Australian Government
Department of the
Prime Minister and Cabinet

PUBLIC SECTOR DATA MANAGEMENT

July 2015

Public Sector Data Management

ISBN 978-1-925238-11-2 (PDF)
ISBN 978-1-925238-12-9 (PPT)

Public Data Branch
Department of the Prime Minister and Cabinet
1 National Circuit
Barton ACT 2600
Email: datapolicy@pmc.gov.au

Table of Contents

Introduction

The Public Sector Data Management Project was commissioned by the Secretary of the Department of the Prime Minister and Cabinet (PM&C) to deliver a roadmap to unlock the potential of public sector data to drive innovation, efficiency, productivity and economic growth.

Project scope

In April 2015, the Secretary of PM&C commissioned an in-house Australian Public Service (APS) study on how public sector data can be better used to achieve efficiencies for government, foster the digital economy and be helpful outside government to lift productivity and growth.

The Project involved delivering a roadmap by mid-2015 on ways to maximise the potential of public sector data.

Project team

The Project was overseen by Secretaries Michael Thawley, Jane Halton and Drew Clarke with guidance from an inter-agency Steering Committee.

PM&C Deputy Secretary,
Dr Heather Smith, led the Project with a small team comprising representatives from PM&C and the:

- Department of Communications (Communications);
- Department of Finance (Finance);
- Australian Bureau of Statistics (ABS); and
- Australian Taxation Office (ATO).

Complementary agendas

Improving the management of public sector data complements APS work already underway and forms part of a broader transformative agenda that will improve the digital economy in Australia.

Government holds a vast amount of data, with even more being created through the Digital Transformation Office (DTO). As the National Broadband Network is rolled-out and connectivity improves, data's value will continue to increase.

Communications is surveying private sector data users to understand their priorities. The ABS is linking key datasets to gain policy insights and the ATO is developing a longitudinal personal tax sample file. Finance manages data.gov.au, providing access to datasets.

Next steps

1: Secretaries Thawley, Halton and Clarke consider report findings.

2: Secretary Thawley presents report outcomes to the Secretaries Board.

3: PM&C (Project Office) scopes first phase of recommendations.

Executive Summary

By making the most of its data, the Commonwealth could grow the digital economy and improve people's lives by transforming how policies and services are delivered.

Data volumes are growing exponentially and so too is the potential value of this data. The Commonwealth needs to invest in new capabilities and early action to embrace the opportunities that data offers.

Linking and sharing public sector data can create opportunities that neither government nor business can currently envisage.

Data is under-utilised in the APS

Currently, the Commonwealth's capacity to fully derive value from public sector data is constrained by competing priorities and the lack of an overarching strategy.

- There is no clear mandate for the Commonwealth to use and release public sector data.
- There are barriers (perceived and real) to sharing data within the Commonwealth and with jurisdictions to improve policy and service delivery.
- The APS lacks sufficient incentives, skills and organisational arrangements to capitalise on its data.
- The Commonwealth does not have a strong culture of publishing data to foster economic opportunities.

Australia's capacity to remain competitive in the digital economy is contingent upon its ability to harness the value of data.

Action is needed to meet the Government's 2013 election commitment to increase access to useful public sector data.

Building capability is critical to unlocking this value. Now is the right time to invest in data capability given the Commonwealth's focus on innovation and STEM (science, technology, engineering and math).

Learning from international experience

Australia lags the United States (US) and United Kingdom (UK) in releasing public data for business, and lags New Zealand (NZ) in the application of data to policy design. All three countries made an upfront investment to drive data policy with a top-down mandate from Ministers.

Some countries started with a clear mandate to release data, while others focused on sharing and analysing it – there is no optimal sequence. Sustained action and commitment was key.

A lot can be done now within existing policy settings, so it is therefore proposed that reform starts with several projects that demonstrate value, uncover barriers, and lead to better designed policy and services that improve people's lives.

The recommendations in this report aim to change the way government does business by setting up the right frameworks, systems and capability to use, share and value data.

Build confidence and momentum by learning and doing (1 – 6 months)

Early wins are important to build confidence. **Several data-driven projects** that reflect government priorities and break down data barriers will demonstrate why government needs to change the way it uses data.

Building trust with the public is key, so engagement work will occur in parallel.

Systems will be set up that support **horizontal collaboration** across the APS and forge partnerships with the non-government sector, because the intersection of datasets is where much of the value lies.

Systematise the use and release of data (1 – 18 months)

Adaptive change is needed to build capability and keep up with new technologies, tools, and analytic techniques.

A platform will be developed that supports trusted user access, while also remaining flexible for future advancements.

Agency ownership will be important, supported by governance mechanisms with strong leadership.

A lifecycle approach to data is required to ensure data is maintained and accessible beyond the point of collection.

Tailored approaches are needed for different types of data

A careful, staged approach to implementation is required to adequately address the different risks, governance requirements and other considerations that each type of data poses.

Personal data	*E.g. Individual citizen health records*		**Agency access** – for service delivery
Research data	*E.g. De-identified, linked health & welfare records*		**Restricted sharing** – de-identified data securely shared with trusted users for research
Open data	*E.g. Aggregated or non-sensitive data such as Centrelink office locations*		**Publicly available** – e.g. through data.gov.au or the ABS Table-builder
Security data	*E.g. Metadata collected for intelligence*		**Closed access** – not in scope

Note: A glossary of terms is provided at Appendix 1, page 39

Proposed approach

Vision
What are we trying to achieve through public sector data?

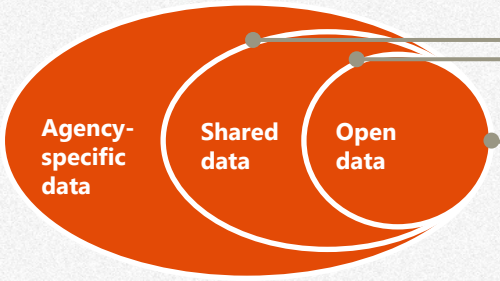
Improve service delivery
Share and analyse data to deliver better services for citizens that are more efficient for government.

Develop more effective government
Data integration and new analytical capabilities strengthen our evidence base for policy development and public administration.

Drive the digital economy
Public sector data is available to support private sector innovation and productivity.

Framework
How will we manage different types of public sector data?

- **Survey data**
- **Administrative data**
- **Other data**



Unit-record level, identifiable data that is **shared on secure networks** and accessed by automated agency processes and trusted users within government agencies for improved service delivery.

Unit-record level, de-identified data, **integrated via a secure network** and analysed by trusted users within government agencies and research institutions to improve policy development.

De-identified data, **published via the Commonwealth's Open Data portal** and used by the private and social sectors to add value, including through enabling the digital economy.

Barriers
Challenges to data management

Social license – privacy provisions in legislation and community expectations limit the use of data.

Limited capability – there is a shortage of data analytic capabilities.

Data management – data is hard to find and difficult to use.

Variable policy settings – there is no clear mandate for data policy in the APS and an absence of good, well-integrated governance arrangements.

Culture – there is not a culture of releasing data or widespread use of data for administration, policy development, service delivery or regulatory functions.

Roadmap to deliver transformation
What will we do to make better use of public sector data?

1-6 Months: Build confidence and momentum

1. **Secretary of PM&C signals data is a priority**
2. **Commission several high-value projects**
3. **Build external partnerships**
4. **Publish readily available non-sensitive datasets**
5. **Build data and analytics capability**
6. **PM&C coordinates projects and progresses APS data policies and governance (covering recommendations 7-8)**

1-18 Months: Systematise the use and release of public sector data

7. **Implement a data policy framework that includes:**
   a. **a public policy statement**
   b. **a simple governance model for data policy**
   c. **a requirement for evidence-based policy**
8. **Build and maintain public trust**
9. **Establish a trusted-access model for sharing integrated data**
10. **Create and publish a searchable whole-of-government data catalogue**
11. **Develop a Commonwealth Government high-value dataset framework**
12. **Publish data management standards**
13. **Establish a consistent and transparent approach to user charging**
14. **Create a legislative environment that supports data use while maintaining privacy, building on a possible Productivity Commission inquiry**
15. **Promote innovation in public administration**

Recommendations: Public sector data roadmap

It is recommended that a clear mandate and several policy projects be initiated to generate confidence in public sector data over the short term and structural changes be implemented over the longer term to transform the way the APS uses data.

1-6 Months: Build confidence and momentum

1. **Secretary of PM&C signals data is a priority** by promoting the agenda across the APS
2. **Commission several high value projects** focused on key policy questions, proving the value of public sector data and removing barriers
3. **Build external partnerships** to foster demand and encourage the use of public sector data
4. **Agencies publish readily available datasets** on or through data.gov.au, and schedule future releases
5. **Build data and analytics capability** by bolstering existing efforts, partnering externally and investing in pockets of excellence
6. **PM&C coordinates progress** on demonstration projects, develops APS data policies and takes steps to rationalise governance arrangements through a cross-agency team
(see recommendations 7-8)

1-18 Months: Systematise the use and release of public sector data

7. **Implement a data policy and governance framework** that includes:
    a) **a public policy statement** that could be delivered by the PM outlining the narrative and high-level principles
    b) **a simple governance model for data policy** that incentivises and drives consistent transformation through clear ownership, links to external groups, a high level steering group, partnerships with the private sector and data champions within agencies
    c) **a requirement for evidence-based policy** development to motivate data use in policy
8. **Build and maintain trust**, engage with the public to understand the benefits to citizens and address privacy concerns on public sector data
9. **Establish a trusted-access model for sharing integrated data** across agencies
10. **Create and publish a searchable whole-of-government data catalogue** to make government data holdings discoverable
11. **Develop a Commonwealth Government high-value dataset framework** to inform prioritisation and release
12. **Publish data management standards** to simplify the processes of integrating data and publishing on data.gov.au
13. **Establish a consistent and transparent approach to user charging** to provide clarity on how to manage costs
14. **Create a legislative environment that supports the use of data while maintaining privacy** by initially providing guidance and ultimately revisiting legislation through a possible Productivity Commission inquiry
15. **Promote innovation in public administration** through targeted awards and resources

What can be learnt from international experience?

Lessons for Australia

1. **Strong leadership** has driven change from the top down.

2. **Transparent communication** has been critical in forming a new social contract with citizens.

3. **Partnerships** with private and research sectors have enabled the early optimisation of data release and are key to sustained momentum.

4. **Data workers** are needed in government to raise competency.

5. **Agile and responsive** government is needed as technology and consumer needs move quickly and in unpredictable ways.

6. **Privacy** has been important for all countries, but has not dominated the agenda nor stopped government progressing with data projects.

7. **Investment** in resources has been made by all countries to pursue this agenda – the US, UK and NZ deliberately excluded data policy from fiscal consolidation.

