Australian Government

**Department of the
Prime Minister and Cabinet**

The Department of the Prime Minister
and Cabinet's submission to the
**Productivity Commission's Inquiry into
Data Availability and Use**

22 July 2016

Table of contents

Executive summary

The Department of the Prime Minister and Cabinet (the Department) is pleased to make this submission to the Productivity Commission's *Inquiry into Data Availability and Use*.

The sections in this submission align with the topics of the Productivity Commission's issues paper released on 18 April 2016. This submission provides information about key data-related policies, programmes and initiatives at the Commonwealth level; as well as examples of how some policy line areas within the Department are using data to further their policy objectives. This submission also includes suggested areas for consideration by the Productivity Commission as part of its inquiry.

Significant progress has been made to advance the public data agenda within Australia. To date, the Government has:

- released the *Public Data Policy Statement*, which provides a clear mandate for Australian Government entities to optimise the use and re-use of public data;
- published over 8,200 datasets on data.gov.au and made over 4,300 API enabled resources available;
- made high-value datasets openly available, including the Geocoded National Address File and Intellectual Property Government Open Live Data;
- developed the NationalMap, a data visualisation tool for accessing over 2,800 government geospatial datasets;
- invested $75 million in Data61 – Australia's largest data innovation group;
- built partnerships between the public, private and research sectors to leverage public data for social and economic benefit;
- commissioned cross-agency projects to improve public data management;
- begun work on the Multi-Agency Data Integration Project to improve how government entities securely share data; and
- established the Digital Transformation Office, which is leading the digital transformation of government service delivery.

However, there is still work to be done. The Productivity Commission's *Inquiry into Data Availability and Use* provides an opportunity to assess the impact of data policies and initiatives to date and identify the barriers and challenges going forward. In particular, the Department would like the Productivity Commission's advice on the following issues:

- what barriers (legislative and otherwise) currently exist that prevent the collection, sharing, publication and use of public data;
- whether there is merit in pursuing overarching Commonwealth legislation to improve public access to data and data sharing across Australian Government entities;
- how the social, political and economic impacts of the public data policies can be better measured and reported; and
- how different individual preferences regarding privacy can be accommodated when developing data policies and initiatives.

Introduction

The value of public data

Public data is a valuable national resource and access to it for use and reuse can help grow the economy, improve service delivery and transform policy outcomes. The range and depth of public data is enormous and it is important that it is effectively managed to maximise its availability and use.

In its 2013 report, Deloitte conservatively estimated the direct economic benefits of UK public sector data at around 1.8 billion pounds per annum.[1] In 2014, Lateral Economics estimated that the aggregate direct and indirect value of government data in Australia is up to AU$25 billion per annum.[2]

On a global scale, the McKinsey Global Institute has identified that the world-wide annual value of open data could be as much as US$3 trillion.[3]

Value derived by the private sector applying innovation to public data also has the potential to substantially impact economic growth. PwC Australia estimated that in 2013 data-driven innovation added AU$67 billion in new value to the Australian economy, or 4.4 per cent of GDP. However, PwC Australia found that there is substantial room to improve, with an estimated AU$48 billion in potential value from data-driven innovation not yet realised.[4]

A Bureau of Communications Research report of 2016 found that "*while there is little consensus on the magnitude of the economic benefits of open government data sets, it is apparent that they provide substantial current and potential net benefits to the economy and society.*"[5]

Public data governance

The Public Data Branch within the Department has responsibility for data policy across the Australian Public Service, including management of the Government's open data platforms, data.gov.au and the NationalMap.

The Secretaries Data Group and Deputy Secretaries Data Group provide governance for public data initiatives across Australian Government entities. A network of senior Commonwealth officials, referred to as the Data Champions, promote the use, sharing and reuse of data across entities.

---

[1] Deloitte, *Market Assessment of Public Sector Information*, May 2013
[2] Lateral Economics, *Open for Business: How Open Data Can Help Achieve the G20 Growth Target*, June 2014
[3] McKinsey Global Institute, *Open Data: Unlocking Innovation and Performance with Liquid Information*, October 2013
[4] PwC Australia, *Deciding with data: How data-driven innovation is fuelling Australia's economic growth*, PricewaterhouseCoopers for Google Australia, September 2014
[5] Bureau of Communications Research, *Open Government Data and Why it Matters*, February 2016

Public Sector Data Management Report

On 3 December 2015, the Department published the *Public Sector Data Management Report.*[6]

The report found that there are pockets of excellence across the Australian Public Service, with some Australian Government entities undertaking projects that focus on a richer analysis of the outcomes of data integration. However, the report found that this approach is fragmented and is subject to a number of barriers, both perceived and real. These include cultural and legislative barriers and a shortage of data analytics skills and capability across the Australian Public Service.

The report established a roadmap of actions to address identified barriers and improve data-related practices across the Australian Public Service. Implementation of these actions is overseen by the Deputy Secretaries Data Group and an implementation report was published on 22 July 2016.

Public Data Policy Statement

On 7 December 2015, the Australian Government released its *Public Data Policy Statement.*[7]

The *Public Data Policy Statement* provides a clear mandate for Australian Government entities to optimise the use and reuse of public data; to release non-sensitive data as open by default; and to collaborate with the private and research sectors to extend the value of public data for the benefit of the Australian public.

The *Public Data Policy Statement* seeks to bring about long term transformational change in Commonwealth entities' cultures and attitudes to the use and accessibility of public data, in addition to improving the quality of public data made openly available.

Data61

As part of the National Innovation and Science Agenda, the Australian Government committed $75 million over the forward estimates to 2019-20 to ensure Australia builds and maintains a world leading data science capability. Data61 was funded to:

- use data analytics to connect disparate government datasets and publicly release them on open data platforms;
- improve industry cyber security and develop new cyber security architectures;
- build a Data Research Network that will link businesses with data researchers; and
- deliver data analytics training to businesses.

---

[6] The Department of the Prime Minister and Cabinet, *Public Sector Data Management Report*, December 2015
[7] The Australian Government *Public Data Policy Statement*, December 2015

1. High-value public sector data

As outlined in the Australian Government's *Public Data Policy Statement*:

> *"Australia's capacity to remain competitive in the digital economy is contingent upon its ability to harness the value of data… Publishing, linking and sharing data can create opportunities that neither government nor business can currently envisage."*

Increasing the availability and use of this valuable national resource can enable public and private sector efficiency, drive innovation, improve service delivery and transform policy outcomes.

The *Public Data Policy Statement* requires Australian Government entities to *"make high-value data available for use by the public, industry and academia, in a manner that is enduring and frequently updated using high quality standards"*.

The *Public Data Policy Statement* requires that, at a minimum, Australian Government entities will publish appropriately anonymised government data by default:

- on or linked through data.gov.au for discoverability and availability;
- in a machine-readable, spatially-enabled format;
- with high quality, easy to use and freely available application programming interface (API) access;
- with descriptive metadata;
- using agreed open standards;
- kept up to date in an automated way; and
- under a Creative Commons By Attribution licence unless a clear case is made to the Department for another open licence.

The *Public Sector Data Management Report* identified the following categories of high-value data:

High-value data categories	
Companies	Government Processes
Crime	Health
Demographics	Immigration
Education	Infrastructure
Employment	Location
Energy	Science and Research
Environment	Trade
Finance	Welfare

Characteristics of high-value data

The Department submits that high-value data is that which has quantifiable benefit to government entities, business, academia, research and community sectors. High-value data can be:

- **High-value for government** – this data will have potential value and outcomes for government through providing more insights for better decision making, policy development, improved service delivery, and efficiencies.
- **High-value for the private sector and community** – this data will have value for the private sector and the broader community, including enabling innovation, commercialisation and the provision of consumer services, as well as increased entrepreneurship, improved research outcomes, and better business and community decision making.

Other characteristics of high-value data may include:

- **Demand:** Clearly one of the best measures of whether data is high-value for the private sector and community will be requests for access by these sectors. At the Commonwealth level, requests for access to public data can be made via data.gov.au or directly with the government entity that holds the data. If access to data is denied by an entity, users may appeal the decision using the public request functionality available through data.gov.au. Users may also request access to data via data@pmc.gov.au.
- **Quality:** Such as the granularity of the data, frequency of updates, associated metadata, completeness, accuracy, timeliness, reusability, provenance and confidence.
- **Accessibility:** Data provided is discoverable and usable, including provision of data in machine-readable, spatially-enabled formats.

National and international definitions of high-value data

When defining the characteristics of high-value data, examples from other national and international jurisdictions can be drawn upon. For example, G8 countries have identified the following categories for high-value data:[8]

Data category	Example datasets
Companies	Company/business registers
Crime and Justice	Crime statistics and safety
Earth Observation	Meteorological/weather, agriculture, forestry, fishing and hunting
Education	List of schools, performance of schools and digital skills
Energy and Environment	Pollution levels and energy consumption
Finance and Contracts	Transaction spend, contracts let, call for tender, future tenders, local budget and national budget (planned and spent)
Geospatial	Topography, postcodes, national maps and local maps
Global Development	Aid, food security, extractives and land
Government Accountability and Democracy	Government contact points, election results, legislation and statutes, salaries (pay scales) and hospitality/gifts
Health	Prescription data and performance data
Science and Research	Genome data, research and educational activity and experiment results
Statistics	National Statistics, Census, infrastructure, wealth and skills
Social Mobility and Welfare	Housing, health insurance and unemployment benefits
Transport and Infrastructure	Public transport timetables and access points broadband penetration

---

[8] G8, *Open Data Charter and Technical Annex*, June 2013

The NSW Government's *Open Data Policy*[9] prioritises the release of high-value datasets in line with demand from the pubic and industry, as a result of consultation, or where the release of the datasets will contribute to better service delivery in NSW.

