller government agencies (or, depending on the sector, not-for-profit organisations) that have limited capacity to invest in data management.
- They will ensure that any improvements made to datasets in the process of preparing them for sharing or release are not lost once projects are completed. For example, Cancer Council Australia (sub. DR254) suggested that ARAs could store linked data to ensure it can be reused by researchers, without duplication of effort. Further, ARAs may work with data custodians to implement ongoing improvements to datasets (subject to the considerations on further transformation, discussed in chapter 9).

The Australian Government should fund the development of centres of capability on key data management issues within particular ARAs (not all ARAs will require in-house access to specialised data techniques, as long as all have access to each other's capability — Chapter 6 discusses cooperation between ARAs, including contracting out of services).

Beyond recognising the importance of standards, the Commission has taken a technology-neutral approach in this Inquiry and our recommendations do not focus on specific technologies. Such decisions are best left to those who work with data collections, and will differ across time and industries. Working within specific sectors will give ARAs the specialised knowledge to identify the most appropriate technologies and standards to manage and share data particular to their area of expertise.

For example, many stakeholders have raised the issue of APIs (application programming interfaces — see appendix C for a detailed discussion) in relation to the new Comprehensive Right for consumers and the way this technology can improve access to data in a wide range of areas. Technology such as APIs undoubtedly offer benefits in some areas, and could be considered as a good solution to overcome difficulties in data sharing. However, there are other technological solutions for data access, and many more will be developed in the future (Australian Information Industry Association, sub. DR244; AGL Energy, sub. DR251). Forcing the use of APIs among, say, medical practitioners for exchange of consumer data in primary care simply because it is the preferred solution of the finance sector is deeply undesirable.

It will be up to each sector to determine the best technological approach to make more use of its data, enhance data sharing and enable better access for consumers to data collected about them. This will be a crucial element of the data-specification process mandated under the Comprehensive Right. Given the multitude of standards developed for data management and exchange across many sectors, technology should not be used as a barrier to competitive outcomes in the consumer interest. The Australian Competition and

Consumer Commission (ACCC) would have the power to sanction entities that attempt to use technology as an excuse to shirking their obligations to consumers.

RECOMMENDATION 10.4

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

In the event that cooperative 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.

Building skills across the economy

While many discussions on 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 CSIRO's 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. Depending on their existing skill sets and the data they manage, different agencies will be required to take different approaches — some will benefit from short-term consulting type services provided by organisations such as Data61, to find ways to improve their processes and systems and make the most of the knowledge and experience of those people who already work with data; others will require more extensive support.

Nonetheless, data skills will become increasingly important for the entire public sector. The Australian Public Service Commission (2016) has recently begun the implementation of the APS Data Skills and Capability Framework, which aims to improve data skills across the entire public sector. State and Territory Governments have also introduced similar policies.

In response to rising demand for data skills in the workforce, universities and other higher education providers are offering a range of data related courses. According to Universities Australia (sub. 90, p. 6), while challenges do exist, overall the 'introduction of the demand-driven system for university places has helped to meet the growing employment need for highly-skilled graduates. … Recent initiatives to improve the take up of STEM [science, technology, engineering and maths] in schools and opportunities for women to stay in STEM careers will also help meet the workforce needs of the digital economy'.

The Commission is of the view that, as with many other required elements of its recommended framework, the ongoing development of data capabilities across the workforce will be greatly assisted by a reformed system that values data and the insights gained from it, and signals a new, more dynamic stance on data sharing and release. For example, ARAs will have the necessary skills to act as hubs of knowledge on data management practices in their specific sectors. This will allow them to assist and advise other agencies on data curation and other technical aspects of data management (chapter 6). As Australian, State and Territory Governments work to improve the quality of their data, it is likely that expertise will grow across the public sector as a whole.

Skills development should not be restricted to the people working directly with data. In working towards a data-driven economy, governments must be cognisant of the fact that data skills are 'unevenly distributed' in the community (Western Sydney University, sub. 119, p. 1). Therefore, they should put in place initiatives to improve data literacy across the board (Westpac, sub. DR324).

Public education and consultation

Data is a policy topic where misconceptions are common. In the private sector, for example, a global survey of managers in medium and large companies found that although the vast majority use data analytics, only one in ten believes the data is accurate (KPMG 2016). Within the community, studies overseas have found that there is significant awareness of data collection and use by government, but a much more limited understanding of how this data is used in practice (Sciencewise 2014). 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), when this is still far from the reality.

Inquiry participants have raised the need for public education, on both personal data management and broader data issues. For example, the Australian Computer Society (sub. 134, p. 6) recommended:

> implementing a national education and awareness raising campaign which helps form a new social contract with citizens around open data and its implications and benefits. An important broader agenda here is to enable people to understand the value of their own data, and hence empowering consumers on issues of access to and control over data concerning them.

Raising public awareness of data issues can be a beneficial step in promoting meaningful consultation and building social licence:

> Social licence activities should commence as soon as possible … It is important that communication to the public be accessible, digestible, timely and focused on the types of public value that can be generated from public data in a way that resonates with the Australian public. Communication should be a two way conversation and start with a significant lead time prior to changes in data accessibility and use.
>
> As part of this process, social licence activities should include benchmarking of public attitudes for a range of public data management activities. Providing tailored information and engagement activities for different segments of the public based on attitudes to data management will support a trusted, transparent and balanced approach to sharing public data. (Department of Social Services, sub. DR255, p. 2)

Research has repeatedly shown that individuals are willing to share their data when they believe it will be held securely and used for the common good (see, for example, Moore and Niemi 2016). Therefore, any public awareness activities that governments undertake should describe the immense potential of data to improve community wellbeing, while also emphasising data security and privacy aspects.

As the Commission's reforms are implemented, raising community awareness about the new Comprehensive Right and the ways to exercise it will be particularly important. The Office of the Australian Information Commissioner (OAIC, sub. 200) already has a role in educating individuals, businesses and government agencies on privacy matters, while the ACCC (sub. DR331) has similar roles in regard to consumer rights.

In implementing the Commission's data Framework, the ACCC should play a central role in community education (recommendation 5.4), although all data custodians and users, both public and private, will have the responsibility to make consumers aware of their rights. The OAIC might also offer support, clarifying how users can exercise their Comprehensive Right, depending on whether their interest is to protect their privacy or to actively reuse their data. The community may take a more active interest in understanding their data rights, offering benefits to all regulators.

New technologies can be considered to make sure that the messages to the community are delivered effectively (ASIC, sub. DR325). Social media, for example, can be beneficial in building trust and communicating directly with consumers in a timely and cost-effective manner. Government agencies can participate in online communities, or facilitate the creation of new communities for specific target groups. Past experience suggests that such online communities can be very successful in fostering trust, but they require careful planning and delivery to safeguard participants' privacy (Bista, Nepal and Paris 2013; Colineau, Cecile and Dennett 2011).

10.5 Finally, a word on leadership

The success of New Zealand in harnessing its public data assets to create the Integrated Data Infrastructure is often presented as an example of the benefits Australia could reap from data reform (see, for example, Grattan Institute, sub. 12). In this context, there are important insights to be gained from considering key lessons from the New Zealand experience. Apart from the need to continue to invest in technology and to evolve in response to demand, Statistics NZ (sub. 62, p. 4) highlighted the importance of *leadership and political support* in its success:

> The IDI [Integrated Data Infrastructure] story so far has been one of significant success and has provided Statistics NZ with some important learnings which could be helpful pointers when considering integrated data use and access elsewhere:
>
> - The backing by Ministers provided us with valuable support in expanding the IDI — the fact that the IDI enables the recently adopted investment approach has made it an integral part of the institutional setting for policy and evaluation in New Zealand.

In their submissions to this Inquiry, the Queensland and Tasmanian Governments also acknowledged the vital importance of high level leadership in increasing data use.

> Much of this [reform] should be … underpinned by the establishment of a ministerial sponsor to ensure that sharing activities are part of the wider government agenda and are viewed favourably amongst competing priorities. (Queensland Government, sub. 207, p. 10)
>
> High-level leadership is considered essential to overcome a cultural reluctance within the State Service sector to share information. Barriers may arise because of risk aversion, perceptions that legislation prevents sharing, lack of confidence in data quality, or lack of trust. It takes clear, high-level direction and authorisation in order to get data-sharing initiatives up and running. (Tasmanian Government, sub. 205, p. 14)

Improving the use of our data can offer much benefit to the community. However, this benefit will be greatly diminished if data reform becomes entangled in a drawn-out political process and cross jurisdictional inefficiencies. Ministerial leadership at the federal and state levels will be vital to gather the support needed, within political circles and the community, to cut through cross-jurisdictional red tape, overcome the inevitable challenges of introducing reform and ensure implementation of these reforms delivers the full scale of possible benefit.

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.

The terms of reference for the Inquiry — reproduced in the preliminary pages of this Report — was received from the Treasurer 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 336 public submissions (table A.1) during the Inquiry — 211 prior to the Draft Report and 125 in response to the Draft Report. Submissions included:

- 94 from industry associations or representative bodies
- 69 from governments or government agencies
- 59 from private sector businesses
- 58 from academics or research groups
- 38 from individuals
- 18 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 130 businesses, business groups, academics, government agencies and individuals in Australia and overseas (table A.2). Three roundtable discussions have been held throughout the Inquiry (table A.3) — with academics; with Australia Government agencies; and with members of the Business Council of Australia. Public hearings were held in Melbourne on 21 November 2016 and in Sydney on 28 November 2016 (table A.4).

The following public documents have been prepared by the Commission in this Inquiry:

- Issues paper — released 18 April 2016
- Draft Report — released 3 November 2016
- Final Report — delivered to Government on 31 March 2016 (to be publicly released within 25 parliamentary sitting days).

The Commission thanks all those who contributed to the Inquiry.

Table A.1 **Public submissions received**

Participant	*Submission no.*
A Future Beyond the Wall – ARC Linkage Project	82
Aboriginal Health Council of WA	181
Actuaries Institute	206
Adrian Bennett	156
AGL Energy	140, DR251
AgriDigital	DR249
AGW Lawyers & Consultants	136
Alzheimer's Australia	75
ANZ	64, DR231
ANZLIC – the Spatial Information Council	164
AOASG – CCAU	DR272
ARC Centre of Excellence for Children and Families over the Life Course (the Life Course Centre)	87, DR271
Archerfish Consulting	30
Association for data-driven marketing & advertising	178, DR275
ASX	177
Atlas of Living Australia	72
Attorney–General's Department	209, DR334
AURIN – The University of Melbourne	116, DR260
AUSTRAC	185
Australasian Open Access Strategy Group	84
Australian Automobile Association	157, DR262
Australian Bankers' Association	93, DR307
Australian Bureau of Statistics	94, DR285
Australian Business Roundtable for Disaster Resilience and Safer Communities	146
Australian Centre for Financial Studies	103
Australian Centre for Intellectual Property in Agriculture	DR215
Australian Child Rights Taskforce	DR291
Australian Chronic Disease Prevention Alliance	86
Australian Communications Consumer Action Network	54, DR306
Australian Competition and Consumer Commission	DR331
Australian Computer Society	134. DR329
Australian Criminal Intelligence Commission	210
Australian Data Archive	139, DR288
Australian Dental Association	8, DR230
Australian Energy Council	127, DR281
Australian Energy Market Commission	158
Australian Financial Markets Assoc.	57

(Continued next page)

Table A.1 (continued)

Participant	*Submission no.*
Australian Food and Grocery Council	DR284
Australian Government Environmental Information Advisory Group	32
Australian Government Linked Data Working Group	46, DR278
Australian Hotels Association	159
Australian Human Rights Commission	DR220
Australian Indigenous Governance Institute	60
Australian Information Industry Association	DR244
Australian Injury Prevention Network	121
Australian Institute of Health and Welfare	162, DR299
Australian Institute of Health Innovation - Macquarie University	DR229
Australian Institute of Marine Science	171
Australian Institute of Superannuation Trustees	96
Australian Institute of Tropical Health and Medicine	52
Australian Library and information Association	100
Australian Longitudinal Study on Women's Health	109
Australian National Data Service	113, DR245
Australian National University	DR226
Australian Payments Clearing Association	44, DR277
Australian Payments Council	34
Australia Post	174, DR332
Australian Privacy Foundation	142, DR313
Australian Private Hospitals Association	183, DR335
Australian Projections Pty Ltd	122
Australian Property Institute	169
Australian Public Service Commission	48
Australian Research Council	22, DR294
Australian Restructuring Insolvency and Turnaround Association	112
Australian Retail Credit Association	DR247
Australian Securities and Investment Commission	195, DR325
Australian Statistics Advisory Council	25, DR237
Australian Taxation Office	204, DR314
Australian Taxpayers' Alliance	DR279
Australian Unity	95
Azcende	DR274
Beau Johnson	DR212
Brotherhood of St Laurence	186
Bruce Sweeting	144
Bryan Kavanagh	28
Bureau of Meteorology	198, DR322
Cancer Australia	104
Cancer Council Australia	141, DR254
Capital Markets Cooperative Research Centre	76

(Continued next page)

Table A.1 (continued)

Participant	*Submission no.*
Center for Data Innovation	138
Centre for Big Data Research in Health, UNSW	21, DR241
Centre for International Finance and Regulation	9
Centre for Policy Development	11, DR222
Chacko Thomas	125
Chartered Accountants Australia and New Zealand	29, DR336
CHOICE	167, DR328
Chris Doulton	145
Clean Energy Regulator	118
Cleary, Michael	17
Commissioner for Privacy and Data Protection	DR320
Commonwealth Bank of Australia	175
Commonwealth Grants Commission	58
Communications Alliance and the Australian Mobile Telecommunications Association	DR250
Community Insight Australia	83
Complementary Medicines Australia	DR223
Consumer Action Law Centre	81, DR308
Cooperative Research Centre for Spatial Information	43
CoreLogic Asia Pacific	102, DR248
Council of Australian University Librarians	97
CREBP – Bond University	85
CSIRO	161, DR323
Curtin University	41
Customer Owned Banking Association	132, DR273
Data Governance Australia	179, DR333
Data Linkage Branch (Department of Health WA)	13
Data Republic Pty Ltd	176
Datanomics	129, DR300
David Dean	DR214
David McKeague	DR290
Department of Agriculture and Water Resources	37, DR224
Department of Computing and Information Systems: The University of Melbourne	DR303
Department of Employment	18
Department of Foreign Affairs and Trade	202
Department of Health	99, DR330
Department of Immigration and Border Protection	168
Department of Industry, Innovation and Science	69, DR235
Department of Infrastructure and Regional Development	201
Department of Prime Minister and Cabinet	20, DR286
Department of Social Services	10, DR255
Digital Industry Group Inc	DR326
Dr Donna-Louise McGrath	DR276

(Continued next page)

Table A.1 (continued)

Participant	*Submission no.*
Dr Hazel Moir	24
Dun & Bradstreet	135, DR319
Ecosystem Science Council of Australia	38
Energy Consumers Australia	DR316
Envestnet Yodlee	DR280
Equality Rights Alliance	124
Fabric Corporation	DR310
Facebook	172
Federal Chamber of Automotive Industries	45, DR321
Federation of Ethnic Communities' Councils of Australia	16
Federation of Victorian Traditional Owner Corporation	14
Federation University Australia	131
Financial Institutions & Management Advisory	73, DR233
Financial Rights Legal Centre	107, DR289
FinTech Australia	182, DR315
Fish Ranger	DR221
Frederick Douglas Michna	DR228
Geoscience Australia	211, DR238
Google Australia	DR292
Grattan Institute	12
Group of Eight	DR304
Health Geography Study Group – Institute of Australian Geographers	92
Health Informatics Society of Australia	199, DR301
Health Research Institute, University of Canberra	115
Health Services Research Association of Australia & New Zealand	39
Heart Foundation	71
IAG	128
IEEE SSIT Australia	19
Impact Investing Australia	111
Independent Schools Council of Australia	DR257
InFact Decisions	27, DR232
Information and Privacy Commission NSW	DR309
Insurance Council of Australia	66, DR318
Integrated Marine Observing System, University of Tasmania	15
Interactive Games & Entertainment Association	DR267
IoT Alliance Australia	188, DR287
iSelect Limited	DR266
Jan Whitaker	208
John D Mathews	36
John Mills	108
Johnson & Johnson	78
Joint Councils of Social Service Network	170

(Continued next page)

Table A.1 (continued)

Participant	*Submission no.*
Judy Allen and Carolyn Adams	106
Julian Lawrence	DR219
KimMic International	65
Law Council of Australia	165
Law Institute of Victoria	184
Leading Age Services Australia	47
Lisa Doyle	194
Lisa Schutz	55
Lorica Health	26
Lyria Bennett-Moses	DR261
Mark Rennick	155
Medibank Private	98
Medical Software Industry Association	DR297
MLC Life Insurance	DR298
Monash University	133
Name withheld	149
Name withheld	150
Name withheld	151
Name withheld	152
Name withheld	153
Name withheld	154
Name withheld	190
Name Withheld	DR216
National Aboriginal Community Controlled Health Organisation	192
National Archives of Australia	114, DR252
National Australia Bank	DR270
National Centre for Vocational Education Research	DR253
National Committee for Data in Science	DR265
National Computational Infrastructure	189
National Health and Medical Research Council	126, DR243
National Health Funding Body	DR234
National Native Title Tribunal	180
NATSEM - University of Canberra	DR225
NetApp Inc	166
NHMRC Centre of Research Excellence in Offender Health	193
NPS Medicine Wise	74
NSW Bureau of Crime Statistics and Research	23
NSW Government	80, DR327
Office of the Australian Information Commissioner	200, DR236
Office of the Information Commissioner - Queensland	42
Office of the Privacy Commissioner - NSW	173, DR268
Open Data Institute Queensland	70

(Continued next page)

Table A.1 (continued)

Participant	*Submission no.*
Open Source Industry Australia	130
Origin Energy	143, DR269
People with Disability	203
Phoensight	2
Pia Waugh	59
Pirate Party of Australia	DR242
Population Health Research Network	110, DR240
Private Healthcare Australia	DR295
Property Council of Australia	147, DR293
QIMR Berghofer Medical Research Institute	77
Queensland Government	207
Rannila, Jukka	6
Red Energy and Lumo Energy	63
Research Australia	117, DR282
Research Industry Council of Australia	DR263
RIM - Professionals Australasia	DR227
Rosie Williams	DR239
Ryan, Bruce	4
SA NT DataLink	123, DR283
Salinger Privacy	DR218
Sam Toady	1
Sax Institute	56, DR256
Semantic Identity	33
SPUR powered by Landgate	67
Statistics New Zealand	62
Surveying and Spatial Sciences Institute	101
Swipezy Pty Ltd	89
Tasmanian Government	205
Telethon Kids Institute	5
Telstra	88, DR312
The BetterStart Child Health And Development Research Group	51
The Department of the Environment and Energy	120
The George Institute for Global Health	68
The Law Society of NSW	160
The National Liveability Study and Place Health and Liveability	105
The Quantium Group Pty Ltd	187
Timothy Finney	31
T'Mir Julius	DR264
Tyro Fintech Hub	61
Tyro Payments Limited	7
Uber	DR311

(Continued next page)

Table A.1 (continued)

Participant	*Submission no.*
Uniting Church in Australia – Synod of Victoria and Tasmania	137
Universities Australia	90, DR302
University of Canberra: Digital Data and Society Consortium	DR213
University of Melbourne	79, 148, DR305
University of New South Wales	50, DR258
University of Sydney	35, DR258
University of Tasmania	196
University of Western Australia	DR296
VANZI	40
VEDA	163, DR259
Victorian Alcohol and Drug Association	91
Western Sydney University	119
Westpac	197, DR324
White Label Personal Clouds	49, DR217

Table A.2 **Consultations**

.id
Academy of the Social Sciences
AusGOAL
Australian Bankers' Association
Australian Bureau of Statistics
Australian Centre for Financial Studies
Australian Competition and Consumer Commission
Australian Communications and Media Authority
Australian Law Reform Commission
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 Communications and the Arts
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 Securities and Investments Commission
Australian Taxation Office
Australian Unity
Australian Urban Research Infrastructure Network
BioGrid
Bureau of Meteorology
Bureau of Transport, Infrastructure and Regional Economics
Business Council of Australia
Clean Energy Regulator
Clerk of Committees, Department of the Senate
Commonwealth Bank of Australia

(Continued next page)

Table A.2 (continued)

CRC Data to Decisions, Deakin University
CSIRO Data61
Customer Owned Banking Association
Data Republic
Deloitte
Department of the Senate
Dominello, Victor – NSW Minister for Innovation and Better Regulation
exSell Group
Facebook
Federal Chamber of Automotive Industries
Financial Ombudsman Service
Financial Rights Legal Centre
FinTech Australia
Geoscience Australia
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
Oliver Wyman
Open Access Working Group
Population Health Research Network
Quantium
RateSetter
Research Australia
Royal Melbourne Hospital
Sax Institute
Shuetrim, Geoff
Semantic Consulting
SocietyOne
SSIS Data Services
Stanley, Fiona
Telstra
Telstra Health
Torque Solutions / Velocity Frequent Flyer
Tyro Payments

(Continued next page)

Table A.2 (continued)

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

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 Data Service – Matthew Woollard
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
Department of Microbiology and Immunology
Office for Research Ethics and Integrity
Department of Surgery, Austin Campus and Faculty of Medicine, Dentistry and Health Sciences
Deputy Vice Chancellor (Research)
Victorian Comprehensive Cancer Centre; Herman Chair of Cancer Medicine, Medicine, Dentistry and Health Sciences
Head of Melbourne School of Population Health
Architecture, Building and Planning, Lecturer in Urban Analytics
Department of Computing and Information Systems
Research Computation Strategy and Infrastructure Services
Melbourne School of Population and Global Health
Melbourne Networked Society Institute
Melbourne Law School
Melbourne Institute of Applied Economic and Social Research
Pro Vice-Chancellor Research Collaboration and Infrastructure
Australian Urban Research Infrastructure Network

16 November 2016 — Business Council of Australia

AGL
ANZ
Australia Post
Australian Stock Exchange
Cisco
Commonwealth Bank of Australia
EnergyAustralia
Gilbert + Tobin
IAG
IBM
National Australia Bank
Westpac

17 November 2016 — Commonwealth agencies

Attorney-General's Department
Australian Bureau of Statistics
Australian Communications and Media Authority
Australian Institute of Health and Welfare
Australian Securities and Investments Commission
Australian Taxation Office

(Continued next page)

Table A.3 **(continued)**

17 November 2016 — Commonwealth agencies (continued)
CSIRO/Data61
Department of Education and Training
Department of Health
Department of Immigration and Border Protection
Department of Industry, Innovation, and Science
Department of Infrastructure and Regional Development
Department of Prime Minister and Cabinet
Department of Social Services
Legal Consultant
Office of the Australian Information Commissioner

Table A.4 **Public hearing participants**

Melbourne – 21 November 2016
Australian Property Institute
Pirate Party Australia
Infact Decisions and Verifier
Monash University
Australian Urban Research Infrastructure Network
University of Melbourne
Research Australia
Datanomics
Consumer Action Law Centre
Dr Henry T Burley
Cetec Pty Ltd

Melbourne – 28 November 2016
NSW Data Analytics Centre
Choice
Cloud Insurance
Tyro Payments Limited
White Label Personal Cloud
Energy Consumers Australia
Faculty of Dentistry, University of Sydney
Centre for Big Data Research in Health, University of New South Wales
Australian Privacy Foundation (Health Committee)
Fintech Australia
Data Republic
Sax Institute
Salinger Privacy
Association for Data Driven Marketing and Advertising
Association of Market and Social Research Organisations

B What the Commission's framework can achieve

Data is collected and used in nearly every sector of the economy, and as such, nearly every sector stands to gain from the implementation of the Commission's recommended reforms. The pragmatic implications of the Commission's recommendations in some very specific examples of data access are described below.