The UK and US started with open data

The UK and US are both 5 years into their open data agendas. Both have had:

- a focus on accountable policy and cutting waste;
- strong political and senior civil service support;
- a supply-driven approach;
- a focus on building capabilities;
- infrastructure in place to facilitate private sector involvement;
- a recent change in focus to identify the needs of private data users; and
- deeper linkages with government, researchers and industry.

Case Study

Neighborhood Score is a mobile app that uses public sector data released by the San Francisco Mayor's Office. It is designed to provide an overall health and sustainability score, block by block for neighborhoods in San Francisco.

The app combines a variety of health-related data, including measures on mental health, safety, traffic and physical wellbeing. This app advises on livability and identifies success and failures in communities.

NZ focused on better value from the spend, rather than the spend itself

NZ has spent the last six and a half years using data to make "smaller government by better government", with Deputy Prime Minister , the Hon. Bill English driving an agenda, including:

- regular meetings between agency heads and cabinet Ministers to discuss how data is being used;
- a new, central infrastructure (Integrated Data Infrastructure) to securely share unit-level data across government agencies;
- positive competition across agencies;
- projects to demonstrate value;
- budget bids which require data-driven cost-benefit analysis and post-implementation evaluation from 2015 onward; and
- reviewing legislation to move away from confusing and outdated laws which have caused risk aversion.

> "If we can't measure effectiveness, it won't be funded through social investment."
>
> The Hon. Bill English, NZ Deputy Prime Minister

The opportunity

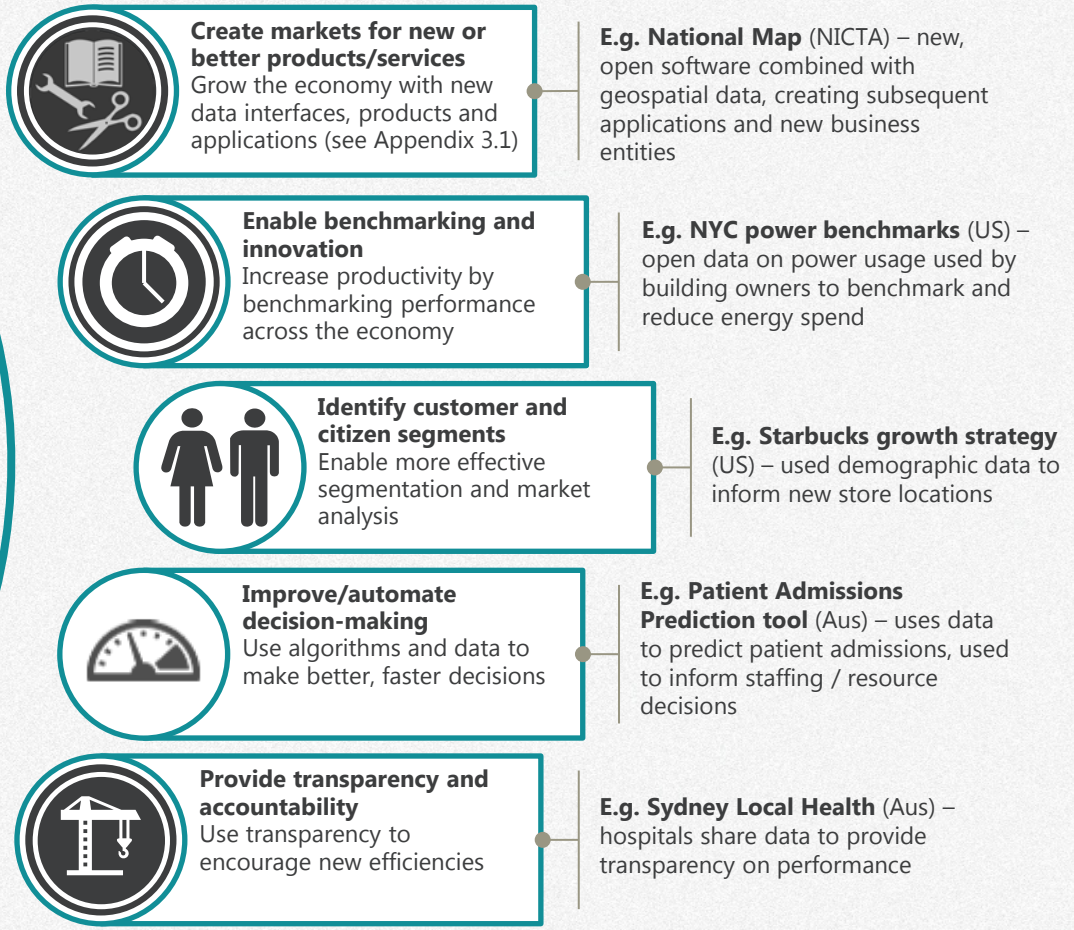
Publishing data will enable innovation outside the APS

In the emerging digital economy, public sector data is a new commodity. While estimated benefits of this data vary, all studies agree there is significant economic value from opening access to public sector data and improving data analytics.

Publish the right public sector data

The Commonwealth's role in fostering innovation includes:

- **publishing the right data** – responding to requests for datasets from the private sector;
- **publishing it in the right way** – use international standards to determine formats and licensing;
- **enabling trusted accessibility** – support appropriate ways to interface with data across the spectrum from personal to open data; and
- **providing channels for feedback** – actively engage with external stakeholders through boards and dialogue.



Commonwealth | **Other Sectors**

Sources: McKinsey, *'Open Data – Unlocking performance through liquid information'*, October 2013 and PricewaterhouseCoopers, *Deciding with Data: How data-driven innovation is fueling Australia's economic growth'*, September 2014

Public sector data can drive citizen-centric government

Public sector data is a valuable economic and social asset. Harnessing the power of data can lead to smarter, citizen-centric services and policy.

> Data is the raw material of the 21st century… Its value is in driving choice and improvements in public services and in inspiring innovation and enterprise that spurs social and economic growth.

Rt Hon. Francis Maude
UK Minister for Trade and Investment
and former Minister for the Cabinet Office
(responsibility for public data)

Opportunities for government

Governments collect, through their interactions with citizens, a huge amount of data.

Data can enable government to improve public services and policy design, bringing wide-ranging social benefits to citizens and improvements in how parts of government are delivered (see Appendix 3.3 for UK case study on service delivery).

Data can unlock ways of helping Australia proactively plan for and respond to some of the biggest challenges facing the nation in areas such as health, social security and indigenous affairs.

As well as helping to develop more effective policies, data can act as a key performance indicator for measuring policy outcomes. It can also enable jurisdictions and non-government organisations (NGOs) to achieve more without additional funding.

Australians are already seeing the benefits of better use of data in service delivery. The introduction of biometric technology at airports has reduced immigration queues and Personally Controlled Electronic Health Records have helped to improve disease management plans.

Public sector data opportunities for the Commonwealth
Develop smarter, evidence-based policy
Deliver better, citizen-centric services
Increase public sector efficiency and reduce overheads
Evaluate programs and benchmark third party providers

State of play

Progress varies by government sector

There are distinctive cultures across the APS with some pockets of excellence.

Security and Law Enforcement

Security, law enforcement and intelligence agencies have built data sharing arrangements through national security imperatives and crises. Intelligence agencies have large databases to share security information.

Social policy

Agencies hold large volumes of data, collected through administration of services and programmes, and surveys. The ATO, the Department of Human Services (DHS) and the Department of Social Services (DSS) are actively working on projects to gain richer analysis of linked data (e.g. longitudinal analysis to understand intergenerational mobility). However, privacy and secrecy concerns (real and perceived) often constrain the ability to link or publish this data.

Science/Resources

Spatial, science and environmental data is actively being published as there are less security, privacy or legislative barriers. Spatial data is the most common type of data on data.gov.au, with Geoscience Australia the largest publisher. The Department of Environment (Environment) has accelerated progress, releasing more than 50 datasets in 12 months.

Some land and spatial data is restricted due to licencing and more could be done to link science data with government administrative data. For example, geocoded national address data could be made openly available to enable innovation and assist with planning and address validation.

Economic

Economic data is largely open. The Reserve Bank of Australia, the Australian Prudential Regulation Authority and the ABS have opened significant datasets. However, where government adds value to data, such as developing adjusted employment data, the product is rarely shared.

Some agencies are currently collaborating with the private sector to share data. Freight and logistics businesses share data with the Department of Infrastructure to inform reporting on the use of national infrastructure, for example, container movements at ports.

States and Territories

Most States and Territories (States) have an open data portal and policy statement. Queensland accelerated its open data agenda under the Newman Government and Western Australia is a leader in linked data. South Australia's open data programme focuses on engaging the data community and building capability. Victoria and New South Wales (NSW) have comprehensive open data agendas.

While States and the Commonwealth are often reluctant to share data, the Australian Institute of Health and Welfare (AIHW) successfully shared data with the Productivity Commission for the Report on Government Services (RoGS). RoGS also led to more state data being published on childcare, education and training; justice and emergency management; health; community services; housing and homelessness.

The data policy landscape is fragmented

Responsibility for whole-of-government data policy and delivery

Delivery on election commitments for Gov 2.0 and Big Data, and data.gov.au – **Finance**

Information and data licensing policy – **Attorney-General's Department**

Election commitment on 'open/big data' – **Communications**

Lead accredited data linking authority, de-identification capability – **ABS**

Digital service delivery transformation – **DTO**

The policy landscape

In 2013, the Government made an election commitment to 'provide expanded access to useful public sector data.'[1]

Progress on open data has been slow within the APS, with:

- policy responsibility spread across various portfolios;
- multiple policy documents in areas such as big data, cloud computing, and open data;
- overlapping governance bodies, working groups, and committees; and
- various legislation restricting the use of data to the purposes for which it is collected.

Achievement against adversity

There have been some improvements in recent years, largely driven by pockets of excellence rather than an institutionalised approach.

The ABS has linked the Census to key datasets, the ATO has increased the tax data available for researchers and there have been significant strides in spatial visualisation and the number of datasets available through data.gov.au, the Open Data Network, NationalMap and the Landsat Archive (see Appendix 3.4).

The DTO is driving the service delivery element of data and service analytics. Larger agencies have begun using data analytic techniques and are developing their big data capabilities.

Outside calls for action

The Commission of Audit recommended the better use of data for policy development, service delivery and fraud reduction. It also recommended the release of de-identified administrative datasets.

Other reviews and reports have made similar calls for action, including the Financial System Inquiry, the 2014 Productivity Commission's Annual Report and the 2013-14 State of the Service Report. The Harper Competition Review is recommending that government and business allow consumers access to their information in an efficient format to improve informed consumer choice.

1. *The Coalition's Policy for E-Government and the Digital Economy,* September 2013

Private users want access to more data

The Open Data 500 Australia, surveying many sectors and comprising large and small private sector organisations, as well as consultations with agency and research organisations, tell us what data is most valuable to users.

> More unit data needs to be released that would enable us to do more longitudinal research: specifically social security census and tax longitudinal data.