In its *Open Data Action Plan,*[10] the NSW Government indicated it will develop a standard checklist for establishing the value of a dataset. The action plan also states that high-value data such as fiscal, extraction industry, environment and pollution, aggregated and statistical data will be prioritised for release.

Examples of high-value data

Spatial data

Spatial data (or location based data) is information about geographic features and boundaries, either current or historic. Spatial data includes information such as addresses, natural or constructed features, co-ordinates, postcodes and statistical areas.

Spatial data can often be made openly available as it is usually free of privacy and legislative barriers. Spatial data is of critical importance to government in the delivery of its functions, including natural disaster mitigation and recovery. When spatial data is integrated with other data, such as de-identified medical or employment data, rich insights can be obtained for decision-making and policy development. Spatial data also has significant value to the private sector, with location based services being a multi-billion dollar market.

ANZLIC – the Spatial Information Council

The Department provides secretariat support for ANZLIC – the Spatial Information Council, which is the peak intergovernmental organisation responsible for developing policies and strategies to promote accessibility and usability of spatial information in Australia and New Zealand.

ANZLIC comprises senior officials from the Commonwealth, State and Territory Governments of Australia and New Zealand who are responsible for coordinating spatial information policy and operational matters within their jurisdictions.

The states and territories collect a range of spatial data, which, when aggregated at the national level, holds significant value for Australia.

ANZLIC's Foundation Spatial Data Framework provides a common reference for the assembly and maintenance of Australian and New Zealand foundation level spatial data in order to serve the widest possible variety of users. Data with similar characteristics are grouped in themes in order to improve the efficiency and effectiveness of information management processes. The Foundation Spatial Data Framework is divided into ten themes: geocoded addressing, administrative boundaries, positioning, place names, land parcel and property, imagery, transportation, water, elevation and depth, and land cover.

---

[9] NSW Department of Finance, Services and Innovation, *NSW Government Open Data Policy*, 2016
[10] NSW Department of Finance, Services and Innovation, *Open Data Action Plan*, 2016

The Foundation Spatial Data Framework is being developed from a conceptual framework to a valuable tool that will have wide-spread application. The Framework's database now contains the details of hundreds of jurisdictional datasets that contribute to the national spatial holdings. Information about where and how to access the spatial data is being collated in a cloud-based system and a prototype is under development.

Australian Geoscience Data Cube

The Australian Geoscience Data Cube is a highly successful prototype capability developed by Geoscience Australia in partnership with the Commonwealth Scientific and Industrial Research Organisation and the Australian National University's National Computational Infrastructure. It organises, calibrates and provides access to almost a petabyte of free satellite data providing a national picture of the Australian landscape at paddock scale for the last 29 years.

Operationalising the Australian Geoscience Data Cube would provide a whole-of-government shared big data analysis capability that will enable Australia to measure and detect changes over time in every 10x10 metre square of the Australian landscape. It will essentially unlock the full value of satellite data over Australia and lower the technical barriers to interrogating these large volumes of data for the national good.

Areas for consideration

There are a number of spatial data priorities that if progressed could potentially provide significant value to the economy, including improved management of land title information, geospatial imagery (such as operationalising the Australian Geoscience Data Cube), spatial data infrastructure and national positioning infrastructure.

Consideration should be given to the costs and benefits of each of these policy areas and models for implementing holistic reforms to achieve efficiency and economic outcomes.

Statistical Spatial Framework

The Australian Bureau of Statistics is addressing the need for, and challenge of, better integration of geospatial and statistical data by developing a Statistical Spatial Framework. The framework provides Australia with a common approach to connecting socio-economic data to a physical location. It also improves the accessibility and usability of geospatially-enabled data, and provides the mechanisms to support the integration of data at both the micro (unit record data) and macro level (aggregate data).

The framework is applicable to many data types, including statistical and administrative datasets. It uses location as a bridge to enable environmental and geospatial data to be connected with statistical and administrative datasets. The spatial integration of these diverse sources of data enriches the research and analysis that can be used to inform decision making.

The framework principles have been endorsed and implemented by a range of government entities, including the Australian Bureau of Statistics. Internationally, there is strong interest in the framework, where it is has been developed into a Global Statistical Geospatial Framework and is being considered for endorsement by the UN Statistical Commission and the UN Committee of Experts on Global Geospatial Information Management in 2016-17.

Geocoded National Address File

Since the *Public Data Policy Statement* was published on 7 December 2015, a number of high-value datasets have been made openly available, including the Geocoded National Address File (G-NAF).

The G-NAF is Australia's authoritative address file, containing more than 13 million records for each physical address in Australia.

The G-NAF is one of the most ubiquitous and powerful spatial datasets which can be used for a wide range of business or operational purposes, such as infrastructure planning, business planning and analysis, logistics and service planning, emergency and disaster response, personal navigation and mapping, fraud prevention, address validation at the point of entry for business and government, and effective government service delivery and policy development.

Denmark found that releasing its national geocoded address data at no charge to end users generated around 62 million Euros for the Danish economy over five years from 2005-2009.

The first release of the G-NAF was published on data.gov.au in February 2016, with updates published quarterly. Since the G-NAF became openly available it has been downloaded almost 1,200 times (as at 14 June 2016).

One of the first innovations that became available was Mappify's 'free geocoding for all Australia'. This free geocoding service converts a list of addresses to coordinates and plots them on a map without the need for third party geocoding services. This provides a simple and practical tool for a broad range of end users (not just data specialists) to use the data within their own businesses and organisations.

Through the use of address data contained in G-NAF, the NRMA has developed a tool called 'Safer Homes'. Safer Homes provides citizens with risk levels associated with water leaks, bush fire, home fires and theft with a specified address. The risk levels are calculated based on data including internal claim statistics, detailed analysis of weather conditions, vegetation and geography; local crime data and local bushfire prone area information. Access to this information enables citizens to have better awareness of and be more proactive about managing risks associated with their property.

Other examples of third party innovation include the G-NAF being used by:

- openaddresses.io, which is a free and open global address collection;
- participants at the Innovate@Locate Hackathon using the datasets to create new and experimental innovations; and

- Addressify.com.au, a cloud based web service that enables Australian address auto-complete verification and validation geocoding, including via API.

Other examples of high-value data

Further examples of high-value public datasets that have been released via data.gov.au since the publication of the *Public Data Policy Statement* include:

- Intellectual Property Government Open Live Data, which is a weekly release of intellectual property information;
- new data from the Australian Securities and Investments Commission, including datasets of registered auditors, Australian financial services authorised representatives, Australian financial services licensees, licensed liquidators, credit licensees and credit representatives;
- data from the Bureau of Infrastructure, Transport and Regional Economics, including data underlying publications on airline performance and road safety;
- the Australian Taxation Office has published corporate tax transparency data. This data has been used widely in Australian media publications; and
- publicly funded research data created as a result of the Bioregional Assessments Programme.

Areas for consideration

As outlined on page 2, there are varying estimates of the value of public data. However, to date there has been little cost/benefit analysis of the impact of releasing high-value datasets into the Australian economy.

As noted in the Productivity Commission's *Issues Paper on Data Availability and Use*, Australia currently ranks 10th in the World Wide Web Foundation's Open Data Barometer. This is largely due to Australia's low ranking for data impact (i.e. political, social and economic outcomes). Greater focus on the outcomes of the public data agenda, rather than just the outputs, could draw attention to priority areas that need to be addressed and help improve Australia's open data ranking internationally.

High-value dataset framework

In accordance with the *Public Data Policy Statement* and under the governance of the Secretaries Data Group, the Deputy Secretaries Data Group and the Data Champions network, Australian Government entities will continue to make non-sensitive public data openly available with the aim of releasing high-value data that is useful for the private and research sectors.

In the Government's policy for *Better and More Accessible Digital Services*,[11] the Government committed to working with research, not-for-profit and private sectors to

---

[11] The Government's policy for *Better and More Accessible Digital Services*, June 2016

identify high value public datasets for release. This will be supported by a public registry of significant non-sensitive datasets yet to be published on data.gov.au.

The Department is developing a *High-Value Dataset Framework* to assist entities and data custodians in identifying high-value datasets for priority release. In accordance with the Government's policy for *Better and More Accessible Digital Services*,[12] the Framework will be developed in collaboration with businesses, academia, research and community sectors. It will also draw on the Productivity Commission's findings regarding characteristics of high-value datasets.

2. Collection and release of public sector data

Public data initiatives and tools

Australian Government entities manage a number of projects and initiatives to encourage the collection, release and use of public data.

data.gov.au

Data.gov.au is the Australian Government's open data platform. It provides an easy way to find, access and reuse public data. The Department works across governments to publish data and improve functionality based on user feedback. The 'toolkit' provided on the data.gov.au platform provides guidance on integrating and publishing data, as well as information on improving published data.

