



Australian Government

Department of the Prime Minister and Cabinet

REVIEW OF AUSTRALIAN GOVERNMENT DATA ACTIVITIES 2018

CONTENTS

EXECUTIVE SUMMARY	2
UNLOCKING RETURNS FROM AUSTRALIAN GOVERNMENT DATA	3
SURVEY RESULTS PROVIDE A PRELIMINARY SNAPSHOT OF HOW DATASETS ARE BEING USED AND MANAGED ACROSS THE AUSTRALIAN GOVERNMENT.....	5
KEY FINDING #1: ACCESS TO PUBLIC SECTOR DATA IS IMPROVING	11
KEY FINDING #2: AGENCIES ARE USING DATA MORE EFFICIENTLY TO PROVIDE AGILE AND EFFECTIVE GOVERNMENT SERVICES	12
KEY FINDING #3: PUBLIC SECTOR DATA SKILLS AND CAPABILITIES ARE IMPROVING	13
KEY FINDING #4: GOVERNMENT DATA PROTECTIONS ARE BUILDING COMMUNITY TRUST AND CONFIDENCE IN HOW PUBLIC SECTOR DATA IS COLLECTED AND USED	14

EXECUTIVE SUMMARY

The Australian Government holds vast amounts of data collected from individuals and businesses, and generated by different government agencies. Efficient collection, use and re-use of data is key to improving the efficiency of government spending and delivering more effective and better targeted —evidence-based — government policies, programs and services.

This Review provides a snapshot of Australian government data activities and related data resources. A preliminary survey of the data activities across the Australian Government found that the use of data is embedded across all areas of policy and program delivery. A number of new analytical and data integration projects are underway, such as the projects arising from the 2017-18 Data Integration Partnership for Australia (DIPA) funding.

Survey information on data governance mechanisms helped inform the development of New Data Sharing and Release Arrangements in the 2018-19 Budget.

The Review identifies four key areas where Government reforms to the public sector data system are yielding improved outcomes:

1. Access to public sector data is improving
2. Agencies are using data more efficiently to provide agile and effective government services
3. Public sector data skills and capabilities are improving
4. Government data protections are building community trust and confidence in how public sector data is collected and used.

Going forward, funding of \$65 million over four years was announced in the 2018-19 Budget, to reform the Australian data system, including establishing a new *Data Sharing and Release Act*, which will act as a circuit breaker to remove existing legislative, and cultural, barriers to data use and re-use. A new National Data Commissioner will administer the new legislation; develop guidance for government entities and the public about data sharing arrangements; and proactively monitor and address risks and broader ethical considerations around data use. A new Consumer Data Right will allow customer to share their transaction, usage and product data with service competitors and comparison services.

This report looks briefly at the importance of Australian Government data, and recent reforms to the Australian Government data system. The report then outlines key findings from PM&C's preliminary survey of Australian Government data activities, followed by a discussion of the four key areas where Government policies are improving outcomes across the public sector data system.



UNLOCKING RETURNS FROM AUSTRALIAN GOVERNMENT DATA

The Australian Government holds vast amounts of data collected from individuals and businesses, and generated by different government agencies.

In the past two decades, developments in computing power have opened up opportunities to use big data more efficiently, by linking information across different datasets and identifying previously indiscernible patterns and preferences. It is an area where the private sector leads, and governments across the world are now working to keep pace.

Better use of data can improve the efficiency of government spending through more effective and better targeted—evidence-based—government policies, programs and services. It can also help solve some of the most complex policy problems and fuel future innovation.

The Government's 2015 *Public Data Policy Statement* recognises 'the data held by the Australian Government is a strategic national resource that holds considerable value for growing the economy, improving service delivery and transforming policy outcomes for the nation'. The Statement commits to: optimise the use and re-use of data; release non-sensitive data as open by default; and collaborate with the private and research sectors to extend the value of public data for the benefit of the Australian public.

Since 2015, a range of measures have been put in place to increase data sharing and release, improve data governance and build stronger data skills and capabilities (see Figure 1 below).