Example 1: Land use and planning data

Land use and planning data is collected and used by many organisations, including private and not-for-profit entities, and local governments. This data is rarely standardised, and only a minority of national datasets can be used freely (SPUR powered by Landgate, sub. 67; ANZLIC — the Spatial Information Council, sub. 164). Sharing of land use and planning data between different organisations is also rare, due to technical and legal limitations (Atlas of Living Australia, sub. 72; AURIN, sub. 116). Improving the discoverability and accessibility of public sector data — which the Commission believes can commence immediately (recommendations 6.1 and 6.2) — will have a substantial effect on data users and custodians in this sector, as well as the broader community:

> For example local councils are usually very data poor when it comes to managing the local environment or planning for emergency mitigation. This is because they lack the financial resources and local data analytics capability to access raw data from state and federal departments. Bringing together data from the [Bureau of Meteorology, Geoscience Australia, the Australian Bureau of Statistics] and other key agencies in a local government data portal with built in analytics could greatly improve services in regional Australia (CSIRO, sub. DR323, p. 15).

Enhanced discoverability of planning information will also assist builders and developers in the private sector, and remove commonplace issues of data access that result in unnecessary costs and inefficiencies. For example:

> A recent example is the activities of two separate state governments who are concurrently, but independently, developing an energy efficiency rating tool for dwellings. … neither jurisdiction is willing to share the data that is informing their work, either with each other, or with the industry.
>
> In addition, there are a range of other public and private sector projects that have collected significant data that should also inform this policy development, but a lack of awareness, or perhaps an unwillingness to engage with these entities, has meant that this has also not been included. The result is two different approaches to the same issue, and an additional layer of red

> tape and complexity for the development industry. This increases the cost of building new dwellings which will lead to either less supply or higher dwelling prices for home buyers. (Property Council of Australia, sub. DR293, p. 2)

Issues such as the one highlighted by the Property Council could be addressed by Accredited Release Authorities (ARAs) facilitating collaborative data projects across jurisdictional borders (in this case, local government) and by agencies maintaining registers of data assets. The Spatial Information Council (ANZLIC, sub. 164) has developed a national foundation data framework that could underpin the work of future ARAs in sectors dominated by spatial data.

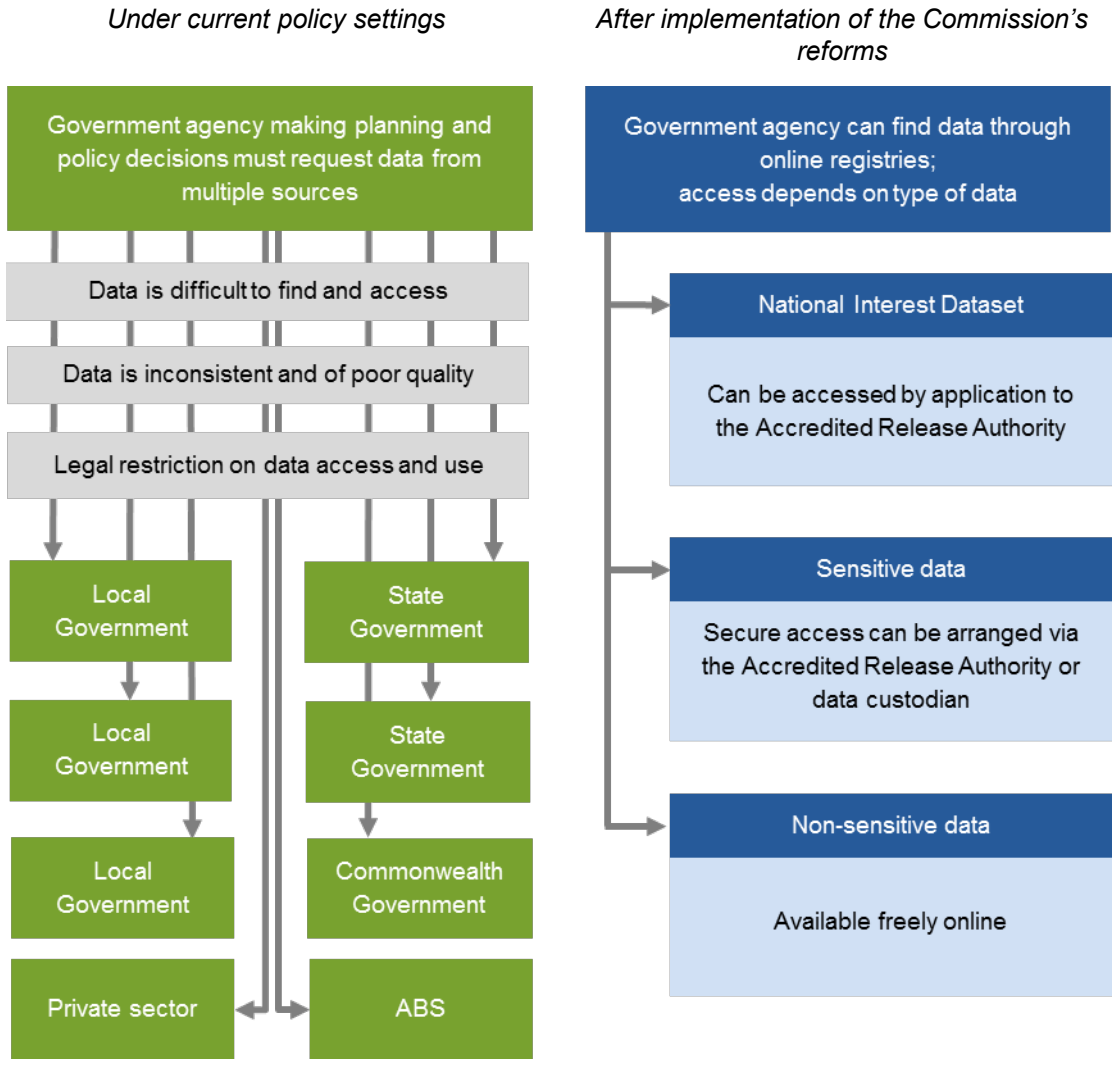
Designating land use and planning datasets as National Interest Datasets (NIDs) and accrediting organisations as ARAs to work with this data would enable progress to occur. Numerous submissions to this Inquiry have suggested that land use and planning datasets are prime candidates to become NIDs. Examples included datasets held within land and property valuation systems in local and state governments and a range of geological and environmental information (AURIN, sub. 116; Geoscience Australia, sub. 211). Once datasets are designated as NIDs, they would be curated and made available to trusted users, overcoming current barriers to access.

Other datasets would be released by data custodians or shared via the ARAs. There are already a number of organisations working to improve the availability and quality of spatial data, using different approaches. For example, the Australian Urban Research Infrastructure Network[21] (sub. DR260, p. 1) submitted that its operating model is one that can provide a 'complete data release and delivery service for researchers and government (local, state and federal).' Given the complexity of land use and planning data, it is likely that a number of entities, including Australian Government and State and Territory Government agencies, will become ARAs, and their specific models of operation will be determined by sectoral consultation. Regardless of the models chosen, the ARAs will support improved data access and use.

Such changes to data policies can become the stepping stones towards implementation of broader reform. For example, the Australian Government's Smart Cities Plan (Department of Prime Minister and Cabinet, sub. 20) relies on using data about our cities to improve infrastructure and planning — but gaps in the available data as well as difficulties in accessing jurisdictional datasets are hampering progress.

[21] The Australian Urban Research Infrastructure Network was established in 2010 with a grant from the Australian Government. Its portal offers researchers and government agencies access to about 1500 datasets and a variety of mapping tools (subs. 116, DR260).

Figure B.1 **Example 1: accessing data for planning and policy decisions**
A simplified description of data access processes



Example 2: Welfare and community services

Improving the experience of individuals receiving assistance

The barriers to data sharing between government agencies have a far reaching effect on some of the most vulnerable groups in the Australian community. As there is limited sharing of data between the multitude of support programs offered, people are required to supply the same information over and over again to apply for benefits. In 2006, the AIHW found that individuals who receive support from multiple agencies (for example, people who access housing support and may also have alcohol and gambling issues) were often required to provide the same information more than 10 times (AIHW 2006). Particularly

for people with complex needs who have to attend multiple services, the need to repeat their story over and over again is a disincentive to seeking assistance (McArthur et al. 2010). More recent reform efforts may have improved the situation, but the Commission was unable to find substantial evidence for this.

Implementing the Comprehensive Right (recommendation 5.1) will give consumers the power to authorise community service providers to transfer personal information provided to them to third parties, which will greatly reduce the burden on individuals. For example, when a person applies for social housing, they will be able to authorise the social housing provider to pass on their information to other support programs that are available from different agencies (such as community support programs provided by not-for-profit organisations or local governments).

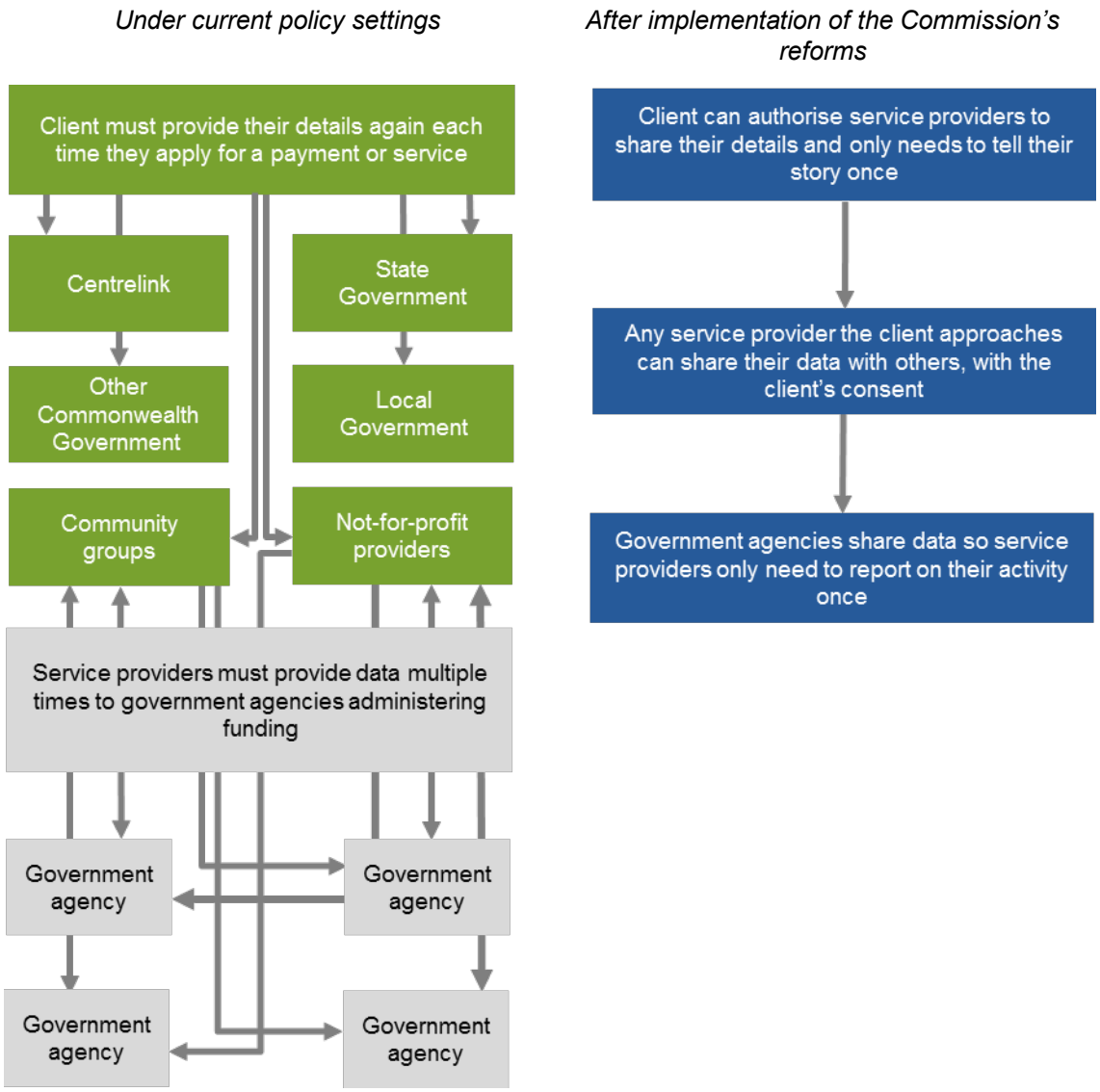
The implementation of these changes does not have to fall on government or not-for-profit organisations alone — private sector service brokers can play an important role in assisting users of multiple community services. The Comprehensive Right can support the development of such brokers, which would be able to link consumers to different support services offered by government and non-government providers. Consumers would be able to authorise brokers to share their information with providers, which could streamline application processes significantly. Such brokers are already emerging in Australia. For example, HubCare is a community services platform developed by an Australian company that offers parents and guardians a central point to manage information about their children who attend early childhood services. The platform can be expanded to health information, welfare payments and other services. Parents can share relevant information with child care providers, and use the platform to manage payments (NICTA 2014).

The benefits of data can reach beyond existing recipients of benefits; it can also increase the take-up rate of welfare benefits for individuals who may not be aware of their eligibility or may be unable to apply. For example, in 2012, the UK Department for Work and Pensions used its own data, as well as data from other parts of government, to identify older people who may be eligible for a means-tested benefit known as the Pension Credit. The Department then contacted these people and encouraged them to apply — most of the individuals who decided to apply went on to receive the benefit (Radford et al. 2012).[22]

In Australia, the same data matching process that is used to identify over-payment of welfare benefits could be used to identify households who are underpaid or who may be eligible for a benefit but are missing out. Research indicates that based on 2008-09 data, over 110 000 families missed out on parenting payments, despite being eligible to receive about $200 a week (Baker 2012).

[22] About 90% of those contacted decided not to apply for the benefit, as they continued to believe they were not eligible, thought the application process was too complex or that the amount on offer was not worth applying for (Radford et al. 2012).

Figure B.2 **Example 2: welfare and community services**
A simplified description of data flows



Such digital 'nudges', where social security administrators use data to directly identify and contact people missing out on benefits or those at risk of over-payment, have the potential to achieve better outcomes for the community (Gregor and Lee-Archer 2016). From the perspective of data policy, the success of a digital nudge in increasing benefit take-up and supporting vulnerable groups can support the social licence of government as it increases its use of data.

Cutting red tape for service providers

From the perspective of the community and not-for-profit organisations providing welfare services, the implementation of the Data Sharing and Release Act (recommendation 8.1)

would bring about a substantial reduction in duplicative reporting requirements. This is currently a problem across many parts of the community services sector:

> There are high levels of duplication in data collection and reporting requirements for [alcohol and other drug] agencies. For instance, where agencies are funded by both Commonwealth and state bodies, they may be required to navigate complex and onerous reporting requirements and to collate service user data multiple times and in varying ways, with duplication of effort and at considerable expense. Mechanisms to simplify these processes are needed so that data can be utilised with greater ease and efficiency. There are also challenges with sharing of data between services, particularly in an environment of increasing competition and marketization of community services including [alcohol and other drug] services. (Victorian Alcohol and Drug Association, sub. 91, p. 2)

The new Act recommended by the Commission will authorise data sharing between agencies (or, indeed, between different branches of the same agency) to enable a significant reduction in red tape. The Australian Government has taken some steps in this direction,[23] and the Commission's recommended reforms will support further progress.

More broadly, data sharing is the foundation of successful reform for Australia's welfare system. In 2015, the McClure Review outlined an ambitious welfare reform plan to improve employment and social outcomes. Along with streamlining welfare payments, one of the key aspects of the reform was the introduction of an investment approach to support services, modelled on a similar policy introduced successfully in New Zealand. The investment approach offers targeted, individualised support to vulnerable groups; however, it cannot work without substantial data sharing across jurisdictions:

> Based on the characteristics of these [target] groups, evidence based interventions would be implemented to increase their chances of sustained employment. This step would be more robust if there is cooperation between the Australian Government, states and territories to include data such as health, justice, housing and child protection data. Initially the use of data from sources such as the income support system, Medicare and the Australian Institute of Health and Welfare would be a first step while jurisdictional arrangements are made. Australia would need a coherent approach to data collection and analysis to support its investments (McClure et al. 2015, p. 128).

The Commission's recommended reforms provide the kind of coherent approach to data management required to increase the effectiveness of the welfare system. It would also facilitate the type of data sharing that is required to identify people who need assistance, and design the most appropriate support systems for them.

[23] For example, in 2014, the Australian Government introduced the Charity Passport, which allows charities to provide their annual reporting data once to the Australian Charities and Not-for-profits Commission, which then makes it accessible to government agencies (ACNC nd).

Example 3: Health services

The challenges of accessing and sharing health information are well documented (appendix E). There are notable difficulties for individuals who are unable to access their own records, for doctors and other health practitioners who have difficulty in accessing information about their patients, for policy makers who lack sufficient data to assess the efficacy of the health system, and for health researchers who wait many years to access data that could uncover new drugs and therapies.

The Commission envisages that key health datasets would likely become NIDs. They would be managed and curated by the health sector ARA, who would also manage access requests from trusted users (recommendation 7.3). In practice, academics who demonstrate that their work has been approved by the appropriate ethics committee would be able to access the data securely, without having to wait for years to achieve approvals from multiple custodians. The data would be de-identified, based on a risk assessment of the specific project the data will be used for. The National Data Custodian will be responsible for providing guidance on best practice de-identification processes, and certify that these processes have been followed (recommendation 6.7).

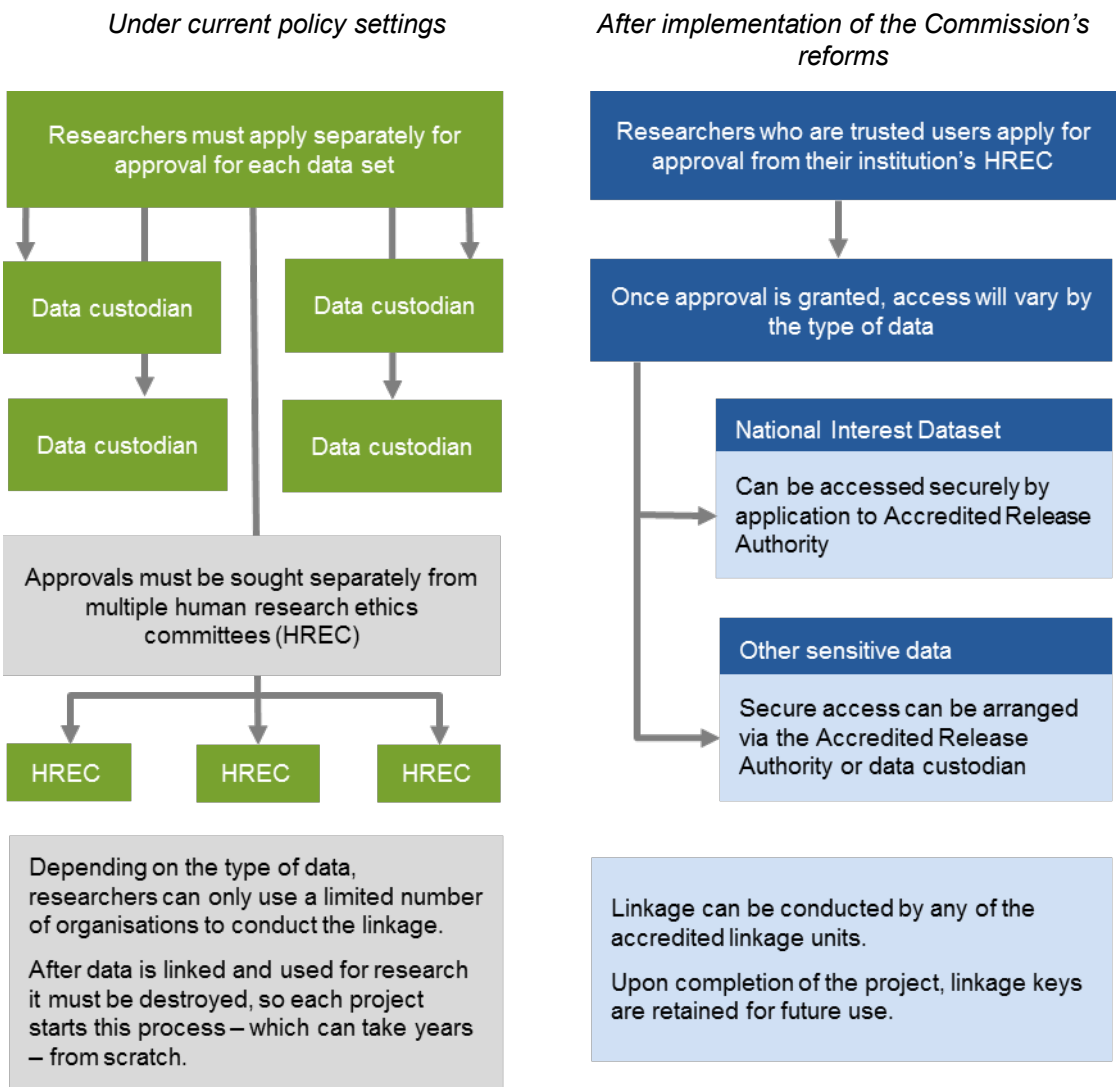
Given the expansive role of States and Territories within the health sector, some NIDs (for example, those including data on the operation of hospitals) would draw on resources from multiple jurisdictions. A network of State-based ARAs would be able to leverage the work of the Australian Institute of Health and Welfare, and implement data standards across various datasets. As the capacity for data linkage increases (recommendation 6.2), the ARAs would facilitate the linking of data from different jurisdictions. Such linkage projects have substantial value for researchers and policy makers alike.

Additional data for inclusion in NIDs may also be sourced from private health insurance companies. Special arrangements would be put in place for the inclusion of privately held datasets in NIDs, including the possibility of compensating the organisations that collect and manage the data (recommendation 7.4).

Interoperability has been highlighted as an important factor impinging on the ability to share health data, at times even between different wards of the same hospital (appendix E). Developments across the health sector, including the introduction of My Health Record, combined with the implementation of the Commission's recommendation to enable transfers of consumer data and increase data sharing, would support greater progress towards interoperability of digital health systems.

Figure B.3 **Example 3: Linking health data for academic research**

A simplified description of data access processes[a]



[a] For a detailed description of data flows in the health sector, see appendix E.

C 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 *technical* and *institutional* frameworks, with a specific focus on data linkage and open data. Chapter 1 provides broader detail on public sector data collection and access, and appendix D aims to sketch the *legislative* frameworks governing Australia's data holdings.

C.1 Key technical concepts in data availability and use

To make data more available and useful requires not only appropriate institutions and policies, but also the implementation of a range of technical measures, from the way the data is stored, to the software that can access and analyse it, and the way it is manipulated to ensure privacy is appropriately protected. The discussion below presents some of the basic technical factors that must be considered when improving data availability and use.

Discoverability

Discoverability refers to the ability to find data. Metadata (or 'data about data') 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. 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 of the data (box C.1).

There are two types of metadata managed by a registry:[24]

- *structural metadata* describes what the data fields mean and what values can be expected. It is critical to machine discovery and data integration. Structural metadata is often known as a data dictionary or an ontology.

[24] A metadata registry is a database used to store and manage metadata definitions (AIHW 2007).

- *content metadata* describes the context of the data, who last created it, when, how, what from (provenance — see below) 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).

Box C.1 How metadata works — and why it matters

The Australian Institute of Health and Welfare (AIHW nd) defines metadata as:

> … the underlying definition or structured description of the content, quality, condition or other characteristics of data.

In practice, metadata is additional information that accompanies data, which allows users to put it in context.

For example, if the dataset contains the figure '19031905', this could represent:

- A date, either a date of birth, death, migration, or service. If it is a date, it could refer to 19 March 1905, or 19 May 1903, or depending on how the data was entered.
- A period of time, such as between the years 1903 and 1905.
- A monetary amount, which depending on the currency could be very large (if, say, it was in Australian dollars) or much smaller (for example, in Indonesian rupiah).
- A population count, or a count of a specific subset within the population.
- An occupation code (for example, such as those used by the US Department of Labor) or a phone number.

Data users rely on the metadata accompanying the figures to understand what variable the figures represent and how it was measured. The metadata can also contain information about how the data was collected and how it relates to other variables in the dataset (AIHW 2007).

Metadata itself can be developed based on accepted standards, which differ depending on the type of data it describes. For example, different metadata standards apply to bibliographical and archival descriptions, record keeping and geo-spatial information (AIHW 2007). Standardising the metadata enables interoperability of data, in effect making it possible for users to find and use data in a clear and structured way.