John Daley
CEO of Grattan Institute

> We want greater granularity across the board and we want data that is complete, current, consistent and authoritative, in a simple way, as and when required.

Mr David Bruce
Managing Director
OMNILINK
Open Data 500 Participant Company

A lot is being done but there is a long way to go

The ABS Census and Labour Force Survey aggregate data is available via subscriptions but some researchers require access to individual level data which is restricted (in the case of Census data) to a 5% confidentialised sample.

Research organisations such as the Grattan Institute, Australian National University (ANU) and the Melbourne Institute have called for unit-record longitudinal social security, labour force and Census data. The ATO is currently developing a longitudinal personal tax sample file for researchers.

Work is currently underway to make national geocoded address data openly available. Making this data openly available will unlock opportunities for government efficiency, industry innovation and competitiveness.

Longitudinal datasets, along with the remaining datasets below, with the exception of data held by jurisdictions (i.e. cadastre – land and property information – and land tenure data) could be delivered, in part, to the trusted research community through the recommendations of this report.

Top 5 data themes	Research sector: Top 5 most requested data	Private sector: Top 5 most requested data
1 Spatial and land	1 Census and Labour Force Longitudinal Data	1 Geo-coded National Address File
2 Socio-economic	2 Geo-coded National Address File	2 Cadastre and Land Tenure
3 Health	3 Personal Income Tax and Business Tax	3 Geo-coded Australian Business Register
4 Transport	4 Health Benefits Data (linked MBS / PBS)	4 Health Benefits Data (linked MBS / PBS)
5 Environment (incl energy)	5 Social Security Payments	5 Census and Labour Force Survey Data

Limited public sector data is being released/accessed

"We were quoted $5,000 for a data request that would have taken a couple of hours to complete."

Commonwealth Agency

The APS has a supply-driven approach to publishing data

The Commonwealth has not engaged widely with private counterparts to understand demand for datasets.

This is partly due to a lack of information about the Commonwealth's own data holdings. Consultations suggest the APS itself does not have a good handle on what data it holds.

Online lists of government data, or 'data catalogues' have helped supply and demand issues and are used in the US, UK, and Canada to allow business and government to search and find data.

Data.gov.au is not front of mind for most of the APS

Data.gov.au is a central point of access to data from across governments. It currently links to 6,700 datasets, compared to 25,461 in the UK and 132,865 in the US. Notably:

- over 75% of the datasets come from just four organisations;
- there are big gaps, including in areas such as health, employment and socioeconomic data; and
- less than 25% of the data is enabled by an Application Programming Interface (code that allows programs to access data).

Most agencies do not release data via data.gov.au as a matter of course. In consultations, many said it simply does not enter their mind to do so.

Consultations indicated that data.gov.au had been under resourced (although some agencies such as DHS also have staff who prepare data for publication on data.gov.au).

User charging arrangements are inconsistent

Publishing data is not costless for agencies. Effort may be required to understand and format data collected for administrative purposes. Data often needs to be reformatted to be useful for potential end-users. However, some Commonwealth agencies are unnecessarily charging for data.

During consultations, researchers and some agencies expressed confusion and frustration about how charging is applied to different datasets and across agencies.

Many expressed a view that data collected by governments has already been paid for by taxpayers, and should be available free of charge wherever possible, particularly where the marginal cost of provision is low.

Privacy fears hinder the use of sensitive data for policy development

Current arrangements are over-cautious and cumbersome

Privacy concerns and cautious interpretation of legislation are holding the APS back from making the most of its data. For example:

- it can take several years and multiple MOUs to establish data sharing arrangements between government agencies;
- when projects occur, often only a few tables are produced from important linked data, which is then destroyed;
- sharing data with the States and Territories is burdensome; and
- the research community is frustrated by the red tape that impedes greater data sharing, the time it takes to be granted access to data and the inconsistency in user charging for data.

"In one case where an important dataset was being linked with the Census, the negotiations on the MOU took up to 18 months while the linking only took two weeks."

"An agency reported having up to 11 MOUs with the same department to access data."

The APS can use tools – available now – to securely share and use the data

- The speed and complexity of the data revolution means government will need to take a careful phased approach to more effectively manage data.
- The highest standards of security and privacy must be upheld. Ensuring this requires models of access where sensitive information is shared only with trusted users and under robust safeguards.
- Over time, this will require putting more obligations on the users of data.

Types of data

Less sensitive: Open access → **More sensitive: Restricted access**

Data type	Aggregate data	Less aggregated, non-identifiable data	Unit-level, personally identifiable data	Unit-level, linked, personally identifiable data
e.g.	Macroeconomic data published by the ABS	De-identified Census data	Raw census data	Linked ATO and DSS data

The APS could get more from data for service delivery

There are examples of service delivery agencies using data to increase their efficiency and effectiveness, but more can be done.

Data is being used within agencies for a range of services

There are good examples of agencies using data to design and deliver services.

For instance, DHS places indicators in Centrelink systems to flag a person's vulnerability and target interventions appropriately.

The ATO is using administrative and third party data to pre-populate tax returns, reducing the amount of information that individuals must separately enter. This reduces compliance costs for taxpayers by $156 million annually.

The DTO will use data to improve the way government information is presented, making services simpler and easier to use. The DTO will set digital design standards for all APS agencies and better link digital, face-to-face and telephone delivery channels.

But service delivery data is often not shared between agencies

The APS is not capitalising on data to evaluate and improve services because the data is often not shared between agencies or with third parties.

For example, when individuals move from one departmental programme to another (such as the Department of Defence to the Department of Veterans' Affairs, or DSS to the Department of Health) there are instances where data and client history is not transferred.

Government programmes, such as health and employment initiatives, are often evaluated but the results are rarely shared with third party providers, such as not-for-profit organisations.

Overall, the lack of data sharing prevents feedback on policy and hinders the potential of data to improve future service delivery.

Capability to use data is patchy across the APS

Case study: How the UK government is building data analytics capabilities

As part of its digital transformation, the UK Government is building data analytics capabilities through four strategies:

1. **build the right environment** by replicating private sector culture;
2. **get the right people in** by hiring high-profile experts from the technology / private sector, with an expectation that they would 'bring in' other experts;
3. **embed policy staff** alongside the makers and 'doers' in 'dream teams' of policy officers, data analysts and visualisation experts; and
4. **make bureaucrats accountable** for progress through a Minister-chaired Board which provides Ministers with means to assess and direct the civil service as it progresses.

Data skills are unique

Data scientists – the 'sexiest job of the 21st Century' according to DJ Patil, the US Chief Data Scientist[2] – require a unique skillset, including:

- fluent coding abilities;
- clear communication skills to communicate findings; and
- 'outside the box' thinking on how data is used.

There is a shortage of data analysts

There is a global under-supply of data and analytics skills which limits the ability to get the most out of data.

Australia and the APS is no different. Ready-for-work graduates with data capabilities are in short supply. During consultations, most agencies expressed a need for more data capabilities.

There are pockets of excellence in APS

There are currently pockets of advanced analytics skills – with the biggest concentrations in the ATO, ABS, Bureau of Meteorology, CSIRO, Geoscience Australia, the Department of Immigration and Border Protection and the intelligence community.

The current approach is ad-hoc

Departments have taken a variety of approaches to building data capabilities. Usually, a 'grass-roots approach' has been taken in the absence of an overarching strategy.

Some, such as the Department of Employment, have pooled their analytic capabilities to service the whole department, while the ATO has developed a Smarter Data Program with staff embedded in core business functions.

Consultations within the APS and with the non-government sector highlighted the importance of moving quickly to build capability.

Recruitment and shared skills are needed

Cross-disciplinary teams with policy, data and visualisation expertise are needed to drive innovation. The APS needs to:

1. embed skills in teams by recruiting data analysts and nurturing existing capabilities through learning and development; and
2. share skills/people across agencies so that data analysts become a shared resource across government.

2. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/, October 2012

Public sector data roadmap

The APS should take action in two ways

In the long term, data has the potential to transform the way the APS delivers parts of public administration – however this relies on a concerted upfront effort and the right foundations.

1-6 Months:
Build confidence and momentum

The Secretary of PM&C, with the support of Secretaries, could promote the importance of better data management; encourage collaboration across portfolios; and encourage a 'reasonable interpretation' of agency legislation to remove barriers to sharing.

Over the next six months, the APS should use data to problem-solve key policy questions to demonstrate the value of public data and work through barriers to sharing. Work should begin on a coordinated approach to improving data analytics capability and a data policy framework which would include a policy statement and strengthened governance.

The Open Data 500 project, which is asking businesses what data they want, is a catalyst for further engagement with the private sector to stimulate demand for government data. Similarly, the APS should analyse community attitudes and develop an engagement strategy for implementation in the longer term.

1-18 Months:
Systematise the use and release of public sector data

As momentum builds, the APS should make structural changes to the way it works, systematising the use, re-use and release of public sector data.

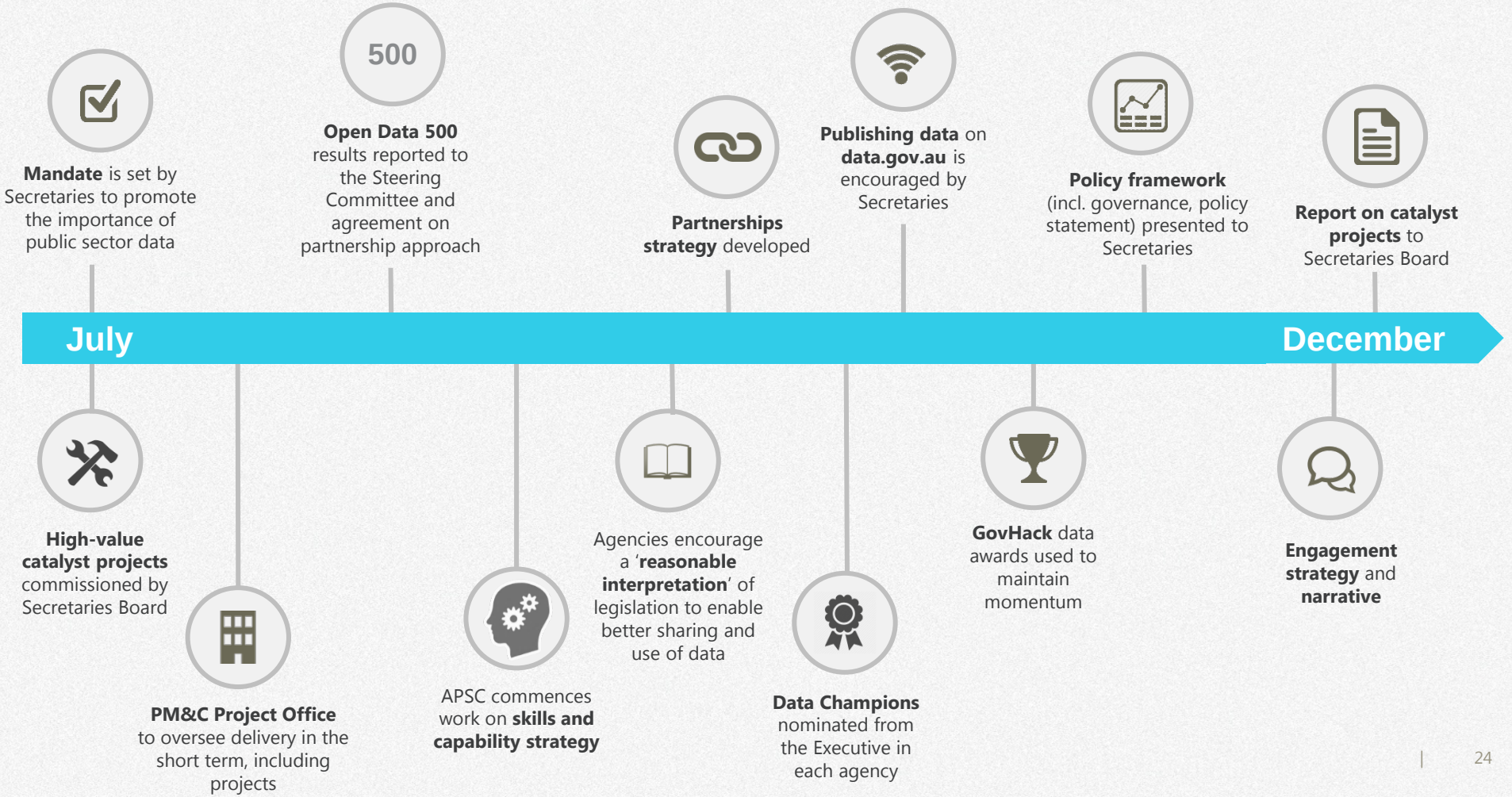
This will involve procedural changes, setting standards, agreeing taxonomies, and streamlining governance. A strong political mandate and formal partnerships with the private sector will be necessary to sustain transformative change.