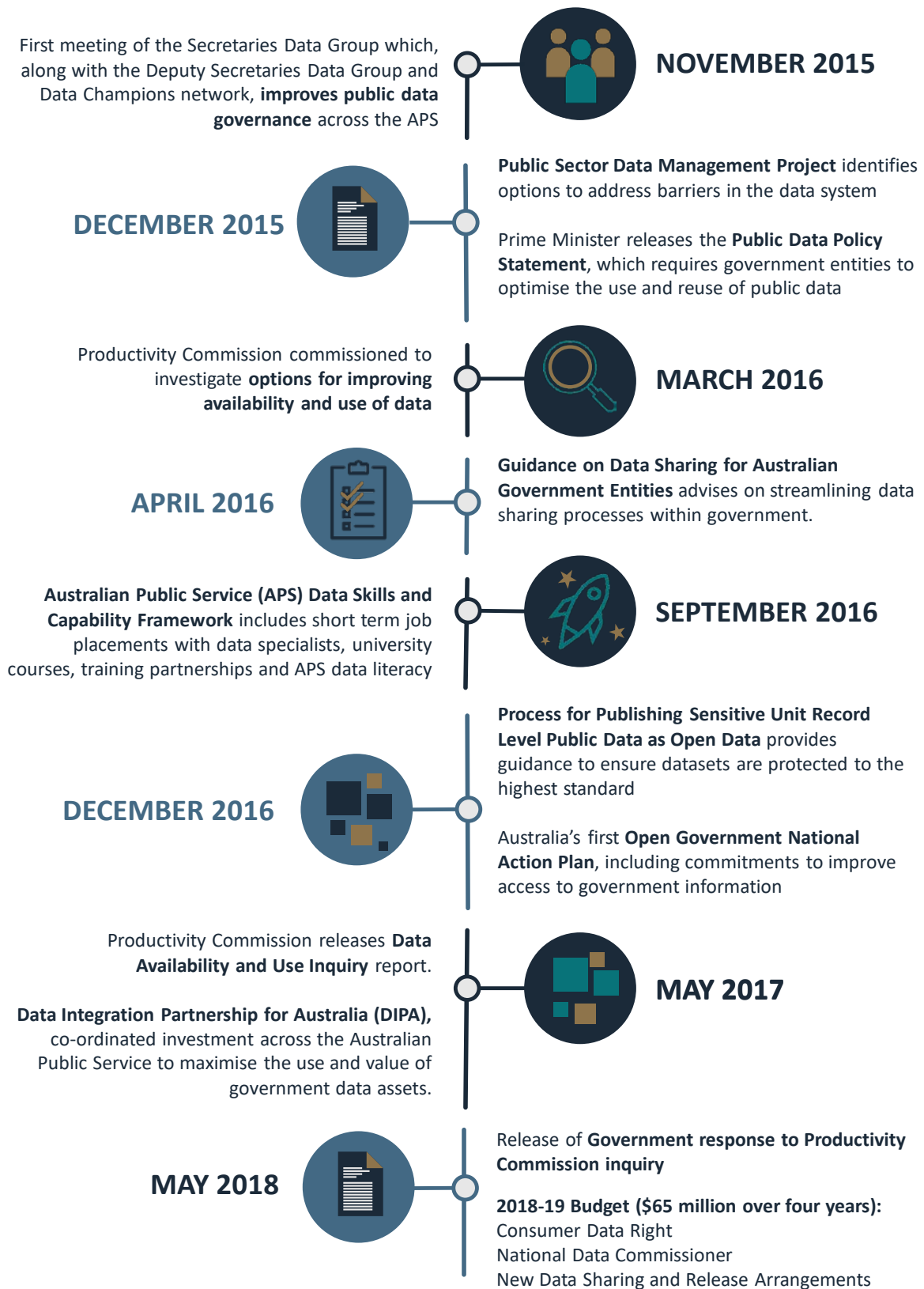
A number of more recent reforms, such as the DIPA (which received funding in the 2017-18 Budget to improve data integration, infrastructure and analysis) are still settling in. The benefits of these reforms will continue to play out in coming years and decades.