In practice, however, individual agencies use different metadata formats and standards. The transition from the current situation to a uniform application of data standards that support interoperability can be very time consuming and entail substantial costs (box C.2).

Box C.2 How BoM resolved interoperability challenges

As part of developing the Australian Water Resources Information System, the Bureau of Meteorology (BoM) faced the challenge of bringing together data from over 260 different organisations on water storage levels. Each of these organisations used different measurement technology and methodology as well as different terminology to describe information about water.

Harmonising water storage was seen by the Bureau as a priority, due to the public interest in the data. The Bureau also believed that the level of complexity in the data would be relatively low, given that it only needed to harmonise information about water levels and volumes. However, the project proved far more complex than originally anticipated.

- There were difficulties in determining consistent definitions of terms, including fundamental notions such as 'what is water storage'. Determining agreed definitions required negotiations with the various organisations supplying the data.
- Extensive engagement with data providers was required to obtain metadata, such as type of water data and units of measurement. The metadata then had to be manually harmonised, to ensure data was comparable.
- There were inaccuracies in the time and spatial location data. For example, 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.
- Even where data was automatically transmitted using agreed protocols, issues such as recording the time measurements were taken proved complex. For example, some providers adjusted their times to reflect daylight savings while others did not.

An extended consultation with agencies led the Bureau to develop an effective water storage diagram. It then allowed the Bureau to work with agencies to ensure the correct provision of data and enable interoperability.

Source: Anderson et al. (2010)

The role of machine learning

Humans can sift through data, organise it and merge datasets together for specific purposes. However, when dealing with very large volumes of data, humans can also introduce errors and the cost of employing sufficient staff to ensure data is accurate will rise rapidly.

Machine learning can overcome these difficulties, and make the process of data management and analysis faster and cheaper. 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. However, to achieve this, there must be sufficient data and metadata for the machines to analyse (Data61 2016a).

Beyond metadata — the benefits of tracing data provenance

Some systems also describe the chain of processing that resulted in the data's metadata — this traceability of resources is called provenance (W3C 2013). If provenance is captured and published at a fine enough scale, it can be used to assist machines in knowing how to automatically combine resources for certain outcomes. It is also important in providing information to users on the best ways to use the data. This is particularly the case where data is not directly observed, but rather derived using statistical techniques (for example, where the Australian Bureau of Statistics (ABS) uses survey data to develop statistics about the entire population). 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).

Accessibility

In determining how data is accessed, different approaches are needed for different types of data. For example, aggregate or non-sensitive data can be open to the public (section C.3), while personal data needs to be protected to prevent risks to privacy (DPMC 2015).

The way data is accessed can be used to manage any possible risks. The CSIRO has identified four broad options for data access 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 released from a system based on less restrictive queries (O'Keefe and Rubin 2015).

Over time, data custodians globally have moved from focusing solely on managing the risk of re-identification,[25] to a broader view of designing access to data. The Five Safes risk assessment framework, developed in the United Kingdom and adopted by the ABS, determines the most appropriate method of data access based on five aspects of risk:

- 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?

[25] Re-identification risk refers to the risk that personal information can be discerned from data, even after identifying information such as names and addresses have been removed (chapter 3).

- 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 Department of Social Services, sub. 10) exploring virtual laboratories or trusted access models, which provide more risky data in a safer environment. For example, a remote access environment is under development at the ABS. Researchers log in 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). Academic researchers can access sensitive data securely through the Sax Institute's Secure Unified Research Environment (SURE), which is also based on trusted user models (box C.3).

Box C.3 The Secure Unified Research Environment (SURE)

The Secure Unified Research Environment (SURE) is a high-powered computing environment developed to help make the 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.

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 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. 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 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. All inputs and outputs are vetted through a 'curated gateway' for compliance and the SURE system records and archives all transactions for future reference.

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 (2016)

Other privacy-preserving technologies, such as confidential computing techniques, are also being developed as technology advances (box C.4). For example, CSIRO has developed a confidential computing platform, which allows knowledge to be extracted from data linkage without sharing raw data values between users (CSIRO 2016).

For this technique to be effective, the data must be clean and comparable, placing responsibility with data custodians to effectively curate their holdings.[26] Moreover, this technology is still in its early stages, and the scope of operations that it can perform are limited. As it is refined and optimised, the breadth of operation that can be securely performed should increase.

Box C.4 Confidential computing techniques

Confidential computing leaves data where it is securely governed and provides cryptographic means to conduct specific analyses. It is a form of restricting queries and output while still using unmodified data. 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.

Source: 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)

Confidentiality

Protecting privacy and confidentiality is an important consideration for any dataset that contains identifiable data, whether it is about people or businesses. When data custodians allow access to this type of data, there are a number of techniques that are used to protect privacy:

- *De-identification:* where the personal identifiers are removed from the data and typically replaced with a token. A person may still be identified if the dataset is combined with other data where the person is identified (re-identification).

[26] Data curation is a broad term used to describe the active and ongoing management of data throughout its lifecycle (CLIR 2014).

- *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.
- *Aggregated confidentialised data:* the data can be further aggregated until at least a small number of people are reported in aggregate in an area.

The process of data de-identification is defined by the Office of the Australian Information Commissioner as removing or altering any information that identifies an individual or is reasonably likely to do so. This generally involves two steps:

3. removing personal identifiers, such as name and address
4. removing or altering other information that may permit the identification of an individual (for example, due to the presence of a rare characteristic) (OAIC 2014).

The underlying purpose of de-identification is to enable some degree of public use of data, while protecting the privacy of individuals whose identifying characteristics appear in the dataset. The objectives of preserving the utility of a dataset on the one hand, and protecting the privacy of individuals on the other, are typically in opposition. The organisation de-identifying data therefore needs to make a decision about what it regards as an acceptable trade-off between these objectives, within the existing regulatory, ethical, and technological environment. In Australia, de-identification is required under the *Privacy Act 1988* (Cth) in certain situations.

There have been significant recent innovations in this area. For example, synthetic datasets make the data safe so that no private information is released. This is achieved using a technology known as Differential Privacy, developed at Microsoft (Dwork 2006). Differential privacy adds random noise to the data but ensures that most trends are reflected accurately when the data is analysed. However, when noise is added, there will be some analyses where the results are unusable. 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.

The need to ensure that privacy is protected when releasing or sharing data can also affect its timeliness. 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 sharing data between agencies.

Storage and cloud computing

Data storage is a general term for archiving data in electromagnetic or other forms for use by a computer or another 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 on demand. In effect, by using cloud computing, users can access computing resources over a network as they need them, rather than building or installing complex and expensive computer systems (Department of Communications 2014). Cloud computing has benefits in terms of flexible use of resources and economies of scale. It can also be beneficial in dealing with confidential data, if it is stored on the cloud and deleted after use.

The infrastructure required may be owned, managed, and operated by the organisation using cloud computing, a third party (including academia or government), or some combination of them, and it may exist on or off the premises of the user. There are four deployment models of cloud computing:

- *Private cloud*. The cloud infrastructure is provisioned for exclusive use by a single organisation comprising multiple consumers (for example, business units).
- *Community cloud*. The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organisations that have shared concerns (for example, security requirements, policy, and compliance considerations).
- *Public cloud*. The cloud infrastructure is provisioned for open use by the general public.
- *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 standardised or proprietary technology that enables data and application portability (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.

Security

Keeping data secure is a crucial consideration in all systems that handle data. There are four main aspects to security measures:

- *Authentication*: identifying who is accessing the data (see below).
- *Authorisation*: providing a defined set of roles or operations an entity can perform on the data. Once the entity's identity is known, it can interact with the service in accordance with authorisation given. This applies both to the data that the user can access and the actions they can perform, and also to the data the service provider can access on the user's device (for example, content on their mobile phone).

- *Encryption*: ensuring only those with the necessary keys can access the data. 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 a designated party.
- *Duplication*: ensuring identical copies of the resource are available elsewhere. By ensuring there are many copies of the content, corruption of a single item can be checked against the others.

Authentication — the foundation of data security

Authenticating identity is the first step in ensuring data security. It is fundamental in allowing access to data and establishing the accuracy of any data provided. For example, people need to authenticate their identity in their interactions with government and the private sector, either online or in person. Sensors and computers also have identifying information (such as serial numbers). Authentication is important as to whether to trust the data reported by the devices.

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.

- 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).
- The Department of Human Services issues public key infrastructure certificates, to allow healthcare organisations to verify their identity, when accessing and transmitting health information online (DHS 2016).
- MyGov allows the creation of an online identity for individuals, which allows them to access multiple services.

The Australian Government's Digital Transformation Agency (DTA) is developing a Trusted Digital Identity Framework (TDIF), to improve the way government agencies work together. It will draw on lessons from overseas systems, such as those used in the US, Canada, New Zealand and the UK. For example, the UK uses a federated identity model, which allows users to verify their identity online through a verification company of their choice. This process needs to be completed only once, and it allows users to access a range of government and private sector services (DTO 2016; GOV.UK 2014).

Data security threats

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 waterhole attacks where messages are sent to access illegitimate web sites often impersonating real ones.[27]
- unauthorised access can lead to malicious software (malware) being installed. Malware can damage devices, steal private encryption keys and recruit machines as bases for other attacks.
- denial of service attacks, which overload a service and stop it from operating.

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. If enough credentials are provided, the verifying authority may accept the fake person as the real one (ACORN 2015).

The incidence and severity of cyber attacks and other data breaches mean that ensuring effective data security arrangements are in place remains a real and pressing concern (chapter 3). Data security is particularly important when storing and transferring files. 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 share information. It creates authenticated, encrypted links between participating agencies to create a secure Virtual Private Network (VPN) across the Internet (Department of Finance 2016). Many academic institutions have secure transfer facilities. One example is SUFEX (box C.5), which is used by the Population Health Research Network (PHRN) to securely transfer files.

For data access and storage, modern security systems rely heavily on public and private key encryption technologies. However, if an entity's private key is stolen, their identity is effectively stolen. Encryption keys form the basis of emerging blockchain technology (box C.6). There are a multitude of suggestions being put forward for utilisation of blockchain technology. The technology shows promise for situations 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 (Data61 2016b).

[27] 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 are computer attack strategy, in which the victim is a particular group (organisation, 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.

Box C.5 **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 and has been designed, implemented and hosted by the Centre for Data Linkage (CDL), a national node of the network.

SUFEX uses the Accellion Managed File Transfer application, which provides various security features, including:

- files are encrypted, in transit and at rest, and cannot be forwarded outside the system
- 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
- logging of recipient email address, time of access and IP address makes it easy to audit activity.

Source: PHRN (2011c)

Box C.6 **Blockchain technology**

Blockchain is the technology that records transactions for the bitcoin cryptocurrency. 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 use 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. Blocks are created independently by different entities in a network, but each block contains a hash (an encoding) of the previous block so the blocks form an immutable chain. Parties in each transaction have immutable evidence as to what happened. Immutability can be breached if more than 50% of participants conspire to change it.

Contracts can be embedded into transactions to trigger the exchange of funds or cancellation of the transaction based on certain conditions. In trade, for example, where sensors are attached to high value items, they can track the state and location of the item during its transit. This data could be transmitted 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.

Source: Bitcoin (2016); Eyers (2016); Smith et al. (2016)

C.2 Data integration and linkage in Australia

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 the part of the integration process that involves creating links between records from different sources based on common features in those sources (NSS 2016f).

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 that was collected for other purposes.

Some researchers argue that linkage can have privacy-preserving benefits — for instance, in the way data linkage is carried out (box C.7).[28] 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). As a result of the data linkage process and the use of coded linkage keys, 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 (Holman et al. 2008). 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 above.

Beyond linkage keys, researchers can use a range of tools to minimise risks to privacy. These tools are based on the 'separation principle', where the individuals working with data are only able to see the information that is relevant to their role. For example, when linking two datasets that include information about medical conditions, those involved in creating the linked dataset will only be able to see the identifying information required to create the links (such as a person's name or date of birth), but they will not see any of the medical information recorded in the dataset. Those analysing the integrated data will be able to see the medical information collated from the datasets, but they will not have access to any identifying information. The result is that no one can see both the linking and analysis data (NSS 2016d).

[28] Other participants in this Inquiry have voiced the view that by creating a richer picture of individuals, linkage can in fact increase the chance of privacy breaches (see, for example, Rosie Williams, sub. DR239).

Box C.7 The process of data linkage

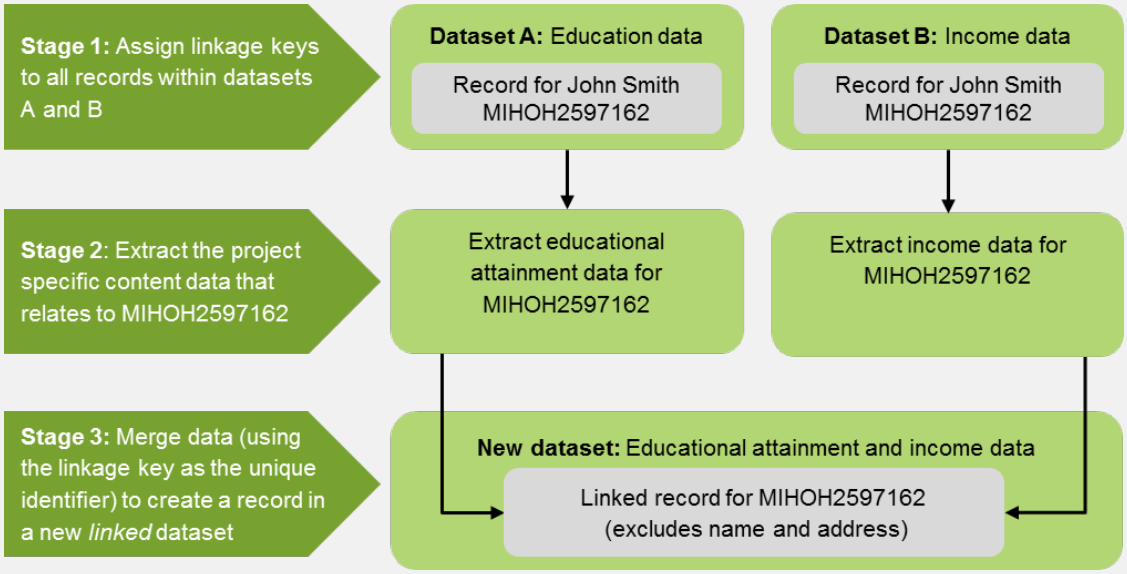
Data linkage creates new datasets by combining existing data. This can have substantial advantages, both in terms of making better use of data that already exists rather than collecting more data from individuals or business; and discovering new patterns in the data.

There are two main techniques for linking datasets: deterministic and probabilistic. In some cases, the datasets to be linked contain a unique identifier for each entity (for example, tax file number). The linkage is done by matching the unique identifier across the datasets — this is also referred to as deterministic linkage.

Probabilistic linkage is based on calculating the likelihood that two records match, when there is no unique identifier. Probabilistic linking is a slower process but generates a higher rate of matches.

Another option that is often used where there are no unique identifiers is the creation of linkage keys. The linkage key is a code created from identifiable information, such as names or addresses, that is included in both datasets to be linked. These linkage keys can be saved and reused across projects. Using linkage keys prevents the need to share identifiable information with researchers, and therefore contributes to minimising privacy risks (NSS nd, 2016e; PHRN 2011b).

Example of deterministic liking using linkage key



Source: NSS nd

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 2016).

Institutional arrangements for integration of 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 C.8). These arrangements only apply to high risk integration of Commonwealth data (box C.9) — 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.

Box C.8 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.
- *Custodian's accountability:* Agencies responsible for source data used in statistical data integration remain individually accountable for their security and confidentiality. Data custodians must 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. Integrating authorities 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. An independent assessment should be carried out to establish 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.
- *Preserving privacy and confidentiality:* Policies and procedures used in data integration must minimise any potential impact on privacy and confidentiality.
  - 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 C.9 Scope of the Commonwealth Government integration arrangements

A project is in scope if it meets **all** of the following criteria:

1) it 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 (2016c)

Where these arrangements do apply, the integration project must be carried out by an accredited Commonwealth integrating authority.

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 C.10).

- An audit by an independent third party to assess the claims made against the eight accreditation criteria
- 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 C.10 Criteria for accrediting an integrating authority

The eight criteria integrating authorities must meet to gain accreditation are:

- ability to ensure secure data management
- demonstrated ability to ensure that information that is likely to enable identification of individuals or organisations is not disclosed to external users
- availability of appropriate skills
- appropriate technical capability
- lack of conflict of interest
- culture and values that ensure protection of confidential information and support the use of data as a strategic resource
- transparency of operation
- appropriate governance and administrative framework.

Source: NSS (2016a)

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 Australian Bureau of Statistics (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 2016b).
- 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 2016b).
- The Australian Institute for Family Studies (AIFS) is the Australian Government's key research body in the area of family wellbeing. Its linkage work focuses on administrative and survey data, relevant to research into children and families (AIFS 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 3.

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 of 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 (NSS 2013).

Data integration using State and Territory Data

Over time, data linkage units have been established in all States and Territories. These units are currently cooperating through the Population Health Research Network (see below). However, they are not accredited to link Commonwealth data.

The WA Data Linkage Branch has been operating within the Department of Health (WA) since 1995. The WA Data Linkage System (WADLS) is capable of securely linking over 400 data collections, through an enduring data linkage system (Data Linkage WA 2016).[29] The Centre for Health Record Linkage (CHeReL), operating the Ministry of Health in New South Wales, also operates an enduring linkage system, within linking data about individuals in NSW and the ACT (CHeReL 2016). Data linkage units have been established by the Departments of Health in Queensland and Victoria (PHRN 2011a).

The linkage units handling data from Tasmania, South Australia and the Northern Territory operate within academic institutions. For example, SA-NT Data Link was established in 2009 and operates within the University of South Australia. It is a collaboration between the Northern Territory and South Australia partners and supports population based data

[29] Unlike data linkages involving Commonwealth data, enduring systems do not destroy linkage keys when projects are completed. These keys can be reused for later projects.

linkage research to inform many areas of policy and service development within South Australia and the Northern Territory, and nationwide (SA NT DataLink 2016).

Linking together these various organisations (as well as the Australian Institute of Health and Welfare), 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. The PHRN is a national network comprising a Program Office located in Perth, a Centre for Data Linkage located at Curtin University in Western Australia, and a secure remote access laboratory located at the Sax Institute in New South Wales (box C.3), in addition to their network of project participants and Data Linkage Units located in each State and Territory.

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 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 (PHRN nd).

C.3 Open data in Australia

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 (table C.1). 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 CSIRO's 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 ABS. It publishes a range of guidance, including on de-identification, and information development plans. The ABS also coordinates statistical activities relating to collection, compilation, analysis and distribution of statistics.
- The Digital Transformation Agency (formerly the Digital Transformation Office) 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.

Table C.1 **Components of public sector open data infrastructure**

	Jurisdictions that have implemented this change	*Jurisdictions that have not implemented this change*
	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 – 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 – Freedom Of Information legislation follows a 'push' model: disclose by default, so Freedom Of Information application should be last resort	VIC (although Information Technology Strategy 2016–2020 moots reform of Freedom Of Information legislation), ACT, AUS, NT, TAS, WA
	Institutional change	
Central institution holding and releasing data?	NZ – Integrated Data Infrastructure (IDI) NSW – Data Analytics Centre 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 (CC) licences 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.	

Source: appendix D; Draft Report chapter 1; Draft Report chapter 3; GOV.UK (2010); Government of South Australia (2016); 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)

AUSGOAL develops licences and promotes adoption of Creative Commons licences.

The Australian Government Information Management Office (AGIMO) (now within the Department of Finance) developed policies for data management and standards. The Department of Finance produced guidance on charging for data services, a range of ICT policies and standards including security, authentication and identity management, and a data centre strategy.

- The Australian Signals Directorate (ASD) and the Attorney General's Department (AGD) are responsible for the Protective Security Policy Framework (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 D).
- ANZLIC is the peak organisation for spatial data development in Australia and New Zealand.
- The Council of Australian Governments (COAG) has a role to play in agreeing to the collection of National Minimum Datasets in areas such as health, early childhood education, and disability services.

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. Other groups responsible for data policy include the Open Access Working Group, and the Science Agencies Data Stewardship Working Group (chapter 3).

State and Territory Governments

Each State and Territory Government 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 evidence-based decision making.

In Victoria:

- The Department of Premier and Cabinet administers DataVic, the Victorian Government's open data directory.
- The Department of Treasury and Finance is responsible for data policy, and driving open data in Victoria.

In Queensland:

- The Department of Premier and Cabinet is responsible for Queensland's open data strategy.
- The Queensland Spatial Information Council provides a strategic forum for the spatial information industry. It comprises representatives from the professional, academic, industry and government sectors and the community.

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.

In Tasmania:

- The Office of eGovernment is responsible for building the government's statistical assets and capability through Stats Matter, 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?

A large number of public sector datasets is available via government websites (table C.2). For the Australian Government, data.gov.au provides a registry and repository for open data.[30]

[30] Data is stored in a repository (which provides storage, and a mechanism to read and write) and findable via a registry (a database that manages metadata). In effect, this means the storage of data and searching for data do not have to be done in the same place.

Table C.2 Selected Australian public sector open data

Jurisdiction	*Websites where data can be found*
Commonwealth	• data.gov.au – open data catalogue • National Map • 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 mapping application that allows users to search geospatial databases and display 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 C.2 (continued)

Jurisdiction	*Websites where data can be found*
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 C.2 (continued)

Jurisdiction	*Websites where data can be found*
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 of 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

In addition to open datasets, the data.gov.au catalogue includes unpublished data and data available for purchase. Data.gov.au contains a toolkit, 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 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 (for example, from agencies' individual websites), not just the designated open data ones. 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).

Varying degrees of access to metadata are also observed (box C.11). Some agencies (such as the BoM) have provided access to their metadata via Open Linked Data, while other metadata registries are available only to certain users. For example, the New South Wales Government has established a metadata registry that is accessible by NSW public servants only (Data.NSW nd).

Box C.11 Data may be open — but still difficult to find

While the open data registries established by Australian governments are growing substantially (at the time of writing, data.gov.au included nearly 34 000 datasets, compared with about 1300 a year earlier (data.gov.au 2017b)), locating Australia's open data remains challenging. The datasets released by the City of Ballarat in Victoria provide a case in point.

The City of Ballarat publishes an extensive list of datasets on the various public assets it maintains, including bridges, footpaths, public toilets, playgrounds, BBQs and many more. These datasets are published on data.gov.au in a range of formats. For example, the dataset detailing Ballarat's BBQs is updated daily and published in seven different formats, including one that allows to be linked into external apps (data.gov.au 2017a).

However, a Google search will not display the data available on data.gov.au, and it must be searched for separately on the website. Google will point the user to the City of Ballarat website, which provides a list of BBQs and their location, as well as a link to all the data on Ballarat available on data.gov.au (City of Ballarat nd). The detailed data cannot be accessed through the City of Ballarat website, and none of the data is listed on the Victorian open data registry (Victorian Government nd).

Open access to research data

Data used or generated by academic researchers in the course of their work is increasingly available through open access. Through the National Collaborative Research Infrastructure Strategy (NCRIS), the Australian Government has been encouraging researchers to make their data available and reuse previously published data where available. NCRIS has provided funding for the Population Health Research Network (discussed above) as well as the Australian National Data Service (ANDS), which is responsible for facilitating the re-use of research data (box C.12). There are several types of open access platforms for research data, for example:

- 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 2016b).

Box C.12 The Australian National Data Service

The Australian National Data Service (ANDS) is a partnership led by Monash University, working in collaboration with the Australian National University (ANU) and the CSIRO. Established in 2008, it is funded by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).

ANDS' flagship service is the Research Data Australia discovery portal where users can find, access and reuse 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 (2016a)

Application Programming Interfaces (API)

Some of the open data published by governments in Australia is available via application programming interfaces (APIs). An API is a structured set of functions and procedures that allow machines to communicate with each other quickly, efficiently and reliably. 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. They also 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 below).