As the supply and demand for data increases, more data will be made available through secure access to trusted users.

1. Build confidence and momentum

The next six months will focus on building momentum

*The first six months will be critical to build momentum for data reform through learning and discovery. From the outset, public sector data should be **prioritised** by **Secretaries** by **commissioning key data projects** to be a catalyst for action. Work should occur in parallel to **create partnerships** and **understand community attitudes.***



Secretary of PM&C / Secretaries signal data is a priority

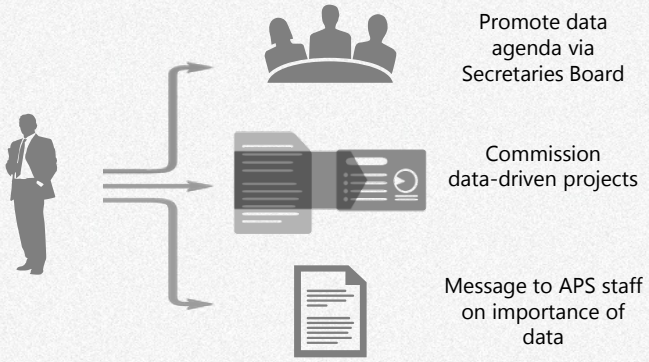
Recommendation 1

The Secretary of PM&C, with the support of the Secretaries of Finance and Communications, sets a mandate to encourage better use of public sector data, and:

a. commissions several data projects (page 26 refers);

b. encourages agencies to take a reasonable interpretation of agency legislation (page 36 refers); and

c. calls for data champions to be identified throughout the APS.



Leadership within the APS

Within the first six months, the Secretary of PM&C, through the Secretaries Board, could set a mandate to better use public sector data and promote its importance.

The Secretaries Board would commission several data projects to give impetus to the policy agenda. The projects would illustrate how data can be used to provide key outcomes for government.

All Secretaries should encourage staff to take a 'reasonable interpretation' of individual agency legislation regarding the use and sharing of data. Training on privacy could be undertaken to support this. Agency heads should also formally recognise senior data champions within their organisations. These data champions should promote the sharing and release of data (including through data.gov.au), and link data technicians with policy experts within their agencies and build external relationships.

A mandate to prioritise data

The key message for the APS is that there is a lot more that can be done with public sector data now. There needs to be less focus on the barriers and more on what can be done within the current policy settings to achieve better outcomes for government.

Other important points include:

- better use of data was a 2013 election commitment;
- agreeing to share data across the APS needs to become the norm rather than the exception; and
- the release of non-sensitive data should be prioritised based on discussions with stakeholders about what they want to access.

Commission several high-value projects focused on key policy questions

Recommendation 2

A coalition of agencies, supported by the Project Office within PM&C, delivers several cross-agency projects within the next several months to build confidence in improving public sector data management and policy innovation.

Projects to catalyse action

The high-value data projects should meet one or more of the following criteria.

- **Be compelling:** align with the Commonwealth's priorities; benefiting the public; and/or using data innovatively to understand a policy or service delivery issue.
- **Address barriers to data sharing:** such as access to state data, interpretation of legislation, linking cross-agency data and private sector partnerships.
- **Be scalable and achievable within short timeframes:** projects should avoid significant funding, resourcing and legislative change.

The projects would serve as 'proof of concept' pilots that help build momentum and capability in the use of data. They will also result in the creation of valuable linked datasets which will serve as an important way of developing data capability.

Focus on demonstrating value

Seven projects are listed at Appendix 4 (starting page 46) that:

Target key outcomes for government

- Address pressing policy issues through innovative linking and analysis
- Improve service delivery by making the APS more customer-centric
- Foster collaboration across the APS and with external parties

Use innovative approaches to achieve results

- Create new linked datasets to derive new insights, and create enduring resources for policy agencies
- Collaborate across the APS in new ways
- Leverage new capabilities through big data analytics

Break down barriers for future work

- Identify and address barriers in sharing data across the APS

Build partnerships to foster private demand for public sector data

Recommendation 3

Building on the findings of the Open Data 500 project and leveraging current relationships and programmes, departments (through their data champions) engage with industry, non-government organisations and state governments on the publication and use of agency data.

This could include:

- participating in high profile annual events, such as GovHack; and
- working with developers to identify how data can be used to create new applications which deliver better services (e.g. integration of government into google search results).

Recommendation 4

Agencies should develop and implement a schedule for the publication of readily available datasets on or through data.gov.au, prioritising data sought by industry, non-government organisations and State and Territory Governments.

A need to understand partners

The Commonwealth should build on the findings of the Open Data 500 to establish relationships with industry counterparts. It will also need to directly engage with universities, technology companies, and NGOs to draw out what data they need, how they plan to use it, and what services will be delivered.

The experience of other countries is that if partnerships are to be meaningful and sustainable, they need to be resourced appropriately.

Profile industry needs

Define industry drivers, how data is being used, and the most efficient means of providing access.

Build internal understanding

Develop strategies for how to deliver on industry needs, including meeting quality, time, and data publishing expectations.

Be consistent to build confidence

Build relationships, regularly publish high-quality data, and be responsive to industry needs.

The US – Partnering to solve policy problems

In the US, the National Science Foundation (NSF) is bringing together regional communities, research and private sector organisations to solve national policy problems.

The Big Data Regional Innovation Hubs programme provides seed funding to regions to encourage collaboration between and across the research sectors in that area.

Regions are asked to provide a single response indicating how a particular policy objective might be achieved or researched. This includes what data will be used, how it will be used, and how multiple organisations will work together collaboratively to solve the problem.

The aim of the hubs is to galvanise local government with NGOs to address pressing issues in their region. For example, addressing drought in the east, crime in the south-west, and policing or crime prevention in the cities.

Build data and analytics capability

Recommendation 5

The Australian Public Service Commission (APSC) and others to collaboratively develop a strategy with government, industry, and academia to build data and analytics capability.

An overarching strategy is needed

A whole-of-government strategy is needed to harness and develop talent and boost the capability of the nation to grow the digital economy. The strategy should encourage a discovery mindset to enable better problem analysis, development of policy solutions, improved service delivery and public sector efficiency.

A whole-of-government strategy would:

Develop a coherent plan

- The APSC should establish a skills capability development programme, drawing on expertise across the APS, including the ATO's Data and Analytics Centre of Excellence.

Harness existing capability

- In the short-term, use data champions to establish multidisciplinary, cross-agency teams to achieve skill-sharing and optimal project outcomes.
- Use entry-level recruitment programmes (e.g. graduates) to target data analysts.

Partner with the private sector

- Develop public-private partnerships to build collective expertise across government and industry to foster collaboration.
- Establish secondments and internships with the private sector to build new skills.
- Identify data analytic functions that could be better undertaken by the private sector, in line with Finance's Contestability Framework.

Work with academia

- Add tertiary courses in data analysis, such as the Finance ICT Skills Programme, to APS learning and development programmes.
- Use scholarships and secondments with academia to build expertise. For example, continue to dedicate part of the Sir Roland Wilson Scholarship programme to data-driven policy research.
- Collaborate with appropriate institutions to develop a curriculum for Australia's data and analytics capability. For example, the APS could regularly sponsor staff to complete the ANU's recently established Applied Data Analytics post-graduate qualifications.

Drive initial scoping work through a central agency

Recommendation 6

PM&C work with agencies to:

a. develop a draft data policy framework including statement from the Prime Minister, governance structures, and possible cabinet requirements;

b. build and maintain trust, engage with the public to understand the benefits to citizens and address privacy concerns on public sector data; and

c. streamline existing APS committee structures.

PM&C will drive progress over the short term

In the initial phase, PM&C will play a coordination role in driving the catalyst projects, drafting the policy framework and developing strategies to build and maintain public trust. Progress could initially be overseen by the existing Steering Committee, established to advise the Project, which includes key agencies such as the ABS.

Existing APS committee structures would be rationalised as soon as possible in consultation with relevant agencies, and the project would then fall under the new governance structure (see page 31), reporting to a Deputy Secretaries Data Group.

Potential model for initial work

PM&C Project Office (with secondees from Finance and Communications)

Oversee catalyst data projects

Work with agency leads to scope projects and report on progress and outcomes.

Build and maintain public trust

Analyse community attitudes, draft narrative for APS and Ministers to build and maintain public trust.

Draft policy framework

Draft framework, including PM's statement and encourage the inclusion of evidence based on data analytics where relevant for policy proposals presented to the Cabinet.

Work to simplify existing governance

Move toward streamlined APS committee structures in consultation with relevant agencies (see page 31).