As at 14 June 2016, there were over 8,200 datasets available via data.gov.au, with over 4,300 API enabled resources. The platform is also linked to other Commonwealth, state, territory and local government open data platforms through search result sharing.

While progress has been made to make more public data openly available, many government entities still do not make their data discoverable through data.gov.au. The Department is working to further increase the number of entities that publish data on data.gov.au, as well as working with entities to publish high-value data.

While information about which entities publish their data is available from data.gov.au, the Department is considering options for communicating this progress more openly, for example, via reporting dashboards or league tables.

NationalMap

The Australian Government has also created the NationalMap, a data visualisation tool for openly available spatial data. The NationalMap is an open source resource designed to facilitate better use of public data, allowing users to visualise and combine geospatial data. This can enable users to better understand and yield valuable insights from spatial data.

---

[12] *Ibid*

The NationalMap is built using open source code which enables users to create applications and new businesses. For example, PropellerAero has used the NationalMap source code to develop a product that enables organisations to process, host, analyse and share drone survey data in the cloud.

TableBuilder

The Australian Bureau of Statistics offers TableBuilder, an online tool for creating customised, confidentialised tables and graphs using Australian Bureau of Statistics microdata. This tool provides users with enhanced data accessibility and analytics.

Public-Private Partnerships

On 2 November 2015, the Department launched DataStart. DataStart is a public-private collaboration that aims to give Australian startups new opportunities to develop sustainable businesses through access to public data.

DataStart is an example of startups, incubators, the corporate sector and government working together to deliver data-driven innovations. The initiative is supported by established Australian incubator, Pollenizer, as well as corporate partners, including Optus, Google, PwC Australia, Data61, Rozetta and the Australian Information Industry Association.

Over 200 applications were received to the DataStart pilot, with cohortIQ announced as the winner in January 2016. cohortIQ is a health startup that uses hospital and open public data to reduce the estimated 235,000 avoidable hospital admissions each year.

cohortIQ is currently undertaking a nine month incubation with Pollenizer to develop and scale their business. Support is provided by the corporate partners and the Public Data Branch within the Department, who provide assistance to the startup in accessing anonymised public data, customer discovery and technical support. Right Click Capital and Pollenizer Investments invested $200,000 in seed funding in cohortIQ.

The private sector also chose to independently award prizes to three other startups:

- Mezo Research: A company passionate about ensuring sustainable natural resource management through advanced data analytics.
- Gemini3: An online platform that uses public data to facilitate job sharing and enable greater flexibility, diversity and retention in the workplace.
- Comployment: A compliance tool which leverages open government data to assist small business.

In the Government's policy for *Better and More Accessible Digital Services*,[13] the Government committed to expanding the DataStart programme to provide more opportunities for startups to collaborate, test their ideas and partner with government.

---

[13] *Ibid*

The Department also continues to participate in the Open Data 500 Global Network. The network has six members, including Australia, Mexico, United States, Italy, Korea and Canada. It meets regularly via teleconference to facilitate peer learning on mechanisms to foster demand and encourage the use of public data.

Areas for consideration

Building on the success of these initiatives, consideration could be given to how government entities can most effectively build public-private partnerships that facilitate knowledge sharing and promote innovation from access to public data.

Government data communities

The Open Data Government Community forum is a bimonthly meeting of government open data practitioners from across Australia. The forum connects operational staff across jurisdictions to share experiences and disseminate information. On average, about 120 people from across multiple Australian jurisdictions participate in the forum.

Data use and sharing by Australian Government entities

Work is also being undertaken to ensure Australian Government entities use their data collections more efficiently and effectively.

Digital Continuity 2020 Policy

The National Archives of Australia's Digital Continuity 2020 Policy enables the integration of information governance principles and practices into the work of government entities and their governance arrangements to:

- optimise the delivery of government programmes and services;
- enable information reuse for economic and social benefits; and
- protect the rights and entitlements of Australians.

The policy promotes a consistent approach to information governance across Australian Government entities. It applies to government information, data and records, as well as systems, services and processes, including those created or delivered by third parties on behalf of Australian Government entities.[14]

Cross-agency data projects

Several high-value cross-agency projects have been commissioned across Australian Government entities to demonstrate the value of public data, uncover barriers to use and enable better designed policies and services. The cross-agency projects include:

---

[14] The National Archives of Australia, *Digital Continuity 2020 Policy*, October 2015

- **Early interventions for troubled families to reduce the risk of long-term welfare dependency** – led by the Department of Human Services.
- **The better targeting of mental health services** – led by the Department of Health.
- **Improve educational outcomes for Indigenous Australians through better targeting of early childhood interventions** – led by the Department.
- **Analyse drivers of productivity and evaluate industry policy** – led by the Department of Industry, Innovation and Science.
- **Improve environmental decision-making, resource allocation and strategic planning** – led by the Department of the Environment.
- **Analyse freight company data to help plan road infrastructure** – led by the Department of Infrastructure and Regional Development.
- **Produce an online market research tool to make location-based market data readily available to small and medium businesses** – led by the Australian Bureau of Statistics.

These projects target key outcomes for government, use innovative approaches to achieve results and breakdown barriers for future work.

Data61 projects

As part of the National Innovation and Science Agenda, Data61 is working to deliver high-value, transformative data projects, including:

1. **Secure API access for the Multi-Agency Data Integration Project** – led by the Australian Bureau of Statistics

This project will develop software that allows open data platforms to interactively access confidential socio-demographic data from the Multi-Agency Data Integration Project (discussed further in Section 3). The project aims to demonstrate how public datasets can be used to deliver customised queries using standard systems while maintaining privacy. The project will reinforce the 'open by default' principle of access to data.

2. **World leading Public Data infrastructure** – led by the Department

Data61 will expand the existing data.gov.au and NationalMap infrastructure to maximise the discoverability and reuse of high-value open data in government, industry and the community sectors by streamlining publishing, improving data quality and enabling better search and discovery.

3. **Open Data Access Infrastructure** – led by the Department of Social Services

This project will develop software and a user interface for use by data researchers to access de-identified social security and family payments at the unit record level and longitudinal data surveys. This project will optimise the value of the Department of Social Services' Investment Approach data by making the data ready and remotely accessible for researchers, while complying with legislated requirements for using protected information.

4. **Synthetic Social Security Payments Dataset** – led by the Department of Social Services

This project will develop algorithms that enable the safe release of a synthetic representation of a dataset covering 15 years of social security and family payments, linked at the individual level. This synthetic dataset could be made openly available to researchers, as it removes the connection between real individuals and their data, alleviating confidentiality risks while retaining the overall characteristics of the original dataset.

5. **Significantly improving our understanding of Australian firms** – led by the Department of Industry, Innovation and Science

This project will improve the Expanded Analytical Business Longitudinal Database (discussed further in Section 3), including its scope and accessibility and thereby to provide data for government and private industry to better understand Australian firms and industry. This project will expand the range of data accessed via the Expanded Analytical Business Longitudinal Database, and the use of technology solutions to enhance access and security (including access by researchers) and facilitate maintenance of the Expanded Analytical Business Longitudinal Database in the medium term.

6. **Regulation as a platform – business tools to reduce cost burdens** – led by the Department of Industry, Innovation and Science

This project will develop and maintain an open business regulation platform that acts as an official open database of business regulation, in a format the ICT industry can leverage, to act as a digital platform for next generation business tools and services. This will support an open market for the development of these advanced tools for business and enable the use of analytics and advanced computation to guide reform and improve whole of regulatory system design.

Barriers to sharing and publishing data

On 11 December 2015, the Department held a workshop to explore legislative and other barriers to sharing and publishing data. The event was attended by 80 participants from 34 government entities. The main barriers identified included:

- **Cultural:** Embedded behaviours and aversion to risk preventing the uptake of open data policies.
- **Resourcing or procedural:** Lack of Australian public servants with the requisite data skill and capability, as well as poor and time-consuming internal or external processes.
- **Legal and regulatory:** Legislative and regulatory restrictions preventing the sharing and releasing of data.
- **Governance and policy:** Overcomplicated and time-consuming data sharing arrangements, both within and between government entities.
- **Infrastructure:** Outdated and siloed technological systems, which do not support interoperability.

> **Areas for consideration**
>
> Building on the intent of the *Government Information (Public Access) Act 2009*, the NSW Parliament passed the *Data Sharing (Government Sector) Bill 2015* to enable data sharing across government entities and support the functioning of the NSW Data Analytics Centre.
>
> While there are clearly different considerations at a Commonwealth level, investigation into the effectiveness and impact of the NSW legislation would be valuable for determining whether there is any justification for pursuing overarching Commonwealth legislation to improve public access to data and data sharing across Australian Government entities.

Standards

A consistent approach to the use of standards is vital to realising the benefits that the improved use and reuse of public data can bring. Open data standards facilitate data interoperability and accessibility, which further enables government partnerships with public, private and research sectors.

While the *Public Data Policy Statement* commits Australian Government entities to use agreed open standards, the policy avoids being prescriptive and does not name specific standards that entities must use.

> **Areas for consideration**
>
> Given that standards are essential to realising the full value of public data, consideration could be given to how Australian Government entities can best be assisted with adopting and applying standards when making data openly available.
>
> Consideration could also be given to how Australia can better participate at an international level to shape and inform data standards.