As recommended by the Productivity Commission *Data Availability and Use* inquiry in 2017, the Government has also agreed to key future reform areas:

- A National Data Commissioner will support a new data sharing and release framework and oversee the integrity of data sharing and release activities of Commonwealth agencies.
- A legislative package will streamline data sharing and release, subject to strict data privacy and confidentiality provisions. A key function of the new framework will be to *authorise* the sharing and release of data (instead of the existing regimes which *restrict* sharing and release).
- A Consumer Data Right will allow consumers to harness and have greater control over their data held by the private sector.



Figure 1: Recent reforms to the Australian Government data system





SURVEY RESULTS PROVIDE A PRELIMINARY SNAPSHOT OF HOW DATASETS ARE BEING USED AND MANAGED ACROSS THE AUSTRALIAN GOVERNMENT

The Department of the Prime Minister and Cabinet (PM&C) undertook an initial survey of the Australian Government data system, obtaining information from 58 different Australian Government agencies on the extent of their data use, including specific investments, governance activities and benefits derived from better use of data.

The survey aimed to provide an initial picture of the Australian Government data system; preliminary feedback on the effectiveness of DIPA; and an outline of data governance systems across the Australian Government.

Importantly, administrative data (for example, tax and Centrelink data) was generally not captured despite being a large part of the data system overall.

The Australian Government's data system is embedded across all areas of policy and program delivery

Government agencies reported, on average, around a third of their work relies on data. The level of data use varies extensively across agencies. Data and research agencies such as the Australian Institute of Health and Welfare, Australian Bureau of Statistics (ABS), CSIRO and Geoscience Australia are some of the most intense creators and users of data. Work undertaken by these agencies has flow on benefits to other parts of government and to the broader economy – including health and welfare, industry and resource investments, and environmental sustainability and protection.

Agencies identified a broad range of partners across their functional activities, including from both the public and private sectors. Many recognised other Commonwealth agencies as well as state and territory agencies as key collaborators on policy, service delivery, program management, regulatory and technical activities.

Annual government expenditure on data activities exceeded \$2.4 billion in 2017-18

The survey collected information on data expenditure from those areas across the APS that are intensive data users (where more than 40 per cent of their activities rely on data).

The annual expenditure on data related activities across these areas was estimated at \$2.4 billion in 2017-18. This figure excludes costs related to compliance or general administration activities (such as taxation and government payments), which were not the focus of this review.

Data expenditure by all teams across all Australian Government agencies would be expected to be much higher.

Most agencies reported similar splits of expenditure across design, collection, use and governance of data:

- around 10 per cent of expenditure goes to data design;
- around 25 per cent of expenditure goes to data collection;
- around 60 per cent of expenditure goes to data use; and



- around 5 per cent of expenditure goes to data governance activities.

Building a strong evidence-base is key to informing better government decisions

Agencies reported, on average, a quarter of their policy development work relies on data. A number of agencies highlighted more policies are now better informed because they are underpinned by evidence, including:

- Better transport and infrastructure development by using data for targeting future investments and addressing needs, as well improving safety across different modes of transportation.
- Building sustainability in the use of environmental resources through taking a data-driven approach to inform decisions such as the allocation and delivery of water and building efficiencies in the energy use.

Effective data use helps improve service delivery to all Australians

Agencies with a service delivery role reported, on average, around a third of their service delivery work relies on data, including:

- Managing demands for telephony and processing services for health and aged care services.
- Planning resources and workloads for service delivery, including refining business operations to improve outcomes and productivity for delivering social services.
- Using data, combined with user insights, to inform user-centric design decisions and enable delivery of positive user experience outcomes.

Using data to inform compliance is key to ensuring policy goals are met

Agencies with a regulatory and compliance role reported, on average, regulatory and compliance work relies on data over 30 per cent of the time. Effective data collection enables agencies to ensure rules are being adhered to and regulation is meeting its policy aims:

- The Clean Energy Regulator (CER) administers legislated schemes for measuring, managing, reducing or offsetting Australia's greenhouse gas emissions. It collects and uses a range of data to inform government policy making, meet international treaty obligations, inform regulatory decisions, detect and respond to non-compliance and fraud, and support statistical services and data publication. To ensure compliance, these datasets are cross validated against external sources. Additionally, the CER is looking into machine learning to further improve compliance controls.
- The Australian Electoral Commission uses data to investigate electoral fraud, electoral advertising, non-voter prosecutions and multi-voter prosecutions.