2. Systematise the use and release of data

Implement a data policy framework

Recommendation 7

PM&C seeks agreement on and implements the policy framework, with three actions:

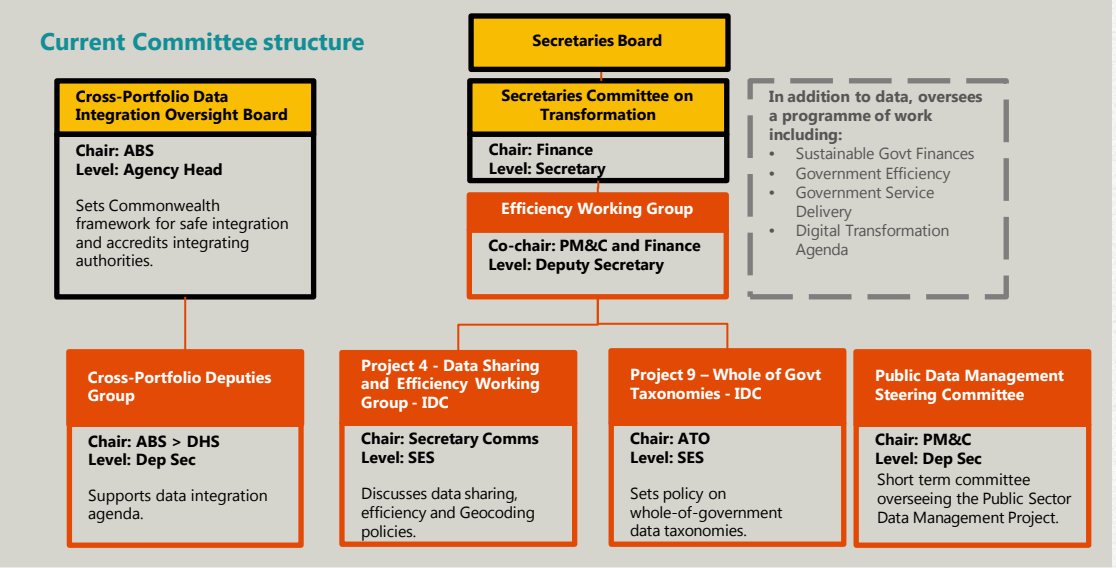
a. The Prime Minister, with Cabinet agreement, sets a clear mandate by publishing a policy statement on the use, sharing and release of public sector data, including principles for using and sharing public sector data.

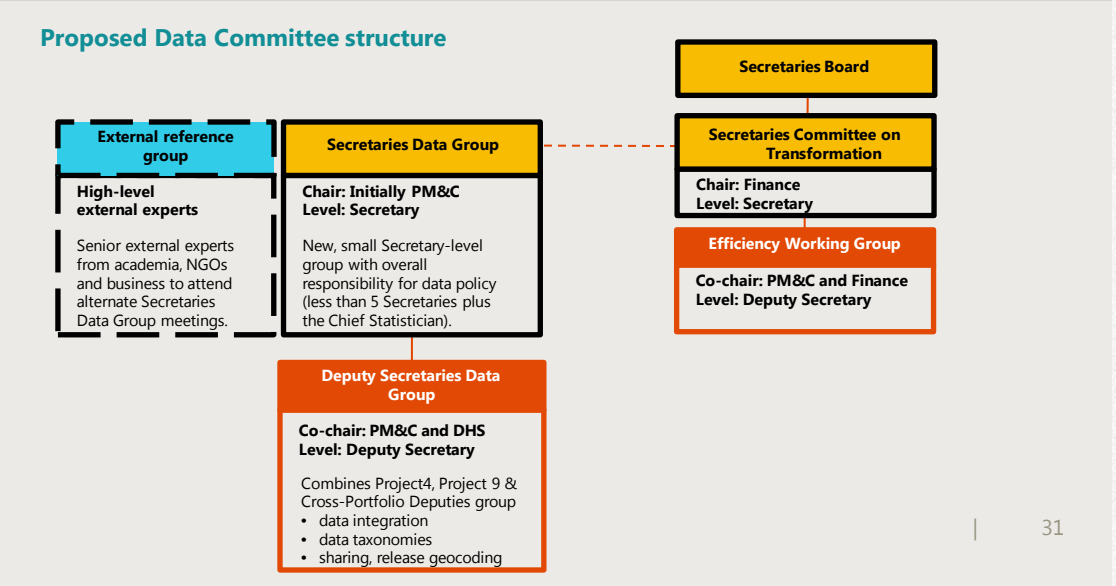
b. The Prime Minister appoints a single Minister or Parliamentary Secretary to be accountable for improving the use and sharing of public sector data, supported by a simplified governance structure, with external representation on a public data reference group.

c. Encourage the inclusion of evidence based on data analytics where relevant for policy proposals presented to the Cabinet.

Proposed APS Committee structure

Governance structures should transition to a model such as the one below, following consultation with relevant agencies.





Engage with citizens and business to build and maintain trust

Recommendation 8

Build and maintain trust in the quality, confidentiality and value of public sector data with government, researchers, businesses, and the public.

Privacy has not dominated elsewhere

UK and NZ research shows the public is cautious but supportive of open data research where it improves people's lives. The public trusts governments to hold their data, but has concerns on how it is used.

In Australia, attitudes vary[3]:

- government is more trusted than most private enterprises[4], however people are less likely to trust government than their friends, family and local councils[5];
- Australians are generally happy to provide personal information in exchange for a service;
- 60% - 70% of people worry about organisations accessing personal information: the older (65 – 69) and younger (18-24) are most concerned[6]; and
- people are most reluctant to share information on their financial situation[4].

Towards informed engagement

The APS needs to build a better understanding of community trust toward public sector data, including perceived values, risks, and attitudes towards privacy and confidentiality. This could be informed by a new ABS survey on community trust.

Engagement should build trust

For the APS to make effective use of data, it is crucial that we have the trust of the public. Strong assurances about data privacy and security based on rigorous adherence to protocols, and demonstrated value are key. A careful, phased approach to progressing the public data agenda is also needed to ensure that any issues or concerns are addressed.

Tailored approaches to engagement

Engagement should be responsive, targeted to key age groups and address the action being taken. Engagement could involve:

- **a clear narrative and talking points** detailing the purpose, value and role of public sector data and clearly explaining what data will not be released;
- **an engagement champion** (Ministerial spokesperson or lead official) to promote benefits and raise awareness;
- **a dedicated website** explaining the agenda, showing success stories, acting as a central point linking to a secure domain, data catalogue, analysis tools;
- **social media** pushing key messages; and
- **targeted articles** on data's role in the digital economy, benefits and myth busting.

3. Price Waterhouse Coopers, *Deciding with data: How data-driven innovations is fuelling Australia's economic growth.,* 2014. *4.* Office Australian Information Commissioner, Community Attitudes to Privacy Survey Research Report 2013. 5. Ernst and Young, *Digital Australia: State of the Nation* , 2014. 6. JWS Research, Edelman, Over the Digital Horizon.

Establish a trusted-access model for sharing integrated data

Recommendation 9 (see Appendices 5 and 6)

Establish a linked Commonwealth dataset with a domain for secure access by government and researchers, providing:

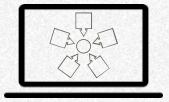
- a collection of linked administrative and survey data that links across time and across sectors;
- data linkage conducted by accredited integration authorities (ABS); and
- streamlined access to aggregate and individual data, while maintaining appropriate protection of individuals' records.

Model for data sharing

1. Identify key public sector datasets

2. Integrate key public sector datasets

3. Secure access to unlock value



Analysing integrated data will produce insights

Linked data can be used by public servants and researchers to better evaluate and design policies.

The task of identifying, integrating and providing access to key datasets could be undertaken by the ABS.

The ABS has already established bilateral agreements with key agencies to link their data with ABS holdings. The ABS is currently progressing a number of flagship data integration projects including a cross-portfolio project involving the ABS, ATO, Health, DHS and DSS. The ABS is also working with ATO to produce linked employer/employee data (LEED) and longitudinal business and personal tax data.

The ABS is currently reviewing its data access policies with a view to making data more accessible to public servants and researchers.

With proper access, this work will create Australia's most valuable economic and social research dataset. Appendix 5 provides further detail.

Integrated data will be accessed through a secure domain

An online domain will provide secure access to this integrated data. The domain will be simple and user-friendly providing easy onsite or online access via remote login (e.g. Citrix). The domain will be:

- **secure**: protected by a firewall so that only authorised users can login;
- **tailored**: access is limited to data fields necessary for research goals;
- **protected**: data will be confidentialised where needed; and
- **monitored**: activity and outputs of users are vetted.

The ABS has onsite access, which could be used initially. Online access will require some software investment. Appendix 6 provides further detail.

Examples already exist

Online technology already exists that can provide researchers with secure, remote access to analyse data. Australia's Secure Unified Research Environment is an example, developed and operated by the Sax Institute (a not-for-profit in health research) with support from Commonwealth and NSW agencies.

Improve the dissemination of public sector data

Recommendation 10

Departments and agencies provide input to a searchable online data catalogue that includes details of non-sensitive major datasets (through metadata), to be published on data.gov.au and updated regularly.

Recommendation 11

Develop a framework for high-value datasets in consultation with data users.

Recommendation 12

Publish whole-of-government data standards and principles, drawing on existing (local and international) guidelines.

Make data discoverable

Data will be made more discoverable through a catalogue that lists and describes the major information stores across the Commonwealth, searchable via data.gov.au and available to all public sector data users.

An initial investment will be required to develop content for the catalogue and there should be a staged approach to populate the content, coordinated by a single agency.

Publish Data standards

In addition to adhering to record keeping guidance, clear data standards are needed, that will:

- outline technical specifications for data; and
- address electronic interchange of data.

The migration to new standards will involve some up-front costs for agencies. These should be outweighed by reduced development costs over time.

Set a framework for high-value datasets

The high-value dataset framework will identify datasets that hold greatest potential to support economic, social and environmental outcomes; and those regularly requested by businesses.

Categories of high-value data are outlined below. States and Territories hold a substantial amount of high-value data, and collaboration with the States and Territories to aggregate and link data is critical.

Companies	Government Processes
Crime	Health
Demographics	Immigration
Education	Infrastructure
Employment	Location
Energy	Science and Research
Environment	Trade
Finance	Welfare

The framework will include details on dataset availability (including release dates), accessibility, and licensing arrangements, as well as prescribing agency responsibility for ensuring the quality and availability of high-value datasets.

Develop consistent and transparent user charging for data

Recommendation 13

Where Australian Government agencies (non-corporate Commonwealth entities) publish data, it should be at no cost to the user unless there are costs associated with tailoring the data for publication.

Goals of user charging for data

The overarching goal is to maximise access to public sector data, without putting an unreasonable financial burden on agencies.

Principles

The Australian Government Charging Framework, overseen by Finance, should provide a consistent and transparent approach to charging for data. This should specify that government entities take into account the following considerations when determining data charges:

- public sector data that is not personally identifiable should be made available 'free by default'; and
- where resources are required to tailor data for specific requests, entities may consider recovering these costs, having regard to whether the data will be made publicly available.

The legislative environment needs to support the use of data

Recommendation 14

Create a legislative environment that supports the use of data while maintaining privacy.