3. Data linkage and integration

Data linkage and integration provides richer insights and enables government and researchers to better design and evaluate policies and improve service delivery. This can deliver efficiencies and economic and social outcomes for Australians.

Examples of data linkage and integration

Cross-jurisdictional projects

As outlined in the *Public Data Policy Statement*, Australian Government entities are engaging with the States and Territories to "*share and integrate data to inform matters of importance to each jurisdiction and at the national level*".

Building on existing cross-jurisdictional data sharing and integration initiatives in a range of areas including crime, environment, health, education and transport; jurisdictions are investigating opportunities for further collaboration, particularly in the areas of social policy, health and education.

In addition, the Council of Australian Governments has agreed to enhance transparency by providing Australian citizens with a greater level of real-time data on how government money is spent and on the outcomes and performance of government initiatives.

Australian Government Longitudinal Data Architecture Review

The key object of the Australian Government *Longitudinal Data Architecture Review* is to inform the Australian Government's future investment in longitudinal data, both administrative and survey based. The work is being led by the Department of Social Services and is overseen by a whole-of-government Steering Committee.

A strategic framework for prioritising investment by the Commonwealth in new and existing longitudinal data is currently being developed and a final report to Government is expected in August 2016.

Multi-Agency Data Integration Project

The Multi-Agency Data Integration Project is led by the Australian Bureau of Statistics and provides an example of government entities working together to improve access to public data and maximise its use.

The purpose of the Multi-Agency Data Integration Project is to create an enduring integrated dataset for research use. The Multi-Agency Data Integration Project is producing a complex dataset that combines data on social security, income tax, Medicare and Census.

Commonwealth data custodians are working together to establish processes for sharing and safe dissemination of linked data. The Multi-Agency Data Integration Project partner entities include the Department of Social Services, Department of Health, Department of Human Services, Australian Taxation Office and the Australian Bureau of Statistics.

Linking Employee-Employer Data

The Australian Bureau of Statistics and Department of Industry, Innovation and Science are developing the Expanded Analytical Business Longitudinal Database and the Linked Employer-Employee Database.

These datasets combine high-value data sources to better support policy development through research and informed discussion.

The Expanded Analytical Business Longitudinal Database integrates administrative data from the Australian Tax Office and the Department of Industry, Innovation and Science with survey data from the ABS for all active businesses in the Australian economy from 2001-2002 to 2013-14.

The Expanded Analytical Business Longitudinal Database has been developed to demonstrate that the value of integrating administrative and firm level data to provide an evidence base for productivity analysis, policy development and evaluation.

The Australian Bureau of Statistics has developed a prototype Linked Employer-Employee Database, which links selected 2011-12 personal income tax data at the employee level with business characteristics (such as industry and turnover) from the Expanded Analytical Business Longitudinal Database. In the future, a longitudinal Linked Employer-Employee Database could be used to inform policy and research questions.

Australian Census Longitudinal Dataset

The Australian Bureau of Statistics created the Australian Census Longitudinal Dataset by linking a five per cent sample of the 2006 and 2011 Censuses to produce nationally representative longitudinal dataset. The Australian Census Longitudinal Dataset is used by universities and policy agencies such as Australian National University, University of Queensland, Department of Employment and Department of Education and Training.

The Australian Census Longitudinal Dataset is a significant national dataset and is an example of the statistical benefits of re-using existing data to create datasets that can be used more widely.

Migrants Settlements and Census Project

The *Migrants Settlements and Census project*[15] was undertaken as part of the Australian Bureau of Statistic's Census Data Enhancement initiative. This initiative aims to improve and expand *"the range of official statistics available to Australian society, and improves the evidence base to support good government policy making, program management and service delivery"*.

The *Migrants Settlements and Census project* integrated Census and immigration data to provide insights into how well permanent migrants to Australia were adjusting to life in a new country – an area where an information gap previously existed. The integration of data from the 2011 Census and the then Department of Immigration and Citizenship's settlement database enabled analysis of migrant Census data via visa category.

The clearer view of existing information enabled government entities to create policies to better target those groups of migrants who may need additional support to settle in Australia. It also enabled questions about migrants that were previously up for debate to be answered with evidence.

Following the conclusion of this project, the Australian Census and Migrants Integrated Dataset was released by the Australian Bureau of Statistics in 2014[16] as part of the Census Data Enhancement initiative.

---

[15]Australian Bureau of Statistics, *Migrants Settlement and Census Project*
[16] Smith, D, Smith, T, Using Linking to Sharpen Policy-Thinking: Early Findings from the Australian Census and Migrants Integrated Dataset, *The Australian Economic Review*, Vol 47, no.2, pp 220-230, May 2014

Integrating authorities

For data integration projects where at least one of the data sources comes from the Commonwealth, an Integrating Authority must be nominated for the project.

The Integrating Authority is responsible for the implementation of the data integration project and the management of the integrated datasets throughout their life cycle. The Integrating Authority ensures that the data is handled in a manner consistent with the requirements of the data custodians and compliant with Commonwealth governance arrangements. The Integrating Authority is also responsible for providing external parties with safe and secure access to the integrated data.

Currently, only three Australian Government entities are accredited integrating authorities. Enabling more organisations to seek accreditation, including state and territory government entities, and research and private organisations, could lead to improved data availability, technical and analytical capability and cross-jurisdictional cooperation.

The Department is working with the Australian Bureau of Statistics to streamline and modernise the governance arrangements for accrediting integrating authorities. This work will involve the establishment of an accreditation subcommittee from the Deputy Secretaries Data Group and the formation of straightforward information about governance structures and guidance for accreditation.

The accreditation process will also consider how to extend arrangements to non-Commonwealth organisations. Noting that private sector organisations may not be bound to the same legal, governance and policy arrangements as Australian Government entities, the possible risks associated with accrediting different types of organisations will need to be considered.

In addition, a review of the *High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes* may need to be considered. These principles were endorsed in 2010 by Portfolio Secretaries as an initiative of the former Cross Portfolio Statistical Integration Committee. A review would assist with determining whether these principles are still current and assess whether more detailed guidelines are required.

4. Access to, and high-value, private sector data

Private sector data

The Strategic Policy and Digital Economy Branch within the Department of Industry, Innovation and Science has policy responsibility for matters involving the broader digital economy, which includes, where possible, driving outcomes regarding private sector data.

With the growing digital economy, data is increasingly being generated by businesses consumers, as well government entities from everyday interactions and online engagement:

- the World Economic Forum noted that 90 per cent of the data in use in 2014 had been created in the previous two years;[17] and
- 67 per cent of Australian CEOs identify data and analytics as the technologies that will deliver the greatest returns in terms of engagement with customers and stakeholders.[18]

Integrating multiple data sources can increase the richness and depth of data; this is true for non-government data as much as it is for public data. Noting that there may be commercial or other sensitivities involved, the private sector should be encouraged to make its data openly available, where possible, so that private sector data can be used to deliver social and economic outcomes. For example, Unearthed's Phase-X Challenge involves BHP Billiton providing its operational data and mine planning scenarios so that innovators can compete to create the best algorithm for optimising open pit mining phases.

Integrating private sector data with public data may also provide significant opportunities. Private sector data could be used to enrich data collected by government entities or assist government with designing policy and delivering services. For example, the Australian Bureau of Statistics has replaced the collection of some pricing data from sources such as supermarkets, restaurants, travel agents and schools with transaction (scanner) data to calculate the Australian Consumer Price Index.

Government entities should continue to engage with the private sector and build public-private partnerships to identify high-value private sector data and facilitate sharing of this data with the broader economy.

Research data

Where appropriate, non-sensitive publicly funded research data should be made open for use and reuse. Improving the availability and quality of data arising from publicly funded research will help researchers, governments, businesses and communities develop high quality research which has the potential to answer questions of national significance. It can also enable the development of new applications, processes and products, open new fields of research, and improve transparency, responsiveness, productivity, efficiency and decision-making.

There are currently a number of impediments to openly releasing publicly funded research data, including:

- **Legislative/regulatory/contractual impediments:** Legislative and regulatory requirements can be barriers to openly sharing publicly funded research data. In addition, in some cases where data is provided to researchers by a third party, the researcher is required to destroy any derivative data once the project has finished. This means that researchers are limited in their ability to build on previous work undertaken using this data.

---

[17] Less Andrade, P, Hemerely, J, Recalde, G, Ryan, P, *From Big Data to Big Social and Economic Opportunities: Which Policies Will Lead to Leveraging Data-Driven Innovation's Potential?,* World Economic Forum, 2014
[18] PwC, *19th Annual Global CEO Survey: Redefining business success in a changing world,* January 2016

- **Financial impediments:** There are costs to managing data and making it openly available in a form that is reliable, interoperable and persistent. Currently, a majority of this cost is borne by researchers and institutions.

- **Technological impediments:** There are also significant costs associated with the infrastructure to support data intensive research, data storage and data sharing. Issues of maturity, reliability and persistence for publicly funded research infrastructures inhibit data management and sharing.

- **Socio-cultural impediments:** These include concerns about privacy, security, cybercrime, and surveillance, and the wider social implications of electronic personal data and ubiquitous sensors. There are also different disciplinary cultures which shape the ways that researchers collect, manage and use data, and their willingness to make it available.