Examples of Australian open data that is available via APIs include tide data from Queensland and land and property spatial data from NSW (Data.Qld 2015; DFSI (NSW) nd; Land and Property Information (NSW) 2016).

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. 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 organisation. 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).

APIs are most useful when data needs to be regularly or frequently accessed or combined with other resources. 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.
- Hybrid: Some APIs are available both externally and internally. The organisation can offer access to some information to the public, and make more available for internal use or to 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 can have read-only or read-write authorisations. Read-only APIs are used when the underlying information is meant to be broadcast. A read-write API allows a consumer to interact with the government and supply information, such as submitting an online form. While useful for timely and repeatable functions, APIs come at a cost. Therefore, they are most effective at scale where there are many users in a community and many programs use the API.

Open access to real time data

Timeliness is an important factor in data quality. 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) 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.

Much non-confidential data has the potential to drastically improve public safety by being released in real-time. For example, while some data relating to potential natural hazards is already published in real-time, ongoing 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 better hazard management strategies. This would be a significant outcome, given that 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 other types of non-confidential data 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. 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, 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.

However, 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)

A common approach to minimising this risk is 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. 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.

D 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.

D.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 D.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 D.2).

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.

Box D.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.

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.

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 D.3) and the Australian Government Department of Prime Minister & Cabinet issued Guidance on Data Sharing for Australian Government Entities (2016a) 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

Box D.2 **Public Sector (Data Sharing) Act 2016 (SA)**

The *Public Sector (Data Sharing) Act 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 Act also establishes 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.

Box D.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 re-use
- 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)

Other relevant policies are as diverse as the Australian Government Smart Cities Plan (DPMC 2016b), and the Digital Transformation Office's (2015) Open Data design guide.

- 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.

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 C).
- 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 State's 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 (VAGO 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) (WALIA 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 C). This work is supported by the Stats Matter strategy, which is a long-term strategy to build Tasmanian Government statistical assets and capability led by the Office of eGovernment (DPC (Tas) 2015).
- 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
- the NSW Government's blank MOU template (currently available on the Department of Finance, Services and Innovation website)
- the MOU between the Australian Federal Police and 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 (2016a) 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.

D.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, all businesses and not-for-profit organisations with an annual turnover of more than $3 million, and some

small businesses.[31] Entities covered by the Australian Privacy Principles (APPs) in 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), most small businesses with an annual turnover of less than $3 million, media organisations, and registered political parties.

Personal, sensitive and credit information

The Privacy Act applies to personal information (box D.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 that is no longer about an identifiable individual or an individual who is reasonably identifiable (section 6).

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 Part IIIA of the Privacy Act, and the Privacy (Credit Reporting) Code 2014 (Version 1.2), appendix E).

Disclosure

Personal information cannot be used or disclosed for a purpose other than that for which it was collected, unless an exception applies, such as where the secondary purpose is related to the primary purpose , and the individual would reasonably expect that use or disclosure (Australian Privacy Principle (APP) 6.2(a)). (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 the collection is consistent with the other APPs, for instance where the collection of personal information is reasonably necessary for one of the businesses functions, and the collection is done fairly and lawfully. The requirement in APP 6 that personal information cannot be used for a secondary purpose is subject to a number of additional exceptions, such as where the individual has consented to the use or disclosure of the information, where the information is required to be disclosed by law, where it is required for

[31] The Privacy Act applies to small businesses that are 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.

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.

Box D.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.

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 entity and the individual.

Agencies are not permitted to charge for access. Organisations are allowed to charge for access 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 D.5).

Box D.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 D.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 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 entity should take reasonable steps to notify an individual, or to otherwise ensure that the individual is aware of certain matters 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 an exception applies, such as where the secondary purpose is related to the primary purpose, 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.
- Disclosure of personal information overseas — APP entities within Australia must take reasonable steps to protect personal information before it is disclosed overseas, except where the entity reasonably believes 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 take reasonable steps to 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 D.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 D.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 D.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 D.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 contravene 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.

Table D.1 **Comparing privacy principles across Australian jurisdictions**

	Cth	*NSW*	*Vic*	*Qld*	*SA*[a]	*WAb*	*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.

Queensland

The *Information Privacy Act 2009* (Qld) contains Information Privacy Principles (table D.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.

The Queensland Government is currently conducting a review of its privacy and freedom of information legislation with a view to, among other things, aligning its privacy principles more closely with the Australian Privacy Principles contained in the Privacy Act 1988.

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 D.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 D.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 D.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.

Northern Territory

The Information Act (NT) sets out Information Privacy Principles (table D.1) for managing personal information.

The NT Information Commissioner is responsible for regulating privacy compliance and handling complaints.

Australian Capital Territory

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 D.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 D.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.

Box D.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.

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)

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.

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 F). 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 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 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.

D.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 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 D.2 covering the laws and information relating to 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 My Health Record (box D.7).

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 D.3, covering information related to health, education and official secrets.

Box D.7 Regulation of My Health 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 My Health Record system, of healthcare identifiers or of other information collected in relation to either the My Health Record system or Healthcare Identifiers Service is subject to civil and criminal penalties.

Healthcare Identifiers Act

A foundation of the My Health Record system is the Healthcare Identifiers Service, which is established under the *Healthcare Identifiers Act 2010* (Cth). The My Health 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 My Health Records Rules to support the operation of the My Health record system. The Rules currently in force are: — requirements for registered entities in the system; and My Health 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 initiated trials of an opt-out registration system in the second half of 2016 (through the My Health Records (Opt-out Trials) Rule 2016, which took effect on 9 February 2016; appendix E). The Health Minister also has authority to implement opt-out nationally but only if the 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 My Health 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 reforms also expressly authorised healthcare provider organisations to upload information to a My Health Record if it includes relevant information about a third party.

Governance

The Independent Advisory Council and Jurisdictional Advisory Committee was abolished in July 2016 and the Australian Digital Health Agency became the My Health Record System Operator in July 2016.

Source: ADHA (2016)

Table D.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: • about	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
• conduct by	Any person	The Statistician or ABS officer past and present	Any person	Any person	Specified persons	An 'entrusted person'
Prohibits • Unauthorised access to information	✓ (s163)	×	✓ (s203)	×	×	×
• 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 D.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 D.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 a Senior Executive Service officer, who is not a direct supervisor of the taxation officer.

Table D.3 **Selected State and Territory secrecy provisions**

Where disclosure of information is an offence (some provisions paraphrased)

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% 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 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, using the following 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 D.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 D.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 D.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. (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. (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 162, 182 and 184 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); section 160 of the *Family Court Act 1997* (WA) (which more or less replicates section 672ZA of the Commonwealth legislation in its entirety); 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 and Lee-Jones (2016) for more detail on the provisions.

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 reporting obligation to disclose is in contrast 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), an information sharing guide for registered medical practitioners, nurses and people in charge of relevant health services in Victoria that outlines the applicable reporting obligations (box D.8).

Box D.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 judgement — 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)

D.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 D.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 D.9).

Box D.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 D.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 D.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 for:									
• Cabinet	✓	×	✓	✓	✓	✓	✓	✓	✓
• Contrary to public interest	✓	×	✓	✓	a	×	✓	✓	✓
• 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 D.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 D.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. 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 D.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 D.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 D.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 D.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.

Australian Capital Territory

The *Freedom of Information Act 1989* (ACT) provides a general right of access to documents held by government agencies (table D.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.

D.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 licences

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 D.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% of government information and data, providing a simple-to-understand and widely used system for the re-use of PSI. Remaining Victorian Government PSI should either not be released, or released under licences tailored specifically for restricted materials.

Box D.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 Government Intellectual Property Rules, and an important government entity in this space is the Australian Governments Open Access and Licensing Framework (AusGOAL).

D.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 D.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 D.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%) are wholly released for public access, while 1.75% 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 (NAA 2016).

Only 0.25% 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 D.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 D.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 D.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.

Box D.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 re-use 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.

Belcher Review recommendations

The recent Belcher Red Tape Review (2015) observed that documents about people's 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's recommendation for a comprehensive review of the FOI Act, which should consider its interaction with both the Archives Act and the Privacy Act (Hawke 2013, Recommendation 1).

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 furthermore to 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*) by changing the calendar day requirement to a business day requirement, in order to reduce the administrative burden; and to provide greater flexibility for the Archives to consult relevant entities on information

- the 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 I*nformation Act 2002*.
- In the ACT, the Territory Records Office operates under the *Territory Records Act 2002*.

D.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 Agency 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 websites (acsc.gov.au, finance.gov.au and dta.gov.au).

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 consists 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.

D.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 3).

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 C.

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 the *Privacy Act 1988* (Cth); the *Australian 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 C.

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 (Department of Health (NSW) 2005); the National Principles of Intellectual Property Management for Publicly Funded Research (ARC et al. 2001); the Australian Research Council open access policy (ARC 2015), and the National Health and Medical Research Council open access policy (NHMRC 2014).
- Requirements surrounding adverse events and research monitoring — for instance, the NHMRC's Monitoring and Reporting of Safety for Clinical Trials Position Statement (2009), 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 policies 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 policies apply, such as the Risk Management Policy and Health Service Directive Research Ethics and Governance issued by Queensland Health.

- Storage and retention of records — in addition to the Australian Code for the Responsible Conduct for Research, specific requirements apply. For example, in New South Wales, these include the *State Records Act 1998* (NSW) and a number of guidelines, such as: 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.

D.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 policies relating to the collection, sharing and dissemination of information in New Zealand, the United Kingdom and the European Union are 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 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 that 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 receiving the information has the necessary research experience, knowledge and skills.

Access to government information

New Zealanders' rights to access government records are contained within the *New Zealand Official Information Act 1982* (NZ OI Act). The express purposes of the NZ OI Act are 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 the appropriate organisation within 10 working days of their receipt and that a decision on whether a request is granted be made available within 20 working days of the request being forwarded. The OI 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 applying to New Zealand's storage of and access to government information are outlined in the Public Records Act 2005, which also specifies the extent of recordkeeping required for information relating to the affairs of central and local government. More specifically, this Act 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 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 declaration stipulates 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 D.5).

The UK Data Protection Act transposes the European Commission's 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 must 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 adequate levels 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 encourage 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 a 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). This directive 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 are 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 consultation 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).32 Prior to the June 2016 referendum on European Union membership, 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 D.14). The result of the referendum (in which the electorate voted to leave the European Union) now means that the Government needs to consider the potential application of the GDPR and how it might interact with the Digital Economy Bill if it is enacted (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

32 At the time of writing, the *Digital Economy Bill 2016* had passed the House of Commons, was under detailed consideration by the House of Lords, and was expected to receive Royal Assent by the end of May 2017 (Jackson 2016; UK Parliament 2017).

in a timely, open and transparent manner; access application processes; and the prohibition of exclusive licences.

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 1). 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' (quasi-autonomous non-Government organisation) 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 Licence — a licence 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 1). 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 D.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, data breach notification and data retention for 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 D.14 The EU General Data Protection Regulation (GDPR)

The GDPR has significant implications for citizens and businesses in the EU. Features include:

- New right to be forgotten — when an individual no longer wants their 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.

Source: European Commission (2015, 2016a)

Access to government information

The European Directive on the re-use 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 re-use of information are non-discriminatory
- charges for re-use be limited to the marginal costs of providing that data, and that charges and other conditions for re-use be pre-established and published
- exclusive arrangements are prohibited (relatedly, licences must not unnecessarily restrict possibilities for re-use, nor be used to stifle competition)
- requests for re-use are processed promptly (20 days for standard cases)

Table D.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	✓ (commenced February 2017)	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 D.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)) [h]	✓ (s6, Principle 3(4)(f)(ii)) [i]	✓ (s1)	✓ Only historical or scientific (Art 4(5), 9(2)[j], 89(1)) [k]	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] Included in the most recent revision of the Privacy Framework as of November 2016. [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 no 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.

E Case Study: Health data

Key points

- Health data collection and use in Australia — by GPs, pharmacies, hospitals, and other healthcare providers — is scattered, disorganised 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 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 healthcare 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 E.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.[33] 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).[34] 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 E.1 What is health information?

The *Privacy Act 1988* (Cth) (OAIC 2016a) 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 2016b).

[33] Examples include the Bettering the Evaluation and Care of Health (BEACH) and Magnet surveys and the Australian Rheumatology Association Database.

[34] 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.

E.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 E.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 E.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.

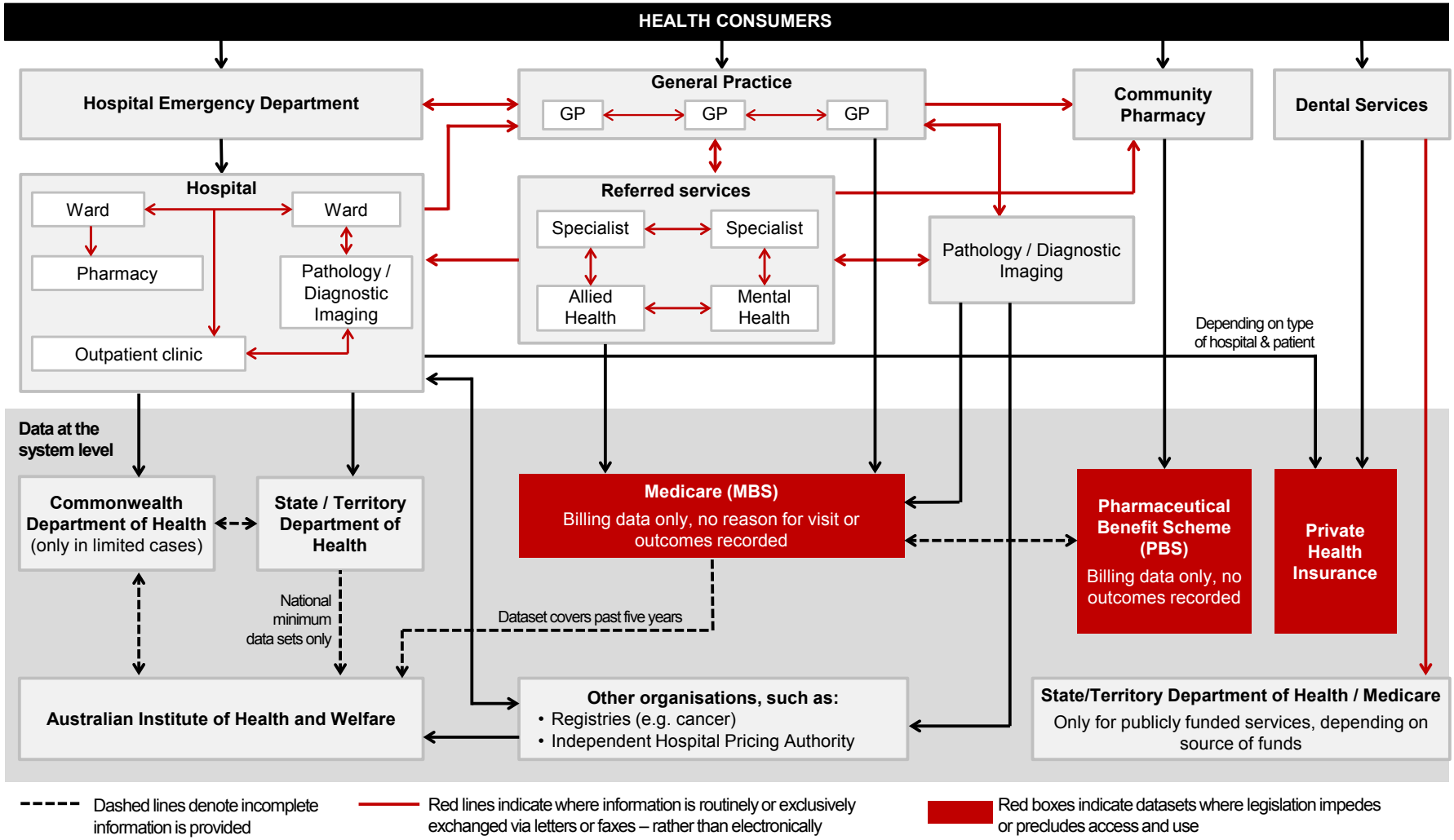
Policy frameworks for aggregate health data are fragmented

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).

Figure E.1 **Data flows in the Australian health sector**



Source: Commission analysis.

Box E.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 Meidcare Benefits Schedule or Pharmaceutical Benefits Scheme 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 (PSA 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 has almost completed the first phase of implementation of the ERRCD system (NSW Government, sub. DR327; XVT Solutions 2016), after achieving complete harmonisation of the state Poisons List schedules with the national Poisons Standard in August 2016. In other states, the Victorian Government allocated $30 million towards implementation of ERRCD the 2016-17 state budget (ABC News 2016; Hennessy 2016), and Western Australia's Department of Health committed $1 million towards ERRCD rollout in early 2017 (Haggan 2017).

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 (ABS), the Australian

Institute of Health and Welfare (AIHW), the National Health and Medical Research Council, and other agencies.[35] 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).[36] 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 receives the data, the ownership of the data remains with the original collecting jurisdiction, which can set publication conditions on the data collected. This 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 E.4).

---

[35] 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).

[36] 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.

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, this situation has not changed substantially in the past eight 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 AIHW 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 healthcare providers).[37] 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 both primary and hospital care, using a variety of existing data collections (NHPA 2015b).

[37] 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 the past year, 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.

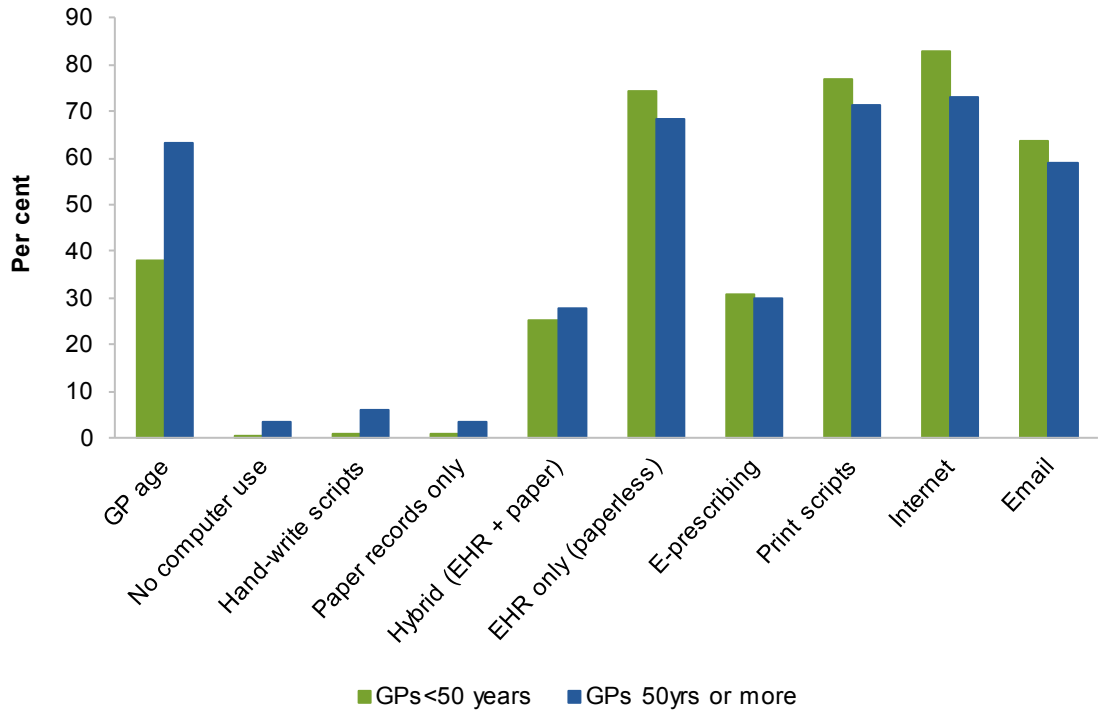
The systems and standards in place

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 E.2).

Figure E.2 **IT use and the age of Australian GPs**



Source: Henderson et al. (2014)

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).

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).

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. 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 is the use of agreed practices and terms for data interpretation (box E.3).

Box E.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: Grieve (2014), as quoted in McDonald (2015b)

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, design 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 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 E.4).

Box E.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 Australian Medical Association 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 … this meant that patients' 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'.

Source: EHSC (2015, p. 23); DHHS (Vic) (2013, p. 31)

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 E.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.

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 National Health Service (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.

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.

Box E.5 Health policy in the United States — implementing data-driven reform

The US healthcare system over the recent past has undergone substantial change, driven primarily by the introduction of the *Patient Protection and Affordable Care Act* in 2010. That Act put 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, included 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 (DHHS (US) nd).

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.

E.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 PCEHR/MHR system, along with other elements of eHealth in Australia (box E.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, where the patient may not be able to provide the information (due to being unconscious or distressed). 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 E.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.

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).

Box E.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 (House of Commons 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, Bristol 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 at a greater level of detail than that offered by the SCR (Healthcare Gateway 2013; London Connect 2013; Pugh 2017).

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.

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 E.3). 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

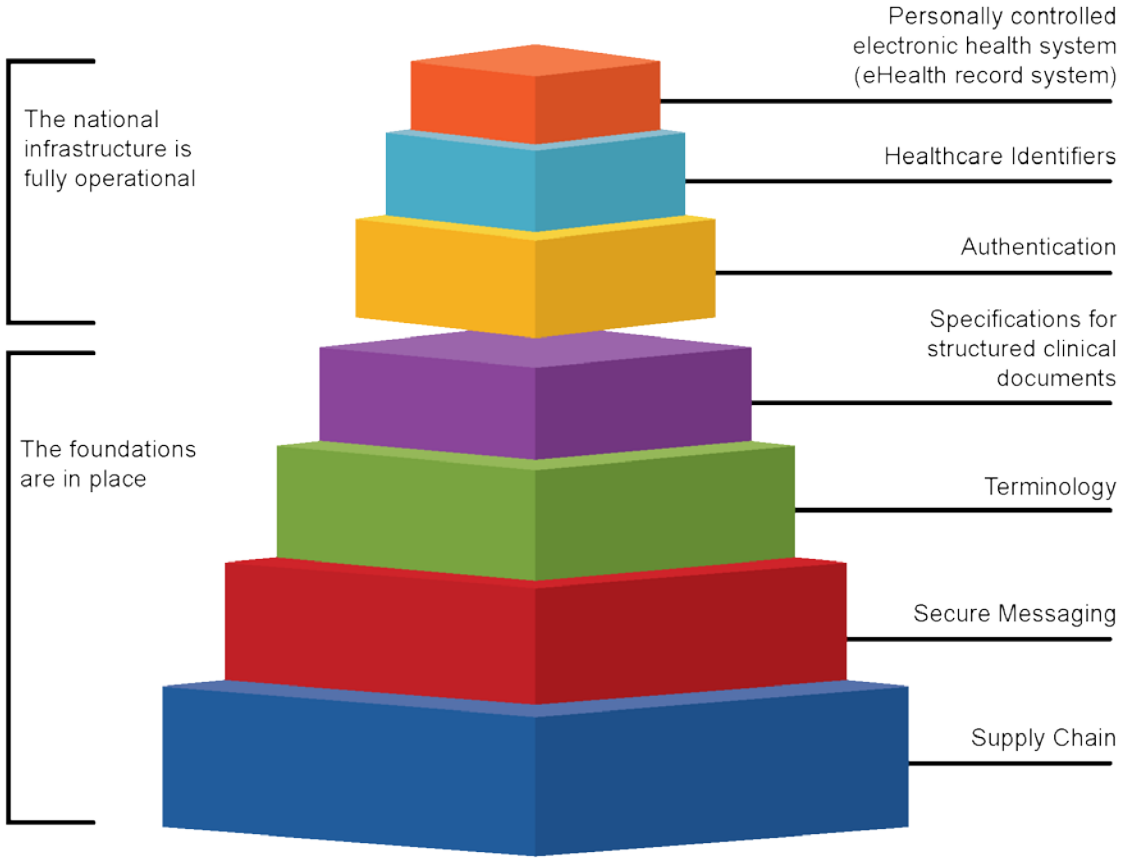
A major milestone in creating the infrastructure for the EHR was the introduction of Individual Healthcare Identifiers in 2010.

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.