In the **short-term**:

- staff should be trained on how to interpret privacy and secrecy laws. Guidelines should be published to support this, including guidelines from the ABS given its experience with de-identification protocols.

In the **medium-term**:

- legislation should be reviewed to identify whether privacy and secrecy laws can be streamlined and modernised to enable data to be better used for policy and research, building on a possible Productivity Commission inquiry.

Think differently about legislation

Significant gains can be made in the short-term by educating staff on how to interpret legislation to share and make better use of data.

This requires a change in mindset for staff to look for ways to make data available within the law. Health is seeking to link extracts of PBS and Medicare Benefits Schedule (MBS) data, to be shared with the States.

Publishing guidelines can also reduce risk aversion. It provides support for staff to take a pragmatic interpretation of the law. For example, the Privacy Guidelines published by the Information Commissioner have enabled staff to release data where it was previously not released due to uncertainty.

The guidelines could indicate that the term 'reasonably identifiable' should take into account:

- who will have access to information;
- what information they have; and
- the likelihood of being able to re-identify information.

Aligning objectives of legislation with the data agenda

In the medium-term, legislation should be reviewed to identify whether privacy and secrecy laws can be streamlined and modernised through an overriding principle-based law to enable data to be better used for policy and research.

If the Commonwealth progresses a Productivity Commission inquiry into data, as recommended by the Financial Systems Inquiry, the terms of reference could focus on the review of legislation.

NZ – Revisiting legislation

As part of their data agenda, in 2012 New Zealand amended its Statistics Act to facilitate access to individual-level data for non-government researchers.

The Hon. Bill English, Deputy Prime Minister said a rewrite of privacy laws had removed barriers to "common sense" solutions.[7]

Changes to the Act included 'guidance and clarity' on its interpretation.[8]

Statistics New Zealand is currently refreshing their legislation with a view to modernisation and better sharing of data.

7. Bill English, Deputy Prime Minister of New Zealand, June 2015 http://www.zdnet.com/article/big-data-to-deliver-nz-budget-shake-up/
8. Legal Counsel, Global Regulatory Affairs, June 2015 http://www.fadv.com/company/blog/entry/articletype/articleview/articleid/127/new-zealand-actively-considering-changes-to-privacy-act.aspx

Promote innovation in public administration

Recommendation 15

Promote innovation in public administration by implementing a suite of mechanisms including:

- Prime Minister's Awards for innovative use of data, providing time and resources to pursue data-driven transformative approaches;
- over time, creating institutional structures with academia, private and public sectors (e.g. leveraging cooperative research centres) to actively support innovation, in line with the UK and NZ;
- shared data laboratory capabilities where innovation can be fostered with minimal outlay; and
- incentivising new ideas for data-driven innovation through grants and/or competitions, with associated funding for trials of new approaches to public administration.

Better data management could help improve public sector administration

To improve public administration, an environment needs to be created that embeds innovation as core business, rather than an 'add-on'.

Examples across the public service include the DTO, the ATO's Smarter Data programme and the approach to shared services.

Innovation is derived through feedback

Research often results in new data which can produce new insights and lead to greater innovation. We therefore need to encourage and incentivise data users to feed their data back into the system so other users can build on their findings to foster innovation.

Structures between public and private sectors need to be institutionalised to drive innovation

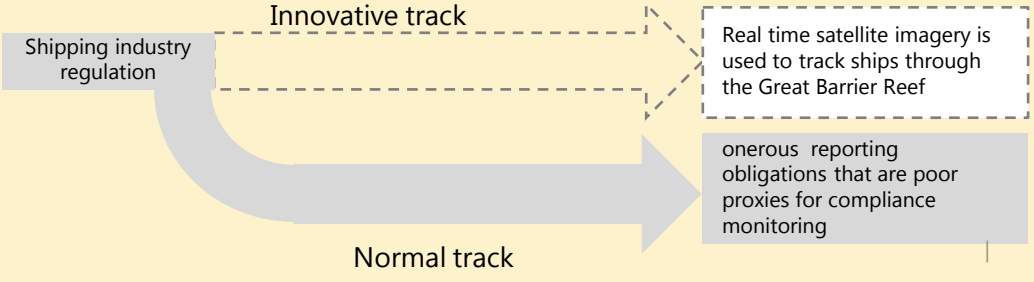
The US and the UK used 'hackathons' early in the data journey to drive private sector innovation and promote open data.

However, those further down the path are focusing on establishing structures between the public and private sectors to institutionalise innovation (e.g. Open Data Institutes in the UK).

The Commonwealth should consider establishing similar structures which bring together data suppliers within the public sector, and data users within the public and private sector. This would require funding. The Queensland Open Data Institute may provide a starting point for partnerships.

Innovation example

Real time satellite images can be used to unburden the shipping industry from reporting obligations that produce less effective outcomes



Appendices

Appendix 1 Definitions

Administrative data	Refers to information collected for delivering public administration – for example to conduct registration, transaction and record-keeping.
Big data	Big data is an evolving term that describes any voluminous amount of data that has the potential to be queried through analysis techniques to arrive at very specific information or answers to questions.
Data sharing	The transfer of data between organisations.
De-identified data	Data relating to a specific individual where the identifiers have been removed to prevent identification of that individual. Otherwise known as Anonymised Data.
High-value data	Data which would have a higher economic value if made available as open data due to its authoritativeness, timeliness, accuracy or other traits.
Linked data	Data created from matching and integration of two or more datasets. This may occur through either an explicit match on unique identifiers, or through a combination of information that gives a high confidence match between the datasets.
Non-sensitive data	Data that is anonymised and does not identify an individual or breach privacy or security requirements.
Open access	Provision of free and unrestricted access to information to the general public.
Open data	Open data refers to making datasets available so that others can use them without restriction on use or redistribution in its licensing conditions.
Personal data	As defined by the *Privacy Act 1988*, data relating to a specific individual where the individual is identified or identifiable in the hands of a recipient of the data.
Public sector data	Data collected or generated by the public service for policy development and public administration. Also known as Public Sector Information.
Re-identified data	Data where the identity of an individual has been ascertained from a de-identified dataset – usually through comparison with other datasets.
Trusted user	A trusted user is someone who is authorised with the appropriate security clearances and confidentiality agreements to access more sensitive public data, such as unit-record administrative data, for purposes such as research or policy development. Examples include public servants and members of research institutions.
Unit-record data	Data that is at the most granular level – for example unit-record patient data would contain personal data about the individual patient.

Appendix 2
Risks and realities of making progress

There are reputational, technical, and financial risks involved in pursuing the public data agenda.

Mitigating risks to the public data agenda

There are five key risks to transforming the way the Commonwealth manages public sector data. Each of these risks can be effectively managed by implementing sensible mitigation strategies.

Risks	Mitigation strategies
Misperception • Risk of community/media outcry from confusing public sector data with broader sensitivities such as meta-data retention, compliance or government transparency	• Understand community attitudes • Implement engagement strategy to build trust and clearly communicate purpose
Adverse policy findings • Risk that information is released which could lead to unfavourable findings about policy effectiveness	• Enable trusted access to linked data in a secure environment where the use of data can be tracked • Develop protocol on releasing findings • Build productive relationships with researchers
Compromised data • Risk that data is hacked by non-state actors or criminals • Risk that the accuracy and consistency of data is not maintained over its life-cycle, leading to flawed analysis	• Increase security against attacks on data systems • Ensure agencies know who has access to data, where it is stored geographically and in what form • Develop guidelines about data management • Prevent copies or release of data outside of agreed protocols and systems
Breach of privacy • Risk that privacy will be breached by matching data from different sources to reveal the identity of individuals – a phenomenon known as the 'mosaic effect'	• Enforce privacy legislation • Standardise de-identification protocols • Use a secure gateway to ensure trusted access only
Inadequate resources • Risk that agencies do not have the resources to curate, link and build the infrastructure needed to better manage public data	• Consider reinvesting savings from better using data within the agency • Prioritise data as part of core government business • Improve skills and capabilities

Appendix 3.1 Case study | Innovation
The US-based Climate Corporation uses public sector data to write crop insurance

Overview

- The Climate Corporation offers weather insurance for farmers based on more than 30 years of published government weather, soil and crop data.
- Today, the company is valued at more than US$1 billion.

What was the context?

- The Climate Corporation was founded in 2006 under the name Weatherbill.
- It offered insurance against weather-related incidents, such as concert cancellations.
- As the Corporation began to turn its attention to farmers, the company found that it needed **more and better data**.
- The Corporation noted the value in weather and climate data published by the US National Oceanic and Atmospheric Administration.

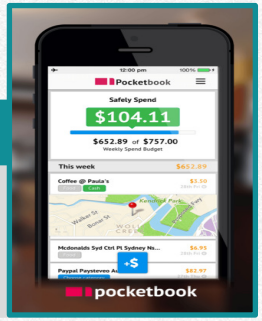
What did they do?

- The Corporation analysed 60 years of crop yield data and 14 terabytes of soil data — **all from government** — to design and price crop insurance.
- The Corporation tailored analysis based on a government network of 159 Doppler radar stations that scan weather data in two million locations.
- The Corporation now adds value to the agriculture industry by:
  - helping farmers manage risk by selling crop insurance; and
  - supporting decision-making on the farm with up-to-the-minute data for field-level monitoring, yield forecasting and crop insights.

What was the result?

- As a demonstration of the value Climate Corporation has created, Monsanto bought the company in 2013 for nearly **US$1 billion**.
- Farmers are benefiting from data-driven decisions and protections against bad weather.

Appendix 3.2 Case study | Innovation
Australian start-up Pocketbook leverages consumer banking data



Overview

- Sydney-based startup Pocketbook created a **consumer finance app** that connects customers to their data across multiple banks, and delivers insights and automation of personal finance tasks.
- Today, the app has over **100,000 Australian users** and delivers valuable targeted analysis based on consumer trends.

What was the context?

- Pocketbook was founded by two Sydney friends in August 2012 with the goal of **simplifying the process of doing taxes.**
- Pocketbook launched an app in 2013, allowing users to connect their bank data together **in real time.**
- Australian financial institutions were beginning to allow third party applications to access data through secure customer-authenticated access, setting an environment for innovation.

What did they do?

- Pocketbook allows users to **sync all of their different bank accounts and credit cards**. It then applies analytics to review users' transactions, categorises them, and predicts recurring bills and payments. It puts these into a budget and bill schedule so users get reminders and alerts.
- By focusing on the user experience, privacy by design, and value-add functions for consumers, Pocketbook has been able to build up a large user base and access significant amounts of consumer-spending data.
- A large user base has enabled Pocketbook to use real life spending data of consumers to create useful insights (de-identified, aggregated, and analysed internally to protect users' security and privacy).

What was the result?

- **Over 16 Australian banks now have data interfaces** that Pocketbook uses, building their user base to over 100,000. They have recently launched a tax calculator app and a tax returns app is in development.
- Insights from Pocketbook's real consumer data are augmenting existing sources such as household expenditure surveys.
- For example, the data shows that subscribers to the US Netflix service are not switching to Australia's Netflix, and Sydney CBD chemists are losing an average $98 spend per customer per month to the new nearby Sephora store.