- **Other ethical considerations:** These include informed consent and its renewal, the protection of rare and threatened resources, and the appropriateness of governance arrangements.

The Open Access Working Group, led by the Department of Education and Training, is currently developing draft principles for *Open Access to Publicly Funded Research Data and Publications.*[19]

Areas for consideration

Consideration should be given to high-value data collected by quasi-private sector organisations, such as Corporate Commonwealth Entities, Commonwealth Companies and Government Business Enterprises.

Examples of Corporate Commonwealth Entities include the Commonwealth Scientific and Industrial Research Organisation, the Australian Broadcasting Corporation, the Australian Postal Corporation (also a Government Business Enterprise) and the National Health Performance Authority. Examples of Commonwealth Companies include NBN Co Limited (also a Government Business Enterprise) and the Australian Institute for Teaching and School Leadership Limited.

These organisations have different obligations to non-corporate Commonwealth entities under the *Public Governance, Performance and Accountability Act 2013* (PGPA Act) and other legislation.

For example, under the PGPA Act, non-corporate Commonwealth entities must govern in a way that is not inconsistent with the policies of the Australian Government. However, corporate Commonwealth entities are only required to comply with specific policies of the

---

[19] The draft principles will address access to research outputs, both data and publications. Given the scope of the Productivity Commission's inquiry, only data-related issues have been addressed in this paper.

Australian Government when the Finance Minister makes a 'government policy order' under subsection 22(1) of the PGPA Act.

These quasi-private organisations often collect and maintain high-value data as part of their normal business operations. The availability and use of this data could have significant benefits for government policy development and service delivery, research outcomes and private sector innovation. An investigation may be required to determine whether there are any legislative or regulatory barriers preventing these organisations from sharing or releasing their data.

5. Consumer access to, and control over, data

Government entities hold a wealth of personal data about individuals. Making this data more accessible to citizens would give individuals enhanced visibility over how their data is used and increased power to decide how it can be used by third parties.

Benefits to citizens from improved access to and control over data

Visibility of data use

Legislative constraints, a lack of focus on how to use data and improve service delivery, and limitations by the use of outdated technologies, have all limited the opportunities for government entities to provide citizens access to data about themselves. The growth of digital services and information access in the private sector has raised expectations for government entities to provide similar access, transparency and choice.

For example, in Estonia citizens can access nearly all of their own data online through a state portal. Through the portal, they can view data about themselves, correct mistakes and even transact with businesses and government. Citizens can see which officials have viewed their data and it is against the law to view someone's data without appropriate reasons.

Control over data to enable access by third parties

There is an opportunity to provide government-held citizen data directly to individuals, so that citizens can decide how their data is used by third parties. This would enable the private sector to design and tailor products and services, particularly apps, to consumers.

For example, 100,000 Australians have taken up the consumer finance app, Pocketbook, since its launch by a Sydney startup in 2013. The app allows users to sync all their bank accounts and credit cards, and uses analytics to review users' transactions and provide predictions and reminders of recurring bills and payments. Such an app was only possible because banks have allowed third parties to access data through secure customer-authenticated access.

This could be achieved by requiring Australian Government entities to publish APIs that allow individuals to control access to their government-held personal information by

third-party applications. This would enable citizen choice in allowing authenticated access to data via a range of government and non-government applications.

The UK's midata programme could be drawn upon as an example of empowering citizen's with control over their own data. The midata programme aims to give consumers access to the information that companies hold about their transactions in a machine-readable and reusable format in order to make it easier to compare the different offers available. A number of new products and services have been created by third parties using midata. For example, gocompare.com is a comparison site offering midata bank account comparison to give customers bespoke results based on their own consumption data.

Trusted Digital Identity Framework

In order to enable citizen control and transparency over their data, a framework whereby an individual can be uniquely identified in a trusted system is required.

The Digital Transformation Office was established in July 2015 and is responsible for leading the transformation of government services to deliver a better experience for Australians.

Part of the Digital Transformation Office's work includes the development of a Trusted Digital Identity Framework. The Framework will establish a set of principles and standards for the use of accredited government and third-party digital identities to enable individuals and businesses to access services more easily.

This will involve establishing a common strategic approach to identity across government and preventing Australian Government entities from investing in unnecessary bespoke solutions. Once implemented, individuals and businesses will no longer have to prove their identity multiple times to government when accessing services.

The Framework will provide users with choice and control in establishing their digital identity. The wider use of digital identities will improve access to services, new products and markets for consumers and industry, and more productive ways of doing business. It will also reduce the burden of managing the ever-growing number of unique digital credentials. The Framework will also take into account privacy and security requirements consistent with the Australian Privacy Principles and existing government security standards.

The establishment of Digital Transformation Office and its work on the Trusted Digital Identity Framework is part of the solution. There are technical, cultural and legislative barriers that must be fully understood before any solution could be implemented.

6. Resource costs of access

Charging for data services

Increased availability of public data enables innovation, especially when it is made openly available at no charge to end users. Evidence indicates that price barriers, even small, deter wider use of data and that making data available at no charge increases take-up, particularly by small and medium enterprises.

In its 2011 *Pricing of Public Sector Information Study*,[20] the European Commission found that reducing the cost of public sector data to zero can increase the number of repeating users by between 1,000 and 10,000 per cent.

Recommendation 13 of the *Public Sector Data Management Report*[21] recommended the establishment of a consistent and transparent approach to user charging to provide clarity on how to manage costs.

On 24 December 2015, the Department of Finance released its *Charging for Data Services Information Sheet*[22] as part of the Australian Government Charging Framework. The Information Sheet indicates that entities can consider charging for the following data services: specialised data collection; provision of specialised data and data analysis services; facilitating specialised access to data; and data support services.

The Information Sheet also advises that Australian Government entities should be aware of the *Public Data Policy Statement*'s requirement that entities *"only charge for specialised data services, and, where possible, publish the resulting data open by default"*.

Developing skills and capability

As the volume and value of data being produced increases exponentially, the demand for data analysts and scientists that have the ability to interrogate and manipulate data also increases. In 2011, the McKinsey Global Institute estimated that by 2018 the United States alone would face a shortage of approximately 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyse big data and make decisions based on their findings.[23]

The Australian Public Service similarly faces a shortage in employees with the requisite data skills and capability. The *Public Sector Data Management Report* identified that ready for work graduates with data capabilities are in short supply in the public service, and during consultations, most entities expressed a need for more data capabilities.

The *Public Sector Data Management Report* recommended that the Australian Public Service Commission and others collaborate to develop a strategy with government, industry and academia to build data and analytics capability.

In response to this, the Department is partnering with the Australian Public Service Commission, other Australian Government entities, and private and research sectors to develop a holistic approach to improve overall data skills and capability across the Australian Public Service. Increasing data literacy will facilitate smarter policy development, better service delivery, public sector efficiency, and better programme evaluation.

---

[20] European Commission, Information Society and Media Directorate-General, *Pricing of Public Sector Information Study: Models of supply and charging for public sector information – final report*, October 2011
[21] The Department of the Prime Minister and Cabinet, *Public Sector Data Management Report*, July 2015
[22] The Department of Finance *Charging for Data Services Information Sheet*, December 2015
[23] McKinsey Global Institute, *Big Data: The next frontier for innovation, competition and productivity*, May 2011

Through this partnership, an *APS Data Skills and Capability Framework* has been developed to empower the Australian Public Service to harness the value of data and increase data literacy across all levels of the Australian Public Service. This includes:

- **Data Training Partnerships:** Data Training Partnerships are an ongoing initiative linking the Australian Public Service to organisations with expertise relating to data. Initial partners include the Open Data Institute Queensland, the Office of the Australian Information Commissioner, and the Data To Decisions Cooperative Research Centre. The Department is working with these partner organisations to make their training and development in data-related expertise available to staff across the Australian Public Service.

- **Australian Public Service Data Literacy Programme:** The Australian Public Service Data Literacy Programme is a suite of initiatives that, once fully implemented, is intended to improve the data capabilities of non-specialist Australian Public Service officers. The programme has been designed by the Australian Public Service Commission in collaboration with the Australian Bureau of Statistics and the Australian Taxation Office and will offer a suite of flexible learning options designed to build core data literacy skills.

- **University courses:** Specialised data analytics courses and subjects will help the Australian Public Service improve its technical data analytics capability whilst boosting the number of future graduates with required data skills. For example, the Australian National University has begun offering a Graduate Diploma in Applied Data Analytics and a Masters of Applied Data Analytics. Australian citizens (including Australian public servants) that have the required qualifications may study these courses as well as individual subjects offered on a non-award basis. While the courses are open to the general public, the ANU courses were developed in close collaboration with the Department of Human Services to ensure public service officers will gain the skills specifically applicable to the public service.

- **Data Fellowships:** The Data Fellowship Programme is an exclusive, competitive programme to provide advanced data training to high performing Australian Public Service data specialists. Up to ten participants a year will be selected to undertake three month research or engineering placements in partner organisations, such as Data61. Participants will bring with them a data-related problem (which may include analytics, foresighting, API development, etc.), which they will work on to develop a solution for their agency.