The Australian Taxation Office uses data for better compliance case selection to better target those likely to have done the wrong thing, as well as assisting taxpayers in preparing complete and accurate tax returns. The pre-filing service provides data for taxpayers to review and include in their tax returns, helping people get it right the first time and reducing the need for follow-up activity.***Effective program management relies on accurate data***

Agencies reported their program and project management activities rely on data more than 20 per cent of the time. This is reflected in agencies being able to measure the effectiveness of programs and projects against key performance indicators:



- Regular monitoring and analysis of work and production data has enabled identification of productivity improvements in granting intellectual property rights. Identifying trends in areas such as seasonal filing volumes, rates of registration and emerging technologies has enabled improved workforce planning and training.
- Data analytics have informed a targeted program of proactive monitoring and inspections of personal insolvency practitioners and personal insolvency administrations to deliver a risk-based, efficient approach to delivering regulatory objectives with minimal impact on the regulated population.
- State of the Service data at the department, division and branch level has helped agencies to better manage workforce, judge employee engagement, gauge manager capability and understand workplace culture and conditions.

Dedicated units across the APS deliver data-specific professional and technical services

Agencies undertaking professional and technical services reported that, on average, over 40 per cent of these activities relied on data. A number of agencies had dedicated areas where most or all of the work relied on data:

- The Health Support and Performance team within the Department of Human Services focuses on data extraction, reporting tools and analysis work.
- The Government Business Analytical Unit within the Department of Finance established through DIPA will conduct analytical projects to support improved public sector performance. It will also implement new tools and capabilities to enable better use of government-centered data.
- The Statistical Services Group within the ABS uses many different data sources to produce economic, social population and environmental statistics.

A range of governance mechanisms help agencies manage data collection, sharing and use

There a range of whole-of-government governance functions to help efficiently manage public sector data

Whole-of-government public sector data governance functions are undertaken by a range of government agencies:

- PM&C is responsible for whole-of-government public sector data strategy;
- The Digital Transformation Agency (DTA) leads the digital transformation of government services;
- The Office of the Australian Information Commissioner (OAIC) regulates the operation of the *Privacy Act 1988*, the *Freedom of Information Act 1982*, and reports to the Australian Government on information policy and practice;
- The Australian Bureau of Statistics is the official statistical authority for the Australian Government and State and Territory governments; and
- The National Archives of Australia is responsible for preserving Australia's archival resources, providing a public right of access to information over 30 years of age (reducing to 20 by 2021), with specific exemptions, and promoting whole-of-government information management.

Individual agencies have their own data governance functions



In addition to these whole-of-government functions, most agencies with significant data use have internal and external boards, frameworks and processes set up to govern the sharing and use of data. Most agencies have at least one internal data governance board, with the 58 agencies reporting a total of 73 internal boards. Almost all agencies also have external facing governance boards, including inter-departmental boards, specialist advisory committees, ethics committees with significant external participation and regular stakeholder consultation forums.

Some departments have made data governance a key organisational priority. For example:

- The Department of Home Affairs has a Data Management division, headed by a Chief Data Officer, to improve the governance and use of data across the department;
- The Department of Human Services has a Chief Citizen Experience Officer to transform the user experience through improving the ICT environment.;
- The Clean Energy Regulator has integrated the governance of business IT systems and data into a single committee, and has specified responsibilities for various data governance positions, including a Chief Data Officer.

While some of these agencies' data governance systems are highly customised, overall there is significant duplication of functions including, in some cases, the creation of data frameworks that are not interoperable between agencies.

Investment in whole-of-government data integration projects is starting to show returns

Data integration brings together data from different sources about the same (or similar) individuals or units. Integrating data leads to meaningful insights that can answer a broader range of questions and help examine problems from multiple angles.

Historically, Australian Government data integration has happened on an ad-hoc basis. DIPA, funded in the 2017-18 Budget, is an initiative to:

- improve the government's data assets and infrastructure
- provide for collaboration across government to use these improvements to deliver better policies, programs and service on complex policy issues, and
- better communicate and engage on data initiatives.