Appendix 3.3 Case study | Service delivery
UK's Community Insight enables housing providers to tailor services based on data

Overview

- Community Insight is a UK web tool that allows social purpose organisations to access publicly available data on a location of their choice.
- It is today used by over 80 housing organisations in the UK to help them understand their local communities.

What was the context?

- Housing providers and organisations needed **social and economic data** on local areas to tailor and target their services.
- The valuable data was publicly available, but not accessible in a centralised, user-friendly way.
- The UK Government, the Housing Associations Charitable Trust (HACT) and Oxford University **partnered** to develop the idea for an online public data tool for community organisations.

What did they do?

- Community Insight developed a community-focused Geographical Information System system that collates and presents vital information on local communities.
- The system provides instant neighbourhood and community profiles for housing providers, drawing on the latest public data sources covering all major social and economic indicators.
- Organisations use the system to target and **prioritise** community investment and **inform** community consultations and planning decisions.
- The system reduces the burdens on data and analytical staff to provide community-focused information.

What was the result?

- The key success of the UK model was the ability for Community Insight to openly access a huge database of constantly updated public sector data.
- The system is now **used by over 80 UK housing associations** and government entities to tailor and target service delivery.
- The Australian Government is working with Community Insight Australia to identify and access data to develop a **similar capability for Australia**.
- Community Insight Australia has developed a draft website which is available at: www.communityinsightaustralia.org

Appendix 3.4 Case study | Policy
The Landsat Archive has been unlocked and used to track changes in land and water

Overview

- Valuable, archived Landsat satellite data has been unlocked by Geoscience Australia and ANU to enable detailed mapping and analysis of Australia's land and water.
- The next step is to extend the use of these datasets and tools to government, researchers and the private sector.

What was the context?

- The Landsat satellites have observed the Earth's surface since 1979, producing thousands of images.
- Because of the costs of data storage, this data has traditionally been stored in raw form on tape storage that had **limited use**.
- Requests for access took Geoscience Australia many months or sometimes years to compile.
- Consequently, this data was used rarely and in a piecemeal and labor-intensive manner.

What did they do?

- The Unlocking the Landsat Archive project was set up in 2011 to help realise the true value and potential of Landsat data.
- This was only possible because in 2008, the US declared that all of their Landsat Data was **open for use by anyone at no cost**.
- The project successfully established a large-scale data processing and science capability for earth observation data in Australia.
- Using the ANU's National Computational Infrastructure to process its archived Landsat data, Geoscience Australia has successfully retrieved 133,000 images.

What was the result?

- Drawing on the Landsat images, Geoscience Australia has **produced maps** of Australia's surface water patterns which provide unique information for flood risk assessment and ecosystem management.
- Image analysis can now be done **more efficiently**, within a matter of hours or days rather than months or years. A CSIRO survey that historically took thousands of hours to complete can now be completed in days.
- The next phase is to apply this rich data to many more government and commercial problems giving them real evidence for their decision-making.

Appendix 3.5 Case study | Policy
New Zealand uses public sector data to curb welfare dependency

Overview

- Since 2010, analytics has been at the heart of New Zealand's welfare reforms.
- By partnering with an analytics firm, the Ministry of Social Development has discovered ways to better target welfare policy to improve outcomes and save money.

What was the context?

- Social welfare accounted for nearly a quarter of New Zealand's gross domestic product.
- The Ministry for Social Development **spent $22 billion a year** providing services to more than a million New Zealanders in need.
- Data showed that 13% of the working population were on welfare – many of whom had been on benefits for a decade – so the agency identified change was needed.

What did they do?

- In 2010, the Ministry for Social Development contracted an analytics company to examine ways of reducing long-term welfare dependency **using big data**.
- By matching and analysing data across several government agencies, the Ministry discovered that more than 70% of its total welfare expenditure was attributable to those who entered the welfare system under 20 years of age.
- The Ministry was able to predict the probability of this population going on to an adult benefit and, in turn, offer targeted services intended to reduce their long-term benefit dependency.
- Such services included being matched with a personal mentor, learning budgeting skills, and receiving more education or training.

What was the result?

- **Employment rose** 9.3% in 2013 and **benefit levels reached a five-year low.**
- The Ministry applied similar analytics to sole parents and, through targeted investments in education and job placement, was able to **reduce sole parents on benefits** by 9.4%.
- After a few years, early results of the analytics approach have been positive, with projected **savings of $1 billion over four years.**

Source: Intelligence Quarterly, Journal of Advanced Analytics, 4Q, 2014, http://www.sas.com/content/dam/SAS/en_us/doc/other1/iq-q414.pdf

Appendix 4.1 | DHS Early interventions for troubled families

Project overview

Objective: earlier interventions to reduce the risk of long-term welfare dependency.

This project would link Commonwealth and state data to enable place-based insights into troubled families.

Problem

What is the policy problem?

International evidence shows the cost of troubled families is disproportionately high compared to average families.

Troubled families are often characterised by the absence of a working adult, children not fully engaged in education and members involved in crime and anti-social behaviour.

Without linked data on troubled families, governments cannot fully understand the risk factors involved to better target policies and design early interventions/investments.

What are the data barriers?

1. Data sharing across governments
    - Sharing data across jurisdictions has been limited by cultural and technical barriers and privacy concerns
2. Interpretation of legislation
    - It is not clear under what legislation this data can be accessed and used for research

Action

What could we do to solve it?

A joint project between the Commonwealth and one or more state agencies could link data on income, family structures, health and welfare by location. At the Commonwealth level, this would involve linking Social Security and related information data, Census data and the Australian Early Development Census.

This dataset would enable place-based insights into troubled families and those at risk of becoming troubled families to identify where services should be targeted.

Who could we work with?

- Commonwealth – ABS (Census), DHS/DSS (income support), Health (MBS/PBS), Education (AEDC). May also involve ATO (personal income data).
- States – state government data about troubled families (NSW Department of Families and Children has expressed interest in being involved with this project).

Outcome

What will the project allow us to do?

By understanding the patterns of behaviour of troubled families, more effective interventions can be developed to help break the cycle of entrenched disadvantage. Identifying troubled families by location would also enable place-based targeting.

Similar projects in NZ have resulted in more targeted services that lead to significant welfare savings.

What could we produce and when?

By 2016, a basic integrated view of troubled families would be developed using Commonwealth data.

By adding variables over time, such as crime and prison data, health data and school attendance data from the Commonwealth and states, a more comprehensive picture could be developed. This is expected to take up to two years.

Appendix 4.2 | Health Better targeting of mental health services

Project overview

Objective: better target mental health services.

The project will link MBS and PBS data with Census data to better assess mental health service provision, including prescribing.

Problem

What is the policy problem?

Around 1.5 million Australians receive over $1.7 billion through the MBS and PBS for mental health services, but large disparities exist. For example, antidepressant use is much higher in low socio-economic groups, and varies greatly between different regions.

Lack of evidence about the causes of these disparities prevents targeting of mental health services to those most in need.

What are the data barriers?

1. Interpretation of legislation
   - Legislative constraints on the use of MBS and PBS would be addressed
2. Data sharing across governments
   - Linking of MBS and PBS data to population characteristics obtained from the Census

Action

What could we do to solve it?

ABS has already linked some data on MBS and PBS services for mental health with Census demographic data.

Further data linkage, including doctor bulk billing and prescribing behaviour, plus more detailed analysis of the resulting dataset will help to explain disparities in prescribing patterns and service provision to inform how services should be targeted.

Who could we work with?

- Health will provide MBS and PBS data, and provide leadership on mental health policy issues.
- The ABS will act as data integration authority to allow MBS and PBS data to be linked with each other and with demographic data.

Outcome

What will the project allow us to do?

The project would allow the Commonwealth to better target its investment to:

- provide mental health care to those most in need, and
- ensure appropriate use of expenditure, and more efficient use of health funds.

What could we produce and when?

Data extraction, linkage and analysis would take around 6 months.

The resulting report would identify the causes of disparities in treatment of mental health issues (e.g. doctor behaviour, access barriers, need), highlighting issues requiring action to ensure equitable and appropriate mental health care.

Appendix 4.3 | PM&C Indigenous early childhood interventions

Project overview

Objective: improve educational outcomes for Indigenous Australians through better targeting of early childhood interventions.

Link Centrelink and state data (i.e. school attendance and attainment, child protection and hospital data), to test specific early childhood experiences and interventions on educational outcomes for Indigenous children.

Problem

What is the policy problem?

By Year 5, 30% of Indigenous children have not met minimum standards in reading and numeracy compared to 6% of non-Indigenous kids.

Governments spend around $4,700 extra per Indigenous student per year in school education*.

The series of life events and circumstances that influence educational attainment are not yet well understood.

What are the data barriers?

1. Interpretation of legislation
   - External release of Centrelink data must first be assessed against legislation and privacy principles
2. Data sharing across governments
   - Jurisdictional barriers have traditionally been difficult to overcome without extensive negotiation processes

Action

What could we do to solve it?

This project will partner with Northern Territory (NT) and South Australian (SA) Governments to link Centrelink data and state government programme data (school attendance and attainment, child protection and hospital data) to create a linked dataset.

Who could we work with?

- Commonwealth – DHS, DSS and Department of Education
- States – NT and SA Government departments
- Academics – Menzies School of Health Research

Outcome

What will the project allow us to do?

This project will inform targeted cross-agency responses to the factors underpinning educational disadvantage such as early childhood interventions, maternal and child health, and family support services.

What could we produce and when?

Phase 1: form partnerships with the data custodians, seek solutions to barriers and seek to extend ethics clearance.

Phase 2: link approved Centrelink data with the linked NT and SA dataset.

Phase 3: use the linked data to answer targeted research questions exploring the factors impacting on educational outcomes and provide a paper to Secretaries with options for policy responses.

*Commonwealth and State expenditure based on 2012-13 data.

Appendix 4.4 | Environment
Better harnessing environmental data

Project overview

Objective: improve environmental decision-making, resource allocation and strategic planning.

The project will re-use environmental data collected for approvals and compliance and improve discoverability of data for greater efficiency, increased value and reduced duplication.

Problem

What is the policy problem?

Over 7,500 development proposals have been considered by the Commonwealth since 2000. Reform processes are establishing an assessment One-Stop Shop. The required environmental data has re-use potential. It is often undiscoverable, may be regathered by multiple proponents at unnecessary cost, and has potential to improve decisions, resource allocation and strategic planning.

What are the data barriers?

Data sharing and access

- Sharing across jurisdictions
- Incorporating industry data
- Maximising the use of existing data

Improving the form and availability of environmental data

Integration and linking of data into existing systems

Action

What could we do to solve it?