In addition, the Data Analytics Centre of Excellence was established by the Australian Taxation Office as a space to build analytics capability across government entities. The purpose of the Centre of Excellence is to enable a common capability framework for analytics, as well as an opportunity to share technical knowledge, skills and tools.[24]

---

[24] Further information is available at: *www.analyticsspace.net.au*

Areas for consideration

Recognising the under-supply of data skills and capability across all sectors of the Australian economy, consideration could be given to initiatives which develop the Australian data analytics skills base beyond the public sector.

The UK Government has released a strategy to develop UK data capability in the commercial, academic and public sectors. The UK strategy focuses on strengthening skills in schools, higher and further education, and continued professional development. It also aims to ensure that the UK's infrastructure and the research and development environment supports data capability.

Initiatives for further exploration could include:
- embedding data skills as a core component of tertiary degree programmes beyond the data speciality degrees;
- establishing graduate or internship programmes for final year university students with private sector or academic organisations to develop data analytics skills;
- leveraging the *Restoring the focus on STEM in schools initiative* coordinated by the Commonwealth Department of Education and Training; and/or
- exploring the possibility of adapting core skills training from the Australian Public Service Commission, Australian Bureau of Statistics and Australian Taxation Office to a wider audience.

7. Privacy protection

As outlined in the Australian Government's *Public Data Policy Statement*, Australian Government entities are required to "*uphold the highest standards of security and privacy for the individual, national security and commercial confidentiality*".

Critical to the public data agenda is earning and maintaining the community's trust, with the implementation of robust privacy safeguards being of great importance.

Tailored approaches are required for different types of data:

- **Non-sensitive data** is data that does not identify an individual or breach privacy or security requirements. This data should be made openly available and as easy to use as possible.

- **De-identified data** is data relating to an individual where the identifiers have been removed to prevent identification of the individual. This data should be shared securely with trusted users to inform research and drive better policy outcomes.

- **Personal data** is data relating to an individual where the individual is identified or identifiable. Sharing of personal data should be limited, it should be securely shared with other government entities to inform policy development and improve service

delivery, and as outlined in Section 5, there are opportunities to make data about citizens more accessible to them.

Technological solutions

Data61 has been working on technological solutions that enable the effective use of data with appropriate privacy and confidentiality safeguards.

Confidential computing

Confidential computing enables valuable insights to be gained from sensitive data from multiple organisations without that sensitive data being accessible to any organisation other than the one that owns the data.

Using partial homomorphic encryption, analysis is undertaken on encrypted data provided by each data custodian. Only the end results or insights are unencrypted and shared. Data analytic techniques that can be used on encrypted data include correlation analysis, classification and prediction, clustering, outlier detection and statistics.

Synthetic data generation

Synthetic data is data that is generated from original data to create a set of plausible values, which preserve some of the patterns contained in the original data. This allows the data to be used and analysed in a way that is realistic but poses less risk of disclosure.

Synthetic data can provide privacy level guarantees while still providing high data granularity, which is useful for analytics.

Trusted-access model for sharing integrated data

As part of the Multi-Agency Data Integration Project, the Australian Bureau of Statistics is leading work to establish a trusted-access model for sharing integrated data. Adapting international best practice, the Australian Bureau of Statistics uses the Five Safe Principles for the assessment of disclosure risk (which is based on a UK model).

The Five Safe Principles include: safe people, safe project, safe setting, safe data and safe output. Applying the Five Safe Principles ensures access to sensitive data can be granted in an appropriately secure environment.

Office of the Australian Information Commissioner Guidelines

On 17 May 2016, the Office of the Australian Information Commissioner published the consultation draft of the *Guide to big data and the Australian Privacy Principles*,[25] which outlines the key privacy requirements for big data projects, including how personal information may be collected, shared and used by entities subject to the requirements of the *Privacy Act 1988*.

---

[25] Office of the Australian Information Commissioner, *Consultation draft: Guide to big data and the Australian Privacy Principles*, 17 May 2016

The guide recognises the potential for big data analytics to bring about enormous social and economic benefits. The guide has been developed to facilitate big data activities while protecting personal information.

Engagement with the public

Key to building and maintaining community trust is communicating the benefits of public data and the measures in place to address security and privacy concerns.

In line with Recommendation 6 of the *Public Sector Data Management Report*, the Department is leading work with other Australian Government entities to "*build and maintain trust, engage with the public to understand the benefits to citizens and address privacy concerns on public sector data*".

To date, engagement has involved:

- data.gov.au and the Department's website provide a central point of information about the public data agenda, including data-related policies and initiatives;
- social media is used to communicate key messages through the following Twitter accounts: @datagovau, @ANZLIC and @dpmc_gov_au. Citizens are also engaged through the data.gov.au blog;
- roadshows have been conducted across the country for a number of data-related initiatives to engage with the public at a grass-roots level; and
- presentations on the public data agenda, including at the Locate16 Conference in Melbourne in April 2016, the Sydney CeBIT 2016 Conference in May 2016, and the KPMG Public Sector Data and Analytics Roadshow in May 2016.

Areas for consideration

Different people have different preferences regarding the protection of their privacy and their data being shared as a trade-off for convenience and better service delivery.

Some individuals want their privacy protected at all costs and are opposed to their data being used and/or shared without their consent. At the other end of the spectrum are individuals willing to sacrifice some of their privacy protections if it means more convenient and streamlined services (both from government and the private sector).

Consideration could be given to how to cater for these different preferences, including opt-out schemes for sharing and use of data. For example, developments in game theory and mathematics are being used to help consumers find suitable financial products to meet their goals and preference in a tailored way. Potentially, this methodology could be used to measure the trade-off on decisions between privacy versus convenience with regard to individuals agreeing to data about themselves being shared between government entities.

8. Other restrictions

On 26 February 2016, the Secretaries Data Group agreed that all Government entities should be encouraged to examine whether legislation is being interpreted within a contemporary context when considering if data can be used and reused. This will be progressed through the Deputy Secretaries Data Group and Data Champions.

The Australian Bureau of Statistics has published a series of information sheets on confidentiality which are available on the National Statistical Service website. The information sheets explain and provide advice on a range of issues around confidentialising both aggregate statistics and microdata. The Australian Bureau of Statistics is also developing further draft Guidelines for *Confidentialising Outputs and Aggregate Statistics* to support Australian Government entities seeking to publish aggregate statistics, whilst protecting the identity and attributes of individuals.

9. Data security

As previously stated, the *Public Data Policy Statement* requires Australian Government entities to "*uphold the highest standards of security and privacy for the individual, national security and commercial confidentiality*".

The risk of a breach of data security can never be eliminated, but in the event that a breach occurs, government entities and many large corporations commit to comply with mandatory data breach notification requirements and the Office of the Australian Information Commissioner's voluntary best practice guidelines for personal information.

The Australian Government has committed to introduce a mandatory data breach notification scheme. Draft legislation (*Privacy Amendment (Notification of Serious Data Breaches) Bill 2015)* was released for public consultation in December 2015. The Government will consider all stakeholder feedback before finalising legislation for introduction as part of its current legislative agenda.

In addition to data security being important for protecting confidentiality, preserving the availability and integrity of data is fundamental to its confident use. As more high-value datasets are made public, and new innovative products and services rely on that data, the importance of appropriately securing the availability and integrity of information is paramount. In many cases the owner of the data will not have visibility about who or how data is being used, increasing the risk that manipulation of data or loss of availability of data could have serious consequences for downstream users.

Australia's Cyber Security Strategy

On 21 April 2016, the Australian Government released its Cyber Security Strategy.[26] The Strategy outlines a range of initiatives to strengthen the protection of Australia's networks and to enable innovation and growth in Australia's economy.

---

[26] The Australian Government, *Cyber Security Strategy*, April 2016

In particular, the Strategy identifies the huge economic potential created by disruptive business models; these are increasingly reliant on access to public data sets and/or big data analytics. While such disruptive business models will open up new business opportunities, many of these depend on adequate cyber security for trust and confidence in underpinning data and for the delivery of online products and services.

Getting cyber security right will mean Australia is a secure and dynamic location for business investment. Relevant initiatives in the Strategy include:

- developing voluntary guidelines on good cyber security practice for Australia's businesses;
- implementing a programme of grants to enable small businesses to have their cyber security tested; and
- establishing a Cyber Security Growth Centre to bring together a national cyber security innovation network so all businesses can benefit from cyber security solutions.

10. Department of the Prime Minister and Cabinet portfolio examples

Indigenous Affairs Group

The Indigenous Affairs Group within the Department regularly analyses changes in outcomes for Indigenous Australians. This requires analysis of all available data including data held by other Commonwealth entities and by the states and territories.

The Indigenous Affairs Group uses data in a wide range of ways from linked datasets through to national data on the *Closing the Gap* targets. The Indigenous Affairs Group also collates data for regional profiles.

High-value public sector data

There a large number of high-value public datasets in Australia including for Indigenous Australians. The *Aboriginal and Torres Strait Islander Health Performance Framework* includes analysis of 65 national data collections and 600 research articles on the health of Indigenous Australians including an analysis of the policy implications of the findings. This project includes the public release of over 2,000 tables on the Australian Institute of Health and Welfare website.[27]

Data quality

One of the major factors that impedes the public release of data is data quality. If data is of poor quality, releasing that data publicly is unlikely to be helpful. The high-value administrative data collected by government entities is a result of extensive work to agree to definitions and metadata. This particularly applies to the national minimum datasets that are collated by the Australian Institute of Health and Welfare, as well as the data on National Assessment Program – Literacy and Numeracy (known as NAPLAN) and school attendance collated by the Australian Curriculum, Assessment and Reporting Authority.