Figure E.3 **The building blocks of eHealth in Australia**



Source: NEHTA (2016))

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 2015). The *PCEHR Act 2012* was amended in November 2015 to become the *My Health Record Act 2012* and reflect these major changes (Parliament of Australia 2015). Some technical features of the previous system were rolled into the new My Health Record (box E.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 E.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.[38] 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 restored 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; OAIC 2011). Meanwhile, there is concern that this precise feature of the record may result in healthcare providers relying on incomplete information, reducing its efficacy and raising questions of liability for healthcare providers (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)

[38] Section 64 of the *My Health Record Act 2012* (Cth) 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.

In a submission on the amending legislation discussion paper, the OAIC considered the privacy issues arising from an opt-out model (2015b). 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 APPs, the model was to be implemented in the most privacy-enhancing way possible — including giving individuals maximum opportunity to exercise their right to opt-out, and ensuring that existing personal controls over information within records were not diminished (OAIC 2015b, p. 4).

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 (MinterEllison 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 (for those who had not opted out) in June. The trial concluded at the end of October, after five months (though the Explanatory Memorandum for the amending Act gave up to nine months for the trials to take place (Parliament of Australia 2015, p. 11)), and just over 970 000 unique My Health Record accounts were created during the trial (Departmental Officer, Department of Health, pers. comm., 21 February 2017).

At the time of writing, the Evaluation Report on the trial's outcomes had not yet been released to the public. However, statistics from the Department of Health indicated that the actual proportion of people opting out of MHR creation in the trial areas was just 1.9% (Departmental Officer, Department of Health, pers. comm., 21 February 2017). Additionally, the Secretary of the Department of Health commented in late 2016 that the rate of Shared Health Summary uploads had increased by more than tenfold from April 2016 to October 2016, suggesting that an expansion of the opt-out system to cover all Australians would likely be of significant benefit (Arnold 2016).

If the Department of Health does decide to pursue expansion of opt-out registration nationwide, no amendment to the *My Health Records Act 2012* (Cth) is necessary: Schedule 1 to the Act already allows the Minister of Health to make rules applying the opt-out registration system to all healthcare recipients in Australia, which may be disallowed by Parliament but otherwise have normal legislative force.

Alongside the later stages of the opt-out registration trial, the Department of Health appointed a team to conduct public consultation on possible secondary uses of MHR data in August 2016. However, in October 2016, this was postponed due to possible overlaps with simultaneous consultations such as that on the National Digital Health Strategy. The consultation process for secondary use of MHR was mooted to resume in early 2017 (Department of Health and HealthConsult 2016).

At present, 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 always be automatically informed if the patient has an MHR, but may need to specifically search the MHR database for it), this may be impeding the creation of a single 'source of truth' patient record.

E.3 Using health data in research and policy development

The data collected by healthcare providers across Australia, and generated from administrative datasets such as Medicare, is used to produce myriad health indicators.[39] 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 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;

[39] 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).

- 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 E.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 E.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 data quality issues (discussed section E.2 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 healthcare data (which would include the Medicare Consumer Directory):

- 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' (section 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 Benefits Schedule (MBS) 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 (OPC 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).[40]

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:

> 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).

[40] Each Act uses different language in describing how information should be handled. According to DHS, this creates significant confusion (ALRC 2010).

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 reuse 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 community' (AIHW 2015, p. 3).

A further step towards increasing the use of Medicare and PBS data in linkage projects occurred in August 2016, when the Department of Health released a linkable, de-identified sample of MBS and PBS data (including roughly 10% of data points from 1984–2014) on data.gov.au. The data was subsequently removed from the website following partial re-identification of the encrypted Healthcare Provider Identifiers; however, the Department has stated that it will work towards making the data available again in the future, following

further de-identification (Department of Health 2016b, 2016d).[41] These changes may affect the usefulness of the data for researchers in a way that making identifiable data available through a restricted trusted user model would not; 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 datasets to be linked (Holman et al. 2008). Currently, Data Linkage WA is able to create linkages between eight core datasets (seven health datasets, and the WA electoral roll), and over 20 other datasets, 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 datasets 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 (DPC (WA) 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

[41] In light of the sample dataset's partial re-identification by University of Melbourne cryptographers, the Commonwealth Government introduced to parliament the *Privacy Amendment (Re-identification Offence) Bill 2016*, which criminalises the re-identification of de-identified personal information published by Commonwealth entities. At the time of writing, the Bill had not yet passed either House (Parliament of Australia 2017).

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 dataset, 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).

E.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 4).

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 E.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 (BHI), including information on activity and performance measures (such as risk-adjusted mortality data) to assist patients' choice of hospitals and practitioners (BHI 2016; Health Stats 2016).

Box E.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 datasets, 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 2016).

- 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 conducts research activities nationwide. 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).
- BioGrid Australia provides a real-time data linkage platform, using data from 34 hospitals and other healthcare providers in different jurisdictions. This allows researchers to conduct multi-site projects, using de-identified data transferred over a secure network. BioGrid also streamlines the ethical approval processes for researchers using its network; in effect, researchers require only one approval to access data from all sites (BioGrid 2017).

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.

F 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 — 'fintech' businesses — are capitalising on these developments 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 Regulation Authority (APRA), the Australian Securities and Investments Commission (ASIC) and the Reserve Bank of Australia (RBA).
- 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 there are clear public benefits associated with higher levels of participation, at present participation levels are below those of other OECD countries.
- 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 — via a preferred Application Program Interface (API) for example — third party access to financial data about customers, with the customer's consent. A more encompassing Comprehensive Right for consumers to access and share data, beyond just financial data, is addressed in chapter 5.
- When deciding what datasets to make more widely available — for example, to 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 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.

F.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 such as Yodlee) 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 regulated under the *Privacy Act 1998* (Cth) (the Privacy Act) (section F.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 an 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 was through their participation in the credit reporting framework (section F.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 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, 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.

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 the Australian Securities and Investments Commission (ASIC)

- the tax reporting and Self-Managed Superannuation Fund (SMSF) systems, as regulated by the Australian Taxation Office (ATO)
- 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. 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 favoured 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).

F.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.[42] 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

[42] The three main credit reporting bodies in Australia are Veda Advantage, Dun & Bradstreet and Experian.

- any credit defaults (which is the failure to meet legal repayment obligations) in the previous five years
- 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, ASIC-licensed credit providers 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 with 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)

Several Inquiry participants (for example, Westpac, sub. DR324) argued that banks would face risks to their reputation if consumers were allowed to transfer banking data to third parties who then suffer from data breaches or otherwise misuse the data. In a system with transparent processes for data transfer, including a clear point of transfer of custom from one provider to another, it would seem rather unreasonable to the Commission that a provider would be blamed for a previous customer's ill-informed decisions.

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:

- 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 has 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 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 & 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, although the materiality of any advantage might diminish over time as data becomes increasingly available. To the extent that governments can encourage data availability, there could be scope for increased competition and improved consumer outcomes in markets for financial services. Of course, there are also costs associated with greater sharing of data (section F.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 F.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 F.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), the Australian Prudential Regulation Authority (APRA) has collected data for over a decade and, in some cases, made the data publicly accessible, primarily through its statistical publications.
- Under the 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.

Source: APRA (2013, 2015)

F.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, DR233), have called for the mandating of partial comprehensive credit reporting (CCR) (as well as reporting of defaults).[43]

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)

[43] Partial CCR would not include repayment history information.

The incentives for an institution to participate in comprehensive credit reporting can be mixed and quite complex. Participation depends on the perceived net benefits, which will differ between different classes of credit provider.

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 reportedly 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).
- 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. The Customer Owned Banking Association (sub. DR273) argued that a 'materiality' threshold should be put in place for mandated participation, with financial institutions with less than $200 billion in assets retaining the right to participate on a voluntary basis.
- 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 & Bradstreet, sub. 135, Financial Rights Legal Centre, sub. DR289).

- Also, as noted by the Attorney General's Department (sub. DR334, p. 5), there is the question of whether the mandating of participation in CCR could '… raise constitutional issues around the acquisition of property and, if so, would require the Australian Government to pay compensation on just terms to credit providers for compelling them to disclose valuable commercial information'.

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.

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.

On balance, 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. However, noting that the current system is the result of a long reform process, and that there have been no reviews of the current system, it would appear to be premature to consider further reforms, particularly since the current system is underutilised.

Obstacles to greater participation

A range of Inquiry participants (see, for example, the Australian Bankers' Association, subs. 93 and DR307; the Customer Owned Banking Association, subs. 132 and DR273; Westpac sub. DR324) 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.

The OAIC, in response to a request from ARCA, has since clarified how RHI should be reported for borrowers in financial hardship (chapter 5) — this should lead to greater participation in CCR.

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 as ODI and Fingleton (2014)), 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 C). 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.

Several submissions received following the Draft Report called for the government to mandate the introduction of APIs to allow customers to share their data with third parties (for example, the Customer Owned Banking Association, sub. DR273). Westpac (sub. DR324) recommended that the financial services sector develop common API standards to facilitate broad adoption of APIs. Other submissions argued that there would serious risks that need to be addressed prior to APIs being rolled out, and that there is a need for an overarching regulatory regime (for example, the Australian Bankers' Association, sub. DR307; National Australia Bank, sub. DR270). The possibility of disincentives to adopt new technological solutions in the future, if APIs were mandated, was also noted (Australian Information Industry Association, sub. DR244).

The private sector has started to implement 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 F.2). In addition, some Australian financial services firms have already begun to implement APIs for commercial reasons (FinTech Australia, sub. 182; National Australia Bank, sub. DR270).

Box F.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 (though not all consumers would necessarily value targeted offers).

Conversely, some Inquiry participants pointed to risks that could arise from increased access to customer data (chapter 5). 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.

Policy developments in the United Kingdom, European Union and Singapore

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, undertaken in 2014, 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. It also found that the costs to banks of implementing APIs were 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.

The recommendations were subsequently implemented on a voluntary basis.

In 2016, the UK Competition and Markets Authority (2016) found that the UK banking sector was not as competitive or innovative as it could 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.

Recent reforms in Europe, notably 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' 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).

In Australia, the issue of banks sharing data through APIs was raised as part of a recent inquiry into Australia's four major banks by the House of Representatives Standing Committee on Economics (2016). The inquiry was prompted by concerns about the degree of market power of the four major banks, as well as numerous banking scandals in recent years. A number of specific issues were canvassed — including barriers to entry and empowering consumers by mandating APIs (for the four major banks).

The Committee recommended that:

- Deposit Product Providers be compelled to provide open access to customer and small business data by July 2018
- ASIC develop a binding overarching framework to facilitate the sharing of data via APIs
- banking entities be required to publish product terms and conditions (that is, product reference information) in a standardised machine-readable format
- the *Corporations Act 2001* (Cth) be amended to introduce penalties for non-compliance with the recommended reforms.

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 F.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).[44]

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 any financial services providers in Australia making customer data sharable *with their competitors* through the use of APIs. There are, however, several businesses that have established data feeds from the major banks, and who facilitate individuals (and businesses) importing their transaction account data into selected accounting software packages (for example, SSIS Data Services).

Finally, even where consumers are able to download and forward 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 2018 deadline for implementation of PSD2 reforms (although this is a functionality beyond that related to data access).

---

44 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.

Disadvantages of the API approach

There are costs associated with the use of APIs, including IT 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 provide 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.) iSelect Limited (sub. DR266) suggested that the initial cost of establishing an API would be between $10 000 and $50 000, with most being about $20-25 000.

There is also the possibility that standards would be required for efficiently implementing open APIs across industry (to ensure a common language is used by all industry participants).[45] There are costs involved in setting standards, both for the private sector and for governments (the latter may, for example, play a coordination role).

Depending on the how, governments may also be involved in the certification of third parties that could receive the data. Such certification would involve administration costs for governments and compliance costs for businesses.

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.

Weighing up the benefits and costs of mandating the use of APIs

Mandating the use of APIs could require resolution of matters surrounding data standards and responsibilities for data quality and data security. There would also be other issues needing examination — for example, would giving consumers the right to share their data also allow them to 'cherry pick' which transactions they share, or should there be rules in place to prohibit them from doing so?

Assessing the overall (community-wide) costs and benefits of mandating customer access to their financial data is difficult because of the large uncertainties involved.

---

[45] Whether such standards should be mandated is a separate issue (discussed in chapter 5).

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 access and transfer personal data is much broader than the financial sector. And as noted in chapter 2, many opportunities that could arise from data availability are not foreseeable prior to data being made available. As such, the Commission has recommended a new suite of rights for consumers for data that relate to their activities as consumers (chapter 5).

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 consumers. 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 format (with an industry agreed layout/format).[46]

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 machine-readable 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.

---

46 To be clear, this relates only to product reference information, such as the terms and conditions, and fees and charges for a particular product. It does not relate to data that would be transferrable under the Comprehensive Right (chapter 5).

F.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. While there would be implementation costs for financial service providers, and these would be relatively more burdensome for smaller providers, it is apparent that there is significant scope for individual consumers to be better off if they were able to seamlessly and securely share their financial data. 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, although in practice this would depend on the scope of data transferrable.

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.

G 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 range of data sources 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 the data that they generate.

G.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 message 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 — particularly popular are fitness trackers, which capture an array of data about the wearer's physical activities. 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 smartphones 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 users to make payments without swiping their card — instead, users can wave their card or device in front of a reader to complete a transaction. MasterCard has similar technological capabilities offered under MasterCard Contactless. These devices use near field communication technology, enabling devices 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 (Telsyte 2016). 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 1, smart electricity meters and refrigerators equipped with sensors that allow them to perform tasks — such as monitoring usage and identifying 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 is 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 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, but recently it is maps and navigation apps that see the highest downloads (table G.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 G.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.

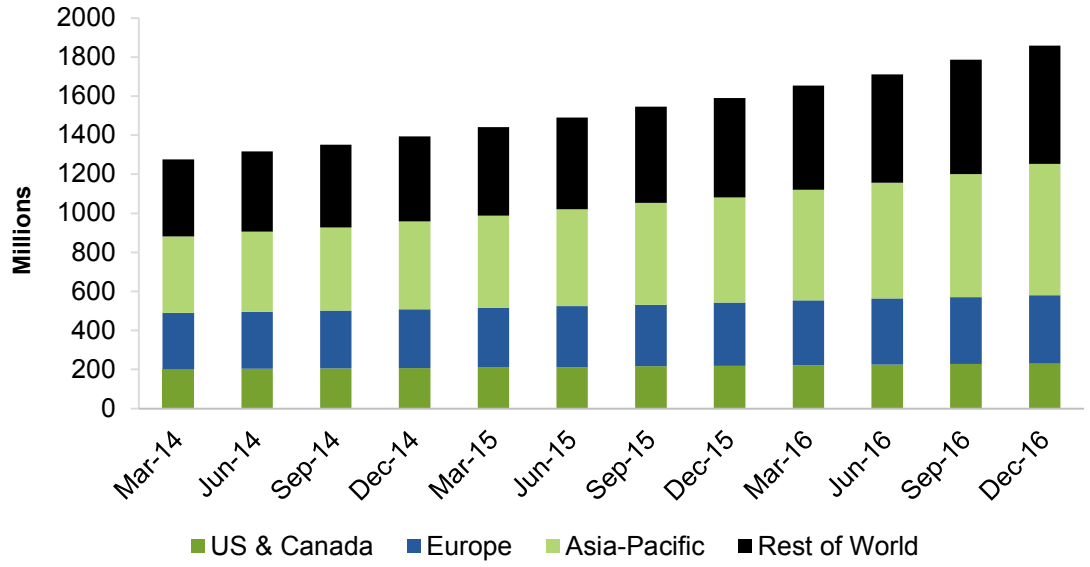
Data generated by devices on the IoT varies with the nature of the device. Typically (with the exception of wearables), data may reflect the activity, location, status (on or off) or performance of a device. Such data may reveal information about the individual person or business who owns the device. 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 services are web-based environments that readily allow users to create, publish, and share content amongst each other. Approximately 69% of Australian consumers, nearly 80% of large businesses and about half of small and medium-sized businesses have a social media presence (Sensis 2016). Facebook is currently the most prominent social media company, with over 1.6 billion monthly active users (figure G.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 G.2). Consumers use social media primarily to catch up with family and friends; businesses use it primarily for two-way communication with clients and contacts (Sensis 2016).

Figure G.1 **Facebook monthly active users (global)**



Source: Facebook (2017)

Table G.2 **Social networking sites used by Australian consumers and businesses**

2016, percentage of survey respondents who used social networking[a]

Site	*Consumers*	*Small business*	*Medium businesses*	*Large businesses*
Facebook	95	89	79	89
LinkedIn	24	22	56	59
Instagram	31	12	28	42
Google+	10	10	10	19
Twitter	19	24	43	61
Pinterest	11	2	11	11
Blog	na	4	4	9
Snapchat	22	0	0	3
Tumblr	5	na	na	na
Vine	3	na	na	na
Yelp	1	2	0	1
Foursquare	1	na	na	na

[a] The survey base comprised 800 consumers and 1100 businesses (of which 900 were small, 100 medium, and 100 large). Of these, 69% of consumers (551 consumers) used social networking; 79% of large businesses (79 large businesses) used social networking, while 54% of medium businesses (54 medium businesses) used social networking, and 48% of small businesses (432 small businesses) used social networking.

Source: Sensis (2016)

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 you have granted), including specific geographic locations
- information about you from third-party websites, apps and partners of Facebook (Facebook 2016c, 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 2016c).

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 2016b). 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 3 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 G.1).

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 G.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.

G.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 97% 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 2017).

Box G.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.

Source: BBC (2014); EC (2016a); Microsoft (2016a)

Box G.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', because 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.

Source: Facebook (2016a, 2016b)

In discussing how this information is used and shared, Facebook explains:

> When we have location information, we use it to 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. (2016c, 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. (2016c, 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 G.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 G.3).

Table G.3 **Attitudes to advertisements on social networking sites**
2016, percentage share of respondents

Statement	*Agree*	*Neutral*	*Disagree*
I'm turned off by companies or brands that advertise on social network sites	32	34	34
I take no notice of the ads on social network sites	53	20	27
I ignore sponsored posts from businesses I do not follow	67	16	17
I sometimes click on ads I see on social network sites to find out more	43	13	44
I like sponsored posts from businesses I follow on social networks	27	23	50
I'm quite happy to see ads on social network sites	34	25	41

[a] Base is total number of social media users in survey = 544.

Source: Sensis (2016)

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 (Twitter 2016a). 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 in 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 trialling the 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. Because 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 G.3).

Box G.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 smartphone, 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 (2015) 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. This alert goes via 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 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 would 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 the insurance firm 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 datasets 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 in the United States of ResearchKit, 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, ultimately aiming for publication as an open dataset, to be accessible to 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, activity levels, medications taken, and foods consumed (as well as calculating their carbohydrate intake). 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 can potentially be used 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 and 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 maximising awareness 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 to monitor 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 continuously collect data 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 technologies to collect data on environmental variables and critical processes. The company uses sensor network technology in its plants to collect data and monitor (in real time) light intensity, the air temperature around printing machines and pipes, and CO2 concentration in areas occupied by workers. 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 collects 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.

G.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 statements 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 2015)

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 2015)

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. (2016d, 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. (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 for the purposes of furthering the commercial interests of the platform in question, and in some cases, associated third parties, usually without compensation to the creator of the content[47]. 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 characteristics of services, such as 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. It was argued that, coupled with data mining techniques, this may create risks to users, including the possibility of events such as employment and insurance discrimination and unwanted marketing.

[47] 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 selves, resulting in complete profiles of their 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 swathes 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, because 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 fully anonymising 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. 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 from those relevant to personal data transmitted over the Internet and mobile phones — for example, data collected may include financial data and precise geolocation. 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 G.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 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 2016, 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 G.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, the 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 Privacy 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 is likely to be used.

Navmii, a navigation app, stipulates in its privacy policy that it shares data with data processors, 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 to 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 policies

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 only 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 from 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 (European Commission 2016b).

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 is no longer necessary in relation to the purposes for which it was 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 (European Commission 2016b).

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 2016c).

The Committee also pointed out that exercising the right to be forgotten 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 it is 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 must 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 ALRC 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). It was 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 ALRC also noted that, given the ease with which information can proliferate and be disseminated 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 ALRC 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 News (Australian Broadcasting Corporation News) 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).

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.

—— 2015, *National Health Survey: First Results, 2014–15*, Cat. 4364.0.55.001, Canberra.

—— 2016a, *About the Australian Bureau of Statistics*, Canberra, http://www.abs.gov.au/about?OpenDocument&ref=topBar (accessed 20 October 2016).

—— 2016b, *Data Integration Initiatives*, February, 1015.00, Information Paper: Transforming Statistics for the Future, Australian Government, Canberra.

—— 2016c, *Internet Activity, Australia, June 2016*, Cat. 8153.0, Canberra.

—— 2016d, *Organisations approved to use CURF Microdata*, Canberra, 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, *Submission to the Productivity Commission Inquiry into the Regulation of Australian Agriculture*, Canberra.

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.

Accenture 2015a, *Revised Payment Services Directive (PSD2) Everyday Payments — Accenture*, Sydney, https://www.accenture.com/au-en/insight-everyday-payments-europe (accessed 8 July 2016).

—— 2015b, *The Digital Disruption in Banking: Demons, Demands and Dividends*, Sydney, 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?*, Sydney, https://www.accenture.com/au-en/insight-future-fintech-banking (accessed 3 October 2016).

—— and General Electric 2014, *Industrial Internet Insights Report for 2015*, Sydney, 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; Report for ANZLIC — the Spatial Information Council, Sydney, http://www.crcsi.com.au/assets/Resources/7d60411d-0ab9-45be-8d48-ef8dab5abd4a.pdf (accessed 9 May 2016).

ACMA (Australian Communications and Media Authority) 2012, *Location services, personal information and identity: Exploratory community research*, Australian Government, Canberra, 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, Australian Government, Canberra, 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, Australian Government, Canberra.

—— 2015, *Communications report 2014–15*, Australian Government, Canberra.

ACNC (Australian Charities and Not-for-profits Commission) nd, *Charity Passport*, http://acnc.gov.au/ACNC/About_ACNC/Redtape_redu/Charity_Passport/ACNC/Edu/Charity_Passport.aspx (accessed 16 January 2017).

ACORN (Australian Cybercrime Online Reporting Network) 2015, *Attacks on computer systems*, 25 November, Australian Government, Canberra, 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, Australian Government, Canberra, http://www.safetyandquality.gov.au/atlas/ (accessed 18 May 2016).

ACT Government 2015, *Proactive Release of Data (Open Data) Policy*, December, Canberra.

—— 2016, *About Access Canberra*, Access Canberra, https://www.accesscanberra.act.gov.au/app/answers/detail/a_id/1782/~/about-access-canberra (accessed 23 February 2017).

ACT Health 2017, *National Mutual Acceptance (NMA) Scheme*, http://health.act.gov.au/research-publications/research/research-ethics-and-governance-office/national-mutual-acceptance-nma (accessed 27 February 2017).

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, Sydney, 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) 2016, *Legislation: Changes to legislation*, 3 April, Australian Government, Canberra, https://myhealthrecord.gov.au/internet/mhr/publishing.nsf/Content/legislation (accessed 20 October 2016).

AEC (Australian Electoral Commission) 2016, *Supply of elector information for use in medical research*, August, Australian Government, Canberra, 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*, Brisbane, https://aehrc.com/research/case-studies/making-health-records-more-accessible/ (accessed 10 June 2016).

AGD (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.

—— 2015, *Data Retention*, Australian Government, Canberra, https://www.ag.gov.au/dataretention (accessed 1 March 2016).

—— 2017a, *Access to telecommunications data in civil proceedings*, Attorney-General's Department, https://www.ag.gov.au/Consultations/Pages/Access-to-telecommunications-data-in-civil-proceedings.aspx (accessed 9 March 2017).

—— 2017b, *Protective Security Policy Framework — Home*, https://www.protectivesecurity.gov.au/Pages/default.aspx (accessed 16 January 2017).

AGIMO (Australian Government Information Management Office) 2006, *Australian Government Information Interoperability Framework*, April, Australian Government, Canberra, 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 Government, Canberra, https://aifs.gov.au/about-us/what-we-do (accessed 20 October 2016).