The department is making spatial and non-spatial environmental data open access and is working with the jurisdictions to do the same. Better tools are required to improve usability of this data.

Work is also required to make environmental data provided as part of development proposal, monitoring and compliance processes discoverable and re-usable by the broader economy.

Who could we work with?

- State environment agencies
- Other Commonwealth agencies including Geoscience Australia and CSIRO
- Business and other proponents of development
- Organisations and institutions that store environmental data

Outcome

What will the project allow us to do?

- Reduce duplication and transaction costs
- Improve data availability and streamline decision-making
- Gain greater value from existing data holdings
- Improve understanding of cumulative impacts and inform strategic planning

What could we produce and when?

Further release of environmental data in open and accessible forms – end 2015

Guidelines for publishing proponent data and arrangements in place with data repositories – mid 2016

Improved tools for the dissemination and integration of data – mid 2016

Better re-use of proposal, monitoring and compliance data – 2017

Appendix 4.5 | Infrastructure
Analyse freight company data to plan road infrastructure

Project overview

Objective: identify road infrastructure bottlenecks and priorities by analysing GPS and logistics data collected by freight companies using 'big data' analytics tools.

Problem

What is the policy problem?

Total road freight in Australia measures over 200 billion tonne kilometres per annum, but there are few regular statistics on where road freight moves.

Understanding freight movements is important to infrastructure planning and investment decisions.

What are the data barriers?

- Obtaining sufficient industry cooperation, most particularly overcoming commercial confidentiality concerns

Action

What could we do to solve it?

- Data collected electronically by road freight operators (via GPS, telematics and logistics systems), suitably aggregated, would provide a source of detailed, timely road freight data.
- Cooperation between industry and agencies would allow publication of regular statistics on road freight movement throughout Australia.

Who could we work with?

- Department of Infrastructure and Regional Development (BITRE)
- ABS
- the 10-20 largest transport and logistics industry firms

Outcome

What will the project allow us to do?

- Provide information about key freight-impeding infrastructure bottlenecks
- Provide data needed to introduce productivity-enhancing innovation within the private sector
- Provide more frequent and up-to-date statistics on road freight movements across Australia

What could we produce and when?

- An interim report on agency/industry data collaboration efforts (December 2015).
- A report into the feasibility of 'big data' for freight transport statistics, including experimental statistics from pilot studies (December 2016).

Appendix 4.6 | Industry Analyse productivity in firms & evaluate policy

Project overview

Objective 1: analyse drivers of productivity and evaluate industry policy by interrogating a new longitudinal database that links firm level data from the ABS and ATO (Expanded Analytical Business Longitudinal Database (EABLD)).

Objective 2: enhance the EABLD by adding further data about businesses and introducing data on employees to create a Linked Employer-Employee Database

Problem

What is the policy problem?

There is a major gap in our understanding of how firm-level investment decisions impact on national aggregate productivity and employment.

The Commonwealth also has limited understanding of the effectiveness of the approximately $10 billion annual investment to improve business performance and increase productivity.

What are the data barriers?

A new longitudinal dataset established by the ABS on behalf of the Department of Industry (EABLD) has recently overcome barriers associated with linking firm-level data between ATO and ABS in a systematic longitudinal, integrated database.

Plans to enhance the EABLD with additional data, including intellectual property, and personal income tax, will need to address privacy concerns and technical issues around data structures.

Action

What could we do to solve it?

Conduct a research program using the EABLD, including:

- determining the factors that contribute most to productivity (skills, innovation, management)
- assessing the nature of innovation and R&D across Australian firms

Enhance the EABLD with additional data about businesses and employees to create a Linked Employer-Employee Database (LEED)

Who could we work with?

- Department of Industry
- ABS
- ATO
- IP Australia
- Australian Financial Security Authority

Examine opportunities to link data through:

- Standard Business Reporting
- Single Touch Payroll (STP)
- AUSkey

Outcome

What could we produce, and when?

- The EABLD currently has firm level longitudinal data (2001-02 to 2012-13) from ABS and ATO covering business characteristics and financial information.
- By December 2015, policy papers on innovation and R&D spillovers between firms.
- Progressive expansion of EABLD to create a LEED.

What will the project allow us to do?

Creation of an enduring firm level statistical asset will increase the capability and capacity of government to undertake firm level analysis of competitiveness, productivity and employment to target expenditure of government funds in its efforts to improve business performance and increase productivity.

Appendix 4.7 | ABS/Treasury
An online market research tool for SMEs

Project overview

Objective: make location-based market data readily available to small and medium businesses. This will be provided in a web-based form that can also be used by third party developers to build new services for business.

Problem

What is the policy problem?

96% of all Australia's businesses are small businesses, employing over 4.5 million people and producing over $330 billion of our nation's economic output per year.

SMEs often lack the market information to make good business decisions. Many do not invest in appropriate market research.

ABS and private sector data could be combined and made available to SMEs in a simple format providing low-cost, basic market research.

What are the data barriers?

Access to non-government administrative data e.g. real estate data.

The cost of putting a tool together.

Data will need to be provided in an easily digestible format for SMEs and also in a format for developers to create new services by merging public and private datasets.

Action

What could we do to solve it?

This project will develop a simple, online tool which SMEs can use to access location-specific information on:

- citizen demographics;
- businesses – competitors, complimentary businesses, performance;
- commercial real estate; and
- transport infrastructure.

ABS regional profiles for selected geographical areas would form the basis of the tool.

Who could we work with?

- Real Estate Institute of Australia / Property Council of Australia
- SMEs will advise upon the design of the website

Outcome

What will the project allow us to do?

The project will enable:

- SMEs to interrogate a number of datasets for improved decision making; and
- developers to re-use the open datasets to create services for business.

We will derive direct benefit from the website by:

- tracking data use/downloads and identifying common trends/threads; and
- linking this administrative data to other linked data projects for re-use.

What could we produce and when?

An online data prototype for SMEs could be developed by June 2016, using ABS data for selected regions.

The data and geography would be extended over time, including private sector data.

Appendix 5 Data can be linked by the ABS, but the challenge is access by others

1. Legal basis has been determined

Under current legislation the ABS has **linked key datasets to ABS data,** creating valuable research assets.

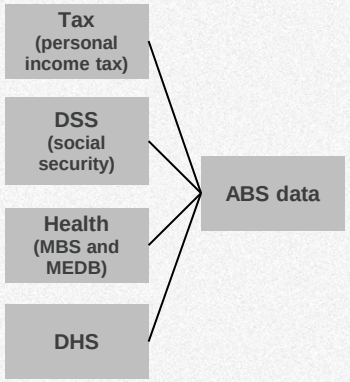
To date, key outputs have been research publications and interactive tables.

Some agencies also work within the ABS on their own research projects.

2. Bilateral linking has occurred

ABS began linking key datasets from each agency with ABS data in 2006.

Bilateral linkages do not unlock the full potential.



3. Multi-agency linking project is underway

The ABS and relevant agencies are working to link key datasets across the ABS, DSS, Health, DHS and the ATO.

This will create a non-identifiable data asset to help agencies and researchers respond to nationally important policy and service delivery questions.

The initial phase is expected to be completed within **six months** and will build over time.

4. Past policies have restricted access

To get the full value of the data, APS staff and researchers need access to the data.

However, past policies have **prevented the ABS from giving researchers and agencies** convenient access to the linked data.

Current approaches have improved the situation.

5. APS Capability Review is improving data access

The ABS **is currently modernising its policies and approaches** to collaborative data sharing, access and use.

A review of relevant policies and practices is expected to be **completed toward the end of 2015**.

Integration

The issue is improving access

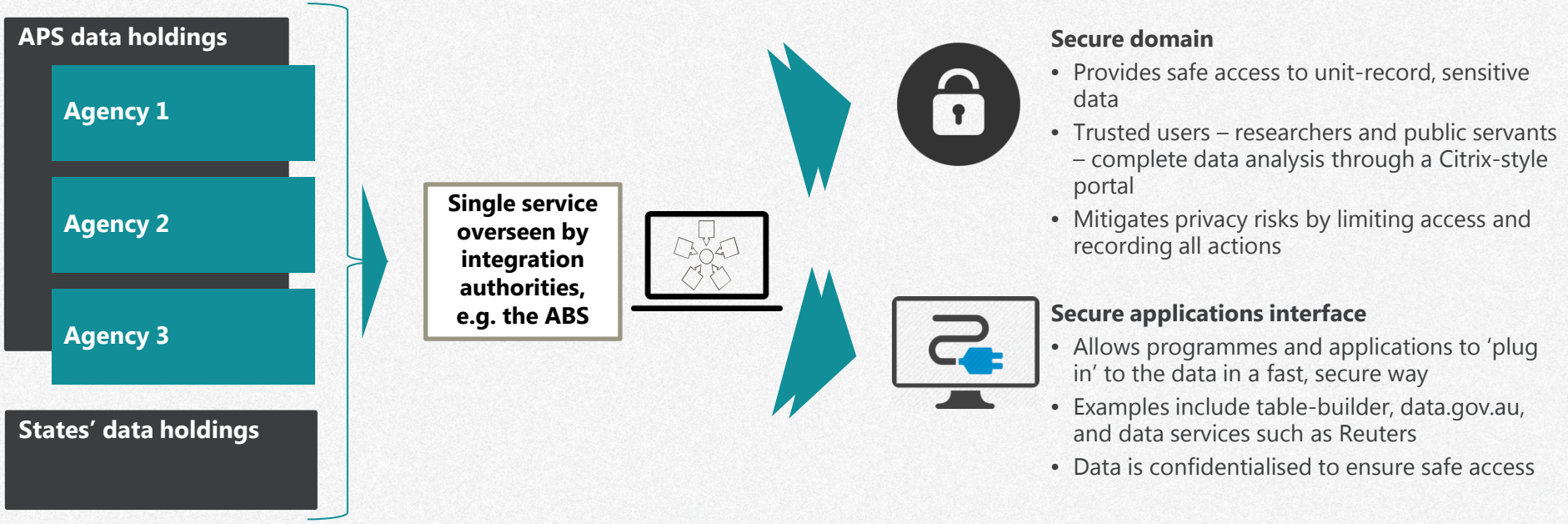
Appendix 6
A trusted-access model

Sources of data… → **…brought together…** → **… and shared securely**

- **Agencies maintain their own datasets, on their own data systems**
- **Common protocols improve usability/sharing of data**

- **Centralised approach streamlines data linking and access**
- **Integration authorities oversee linking to address ethical / practical issues**

- **Tiered access manages privacy risks while enabling the right users to access and use the data they need**



Secure domain

- Provides safe access to unit-record, sensitive data
- Trusted users – researchers and public servants – complete data analysis through a Citrix-style portal
- Mitigates privacy risks by limiting access and recording all actions

Secure applications interface

- Allows programmes and applications to 'plug in' to the data in a fast, secure way
- Examples include table-builder, data.gov.au, and data services such as Reuters
- Data is confidentialised to ensure safe access