---

[27] The Australian Government, *The Aboriginal and Torres Strait Islander Health Performance Framework*, 2014

While there are still significant issues with the quality of data for Indigenous Australians, considerable progress has been made in recent years. These data improvements have come from a concerted effort to improve data quality and consistency.

Through the *National Indigenous Reform Agreement*, all jurisdictions have agreed to a set of enhancements to national data collections. The main focus of this data development plan was to improve the quality of Indigenous identification across a range of administrative data collections. For example, all states and territories agreed to enhance the *Perinatal National Minimum* by including a data item on the Indigenous status of the baby (previously only the Indigenous status of the mother was included) and data items on smoking during pregnancy and antenatal care. There is now perinatal data on the Indigenous status of the baby for all jurisdictions, smoking during pregnancy data available for seven jurisdictions and antenatal care data available for four jurisdictions.

Almost all deaths in Australia are registered. However, the Indigenous status of the deceased is not always recorded correctly which leads to an under-estimate of Indigenous deaths in Australia. The Australian Bureau of Statistics and the Australian Institute of Health and Welfare have been working with jurisdictions and funeral directors to improve the quality of Indigenous identification in mortality data for many years.

Up until 1998 only three jurisdictions had sufficient quality Indigenous identification in deaths data to include in national reports (Northern Territory, South Australia and Western Australia). Queensland only introduced the Indigenous status data item in the collection in 1996 and from 1998 New South Wales and Queensland Indigenous mortality data was considered of sufficient quality for inclusion. There is still work to be done to bring the other jurisdictions up to the level of quality required for reporting Indigenous mortality rates and the gap with non-Indigenous Australians.

Another example of data consistency has been the work undertaken across jurisdictions to allow a nationally consistent school attendance rate to be published. In the Prime Minister's 2016 *Closing the Gap Report*,[28] a nationally consistent school attendance rate was published for Indigenous and non-Indigenous students for the first time. Prior to 2016, data on school attendance for both Indigenous and non-Indigenous students was published for each jurisdiction and a national figure was not available given differences in definitions across jurisdictions.

National initiatives through the Council of Australian Governments have led to significant data improvements in the quality of data for Indigenous Australians but there is still considerably more work to be done. For example, there is only limited crime data in Australia published for Indigenous Australians. The Australian Bureau of Statistics publication *Recorded Crime – Victims (cat. no. 4510.0)* only includes data for Indigenous Australians for New South Wales, Queensland, South Australia and the Northern Territory. The Australian Bureau of Statistics publication *Recorded Crime – Offenders (cat. no. 4519.0)* also only includes data for these jurisdictions, with data for the Australian Capital Territory

---

[28] The Australian Government, *Closing the Gap Report*, 2016

included only since 2013-14. For both publications information on Indigenous status is not of sufficient quality in other jurisdictions for data to be published by the ABS.

Data about regional areas

The Department of Prime Minister and Cabinet's Indigenous Affairs Group often prepares regional profiles to inform policy development work and regional data is currently being collated for the *Empowered Communities Initiative*. It can be challenging to gather regional data even within government entities as it can be time consuming to obtain permission both to access and publicly release data.

While the Australian Institute of Health and Welfare holds data from a range of national minimum datasets, the Australian Institute of Health and Welfare is not able to release regional data from these holdings without the permission of various data custodians in each state and territory. At times it can be a lengthy process to obtain permission to release regional data for Indigenous Australians from data custodians and the restrictions that jurisdictions may put around data release are not always consistent. It is acknowledged that state and territory data custodians often receive many requests for data and they need to manage these requests with limited resources. If more regional and place-based data was regularly released then considerable time could be saved.

However, it should be noted that several jurisdictions regularly publish useful regional data. For example, several states and territories (New South Wales, Victoria, Queensland, South Australia, Western Australia and the Northern Territory) regularly publish regional police data on crime (usually at the local government area level). While this data is not normally split by Indigenous status, comparative data on crime levels across a state or territory is useful. In the case of remote Indigenous communities where the vast majority of people are Indigenous, overall crime data is usually all that is required.

A large range of regional data is already published by state and territory and Commonwealth entities; however, there would be value in making this data more accessible and findable. In addition, ideally regional data would be available against agreed geographies such as the Australian Bureau of Statistics SA1 and SA2. Data by postcode is not ideal as postcodes cover large areas in remote Australia and postcodes do not provide a sound basis for the publication of regional data. Postcodes have not been designed to facilitate sound regional data analysis.

There is an increasing desire from Indigenous communities to have access to information on all the government services and programmes that are provided to support Indigenous people in particular locations. Providing this sort of information is time consuming as there is currently no systematic collation of information on what programmes and services are provided in each location. In the past, providing this sort of information for grant programmes was often challenging as the service location was not always recorded in grant management systems. In the case of the Department, grants and programmes are now geocoded making information available about grants and services funded within different locations.

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

Data linkage

Data linkage is critical to effective evaluation and analysis. While it is useful to publish data on outcomes this data alone does not provide information on policy impact or effectiveness as outcomes can change due to factors completely unrelated to a policy or programme. In addition, often the outcome from a policy or programme will not be observed for several years. For that reason data linkage can be critical. For example, it is well known that the value of preschool mainly comes when children are older so if it was possible to compare outcomes at school (and after children finish school) for individuals who did and did not participate in preschool then it would be possible to estimate the impact of preschool participation on outcomes. Looking at data collated by preschools themselves does not itself provide comprehensive information on programme impact.

By linking data for individuals across time, much better evaluation and analysis is possible. Longitudinal data has many major advantages over simple cross-sectional data as with longitudinal data each individual effectively acts as their own control group. This in turn allows a large source of heterogeneity to be removed which in turn allows for much more effective modelling of programme impact. Australia has built a good set of longitudinal surveys including the *Longitudinal Survey of Indigenous Children*; however, by their nature surveys can be limited due to sample size.

There is a potential to better utilize de-identified data linkage files. While the Commonwealth is embarking on several high-profile data linkage projects it is important to acknowledge that several states and territories have also undertaken a considerable amount of data linkage. This is important because if data linkage projects are to reach their full potential it is important that Commonwealth and state and territory data be linked.

SA-NT DataLink has made progress in linking data and making that data available in a secure environment both to policy makers and researchers. The Department is currently working with SA-NT DataLink to link Commonwealth data on the receipt of income support payments with a subset of data already linked for the Northern Territory. Opportunities are being investigated for the most effective way of linking this data. If this proposed link is successful, options will be investigated for linking Commonwealth income support data with the other data in SA-NT DataLink in an enduring way. This project is a component of the high-value cross-agency project to improve educational outcomes for Indigenous Australians through better targeting of early childhood interventions, which is being led by the Department.

The Department has a contract with the Menzies School of Health Research for analysis of the existing linked de-identified dataset in the Northern Territory created through SA-NT

DataLink. Under the contract, the Menzies School of Health Research will analyse the impact of a range of programmes on subsequent outcomes for Indigenous Australians. With such a large linked dataset it is possible to create control groups using existing data as long as it is possible to identify who has and who has not participated in various programmes. This in turn reduces the need to collate data for specific evaluations and makes for a much more efficient and effective evaluation process.

Current privacy legislation varies between jurisdictions and the Commonwealth and these variations add a complicating factor. A common agreed approach relating to privacy legislation between the Commonwealth and the states and territories for linked datasets would simplify current arrangements while ensuring that privacy is protected.

Analytical skills

Large datasets are increasingly becoming available both the policy makers and researchers. While well organised datasets with good metadata may reduce the need for some ad hoc data queries, greater data availability will increase the need for data and analytical skills. For example, to fully realise the value of longitudinal linked datasets, it will be important that government entities have staff who are able to undertake more sophisticated approaches such as regression analysis to inform policy analysis and advice.

Office for Women

Gender disaggregated data is essential to the consideration and implementation of policy across all areas of government. There are a large range of data sources that collect gender disaggregated data for various reasons, as outlined below.

The Office for Women considers the mainstreaming of gender equality to be a high priority. The availability and use of quality data sources is essential to ensuring that gender impacts are considered by policy entities in the development of policies or programmes for Australia.

The Office for Women uses a variety of data sources in order to undertake its work and drive policy that directly benefits girls and women. This includes monitoring the impact of policies and programmes on women and girls, as well as the development of new policy from a whole-of-government perspective. Data is sourced on subject matter including:

- women's workforce participation;
- women's leadership;
- educational attainment (including subject choice);
- international comparators;
- women's safety;
- women's economic security;
- financial literacy; and
- diversity.

The data sources most frequently used by the Office for Women are outlined below.

Australian Bureau of Statistics Gender Indicators[29]

Gender Indicators Australia, developed and maintained by the Australian Bureau of Statistics, provides a summary of gender-specific data in six areas of social concern for gender equality, including economic security, education, health, work and family balance, safety and justice, and democracy, governance and citizenship.