AIHW (Australian Institute of Health and Welfare) nd, *About metadata*, http://meteor.aihw.gov.au/content/index.phtml/itemId/268284 (accessed 16 February 2017).

—— 2002, *Australia's Health 2002*, Cat. AUS 25, Australian Government, Canberra.

—— 2006, *Cutting the red tape: preliminary paper detailing the problem of multiple entry and reporting by service providers*, http://www.aihw.gov.au/publication-detail/?id=6442467913 (accessed 16 January 2017).

—— 2007, *A guide to data development*, Cat. HWI 94, Australian Government, Canberra.

—— 2014a, *Australia's Health 2014*, Cat. AUS 178, Australian Government, Canberra.

—— 2014b, *Data Governance - in-brief*, Australian Institute of Health and Welfare, Canberra.

—— 2014c, *Data Governance Framework 2014*, Australian Government, Canberra, http://www.aihw.gov.au/data-governance-framework/ (accessed 24 August 2016).

—— 2015, *Submission to the Senate Select Committee on Health*, Australian Government, Canberra, http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Health/Health/Submissions (accessed 26 May 2016).

—— 2016a, *About*, Australian Government, Canberra, http://www.aihw.gov.au/about/ (accessed 22 September 2016).

—— 2016b, *Annual Report 2015-16*, October, Australian Institute of Health and Welfare, Canberra.

—— 2016c, *Cancer registration in Australia*, Australian Government, Canberra, http://www.aihw.gov.au/cancer-registration-in-australia/ (accessed 9 June 2016).

—— 2016d, *Custom Data Request Service*, Australian Government, Canberra, http://www.aihw.gov.au/custom-data-request-service/ (accessed 27 September 2016).

—— 2016e, *Privacy of data*, http://www.aihw.gov.au/privacy-of-data/ (accessed 23 December 2016).

—— 2017, *Committees*, http://www.aihw.gov.au/committees/ (accessed 17 January 2017).

Allianz Australia 2016, *Privacy Policy*, https://www.allianz.com.au/about-us/privacy/ (accessed 15 June 2016).

—— nd, *Use of telematics in vehicles*, Allianz Australia, https://www.allianz.com.au/car-insurance/news/use-of-telematics-in-vehicles (accessed 20 March 2017).

ALRC (Australian Law Reform Commission) 2003, *Essentially Yours: The Protection of Human Genetic Information in Australia*, Final Report, 96, Australian Government, Sydney, http://www.alrc.gov.au/publications/report-96 (accessed 20 October 2016).

—— 2008, *For Your Information: Australian Privacy Law and Practice*, Final Report, 108, Australian Government, Sydney, http://www.alrc.gov.au/publications/report-108 (accessed 8 April 2016).

—— 2010a, *Family Violence — A National Legal Response*, Final Report, 112, Australian Government, Sydney, 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*, Final Report, 112, Australian Government, Sydney, http://www.alrc.gov.au/publications/report-112 (accessed 20 April 2016).

—— 2014, *Serious Invasions of Privacy in the Digital Era*, FinalReport, September, 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*, Media Release, Canberra, 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*, Canberra, https://ama.com.au/submission/submission-ama-submission-pcehr-review (accessed 16 June 2016).

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.

—— 2015, *Administration of the Australian Childhood Immunisation Register*, Audit Report 46 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' Association Conference 2010, pp. 1–19, http://aha.net.au/article/harmonising-australias-water-information-reflection-on-first-steps/ (accessed 17 March 2017).

Andrews, D. 2016, *New Digital Start For Victorian Government*, Media Release, Victorian Government, Melbourne, http://www.premier.vic.gov.au/new-digital-start-for-victorian-government/ (accessed 25 July 2016).

ANDS (Australian National Data Service) 2016a, *About us*, Canberra, http://www.ands.org.au/about-us/what-we-do (accessed 16 October 2016).

—— 2016b, *Analysing bee data to protect crops*, http://www.ands.org.au/news-and-events/dataimpact/data-impact-stories/analysing-data-from-bees-to-protect-pollination (accessed 5 January 2017).

—— 2016c, *Data storage*, http://www.ands.org.au/guides/data-storage (accessed 5 October 2016).

ANZ (Australia and New Zealand Banking Group) 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.

ANZOD and ANZDATA (Australia and New Zealand Organ Donation Registry and Australia and New Zealand Dialysis and Transplant Registry) 2013, *De-identified Data Use Agreement*, ANZDATA.

APF (Australian Privacy Foundation) 2011, *The PCEHR: Checklist of Privacy Concerns (Discussion Draft)*, Canberra, 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) 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)*, 23 June, Australian Government, Sydney, http://www.apra.gov.au/CrossIndustry/Documents/150622-LTI-Public-disclosure-for-prudential-purposes-for-insurers-June-2015.pdf (accessed 4 October 2016).

ARCA (Australian Retail Credit Association) nd, *Principles of Reciprocity and Data Exchange*, Melbourne, http://www.arca.asn.au/focus/principles-of-reciprocity-data-exchange-prde.html (accessed 4 February 2016).

—— 2014, *Additional Submission to the Australian Financial System Inquiry*, September, Melbourne, http://fsi.gov.au/files/2015/03/arca.pdf (accessed 19 May 2016).

ARC, ATTICA, AVCC and DETYA (Australian Research Council, The Australian Tertiary Institutions Commercial Companies Association, The Australian Vice-Chancellors' Committee and Department of Education, Training and Youth Affairs) 2001, *National Principles of Intellectual Property Management for Publicly Funded Research*, Australian Government, http://www.arc.gov.au/

national-principles-intellectual-property-management-publicly-funded-research (accessed 15 March 2017).

ARC (Australian Research Council) 2015, *ARC Open Access Policy*, Australian Government, http://www.arc.gov.au/arc-open-access-policy (accessed 18 March 2016).

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.

Arnold, H. 2016, 'Trials indicate increasing patient confidence in My Health Record', *Australian Financial Review*, 6 November, http://www.afr.com/news/special-reports/future-of-healthcare/trials-indicate-increasing-patient-confidence-in-my-health-record-20161104-gsic1h (accessed 17 February 2017).

AusGOAL 2011, *AusGOAL — Australian Governments Open Access and Licensing Framework*, Australian Government, Canberra, http://www.ausgoal.gov.au/ (accessed 25 July 2016).

AUSTRAC (Australian Transaction Reports and Analysis Centre) 2014, *About the Anti-Money Laundering and Counter-Terrorism Financing Act 2006*, 31 October, Australian Government, http://www.austrac.gov.au/businesses/legislation/amlctf-act (accessed 17 May 2016).

—— 2016, *AUSTRAC: Australia's financial intelligence unit*, Australian Government, 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*, Australian Government, Canberra, https://submissions.education.gov.au/Forms/National-Research-Infrastructure-Capability-Issues-Paper-Submissions/Documents/Australian%20Data%20Archive.pdf (accessed 15 October 2016).

Australian Public Service Commission 2016, *APS data literacy*, http://www.apsc.gov.au/learn-and-develop/aps-data-literacy (accessed 12 January 2017).

Bajkowski, J. 2016, *NSW launches its own government digital identifier*, *Government News*, 4 February, http://www.governmentnews.com.au/2016/02/nsw-launches-its-own-government-digital-identifier/ (accessed 30 January 2017).

Baker, D. 2012, 'Match Making: Finding People Missing Out on Government Assistance', *Journal of Economic and Social Policy*, vol. 15, no. 1, http://epubs.scu.edu.au/jesp/vol15/iss1/3 (accessed 6 February 2017).

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.

Barnett, A., Campbell, M., Shield, C., Farrington, A., Hall, L., Page, K., Gardner, A., Mitchell, B. and Graves, N. 2016, 'The high costs of getting ethical and site-specific approvals for multi-centre research', *Research Integrity and Peer Review*, vol. 1, no. 16, https://researchintegrityjournal.biomedcentral.com/articles/10.1186/s41073-016-0023-6 (accessed 27 February 2017).

Basu, T. 2015, *Ashley Madison Slammed With $578 Million Lawsuit*, Time, http://time.com/4007374/ashley-madison-578-million-lawsuit-canada/ (accessed 16 March 2017).

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).

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. 2016, *Your Internet Privacy Should Be Up For Sale*, Forbes, http://www.forbes.com/sites/omribenshahar/2016/08/08/your-internet-privacy-should-be-up-for-sale/ (accessed 22 March 2017).

Berners-Lee, T. 2017, *Three challenges for the web, according to its inventor*, World Wide Web Foundation, http://webfoundation.org/2017/03/web-turns-28-letter/ (accessed 16 March 2017).

Betters, E. 2014, *Better be polite: Here's how to find out your passenger rating on Uber*, Pocket-lint, http://www.pocket-lint.com/news/129247-better-be-polite-here-s-how-to-find-out-your-passenger-rating-on-uber (accessed 30 January 2017).

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*, Tata Consultancy Services, http://www.tcs.com/resources/white_papers/Pages/Business-Value-Big-Data-Insurers.aspx (accessed 27 June 2016).

BHI (Bureau of Health Information (NSW)) 2016, *Bureau of Health Information — Home*, NSW Government, Sydney, http://www.bhi.nsw.gov.au/ (accessed 10 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.

BioGrid 2017, *About BioGrid Australia*, BioGrid, https://www.biogrid.org.au/page/3/about-us (accessed 17 March 2017).

Bista, S., Nepal, S. and Paris, C. 2013, 'The Human Touch of Government Services', presented at 21st Conference on User Modeling, Adaptation, and Personalization, Rome, Italy, 10 June, http://ceur-ws.org/Vol-997/pegov2013_paper_3.pdf (accessed 3 February 2017).

Bitcoin 2016, *Open source P2P money*, https://bitcoin.org/en/ (accessed 16 October 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.

Britt, H., Miller, G. and Henderson, J. 2015, *General practice activity in Australia 2014–15*, Family Medicine Research Centre, University of Sydney, Sydney, http://ses.library.usyd.edu.au/bitstream/2123/13765/4/9781743324530_ONLINE.pdf (accessed 9 May 2016).

——, ——, ——, Bayram, C., Harrison, C., Valenti, L., Pan, Y., Charles, J., Pollack, A., Wong, C. and Gordon, J. 2016, *General practice activity in Australia 2015–16*, Family Medicine Research Centre, University of Sydney, Sydney, https://ses.library.usyd.edu.au/bitstream/2123/15514/5/9781743325148_ONLINE.pdf (accessed 23 February 2017).

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.

Bruce, D. and Bruce, J. 2015, *Transformation Index Monitor: Baseline Report*, Australian Government Digital Transformation Office, Canberra.

Bruno, R. 2013, 'Real time monitoring of opioid prescriptions: DORA and her big brother ERRCD', presented at 2013 National Drug Trends Conference, Melbourne, 15 October.

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).

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).

Cantwell, E. and McDermott, K. 2016, *Making Technology Talk: How Interoperability Can Improve Care, Drive Efficiency, and Reduce Waste*, May, Healthcare Financial Management Association, http://medicalinteroperability.org/wp-content/uploads/2016/04/Making-Technology-Talk_HFM-reprint_May2016.pdf (accessed 11 May 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, United States.

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*, vol. 41, no. 5, 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).

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 Office (US)) 2008, *Evidence on the Costs and Benefits of Health Information Technology*, United States Congress, Washington D.C.

CDC (Centres for Disease Control and Protection) 2016, *Flu Activity & Surveillance*, National Center for Immunization and Respiratory Diseases, US Government, United States, http://www.cdc.gov/flu/weekly/fluactivitysurv.htm (accessed 19 October 2016).

CESSDA (Consortium of European Social Science Data Archives) 2016, *CESSDA — History*, https://cessda.net/About-us/History (accessed 2 February 2017).

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, 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*, Sydney, http://www.cherel.org.au/completed-projects (accessed 27 September 2016).

—— 2016b, *Master Linkage Key (MLK)*, Sydney, http://www.cherel.org.au/master-linkage-key (accessed 25 July 2016).

Chirgwin, R. 2017, *Privacy watchdog to probe Oz gov's right to release personal info 'to correct the record'*, The Register, https://www.theregister.co.uk/2017/02/28/australian_information_commissioner_to_investigate_centrelink_doxing/ (accessed 1 March 2017).

CHOICE 2014, *Submission to the Competition Policy Review Issues Paper*, Sydney, https://www.choice.com.au/consumer-advocacy/policy-submissions (accessed 20 October 2016).

Chui, M., Farrell, D. and Jackson, K. 2014, *How government can promote open data*, April, McKinsey Global Institute, Washington D.C., http://www.mckinsey.com/industries/public-sector/our-insights/how-government-can-promote-open-data (accessed 18 July 2016).

City of Ballarat nd, *Public BBQs*, http://www.ballarat.vic.gov.au/lae/parks-and-playgrounds/public-bbqs.aspx (accessed 6 March 2017).

City of New York 2013, *Mayor Bloomberg And Fire Commissioner Cassano Announce New Risk-based Fire Inspections Citywide Based On Data Mined From City Records*, http://www1.nyc.gov/office-of-the-mayor/news/163-13/mayor-bloomberg-fire-commissioner-cassano-new-risk-based-fire-inspections-citywide#/0 (accessed 17 March 2017).

CLEDS (Vic) (Commissioner for Law Enforcement Data Security (Vic)) 2013, *Social Media and Law Enforcement*, July, Victorian Government, Melbourne.

Cleeland, A. 2016, 'Farmer launches MobTracker app', *North Queensland Register*, 19 December, http://www.northqueenslandregister.com.au/story/4352970/farmer-launches-mobtracker-app/ (accessed 24 January 2017).

CLIR (Council on Library and Information Resources) 2014, *Data curation*, Washington D.C., https://www.clir.org/initiatives-partnerships/data-curation (accessed 21 October 2016).

CMA (UK) (Competition and Markets Authority (UK)) 2016a, *Retail Banking Market Investigation*, Final Report, August, UK Government, London.

—— 2016b, *Retail Banking Market Investigation: Overview*, 9 August, UK Government, London, 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)*, June, United States, https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ACO/index.html?redirect=/ACO/ (accessed 10 June 2016).

—— 2016, *Physician Quality Reporting System*, May, United States, 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*, 7 November, ACT Government, Canberra, http://www.cmd.act.gov.au/smartparking/home (accessed 9 August 2016).

COAG (Council of Australian Governments) 2013, *National Health Information Agreement*, Canberra, http://www.coaghealthcouncil.gov.au/Portals/0/130920_NHIA_revised_SCoH_FINAL_1.pdf (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).

Colineau, N., Cecile, P. and Dennett, A. 2011, 'Exploring the use of an online community in welfare transition programs', presented at the *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, http://dl.acm.org/citation.cfm?id=2305316.2305399 (accessed 3 February 2017).

Comite, F. 2015, *Some doctors DO want your Fitbit data*, 3 September, Venturebeat, http://venturebeat.com/2015/09/03/some-doctors-do-want-your-fitbit-data/ (accessed 25 May 2016).

Confluence nd, *About Phishing Attacks*, https://confluence.biola.edu/display/itservices/About+Phishing+Attacks (accessed 20 October 2016).

Conroy, P., Milano, F., Narula, A. and Singhal, R. 2014, *Building consumer trust: Protecting personal data in the consumer product industry*, Deloitte University Press.

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).

Cooper, H. 2017, *'Grave concerns' metadata could be used in civil, divorce cases*, ABC News, http://www.abc.net.au/news/2017-01-05/telco-industry-pushes-for-metadata-collection-changes/8162896 (accessed 9 March 2017).

Costa, A., Deb, A. and Kubzansky, M. 2016, *Big Data, Small Credit*, Omidyar Network, California, https://www.omidyar.com/insights/big-data-small-credit (accessed 25 May 2016).

Cowan, P. 2017a, *Qld govt jumps on data analytics bandwagon*, iTnews, http://www.itnews.com.au/news/qld-govt-jumps-on-data-analytics-bandwagon-449808 (accessed 6 February 2017).

—— 2017b, *Victoria to get its first chief data officer*, iTnews, http://www.itnews.com.au/news/victoria-to-get-its-first-chief-data-officer-447787 (accessed 1 February 2017).

Cowling, D. 2016a, *Facebook Reaches 15 Million Australian Users*, SocialMediaNews.com.au, http://www.socialmedianews.com.au/facebook-reaches-15-million-australian-users/ (accessed 17 May 2016).

—— 2016b, *Social Media Statistics Australia — November 2016*, Social Media News, https://www.socialmedianews.com.au/social-media-statistics-australia-november-2016/ (accessed 16 January 2017).

Coyne, A. 2015, *Customer data stolen in Kmart Australia hack*, iTnews, http://www.itnews.com.au/news/customer-data-stolen-in-kmart-australia-hack-409944 (accessed 14 March 2017).

CPP (Commission for the Protection of Privacy) 2015, *Recommendation no. 04/2015 of 13 May 2015*, 13 May, European Commission, Brussels.

CPSIC (Cross Portfolio Statistical Integration Committee) 2010, *High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes*, National Statistical Service, Canberra, 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*, 23 July, Brisbane, http://creativecommons.org.au/learn/licences/ (accessed 11 October 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*, Canberra, https://confluence.csiro.au/public/PROMS (accessed 20 October 2016).

—— 2016, *Privacy: Keeping data confidential*, Canberra, 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.

Data61 2016a, *Automating Data Integration with Machine Learning*, CSIRO, Canberra, https://a.confui.com/public/conferences/575eb5b495c1bfbaaf00001f/topics/57dce7ed8d3ed20182000035/slides (accessed 1 October 2016).

—— 2016b, *Data61 and Treasury to examine blockchain technology potential*, CSIRO, Canberra, http://www.csiro.au/en/News/News-releases/2016/Data61-and-Treasury-to-examine-blockchain-technology-potential (accessed 17 October 2016).

Data Linkage WA 2013, *Developmental Pathways Project*, WA Government, Perth, http://www.datalinkage-wa.org.au/projects/developmental-pathways-project (accessed 10 June 2016).

—— 2015, *WA Data Linkage Branch Access and Charging Policy*, December, Version v1.01, Perth.

—— 2016a, *About the Data Linkage Branch*, http://www.datalinkage-wa.org.au/about-data-linkage-branch (accessed 4 January 2017).

—— 2016b, *Data Collections*, WA Government, Perth, http://www.datalinkage-wa.org/data-collections (accessed 10 June 2016).

—— 2016c, *Frequently Asked Questions*, WA Government, Perth, http://www.datalinkage-wa.org.au/about-us/faq#46 (accessed 25 July 2016).

—— 2016d, *Glossary*, WA Government, Perth, http://www.datalinkage-wa.org.au/data-linkage/glossary (accessed 21 October 2016).

—— 2016e, *Projects*, WA Government, Perth, http://www.datalinkage-wa.org.au/projects (accessed 27 September 2016).

data.gov.au 2017a, *Ballarat BBQs*, https://data.gov.au/dataset/ballaratbbq (accessed 6 March 2017).

—— 2017b, *Data.gov.au Site Statistics*, Australian Government, Canberra, http://data.gov.au/stats#res-by-org (accessed 25 January 2017).

Data.NSW nd, *About the Information Asset Register*, NSW Government, Sydney, http://data.nsw.gov.au/iar/pages/about-the-iar (accessed 20 October 2016).

Data.Qld 2015, *1996-2000—Bowen tide gauge archived interval recordings*, Queensland Government, Brisbane, https://data.qld.gov.au/dataset/bowen-tide-gauge-archived-interval-recordings/resource/dd6d88b4-ae2b-42de-916a-b1417feea14e (accessed 16 October 2016).

Davenport, T.H. and Dyché, J. 2013, *Big Data in Big Companies*, May, International Institute for Analytics, 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).

Davis, W. 2016, *Web Users Agree To Anything Online, Study Finds*, MediaPost, http://www.mediapost.com/publications/article/280139/web-users-agree-to-anything-online-study-finds.html (accessed 16 January 2017).

DBIS (UK) (Department for Business, Innovation & Skills (UK)) 2011, *The midata vision of consumer empowerment*, 3 November, UK Government, London, https://www.gov.uk/government/news/the-midata-vision-of-consumer-empowerment (accessed 10 March 2016).

—— 2014, *Review of the midata Voluntary Programme*, UK Government, London, 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) 2016, *Smart Meter Compatible Web Portals*, Victorian Government, Melbourne, http://www.smartmeters.vic.gov.au/interactive-devices/web-portals (accessed 29 March 2016).

Deloitte 2008, *National eHealth Strategy*, Report commissioned by the Australian Government Department of Health, Sydney.

—— 2013, *Market Assessment of Public Sector Information*, Report commissioned by the Department for Business, Innovation and Skills and Cabinet Office, UK Government, May, London, 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*, Report commissioned by the Australian Government Department of Health, Sydney.

—— 2015, *Mobile Consumer Survey 2015 — the Australian Cut*, Deloitte Touche Tohmatsu, Sydney.

Department of Communications 2014, *Cloud computing and privacy*, https://www.communications.gov.au/sites/g/files/net301/f/2014-112101-CLOUD-Consumer-factsheet.pdf (accessed 10 February 2017).

Department of Education (NT) 2016, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*, NT Government, Darwin.

Department of Finance 2013, *Big Data Strategy: Issues Paper*, Australian Government, Canberra.

—— 2014a, *Australian Government Cloud Computing Policy*, October, Australian Government, Canberra, http://www.finance.gov.au/cloud/ (accessed 23 March 2016).

—— 2014b, *Australian Government Cost Recovery Guidelines, Resource Management Guide No. 304 – Third Edition*, Australian Government, Canberra, https://www.finance.gov.au/sites/default/files/australian-government-cost-recovery-guidelines.pdf (accessed 1 October 2016).

—— 2015a, *Charging for Data Services Information Sheet*, Australian Government, Canberra, https://www.finance.gov.au/sites/default/files/charging-for-data-services.docx (accessed 1 October 2016).

—— 2015b, *The independent Review of Whole-of-Government Internal Regulation (Belcher Red Tape Review)*, http://www.finance.gov.au/publications/reducingredtape/ (accessed 16 January 2017).

—— 2016a, *Australian Securities and Investments Commission (ASIC) Registry – FAQs*, May, Australian Government, Canberra, https://finance.gov.au/procurement/scoping-studies/asic-faqs/ (accessed 4 July 2016).

—— 2016b, *Fedlink*, Text, Australian Government, Canberra, http://www.finance.gov.au/collaboration-services-skills/fedlink/ (accessed 5 October 2016).

Department of Health 2014, *Primary Health Care Research, Evaluation and Development (PHCRED) Strategy*, Australian Government, Canberra, http://www.health.gov.au/internet/main/publishing.nsf/Content/pcd-programs-phcred (accessed 9 June 2016).

—— 2015a, *Data Access and Release Policy*, Australian Government, Canberra, 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*, Australian Government, Canberra, 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*, 3 February, Australian Government, Canberra, http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Health/Health/Additional_Documents (accessed 30 May 2016).

—— 2016b, *Data update*, 29 September, Australian Government, Canberra, 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*, 29 March, Australian Government, Canberra, 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) - Data.gov.au*, Australian Government, Canberra, https://data.gov.au/dataset/mbs-sample-10pct-1984-gz (accessed 2 August 2016).

—— 2016e, *MyHealth Record — What you should know before you get one*, Australian Government, Canberra, 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*, Australian Government, Canberra, http://www.health.gov.au/internet/main/publishing.nsf/Content/mr-yr16-dept-dept002.htm (accessed 10 June 2016).

Department of Health and Ageing 2012, *Fact Sheet: Electronic Recording and Reporting of Controlled Drugs*, Australian Government, Canberra.

Department of Health and HealthConsult 2016, *My Health Record Data — Public Consultation*, http://www.myhealthrecorddata.healthconsult.com.au/public-consultations/ (accessed 21 February 2017).

Department of Health (NSW) 2005, *Intellectual Property Arising from Health Research — Policy*, NSW Government, http://www1.health.nsw.gov.au/PDS/pages/doc.aspx?dn=PD2005_370 (accessed 15 March 2017).

Department of Health (WA) 2016, *WA Public Hospital Activity: ED*, WA Government, Perth, http://www.health.wa.gov.au/emergencyactivity/edsv/ (accessed 24 August 2016).

Desai, T., Ritchie, F. and Welpton, R. 2016, *Five Safes: designing data access for research*, Economics Working Paper Series, University of the West of England, United Kingdom.