The collaboration includes analytical units across several priority areas: social, health, and welfare; the environment; and the economy. The co-ordinated investment in the data integration capabilities and assets within government enables new integrated datasets as well as assisting to maintain and develop selected existing datasets.

2017-18 projects are showing promising results including:

- enhancing patient safety through identification of adverse events from medicines (see case study below)
- exploring the effects of changes in gas use by Australian business, and
- understanding the outcomes of government grants to business to better target support.



CASE STUDY: *Enhancing Patient Safety through Identification of Adverse Events from Pharmaceutical Use* – a Data Integration Partnership for Australia (DIPA) Project

Every year the Australian Government expends multiple billions of dollars under the Pharmaceutical Benefits Scheme (PBS) to ensure Australians can access medicines at a reasonable price. These medicines can be life-saving and life-changing. However some medicines or combinations of medicines may lead to poor health outcomes due to adverse events that aren't detected in clinical trials as they occur infrequently (less than 1 in 10,000 people), are due to interactions with other medicines, don't occur until many years after treatment or occur in patient groups excluded from clinical trials (such as pregnant women and the elderly or those with co-morbidities). A well-known example from the past is thalidomide, used for morning sickness in the 1950s and 1960s, which was subsequently found to cause birth defects.

Using linked PBS data, the Social, Health and Welfare Analytical Unit aims to use statistical and machine learning techniques to detect adverse events (e.g. heart failure) from medication use. A number of potential medicines of interest have already been found using PBS data alone and are undergoing further investigation and verification by clinicians. Using current patient datasets will assist in confirming signals that have traditionally only been accessible through patient registries or clinical trials. Machine learning will enable identification of relationships between medicines and particular adverse events in a systematic way. This will allow earlier action to be taken to manage or minimise the occurrence of adverse events leading to cost savings due to avoided hospitalisation and treatment.

The survey did not show evidence DIPA is yet resulting in a catalytic change around Commonwealth data sharing and data integration. To date, it is very largely the projects receiving funding through DIPA that are gaining traction. However, as funding only commenced in 2017-18, it may be too early to expect such paradigm change. There is some anecdotal evidence of new relationships being built across agencies to reduce cultural and governance barriers.

Future directions

This initial survey has provided a much richer picture of the Australian Government data system than has previously been available. The results of this work informed the development of New Data Sharing and Release Arrangements in the 2018-19 Budget.

However, there were limitations to this preliminary survey:

- administrative data — data collected through routine stakeholder engagement, consultation and communications (for example, tax and Centrelink data) was generally not captured despite being a large part of the data system overall
- due to time constraints, the scope of the survey was limited and agencies had little time to provide survey information
- the scope of the survey excluded most corporate Commonwealth entities, Commonwealth companies and those agencies involved in national security
- the survey was self-reported, raising the potential for inconsistent understanding of survey questions (in particular around the key definitional issues of 'what is meant by data', 'what is meant by 'relies on data' and 'what kinds of data are in scope for the survey')
- agencies didn't have records to easily respond to the survey questions



- a point in time survey cannot provide an indication of the evolution of the Australian Government data system.

Future iterations of this survey could incorporate administrative data and with questions updated to reflect feedback from the current review would give a better snapshot of the existing Australian Government data system. A longer timeframe would allow for a more structured questionnaire (with a combination of quantitative and qualitative questions) about what data entities collect, how they use it, what they release and to whom. Two-yearly surveys would also provide information on how the system is evolving in response to policy reforms.



KEY FINDING #1: ACCESS TO PUBLIC SECTOR DATA IS IMPROVING

The Productivity Commission found increased public availability of public sector data could generate social and economic benefits for all Australians through:

- opportunities for businesses to create new products and services, enhance existing ones, and introduce new business models
- better monitoring and use of resources
- opportunities for new research and development.