This data is used to undertake analysis of gender-specific trends in the policy domains of the Office for Women. Used as a quick reference guide, the Gender Indicators are a key source of information for government policy making.

Gender Pay Gap data

The gender pay gap is a complex problem that has many contributing factors. There are multiple sources and measurements of the pay gap, all of which are used for different purposes. The Office for Women monitors the following sources:

- **Average Weekly Earnings:[30]** Average Weekly Earnings is a quarterly survey conducted at the business level and has a long-time series (electronic series available from 1983). The survey size is approximately 5,500 employees. This survey reports on earnings based on a full-time equivalent employee, which when considered from a gender perspective can be problematic given the high proportion of women that work part-time. However, it is a widely recognised and reported measure of the gender pay gap in Australia and in this regard is useful for monitoring the rate of change in the pay gap over time.

  Since May 2010, the series has included an Average Weekly Cash Earnings component, which includes salary sacrificed amounts. This is an important inclusion from a gender perspective as many industries that are female dominated (for example the social and community services sector) may be eligible for salary sacrifice arrangements, thus boosting weekly income. However, there can be some large fluctuations in contributions throughout the year, so this component is only assessed via a time series comparison.

- **Employee Earnings and Hours:[31]** The Employee Earnings and Hours series is a two-yearly business survey which collects information at the individual employee level. The survey size is approximately 9,000 employers and 57,000 employees. It includes employment characteristics, such as full-time or part-time, hours paid for, security of job (permanent, fixed-term, casual), managerial and non-managerial, method of setting pay, salary sacrifice and occupation. Given the frequency of collection, this series is useful for long-term measurement of the pay gap, but not for regular monitoring.

---

[29] Australian Bureau of Statistics, *Gender Indicators*, Cat No. 4125.0
[30] Australian Bureau of Statistics, *Average Weekly Earnings*, Cat. No. 6302.0
[31] Australian Bureau of Statistics, *Employee Earnings and Hours*, Cat. No. 6306.0

- **Employee Earnings, Benefits and Trade Union Membership:**[32] This annual household survey supplement to the Labour Force Survey reports on earnings by employment characteristics and socio-demographics, including full-time or part-time, hours worked, age, relationship in household, educational attainment, marital status, geography, country of birth, occupation, leave entitlements and tenure. This is useful for undertaking specific cohort analysis.

- **Workplace Gender Equality Agency Dataset:**[33] The Workplace Gender Equality Agency Dataset is compiled from an annual business collection from non-public sector employers in Australia with 100 or more employees. Collection in this format commenced in 2013-14. Measurement of the gender pay gap is based on total remuneration, including base salary, bonuses, superannuation and other payments. It can be broken down by management and non-management positions, permanency status, standardised occupations, and industry. It is useful for industry and cohort analysis and presents a different measurement of the pay gap.

Labour force participation rates

Labour force participation rates are monitored using the Australian Bureau of Statistics' Labour Force survey, released monthly.[34] Seasonally adjusted data is analysed to monitor the participation rate of men and women, as well as unemployment, underemployment and part-time work rates. Cohort analysis using the survey is also used for family characteristics (e.g. parents with children under five years), culturally and linguistically diverse women, Indigenous women, women with disability, carers, and youth and mature age cohorts.

Labour force participation rates are also used in a range of other policy settings, meaning that data from multiple sources needs to be considered and compared. For example, there are environmental, attitudinal and cultural considerations. This includes matters such as:

- the cost and availability of child care, including for outside of school hours care and pre-school hours care; the use of informal child care;[35]
- the proportion of non-public sector employers that offer support for employees returning to work, for example breastfeeding facilities, employer subsidised child care;[36]
- take up rates of flexible work arrangements by men and women;[37]
- superannuation account balances and coverage;[38] and
- average retirement income and source of income.[39]

---

[32] Australian Bureau of Statistics, *Employee Earnings, Benefits and Trade Union Membership*, Cat. No. 6310.0
[33] Workplace Gender Equality Agency, *WGEA Data Explorer*
[34] Australian Bureau of Statistics, *Labour Force, Australia*, Cat. No. 6202.0
[35] Australian Bureau of Statistics, *Childhood Education and Care Survey*, Cat. No. 4402.0
[36] Workplace Gender Equality Agency, *WGEA Data Explorer*
[37] Workplace Gender Equality Agency, *WGEA Data Explorer*
[38] Australian Bureau of Statistics, *Gender Indicators*, Cat. No. 4125.0
[39] Data available on request from the Australian Bureau of Statistics

However, more data and research would be beneficial to identify attitudinal factors that inhibit workforce participation rates, such as:

- unconscious bias in the workplace;
- employer perceptions of the difficulties and cost of providing flexibility;
- the normalisation of men taking parental leave and accessing flexible working arrangements;
- achieving a more equitable distribution of unpaid work; and
- the constraint of women's work choices by ingrained cultural norms.

Women's leadership

The Office for Women uses a range of statistics to monitor women's leadership, including political and board appointments. In addition, the Office for Women also collects data on the gender balance of Australian Government boards (including the number of boards, gender balance of all board positions, and gender balance of Chair and Deputy Chair positions). This is taken from AusGovBoards and confirmed with Australian Government entities. Data is released publicly on an annual basis but is monitored on a real time basis by the Office for Women.

The Office for Women also collects resumes and personal data of women seeking appointment to Government boards. These are stored in the BoardLinks candidate database. Access to this database is restricted to authorised Government users for privacy reasons.

In terms of monitoring women's leadership and board appointments in the non-government sector, the Office for Women uses the following data sources:

- Australian Institute of Company Directors statistics on ASX 200 Board appointments; and
- business operators and contractors by sex.[40]

It is unclear whether there are data sources currently available for monitoring the percentage of startup businesses led by women. Collecting data in this area may potentially address data gaps for women in businesses and women's entrepreneurship, which is a recognised data gap identified by the Organisation for Economic Cooperation and Development for Australia and many other countries.

Other data gaps for women's leadership include:

- regularly collated data on gender balance of serving elected members at local government level. This data is available but not collated into a single source and may differ in terms of measurement between jurisdictions;
- data on the number of women standing for election and their success rate; and

---

[40] Australian Bureau of Statistics, *Labour Force, Australia,* Cat. No. 6202.0

- an updated comprehensive study on the gender balance of not-for-profit boards. This study is warranted given this is where many people commence their board careers. Gender disaggregation at this level is therefore important to measure the flow to non-government and government boards.

International comparators

The Office for Women closely monitors international gender disaggregated data for the purposes of examining conditions in other countries that could potentially influence policy settings in Australia. Australia supports this international collaboration through the provision of data for international comparison purposes to organisations such as the Organisation for Economic Cooperation and Development, United Nations and the Asia Pacific Economic Forum.

In terms of the work to support international engagement and collaboration, government entities collaborate in order to ensure effective implementation and monitoring. For example, work is underway for the domestic implementation of the Sustainable Development Goals and how these can be measured using the data sources available. A data gap in this area relates to the average weekly hours of unpaid domestic and care work disaggregated by sex, age and location.

A common issue encountered for international data is jurisdictional availability and measurement as these can differ between states and territories making the consistent reporting of Australian data difficult.

Cities Taskforce

The Cities Division within the Department is responsible for implementing the Australian Government's Smart Cities Plan, released on 29 April 2016.[41] The Smart Cities Plan includes a range of actions to improve the productivity and liveability of Australia's cities through smart investment, smart policy and smart technology.

The key mechanism for implementing the Smart Cities Plan will be City Deals. City Deals are agreements between the Australian Government and state, territory and local governments. The Government has made a commitment to establish early City Deals in Western Sydney, Townsville and Launceston.

Each City Deal will be unique but will benefit from high quality public data in areas such as transport, education, the environment and economic output. This information will help City Deals partners to:

- identify the highest value investment and reforms;
- set achievable objectives and targets; and
- measure the performance of the city against these objectives over time.

---

[41] The Australian Government, *Smart Cities Plan*, April 2016

The necessary data is not always collected, made publicly available or reported at the right level to inform decision-making for cities. Through City Deals the Government will work with partners in state, territory and local governments and the private sector to, amongst other things, identify key data needs and actions to make this data available over time.

The Government has also recently announced a $50 million competitive Smart Cities and Suburbs Program[42] for local governments to apply innovative technology based approaches to improve the liveability of cities and suburbs. Opening up council data and information, business processes and technologies to third parties will create a platform for the identification of urban problems and contestable approaches to solutions.

---

[42] The Government's policy for *Smart Cities*, June 2016

Appendix 1: Glossary

Administrative data	Refers to information collected for delivering public administration – for example to conduct registration, transaction and record-keeping.
Anonymised data	Data relating to a specific individual where the identifiers have been removed to prevent identification of that individual. Otherwise known as de-identified data.
API	Application programming interfaces are standardised codes and protocols for using data.
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.
Data integration	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.
MADIP	Multi-Agency Data Integration Project
Non-sensitive data	Data that is anonymised / de-identified 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 data	Data collected or generated by the public service for policy development and public administration. Also known as Public Sector Information.
Synthetic data	Data that is generated from original data to create a set of plausible values, which preserve some of the patterns contained in the original data. This allows the data to be used and analysed in a way that is realistic but poses less risk of disclosure.
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.