DFSI (NSW) (Department of Finance, Services and Innovation (NSW)) nd, *Web Services: Land and Property Information*, NSW Government, Sydney, http://spatialservices.finance.nsw.gov.au/mapping_and_imagery/lpi_web_services (accessed 16 October 2016).

—— 2015, *Digital+ 2016 - NSW Government ICT Strategy*, NSW Government, Sydney, https://www.finance.nsw.gov.au/ict/sites/default/files/resources/Digital_Strategy_2016_20151125.pdf (accessed 14 July 2016).

—— 2016a, *NSW Data Analytics Centre*, NSW Government, Sydney, https://www.finance.nsw.gov.au/nsw-data-analytics-centre (accessed 25 July 2016).

—— 2016b, *NSW Government Open Data Policy*, NSW Government, Sydney.

DHHS (US) (Department of Health and Human Services (US)) nd, *Consumer Health Data Aggregator Challenge*, US Government, Washington D.C., https://www.challenge.gov/challenge/consumer-health-data-aggregator-challenge/ (accessed 10 June 2016).

—— 2016a, *HHS announces major commitments from healthcare industry to make electronic health records work better for patients and providers*, February, US Government, Washington D.C., 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).

—— 2016b, *Unlocking data*, April, US Government, Washington D.C., 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*, Victorian Government, Melbourne, https://www2.health.vic.gov.au:443/about/publications/researchandreports/ministerial-review-of-victorian-health-sector-information-and-communication-technology (accessed 20 May 2016).

—— 2015, *Streamlining ethical review*, https://www2.health.vic.gov.au/about/clinical-trials-and-research/health-and-medical-research/streamlining-ethical-review (accessed 27 February 2017).

DHS (Department of Human Services) 2015, *Annual Report 2014-15*, Australian Government, Canberra, https://www.humanservices.gov.au/corporate/annual-reports/annual-report-2014-15 (accessed 10 June 2016).

—— 2016, *Public Key Infrastructure*, Australian Government, Canberra, 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, McKinsey Global Institute, Brussels, http://www.mckinsey.com/industries/financial-services/our-insights/cutting-through-the-noise-around-financial-technology (accessed 3 October 2016).

Directivity, Citrus and First Point Research 2015, *For love or money: 2015 consumer study into Australian loyalty programs*, Melbourne, http://www.theloyaltypoint.com.au/for-love-or-money-2015 (accessed 25 July 2016).

DJAG (Qld) (Department of Justice and Attorney-General (Qld)) 2016, *2016 Consultation on the Review of the Right to Information Act 2009 and Information Privacy Act 2009*, Consultation Paper, Queensland Government, Brisbane.

Dobbin, M. 2014, 'Pharmaceutical drug misuse in Australia', *Australian Prescriber*, vol. 37, no. 3, pp. 79–81.

Dobbs, R., Manyika, J. and Woetzel, J. 2015, *No Ordinary Disruption: The Four Global Forces Breaking All the Trends*, PublicAffairs, New York.

DoseMe nd, *DoseMe - World's First Precision Dosing Software for Clinical Practice*, DoseMe, https://doseme.com.au/introducing-doseme (accessed 27 February 2017).

DPC (SA) (Department of Premier and Cabinet (SA)) 2013, *Open Data Declaration*, September, SA Government, Adelaide, http://digital.sa.gov.au/resources/topic/open-data/open-data-declaration (accessed 20 October 2016).

DPC (SA) (Department of Premier and Cabinet (SA)) 2016, *SA Government Digital Achievements*, Office for Digital Government, https://digital.sa.gov.au/resources/topic/digital-government/sa-government-digital-achievements (accessed 30 January 2017).

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 (Vic) (Department of Premier and Cabinet (Vic)) 2016, *Information Technology Strategy for the Victorian Government: 2016-2020*, Victorian Government, Melbourne.

DPC (WA) (Department of Premier and Cabinet (WA)) 2016, *Data Linkage Review*, May, WA Government, Perth, https://www.dpc.wa.gov.au/Consultation/Pages/Data-Linkage-Review.aspx (accessed 30 May 2016).

DPMC (Department of Prime Minister and Cabinet) 2015a, *Open Government Partnership: Public consultation for the Australian Government's National Action Plan now open*, Text, 17 November, Australian Government, Canberra, https://www.dpmc.gov.au/news-centre/government/open-government-partnership-public-consultation-australian-government%E2%80%99s-national-action-plan-now-open (accessed 27 September 2016).

—— 2015b, *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, *Australia's First Open Government National Action Plan 2016-18*, http://ogpau.pmc.gov.au/2016/12/07/australias-first-national-action-plan-submitted (accessed 6 January 2017).

—— 2016b, *Data Skills and Capability in the Australian Public Service*, Australian Government, Canberra.

—— 2016c, *Guidance on Data Sharing for Australian Government Entities*, Australian Government, Canberra.

—— 2016d, *High-Value Data Roundtables Commence*, Department of Prime Minister and Cabinet, https://www.dpmc.gov.au/news-centre/public-data/high-value-data-roundtables-commence (accessed 1 February 2017).

—— 2016e, *Smart Cities Plan*, 29 April, Australian Government, Canberra.

DTF (Vic) (Department of Treasury and Finance (Vic)) 2012, *DataVic access policy*, August, Victorian Government, Melbourne, 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*, August, Victorian Government, Melbourne, 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) 2015, *Open data*, 21 July, Australian Government, Canberra.

—— 2016, *Digital Identity - early days in the Discovery process*, Australian Government, Canberra, https://www.dto.gov.au/blog/digital-identity-early-days-in-the-discovery-process/ (accessed 10 October 2016).

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).

Easton, S. 2016, *Australia expands Open Government plan, readies for international scrutiny*, The Mandarin, http://www.themandarin.com.au/73454-australia-expands-submits-open-government-action-plan/ (accessed 11 January 2017).

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.

eHealthNT 2011, *The Origins of eHealthNT*, 5 January, NT Government, http://ehealthnt.nt.gov.au/About_Us/Origins_of_eHealthNT/index.aspx (accessed 24 May 2016).

EHSC (WA) (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, WA Government, 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).

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*, Australian Government, Canberra, 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*, Brussels.

—— 2015, *Questions and Answers - Data protection reform*, Press release, 21 December, Brussels, 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, Brussels, 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*, Brussels, 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).

European Commission 2016d, *Reform of EU data Protection Rules*, Brussels, http://ec.europa.eu/justice/data-protection/reform/index_en.htm (accessed 18 February 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).

Eyers, J. 2016, 'Banks planning big data deals to target customers', *Australian Financial Review*, http://www.afr.com/technology/banks-planning-big-data-deals-to-target-customers-20160521-gp0pnq (accessed 22 June 2016).

Facebook 2015, *Terms of Service*, https://www.facebook.com/terms (accessed 9 March 2017).

—— 2016a, *About Social Plugins*, https://www.facebook.com/help/443483272359009/ (accessed 17 March 2017).

—— 2016b, *Cookies & Other Storage Technologies*, California, https://www.facebook.com/policies/cookies/# (accessed 30 May 2016).

—— 2016c, *Data Policy*, https://www.facebook.com/policy.php (accessed 15 March 2017).

—— 2016d, *Facebook Q1 2016 results*, California, http://investor.fb.com/results.cfm (accessed 17 May 2016).

—— 2016e, *Suicide Prevention*, California, https://www.facebook.com/help/suicideprevention (accessed 30 August 2016).

—— 2017, *Facebook Annual Report*, 2 February, Facebook, California, https://investor.fb.com/financials/?section=secfilings (accessed 15 February 2017).

Fidor nd, *Fidor API Reference*, http://docs.fidor.de/#introduction (accessed 8 July 2016).

Fielding, R. 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*, Report prepared for the Australian Government Treasury, February, Sydney, http://www.fintechaustralia.org.au/single-post/2016/01/28/FINTECH-AUSTRALIA-SUBMISSION-FinTech-Australia-Recommendations-for-Reform-of-the-Australian-Financial-Services (accessed 15 May 2016).

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)*, University of Sydney, Sydney, http://sydney.edu.au/medicine/fmrc/beach/index.php (accessed 9 June 2016).

Forsyth, L., Armytage, P. and Lawrence, R. 2016, *Unlocking the Future: Maranguka Justice Reinvestment Project in Bourke (Preliminary Assessment)*, September.

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, US Government, Washington D.C.

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).

Garfinkel, S.L. 2015, *De-Identification of Personal Information*, NISTIR 8053, October, National Institute of Standards and Technology, U.S. Department of Commerce, Washington D.C.

Garmin 2016, *Privacy Statement*, 11 January, http://www.garmin.com/en-AU/legal/privacy-statement (accessed 23 May 2016).

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*, Australian Government, Canberra, http://www.sydneycoastalcouncils.com.au/sites/default/files/coovermar_ga_slides.pdf (accessed 16 September 2016).

Gleit, N., Zeng, S. and Cottle, P. 2014, *Introducing Safety Check*, Facebook, California, 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).

Gluckman, P. 2015, *The intended and unintended consequences of e-research: why scientists must engage openly with the community*, E--Research 2020 Workshop, office of the prime minister's chief science advisor, Wellington, New Zealand.

Google nd, *Google Flu Trends*, https://www.google.org/flutrends/about/ (accessed 30 August 2016).

—— 2016, *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).

Government 2.0 Taskforce 2009, *Engage: Getting on with Government 2.0*, Commissioned by the Department of Finance, Australian Government, Canberra, http://www.finance.gov.au/archive/publications/gov20taskforcereport/ (accessed 7 April 2016).

GOV.UK 2010, *Letter to government departments on opening up data*, UK Government, London, https://www.gov.uk/government/news/letter-to-government-departments-on-opening-up-data (accessed 16 June 2016).

Gregor, S. and Lee-Archer, B. 2016, 'The digital nudge in social security administration', *International Social Security Review*, vol. 69, pp. 63–83.

Grieve, G. 2014, 'Health Intersections — Architectural Approaches for Exchanging Information', http://www.slideshare.net/informaoz/grahame-grieve-health-intersections (accessed 17 February 2017).

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*, IoTHub, Sydney, 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).

Haggan, M. 2017, *WA moves to cut doctor shopping*, AJP — Australian Journal of Pharmacy, https://ajp.com.au/news/wa-moves-cut-doctor-shopping/ (accessed 28 February 2017).

Hamilton, B. and O'Dowd, K. 2017, *It's nothing personal… Federal Court finds that 'personal information' must be information 'about an individual'*, Hall & Wilcox, http://www.hallandwilcox.com.au/its-nothing-personal-federal-court-finds-that-personal-information-must-be-information-about-an-individual/ (accessed 1 February 2017).

Harper, I., Anderson, P., McCluskey, S. and O'Bryan, M. 2015, *Competition Policy Review Final Report (Harper Review)*, 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, Australian Government, Canberra, https://www.ag.gov.au/consultations/pages/reviewoffoilaws.aspx (accessed 15 September 2016).

Head, B. 2016, 'Woodside retains corporate memory using cognitive computing', *Australian 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).

Health Stats 2016, *Health Stats NSW*, NSW Government, Sydney, http://www.healthstats.nsw.gov.au/ (accessed 10 June 2016).

Healthcare Gateway 2013, *Urgent care record sharing in Cumbria*, http://www.healthcaregateway.co.uk/case-studies/urgent-care-record-sharing-in-cumbria (accessed 3 June 2016).

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, Victoria.

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, Washington D.C.

Heydon, G. and Zeichner, F. 2015, *Enabling the Internet of Things for Australia*, October, Communications Alliance, Sydney.

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*, 29 July, 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*, University of Western Australia, Perth, 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, B. 2011, *Citizens' engagement in policymaking and the design of public services*, Research Paper, 1, 2011–12, Parliament of Australia, Department of Parliamentary Services, http://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp1112/12rp01 (accessed 9 January 2017).

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, Melbourne.

House of Commons Committee of Public Accounts 2013, *The dismantled National Programme for IT in the NHS*, 19, Session 2013-14, UK Government, London.

House of Representatives Standing Committee on Economics 2016, *Review of the Four Major Banks (Second Report)*, text, 15 September, Canberra, http://www.aph.gov.au/Parliamentary_Business/Committees/House/Economics/FourMajorBanksReview2 (accessed 3 March 2017).

House of Representatives Standing Committee on Health 2016, *Report on the Inquiry into Chronic Disease Prevention and Management in Primary Health Care*, 5 October, Australian Government, Canberra, 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).

Iansiti, M. and Lakhani, K.R. 2014, 'Digital Ubiquity: How Connections, Sensors, and Data Are Revolutionizing Business', *Harvard Business Review*, 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) 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 2016a, *Declaration on Open and Transparent Government*, 20 September, NZ Government, Wellington, New Zealand.

—— 2016b, *New Zealand Government Open Access and Licensing framework (Version 2)*, NZ Government, Wellington, https://www.ict.govt.nz/guidance-and-resources/open-government/new-zealand-government-open-access-and-licensing-nzgoal-framework/nzgoal2/ (accessed 20 October 2016).

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).

Iedema, R., Allen, S., Britton, K. and Hor, S. 2013, 'Out of the frying pan? Streamlining the ethics review process of multisite qualitative research projects', *Australian Health Review*, vol. 37, no. 2, pp. 137–139.

Infrastructure Australia 2016, *Australian Infrastructure Plan*, February, Australian Government, Canberra.

Involving People in Research 2017, *Involving People in Research - Consumer & Community Health Research Network*, http://www.involvingpeopleinresearch.org.au/find-out-more/our-resources/the-barriers-report (accessed 9 January 2017).

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).

Jackson, M. 2016, *New UK Digital Economy Bill Details Broadband USO and Internet Changes*, ISP Review UK, http://www.ispreview.co.uk/index.php/2016/07/new-uk-digital-economy-bill-details-broadband-uso-internet-changes.html (accessed 17 March 2017).

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).

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: the Finsia Journal of Applied Finance*, no. 3, p. 44.

Johnston, A. 2017, *Mobiles, metadata and the meaning of 'personal information'*, Salinger Privacy, http://www.salingerprivacy.com.au/2017/01/19/federalcourtdecision/ (accessed 1 February 2017).

Jolly, R. 2011, *The eHealth revolution — easier said than done*, Department of Parliamentary Services Research Papers 2011–12, 3, Social Policy Section, Australian Government, Canberra.

Jones, K.H., Laurie, G., Stevens, L., Dobbs, C., Ford, D.V. and Lea, N. 2014, 'The other side of the coin: Harm due to the non-use of health-related data', *International Journal of Medical Informatics*, vol. 97, pp. 43–51.

Kantor, L. and Bhunia, P. 2016, *High Value Open Data in Australia — Quality, Availability and Use*, OpenGov Asia, Singapore, 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*, Brookings Institution, Washington D.C., 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.

Kirk, J. 2015, *Premera, Anthem data breaches linked by similar hacking tactics*, Computerworld, http://www.computerworld.com/article/2898419/data-breach/premera-anthem-data-breaches-linked-by-similar-hacking-tactics.html (accessed 14 March 2017).

Kitchin, R. 2014, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, Sage, London.

KPMG 2016, *Building trust in analytics*, https://home.kpmg.com/au/en/home/insights/2016/10/building-trust-in-analytics.html (accessed 9 January 2017).

KPMG International 2016, *Creepy or Cool? Staying on the Right Side of Consumer Privacy*, Amsterdam, The Netherlands.

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 (NSW) 2016, *Information brokers*, NSW Government, Sydney, http://www.lpi.nsw.gov.au/land_titles/information_brokers (accessed 16 October 2016).

Lane, J., Stodden, V., Bender, S. and Nissenbaum, H. 2014, *Privacy, Big Data, and the Public Good*, Cambridge University Press, United Kingdom.

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.

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).

Ley, S. 2015, *Patients to get new myHealth Record: $485m 'rescue' package to reboot Labor's e-health failures*, Media Release, 10 May, Australian Government Department of Health, Canberra.

—— 2016, *My Health Record gets one million more reasons to sign up*, 4 March, Australian Government Department of Health, Canberra.

—— and Bailey, S. 2016, *PSD2 opens the door to new market entrants*, Deloitte Touche Tohmatsu, Sydney, 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*, Working paper, May.

Lockstep Consulting nd, *Privacy Impact Assessment Report: Advanced Metering Infrastructure (AMI)*, Report for the Department of Economic Development, Jobs, Transport and Resources, Victorian Government, Melbourne.

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, London.

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.

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).

MacGibbon, A. 2016, *Review of the Events Surrounding the 2016 eCensus*, 13 October, Department of the Prime Minister and Cabinet, http://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;query=Id%3A%22publications%2Ftabledpapers%2Fa41f4f25-a08e-49a7-9b5f-d2c8af94f5c5%22 (accessed 11 January 2017).

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).

Mandel, R. and Noyes, E. 2013, 'Beyond the NWS: Inside the Thriving Private Weather Forecasting Industry', *Weatherwise*, vol. 66, no. 1, pp. 12–19.

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, McKinsey Global Institute, Brussels, 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.

McArthur, M., Thomson, L., Winkworth, G. and Butler, K. 2010, *Families' experiences of services*, Occasional Paper, 30, Department of Families, Housing, Community Services and Indigenous Affairs, Canberra.

McCandless, D. 2017, *World's Biggest Data Breaches & Hacks*, Information is Beautiful, http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/ (accessed 7 March 2017).

McClure, P., Sinclair, S. and Aird 2015, *A New System for Better Employment and Social Outcomes*, Report of the Reference Group on Welfare Reform to the Minister for Social Services, February, https://www.dss.gov.au/review-of-australias-welfare-system (accessed 16 January 2017).

McCoach, L. and Landy, D. 2014, 'Australia', *The Banking Regulation Review*, Fifth Edition, Law Business Research, Sydney.

McColl, R. 2010, *Freedom of Information — A New Paradigm (The 2010 Whitmore Lecture)*, Council of Australasian Tribunals (NSW Chapter), Sydney.

McDonald, K. 2012, 'Real-time access to controlled drugs data from July', *Pulse+IT*, 20 February, 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*, 3 December, 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*, 20 March, 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, 'Why is eHealth interoperability so hard?', *Pulse+IT*, 5 February, http://www.pulseitmagazine.com.au/component/content/article?id=2271:why-is-ehealth-interoperability-so-hard (accessed 6 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).

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).

McLeods Barristers & Solicitors 2014, *Deed of agreement for data sharing*, 15 December, Perth.

Medibank Private 2015, *Medibank Privacy Policy*, May, Melbourne.

Melbourne Institute nd, *Organisations with Organisational Deed of Licence*, University of Melbourne, Melbourne, 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, http://dx.doi.org/10.6028/NIST.SP.800-145 (accessed 17 March 2017).

Meyer, M., Niech, C. and Eggers, W.D. 2015, *Anticipate, sense, and respond: Connected government and the Internet of Things*, Deloitte Touche Tohmatsu, Sydney, http://dupress.deloitte.com/dup-us-en/focus/internet-of-things/iot-in-government.html (accessed 12 September 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.0001ioe8an8f0fnwyfm2cvyrpix0x (accessed 12 September 2016).

MinterEllison 2015, *Privacy Impact Assessment Report: PCEHR System Opt-Out Model*, 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, D. and Niemi, M. 2016, *The Sharing of Personal Health Data – A Review of the Literature*, Report prepared for the Data Futures Partnership, 29 June, http://datafutures.co.nz/assets/Uploads/The-Sharing-of-Personal-Health-Data-Sapere-FINAL.pdf (accessed 6 January 2017).

Morey, T., Forbath, T. and Schoop, A. 2015, 'Customer Data: Designing for Transparency and Trust', *Harvard Business Review*, May, https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust (accessed 4 June 2016).

Morgan, L. 2016, *List of data breaches and cyber attacks in 2016 – 3.1 billion records leaked*, *IT Governance Blog*, 12 December, https://www.itgovernance.co.uk/blog/list-of-data-breaches-and-cyber-attacks-in-2016-1-6-billion-records-leaked/ (accessed 13 March 2017).

Murray, D., Davis, K., Hewson, C. and McNamee, B. 2014, *Financial System Inquiry: Final Report (Murray Inquiry)*, 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.

NAA (National Archives of Australia) nd, *Normal administrative practice*, Australian Government, Canberra, http://www.naa.gov.au/records-management/agency/keep-destroy-transfer/nap/index.aspx (accessed 20 October 2016).

—— 2015a, *Digital Continuity 2020 Policy*, October, Australian Government, Canberra, 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*, May, Australian Government, Canberra, 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*, Australian Government, Canberra, 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 Catholic Education Commission 2016, *Subsmission to the National Education Evidence Base Inquiry*, http://www.pc.gov.au/__data/assets/pdf_file/0008/199682/sub049-education-evidence.pdf (accessed 20 March 2017).

National Weather Service 1991, *Policy Statement on the Weather Service - Private Sector Roles*, US Government, Washington D.C., 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).

NEHTA (National Electronic Health Transition Authority) 2016, *Features of the My Health Record system: Clinical Documents*, Australian Government, Canberra, 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)*, Australian Government, Canberra, 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*, *Smart Data Collective*, 28 June, http://www.smartdatacollective.com/michelenemschoff/206391/quick-guide-structured-and-unstructured-data (accessed 10 May 2015).

NHHRC (National Health and Hospitals Reform Commission) 2009, *A Healthier Future For All Australians*, Final Report, NHHRC, Canberra.

——, 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*, Australian Government, Canberra, 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*, Australian Government, Canberra, https://www.nhmrc.gov.au/guidelines-publications/r39 (accessed 20 October 2016).

NHMRC (National Health and Medical Research Council) 2009, *Position Statement: Monitoring and Reporting of Safety for Clinical Trials*, Australian Government, https://www.nhmrc.gov.au/guidelines-publications/e112 (accessed 15 March 2017).

—— 2012, *The National Certification Scheme of Institutional Processes related to the Ethical Review of Multi-Centre Research*, Australian Government, Canberra, https://hrep.nhmrc.gov.au/certification? (accessed 18 July 2016).

—— 2014a, *Guidelines approved under Section 95A of the Privacy Act 1988*, Australian Government, Canberra, https://www.nhmrc.gov.au/guidelines-publications/pr2 (accessed 1 June 2016).

—— 2014b, *NHMRC Open Access Policy*, 9 December, Australian Government, Canberra, http://www.nhmrc.gov.au/grants-funding/policy/nhmrc-open-access-policy (accessed 4 April 2016).

—— 2014c, *NHMRC's Policy on the Dissemination of Research Findings*, Australian Government, Canberra, http://www.nhmrc.gov.au/grants-funding/policy/nhmrc-open-access-policy (accessed 21 March 2016).

—— 2015, *NHMRC Funding Rules 2015*, Australian Government, Canberra.

—— 2016a, *Good Practice Process for Site Assessment and Authorisation Phases of Clinical Trial Research Governance*, https://www.nhmrc.gov.au/research/clinical-trials/development-good-practice-process-site-assessment-and-authorisation-clinica (accessed 28 February 2017).

—— 2016b, *Human Research Ethics Committees*, Australian Government, Canberra, https://hrep.nhmrc.gov.au/certification/hrecs (accessed 18 July 2016).

—— 2016c, *National Approach to Single Ethical Review*, Australian Government, Canberra.

NHPA (National Health Performance Authority) 2015a, *About us*, Australian Government, Canberra, http://www.nhpa.gov.au/internet/nhpa/publishing.nsf/Content/About-us (accessed 10 June 2016).

—— 2015b, *National Health Performance Authority — Data Plans*, Australian Government, Canberra, 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*, 20 May, UK Government, London, 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).

NICTA (National Information and Communication Technology Australia) 2014, *New models for Digital Government: The role of service brokers in driving innovation*, http://apo.org.au/node/42458 (accessed 2 March 2017).

Nike 2016, *Nike Sustainable Business Report*, http://www.nikeresponsibility.com/report/ (accessed 9 March 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, *Deterministic linking and linkage keys*, Data Linking Information Sheet Three, http://www.nss.gov.au/nss/home.nsf/pages/Data%20integration%20-%20data%20linking%20information%20sheet%20three (accessed 15 February 2017).