Existing measures to increase the public release of non-sensitive information include:

- data.gov.au provides an easy way to find, access and reuse public sector datasets. (see Box 1 below)
- The Foundation Spatial Data Framework has been established, through a collaboration of the Commonwealth and the states, to make national spatial data open and free to use by the economy
- New *Data Sharing and Release* legislation will facilitate greater release of non-sensitive datasets and greater risk-based sharing of identifiable data
- The new National Data Commissioner will be the trusted overseer of a new data sharing and release framework, allowing Australia to realise the full potential of data while maintaining public trust in the data system
- The DTA is developing a data discovery tool to help Commonwealth agencies identify data holdings across agencies. This project will also help build the foundation for creating a public registry of significant non-sensitive Australian Government datasets (both those published on data.gov.au and those yet to be released).

Box 1: data.gov.au

data.gov.au provides an easy way to find, access and reuse public sector datasets.

data.gov.au provides a central catalogue to discover public sector data. It also provides hosting for tabular, spatial and relational data with hosted APIs and the option for agencies to link data and services hosted by other government sources.

Since data.gov.au was established in 2013-14, there has been strong and ongoing growth in the number of datasets on the website. Improving the quantity and quality of the government data available on the data.gov.au is an ongoing process.

The number of items shared on data.gov.au has jumped from fewer than 1,300 at the beginning of 2016 to over 26,000 in April 2018.

More than 130 datasets were added in the first four months of 2018.



KEY FINDING #2: AGENCIES ARE USING DATA MORE EFFICIENTLY TO PROVIDE AGILE AND EFFECTIVE GOVERNMENT SERVICES

Access to, and efficient use of, high quality data is critical to the efficient and effective delivery of government services. Our preliminary survey found, on average, around a third of the service delivery work of agencies with a service delivery role relies on data.

Accurately administering government payments requires (and generates) significant volumes of data. Governments rely on data to direct programs to where they're needed, deliver services efficiently and effectively, and identify where programs are being breached or misused. Effective use of datasets is also critical to assess the effectiveness of existing government services.

Combining existing datasets results in a more complete, informed picture of the effectiveness of current government programs, and can help answer complex questions about society, our environment and the economy that single datasets alone cannot answer.

Existing measures to increase the returns from existing government datasets include:

- The Data Integration Partnership of Australia (DIPA) (see Box 2 below)
- The Behavioural Economics Team of the Australian Government (BETA) has data analysts and behavioural scientists working to tackle cross-cutting policy and operational issues
- Senior officials within Government agencies are acting as Data Champions, promoting data use, sharing and reuse within their organisations
- New *Data Sharing and Release* legislation will facilitate further sharing of datasets.

Box 2: Data Integration Partnership for Australia (DIPA)

DIPA is an investment to maximise the use and value of the Government's data assets. Through data integration and analysis, DIPA is creating new insights into complex and important policy questions.

DIPA will: improve our technical data infrastructure and data integration capabilities; preserve individuals' privacy and ensure the security of sensitive data; improve our data assets in important areas such as health, education and social welfare; and maximise the use of these assets through data integration and analysis.

DIPA is increasing the use and value of the Multi-Agency Data Integration Project (MADIP) and the Business Longitudinal Analysis Data Environment (BLADE).

MADIP has linked existing Medicare, government payments, personal income tax and 2011 Census data. MADIP projects have identified that lower income households typically make greater use of health services associated with some chronic medical conditions and have poorer health outcomes than households in higher socioeconomic areas. This has highlighted possibilities for improving Australians' access to the health care system.

BLADE is an analytical tool, consisting of linked datasets, that helps researchers analyse business performance, dynamics, demography and characteristics. Data from BLADE has been used to show Australian businesses demonstrate superior growth while they are preparing to become exporters; but this growth falls back once they start exporting. This can assist in the design of better support for Australian exporters and non-exporters.



KEY FINDING #3: PUBLIC SECTOR DATA SKILLS AND CAPABILITIES ARE IMPROVING

Strong public sector data literacy is key to supporting evidence-based decision making, developing more efficient government policies and delivering services that meet the needs of Australian people.

This requires strong foundational data skills across the public service, so that data use is embedded at all stages in the policy cycle. It is also important for the public service to have more specialised data analysts, data scientists and data architects who can ensure data is collected, used and shared to maximise the returns from the government's existing datasets.