—— 2009, *A Good Practice Guide to Sharing Your Data With Others*, 1 November, Australian Government, Canberra, http://www.nss.gov.au/nss/home.nsf/NSS/E6C05AE57C80D737CA25761D002FD676?opendocument (accessed 10 March 2016).

—— 2010, *High level principles for data integration involving Commonwealth data for statistical and research purposes*, Australian Government, Canberra.

—— 2013, *Risk assessment guidelines*, December, Statistical Data Integration Involving Commonwealth Data, Australian Government, Canberra.

—— 2015, *Rights, responsibilities and roles of data custodians*, Australian Government, Canberra, http://www.nss.gov.au/nss/home.NSF/533222ebfd5ac03aca257110000 44c9e/59fd060543b4e9e0ca257a4e001eacfe/$FILE/Rights,%20responsibilities%20and %20roles%20of%20data%20custodians_Dec2013.pdf (accessed 18 July 2016).

—— 2016a, *Accreditation*, A Guide for Data Integration Projects Involving Commonwealth Data for Statistical and Research Purposes, Australian Government, Canberra, https://statistical-data-integration.govspace.gov.au/ topics/accreditation/ (accessed 10 August 2016).

—— 2016b, *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).

—— 2016c, *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).

—— 2016d, *Scope of the Commonwealth Arrangements*, https://statistical-data-integration.govspace.gov.au/topics/scope-of-the-commonwealth-arrangements/ (accessed 20 October 2016).

—— 2016e, *The separation principle*, Australian Government, Canberra, http://statistical-data-integration.govspace.gov.au/topics/applying-the-separation-principle/ (accessed 22 July 2016).

—— 2016f, *What is statistical data integration?*, A guide for data integration projects involving Commonwealth data for statistical and research purposes, Australian Government, Canberra.

NSW Government 2016, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*, Sydney.

NZDFF (New Zealand Data Futures Forum) 2014, *Harnessing the economic and social power of data*, NZ Government, Wellington, https://www.nzdatafutures.org.nz/sites/default/files/NZDFF_harness-the-power.pdf (accessed 1 June 2016).

OAIC (Office of the Australian Information Commissioner) 2011a, *Freedom of information — The information publication scheme for Australian Government agencies*, Australian Government, Sydney, 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*, Australian Government, Sydney.

—— 2013a, *Community Attitudes to Privacy Report: 2013*, Australian Government, Sydney.

—— 2013b, *Open Public Sector Information: From Principles to Practice*, Australian Government, Sydney, 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, Australian Government, Sydney.

—— 2014b, *Data breach notification guide: A guide to handling personal information security breaches*, August, Australian Government, Sydney.

—— 2014c, *Privacy business resource 4: De-identification of data and information*, Australian Government, Sydney.

—— 2014d, *Privacy fact sheet 38: Hardship assistance and your credit report*, May, Australian Government, Sydney, 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*, May, Australian Government, Sydney, 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*, Australian Government, Sydney.

—— 2015b, *Australian Privacy Principles Guidelines*, 1 April, Australian Government, Sydney, 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*, Australian Government, Sydney.

—— 2016a, *Determinations*, Australian Government, Sydney, https://www.oaic.gov.au/privacy-law/determinations/ (accessed 16 October 2016).

—— 2016b, *Does my business have privacy obligations in relation to consumer credit reporting under the Privacy Act?*, Australian Government, Sydney, 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).

—— 2016c, *Government data-matching*, Australian Government, Sydney, https://www.oaic.gov.au/privacy-law/other-legislation/government-data-matching (accessed 4 July 2016).

—— 2016d, *Privacy Act*, Australian Government, Sydney, https://www.oaic.gov.au/privacy-law/privacy-act/ (accessed 31 March 2016).

—— 2016e, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*, Australian Government, Sydney, http://www.pc.gov.au/inquiries/current/education-evidence/submissions (accessed 4 July 2016).

—— 2016f, *What is health information?*, Australian Government, Sydney, 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 social networking services*, 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 (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.

ODI (Open Data Institute) 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, Paris, http://dx.doi.org/10.1787/9789264193505-en (accessed 12 May 2016).

—— 2014, *OECD Public Governance Reviews Open Government in Latin America*, OECD Publishing, Paris.

—— 2015a, *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris.

—— 2015b, *OECD Reviews of Health Care Quality: Australia 2015*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264233836-en (accessed 18 May 2016).

—— 2015c, 'Open government data', *Government at a Glance*, Organisation for Economic Co-operation and Development, Paris, pp. 150–151, http://www.oecd-ilibrary.org/content/chapter/gov_glance-2015-48-en (accessed 27 September 2016).

OGCIO (WA) (Office of the Government Chief Information Officer (WA)) 2015, *Whole of Government Open Data Policy*, 3 July, WA Government, Perth,

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) 2017, *About the Open Government Partnership*, Open Government Partnership, http://www.opengovpartnership.org/about (accessed 16 September 2016).

OIC (Qld) (Office of the Information Commissioner – QLD) 2012, *IP addresses, Google Analytics and the privacy principles*, Queensland Government, Brisbane, 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, Queensland Government, Brisbane, 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 (Office of the Privacy Commissioner) 2008, *Privacy Guidelines for the Medicare Benefits and Pharmaceutical Benefits Programs*, Australian Government, Sydney, https://www.legislation.gov.au/Details/F2008L00706 (accessed 10 June 2016).

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, Canadian Government, Gatineau, Quebec.

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).

Palmer, D. 2015, *Australia recommits to Open Government Partnership*, *Delimiter*, 18 November, https://delimiter.com.au/2015/11/18/australia-recommits-open-government-partnership/ (accessed 3 February 2017).

Parkin, E. 2016, *A paperless NHS: electronic health records*, House of Commons Briefing Paper 07572, 25 April, UK Government, London.

Parkopedia nd, *About Parkopedia*, http://au.parkopedia.com/about-us/ (accessed 19 July 2016).

Parliament of Australia 2015, *Health Legislation Amendment (eHealth) Bill 2015*, Explanatory Memorandum, 17 September, Australian Government, Canberra.

—— 2016, *Whistleblower protections in the corporate, public and not-for-profit sectors*, http://www.aph.gov.au/Parliamentary_Business/Committees/Joint/Corporations_and_Financial_Services/WhistleblowerProtections (accessed 23 February 2017).

—— 2017, *Parliamentary Business: Privacy Amendment (Re-identification Offence) Bill 2016*, http://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=s1047 (accessed 17 February 2017).

Parliament of South Australia 2016, *Public Sector (Data Sharing) Bill 2016*, 4 August, 146, SA Government, Adelaide.

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*, Australian Government, Canberra.

PC (Productivity Commission) 2001, *Cost Recovery by Government Agencies*, Report 15, 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, *Mineral and Energy Resource Exploration*, Australian Government, Canberra.

—— 2013c, *Regulator Engagement with Small Business*, Australian Government, Canberra.

—— 2013d, 'Using administrative data to achieve better policy outcomes', *Annual Report 2012–13*, Australian Government, Canberra.

—— 2014a, *Childcare and Early Childhood Learning*, Report no. 73, Canberra.

—— 2014b, *Natural Disaster Funding Arrangements*, Report no. 74, Australian Government, Canberra.

—— 2014c, *Public Infrastructure*, Report no. 71, Australian Government, Canberra.

—— 2015a, *Business Set-up, Transfer and Closure*, Report no. 75, Canberra.

—— 2015b, *Efficiency in Health*, Commission Research Paper, Canberra.

—— 2015c, *Housing Assistance and Employment in Australia*, Commission Research Paper, Canberra.

—— 2016a, *Data Availability and Use: Draft Report*, Australian Government, Canberra.

—— 2016b, *Digital Disruption: What do governments need to do?*, Commission Research Paper, Australian Government, Canberra.

—— 2016c, *Intellectual Property Arrangements: Draft Report*, Australian Government, Canberra.

—— 2016d, *National Education Evidence Base: Draft Report*, Australian Government, Canberra.

—— 2017, *Report on Government Services 2017: Volumes A–G*, Report on Government Services, Australian Government, Canberra.

PCU (Pulse Credit Union) nd, *Pulse Credit Union Limited Privacy Policy*, http://www.mucu.com.au/about/privacy.html (accessed 4 July 2016).

PERC (Policy and Economic Research Council) 2012, *Credit Impacts of More Comprehensive Credit Reporting in Australia and New Zealand*, Durham, NC, 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, *Electronic Recording and Reporting of Controlled Drugs (ERRCD)*, Fact Sheet, Canberra, 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, *Funding Sources*, http://www.phrn.org.au/about-us/who-is-involved/funders/ (accessed 15 February 2017).

—— 2011a, *About us*, Perth, http://www.phrn.org.au/about-us/ (accessed 20 October 2016).

—— 2011b, *How Is Data Linked*, Perth, http://www.phrn.org.au/about-us/data-linkage/how-is-data-linked/ (accessed 10 June 2016).

—— 2011c, *Linkage and Security*, Perth, http://www.phrn.org.au/about-us/data-linkage/linkage-and-security/ (accessed 16 October 2016).

—— 2016, *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*, Perth, http://www.phrn.org.au/media/80968/phrn_mrff-priorities-submission-_v10.pdf (accessed 20 October 2016).

Pillar, P. 2013, *The Pendulum of Opinion on Security and Privacy*, The National Interest, http://nationalinterest.org/blog/paul-pillar/the-pendulum-opinion-security-privacy-8567 (accessed 2 February 2017).

Pollock, R. 2008, *The Economics of Public Sector Information*, University of Cambridge.

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*, 1 November, 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, Washington D.C., 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, London, 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*, 11 August.

Privacy Committee Of South Australia 2015, *Privacy and Open Data Guideline*, SA Government, Adelaide, 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*, Melbourne, https://www.psa.org.au/policies/position-statement-real-time-recording-and-reporting-of-drugs-of-dependence (accessed 20 October 2016).

PwC (PricewaterhouseCoopers) 2013, *Where have you been all my life? How the financial services industry can unlock the value in Big Data*, October, New York, 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*, September, New York, 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*, March, New York, http://www.pwc.com/us/en/consumer-finance/publications/social-media-in-credit-underwriting-process.html (accessed 27 June 2016).

—— 2015b, *The Internet of Things: what it means for US manufacturing*, February, New York.

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*, Sydney, 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*, Sydney, https://www.quantium.com/ (accessed 3 August 2016).

Quigley, R. and Baines, J. 2014, *How to improve your social licence to operate - A New Zealand Industry Perspective*, Prepared for Aquaculture Unit, Ministry fot Primary Industries, MPI Information Paper 2014/05, http://www.aquaculture.org.nz/resource-library/general/how-to-improve-your-social-licence-to-operate/ (accessed 3 February 2017).

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.

Radford, L., Holland, J., Maplethorpe, N., Kotecha, M. and Arthur, S. 2012, *Evaluation of the Pension Credit payment study*, Department for Work and Pensions, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/191749/795and796summ.pdf (accessed 6 February 2017).

Rahimzadeh, V. and Knoppers, B. 2016, 'How mutually recognizable is mutual recognition? An international terminology index of research ethics review policies in the USA, Canada, UK and Australia', *Personalized Medicine*, vol. 13, no. 2, pp. 101–105.

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).

Ritchie, F. and Welpton, R. 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, London, http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/Economics%20Papers%202014/1413.pdf (accessed 13 July 2016).

RMH (The Royal Melbourne Hospital) nd, *Streamlined ethics and governance review*, https://www.thermh.org.au/research/researchers/about-research/streamlined-ethics-and-governance-review (accessed 27 February 2017).

Robert Walters 2013, *Understanding the role of social media to complement attraction strategies*, Robert Walters Whitepaper, Sydney.

Rosenblum, A. 2015, 'Your Doctor Doesn't Want to Hear About Your Fitness-Tracker Data', *Technology Review*, 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, Australian Government 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 2016a, *List of Completed SA NT DataLink Projects by Financial Year*, Darwin.

—— 2016b, *Supporting health, social and economic research, education and policy in South Australia and the Northern Territory*, Darwin, https://www.santdatalink.org.au/ (accessed 20 October 2016).

SA Premier 2016, *Parliament urged to deal swiftly with child protection reform package*, Media Release, 20 September, Government of South Australia, Adelaide, 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*, 22 August, 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).

Sax Institute nd, *The 45 And Up Study Policy On Collection Of Biological Specimens In Sub-Studies*, Sydney, https://www.saxinstitute.org.au/wp-content/uploads/Policy-on-Collection-of-Biological-Samples-in-Sub-Studies.pdf (accessed 6 September 2016).

—— 2016a, *SURE*, Sydney, http://www.saxinstitute.org.au/our-work/sure/ (accessed 10 June 2016).

—— 2016b, *SURE: The Secure Unified Research Environment fact sheet*, Sydney, 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.

Sciencewise 2014, *Big Data - Public views on the collection, sharing and use of personal data by government and companies*, April, London, http://www.sciencewise-erc.org.uk/cms/public-views-on-big-data/ (accessed 11 January 2017).

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).

Senate Economics References Committee 2016, *2016 Census: issues of trust*, November, http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Economics/2016Census/Report (accessed 11 January 2017).

Sense-T 2015, *New AirRater app to help Tasmanians breathe easier - Sense-T - from sensing to intelligence*, http://www.sense-t.org.au/latest-news/news-items/new-airrater-app-to-help-tasmanians-breathe-easier (accessed 27 February 2017).

—— 2016, *Agriculture - Sense-T - from sensing to intelligence*, http://www.sense-t.org.au/projects-and-research/agriculture (accessed 27 February 2017).

—— nd, *Aquaculture - Sense-T - from sensing to intelligence*, http://www.sense-t.org.au/projects-and-research/aquaculture (accessed 27 February 2017).

Sensis 2016, *Sensis Social Media Report 2016*, 1 June, Melbourne.

Shaw, R. 2017, *Australian business struggles to keep up with big data explosion*, iTWire, http://www.itwire.com/data/77288-australian-business-struggles-to-keep-up-with-big-data-explosion.html (accessed 21 March 2017).

Sheng, C. 2013, 'China Merchants Bank Tries Deeper Integration with WeChat', *TechNode*, 18 June, 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).

Shneiderman, B. 2000, 'Designing Trust into Online Experiences', *Communications of the ACM*, vol. 43, no. 12, pp. 57–59.

SIIAA (Spatial Information Industry Action Agenda) 2001, *Positioning for Growth — Technical Report*, Commonwealth of Australia, Canberra.

Silva-Goncalves, J. 2015, *Australians' switching behaviour in banking, insurance services and main utilities*, September.

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, London, 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, Canberra.

South Australian Department for Communities and Social Inclusion 2016, *New BlueBays app to help disability permit holders find a park*, Collection, https://www.dcsi.sa.gov.au/services/latest-news/media-releases-2016/new-bluebays-app-to-help-disability-permit-holders-find-a-park (accessed 5 January 2017).

Srinivasan, U., Rao, S., Ramachandran, D. and Jonas, D. 2016, *Flying blind - Australian Consumers and Digital Health*, Australian Health Data Series: Volume 1, Capital Markets Cooperative Research Centre (CMCRC).

SSCH (Senate Select Committee on Health) 2016, *Big health data: Australia's big potential*, Sixth interim report, May, Australian Government, Canberra, http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Health/Health/Sixth_Interim_Report (accessed 17 May 2016).

Stanford University nd, *Data Use Agreement (DUA) FAQs*, https://privacy.stanford.edu/faqs/data-use-agreement-dua-faqs (accessed 21 December 2016).

Statistics New Zealand 2016, *Integrated Data Infrastructure*, NZ Government, Wellington, 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*, October, Computer and Communications Industry Association, Washington D.C.

Suncorp Group nd, *Suncorp Group Privacy Policy*, Suncorp Group, Brisbane.

Tasmanian Government 2016a, *Submission to the Productivity Commission Inquiry into the National Education Evidence Base*, May, Hobart, http://www.pc.gov.au/inquiries/current/education-evidence/submissions (accessed 6 July 2016).

—— 2016b, *Tasmanian Government Open Data policy*, Hobart, http://www.egovernment.tas.gov.au/stats_matter/open_data/tasmanian_government_open_data_policy (accessed 20 October 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*, 27 October, 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, Melbourne.

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*, Sydney, https://www.parliament.nsw.gov.au/committees/DBAssets/InquirySubmission/Summary/50884/013%20The%20Benevolent%20Society%20.pdf (accessed 6 May 2016).

The Treasury 2016, *Review of tax and corporate whistleblower protections in Australia*, Australian Government Treasury, http://www.treasury.gov.au/ConsultationsandReviews/Consultations/2016/Review-of-whistleblower-protections (accessed 23 February 2017).

—— 2017a, *Australian Government response to the Senate Economic References Committee report: 2016 Census: issues of trust*, http://treasury.gov.au/PublicationsAndMedia/Publications/2017/2016-Census-issues-of-trust (accessed 1 March 2017).

—— 2017b, *Increasing transparency of the beneficial ownership of companies*, Australian Government Treasury, http://treasury.gov.au/ConsultationsandReviews/Consultations/2017/Beneficial-ownership-of-companies (accessed 23 February 2017).

Turnbull, M. 2015, *Australian Government Public Data Policy Statement*, Australian Government Department of Prime Minister and Cabinet, Canberra, 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*, White paper for the OECD, 1 January, PERC (The Policy and Economic Research Council), United States, https://www.researchgate.net/publication/215991947_The_Economic_Consequences_of_Consumer_Credit_Information_Sharing_Efficiency_Inclusion_and_Privacy (accessed 4 October 2016).

Twitter 2016a, *2016 Annual Report*, California.

—— 2016b, *Company*, https://about.twitter.com/company (accessed 18 May 2016).

—— 2016c, *Twitter Privacy Policy*, https://twitter.com/privacy (accessed 5 September 2016).

—— 2016d, *Twitter Terms of Service*, 27 January, Twitter, California.

Uber 2016, *Uber joins with Infrastructure Partnerships Australia to show how cities move*, Uber, https://newsroom.uber.com/australia/uber-ipa-commute/ (accessed 9 January 2017).

UK Data Service 2014, *Annual Report, October 2012-March 2014*, https://www.ukdataservice.ac.uk/about-us/reports (accessed 17 March 2017).

UK Parliament 2017, *Digital Economy Bill 2016–2017*, UK Government, London, http://services.parliament.uk/bills/2016-17/digitaleconomy.html (accessed 20 October 2016).

UN ECE (United Nations Economic Commission for Europe) 2015, *Big Data*, United Nations, Geneva, http://www1.unece.org/stat/platform/display/msis/Big+Data (accessed 1 August 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).

US DoJ (United States Department of Justice) nd, *1663. Protection Of Government Property - Protection Of Public Records And Documents*, https://www.justice.gov/usam/criminal-resource-manual-1663-protection-government-property-protection-public-records-and (accessed 10 March 2017).

VAGO (Victorian Auditor-General's Office) 2015a, *Access to Public Sector Information*, Report no. 2015–16:20, Victorian Government, Melbourne, http://www.audit.vic.gov.au/reports_and_publications/latest_reports/2015-16/20151210-access-to-information.aspx (accessed 23 March 2016).

—— 2015b, *Realising the Benefits of Smart Meters*, Victorian Auditor-General's Report, 2015–16:8, Victorian Government, 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)*, Victorian Government, Melbourne, http://www.vcaa.vic.edu.au/Pages/prep10/naplan/index.aspx (accessed 13 June 2016).

Veda nd, *Comprehensive credit reporting — Your Credit and Identity*, Sydney, http://www.veda.com.au/yourcreditandidentity/comprehensive-credit-reporting (accessed 7 April 2016).

Verizon 2016, *2016 Data Breach Investigations Report*, Melbourne, http://www.verizonenterprise.com/verizon-insights-lab/dbir/2016/ (accessed 5 September 2016).

VHA (Vodafone Hutchison Australia) 2016, *Privacy Policy*, Sydney, http://www.vodafone.com.au/about/legal/privacy (accessed 5 July 2016).

Victorian Government nd, *data.vic.gov.au*, https://www.data.vic.gov.au/ (accessed 6 March 2017).

—— 2016, *Information Technology Strategy — Victorian Government 2016–2020*, Melbourne, https://www.enterprisesolutions.vic.gov.au/wp-content/uploads/2016/05/Information-Technology-Strategy-for-the-Victorian-Government-2016-to-2020.pdf (accessed 14 July 2016).

W3C 2013, *An Overview of the PROV Family of Documents*, W3C Working Group Note, 30 April.

WA Ombudsman 2013, *Guidelines for the Management of Personal Information*, May, WA Government, Perth.

WALIA (Western Australian Land Information Authority) 2015, *Providing access to WA government data*, April, WA Government, Perth, http://data.wa.gov.au/open-data-policy (accessed 10 October 2016).

WALIS (Western Australian Land Information System) 2012, *Location Information Strategy for WA*, WA Government, Perth.

—— nd, *Location Information Access Framework*, WA Government, Perth, http://www.walis.wa.gov.au/projects/location-information-access-framework-liaf (accessed 10 October 2016).

Wallace, N. 2017, *EU's Right to Explanation: A Harmful Restriction on Artificial Intelligence*, Tech Zone 360, http://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm# (accessed 16 March 2017).

Waller, 2013 Matt and Boccasam, P.V. 2013, 'How Sharing Data Drives Supply Chain Innovation', *Industry Week*, 12 August, 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.

Waugh, P. 2013, *New data.gov.au — now live on CKAN*, *Australian Government Department of Finance*, 17 July, https://www.finance.gov.au/blog/2013/07/17/new-datagovau-%E2%80%93-now-live-ckan/ (accessed 27 September 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).

Westpac 2014, *Westpac Privacy Policy*, 12 March, Sydney.

White, S. 2017, 'The stuff of secrets and the career paths they can grow', *The Sydney Morning Herald*, http://www.smh.com.au/business/workplace-relations/the-stuff-of-secrets-20170221-guicq5.html (accessed 28 February 2017).

Wikipedia 2017a, *Ashley Madison data breach*, Wikipedia, https://en.wikipedia.org/wiki/Ashley_Madison_data_breach (accessed 16 March 2017).

—— 2017b, *Phorm*, Wikipedia, https://en.wikipedia.org/w/index.php?title=Phorm&oldid=770255421 (accessed 16 March 2017).

Williams, C. 2016, 'Aim-listed online ads company Phorm goes bust leaving investors £200m out of pocket', London, UK, *The Telegraph*, http://www.telegraph.co.uk/business/2016/04/14/aim-listed-online-ads-company-phorm-goes-bust-leaving-investors/ (accessed 16 March 2017).

Williams, M. 2013, 'IAG's $1.85bn Wesfarmers underwriting bet', *Asia-Pacific Banking & Finance*, 17 December, http://www.australianbankingfinance.com/insurance/iag-s--1-85bn-wesfarmers-underwriting-bet/ (accessed 5 January 2016).

Wittes, B. and Kohse, E. 2017, *The privacy paradox II: Measuring the privacy benefits of privacy threats*, Center for Technology Innovation at Brookings, Washington D.C.

Wong, A. 2017, 'Algorithms need human touch', Sydney, *The Australian*, 24 January, http://www.theaustralian.com.au/business/technology/opinion/complex-algorithms-can-use-a-little-of-that-human-touch/news-story/c6cf107dab1d40037df410d36aa2f3b7?csp=a654e423828dfca3f278065d479291f3 (accessed 24 January 2017).

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) 2016a, *Open Data Barometer Global Report*, 3rd edn, http://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-GlobalReport.pdf (accessed 1 July 2016).

—— 2016b, *Open Data Barometer Methodology — v1.0*, opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-Methodology.pdf (accessed 1 August 2016).

XVT Solutions 2016, *NSW Launch Phase 1 of the Electronic Recording & Reporting of Controlled Drugs System (ERRCD)*, 15 September, https://www.xvt.com.au/nsw-launch-phase-1-electronic-recording-reporting-controlled-drugs-system-errcd/ (accessed 24 January 2017).

Yates, B. and Horvath, C. 2013, *Social License to Operate: How to Get It and How to Keep It*, 2013 Pacific Energy Summit, http://www.nbr.org/publications/element.aspx?id=681 (accessed 3 February 2017).

YouTube 2010, *Terms of Service*, 9 June, https://www.youtube.com/static?gl=AU&template=terms (accessed 18 May 2016).