Existing measures to build public sector data skills and capability include:

- The August 2016 strategy document for data skills and capability in the Australian Public Service (APS) introduced foundational data literacy training for all APS employees and specialised 'data fellowships'. The DTA is leading implementation of this policy suite.
- The Australian Public Service Commission is responsible for an APS Data Literacy program to ensure all APS employees have a minimum foundational level of data literacy.
- A number of agencies (including the Department of Industry, Innovation and Science, the Department of Agriculture and Water Resources, the Department of Health, the Department of Home Affairs and the Department of Human Services) are developing and implementing data management strategies to ensure data training and data tools are readily available to all staff and datasets are well managed and appropriately accessible (see Box 3 below).
- ABS and AIHW are Accredited Integrating Authorities, funded under DIPA to boost their capability and capacity to more effectively integrate data from across the Australian Government. In addition, five analytical units have been funded \$11.2 million over three years under DIPA, funding data analysts drawing on integrated data to improve policy and program outcomes.
- Work is being undertaken through the graduate network to develop initiatives to drive cultural change and incentivise data use.

Box 3: Department of Industry, Innovation and Science (DIIS) Data Management Strategy

The DIIS Data Management Strategy provides a two-year plan for how the department will transition to managing data as an asset by bringing the department's data into a governed, curated state where data is ready for analytics, evaluation and sharing.

The Strategy incorporates the department's vision: As a data user, I can easily access data in a timely manner and trust I can use this data to make decisions. As a decision maker, I can trust the data I am provided to be accurate and a sound basis from which to make decisions.

The Strategy also identifies the processes for taking forward five key goals:

Goal 1: Data is well governed and risks are managed to support the department's work

Goal 2: The department's datasets are discoverable and curated to enable data to be ready for use

Goal 3: All the department's datasets are managed in a standard way over the full data life cycle enabling data to be widely used and reused

Goal 4: More of the department's datasets are shared externally or published

Goal 5: Tools are available for each type of data user, which are well maintained and appropriate to the user's level of expertise and business need.



KEY FINDING #4: GOVERNMENT DATA PROTECTIONS ARE BUILDING COMMUNITY TRUST AND CONFIDENCE IN HOW PUBLIC SECTOR DATA IS COLLECTED AND USED

Community buy-in and social acceptance around data collection and data sharing is critical to maintaining the quantity and quality of government data. Higher levels of trust mean greater willingness to share data and greater enthusiasm for data use. Australian citizens need to see how either they, or the population more broadly, stand to gain from data use.

Building trust takes sustained effort over time. Communication is critical, underpinned by consistency of message across agencies.

Government programs aim to: communicate about data initiatives to demonstrate their benefits and value; provide tools to manage and address data and digital incidents; and engage with the community to ensure concerns are heard and appropriately addressed. Existing measures include:

- Notifiable data breach notification (see Box 4 below)
- Publicly available Data Sharing Agreements to improve transparency around how public data is used and shared
- PM&C has prioritised the development of an overarching, whole-of-government data crisis management protocol to provide for consistent, appropriate best-practice response in the event of a data incident
- The Building Trust communication strategy and ‘campaign-in-a-box’ is a toolkit for agencies to use in creating their own data communication to better communicate with the public about data initiatives. The toolkit includes guidance on key messages, case studies and a crisis communication plan (which complements the data crisis management protocol)
- The OAIC in consultation with PM&C delivered an APS-wide Privacy Code to improve privacy risk management capability to support the public data agenda and build trust
- The 2016 Cyber Security Strategy is helping ensure Australia’s networks and systems are hard to compromise and resilient to cyber attack
- A new National Data Commissioner will proactively monitor and address risks and broader ethical considerations around data use
- A new National Data Advisory Council will advise the National Data Commissioner on emerging risks and ethical issues around data use
- A disclosure risk management framework with data safeguards to be dialled up or down according to the level of risk associated with different types of data and different types of data uses will be implemented through the new *Data Sharing and Release Act*.

Box 4: Notifiable Data Breach scheme

The Notifiable Data Breach scheme mandates that Australian Government agencies and the various organisations with obligations to secure personal information under the *Privacy Act 1988* notify affected individuals and the Australian Information Commissioner, in circumstances where the data breach is likely to result in serious harm.

By reinforcing accountability for personal information protection, the National Data Breach scheme supports greater consumer and community trust in data management.