



Australian Government

Department of Social Services

Serena Wilson PSM  
Deputy Secretary

Mr Peter Harris and Ms Melinda Cilento  
Chairman and Commissioner  
Data Availability and Use Inquiry  
Productivity Commission  
data.access@pc.gov.au

Dear Chairman and Commissioner

### **Data Availability and Use – Productivity Commission Issues Paper**

The Department of Social Services (DSS) welcomes the opportunity to provide comments on the Productivity Commission's *Data Availability and Use* Issues Paper as requested by the Treasurer in April 2016. This submission provides information to assist the Commission in preparing its report on ways to improve the availability and use of public and private sector data.

DSS aspires to be the Australian Government's pre-eminent social policy agency with responsibility for national policies and programmes that aim to improve the lifetime wellbeing of people and families in Australia. The work of the Department in the areas of Social Security, Families and Communities, Housing and Disability and Carers responds to needs across people's lives and encourages independence and participation to support a cohesive society.

DSS produces data as a result of its administrative responsibilities, policy development activities and interactions with the Australian public. In line with the *Australian Government Public Data Policy Statement*, the Department recognises that 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.

In this context, we consider the key data sets for which we are responsible, comprise a public good in the way they support improving wellbeing and we would welcome advice on how to better facilitate access and drive the open data agenda forward.

The Department is committed to optimising the use and reuse of public data; to the release of non-sensitive data as open by default; and to collaborating with private and research sectors to extend the value of public data for the benefit of the Australian public.

As part of this commitment, DSS's vision is for the creation of a shared Data Analytics Hub that links public datasets supplied by Commonwealth custodian agencies and provides an agile service model that responds to users' needs at a whole-of-government level.

The four sections of this submission focus respectively on:

1. DSS data sharing activities
2. The DSS vision for data sharing
3. Managing Commonwealth debt through data sharing, and
4. Responses to questions raised in the Commission's issues paper

We trust you find this information useful, and would be pleased to provide further input during the course of the Commission's Inquiry.

If your Office requires further information, please contact David Dennis, Branch Manager, Policy Evidence Branch

Yours sincerely

Serena Wilson

14 July 2016

Lead by example  
value diversity | respect diversity | celebrate diversity

## Data Availability and Use – Productivity Commission Issues Paper

The following comments refer to issues raised in the *Productivity Commission's Issues Paper on Data Availability and Use* from the Department of Social Services perspective.

### Part 1 – DSS Data Sharing Activities

This section outlines the range of data sharing activities undertaken by DSS to support the use and reuse of public data.

#### 1.1 Data Access Policy Principles

The Department's *Data Access Policy* (August 2014) provides a unified approach to supporting access to DSS payment, programme, grants and survey data, and to standardising processes related to data access. The [Data Access Policy Principles](#) are outlined below:

##### 1.1 Data Access Policy Principles

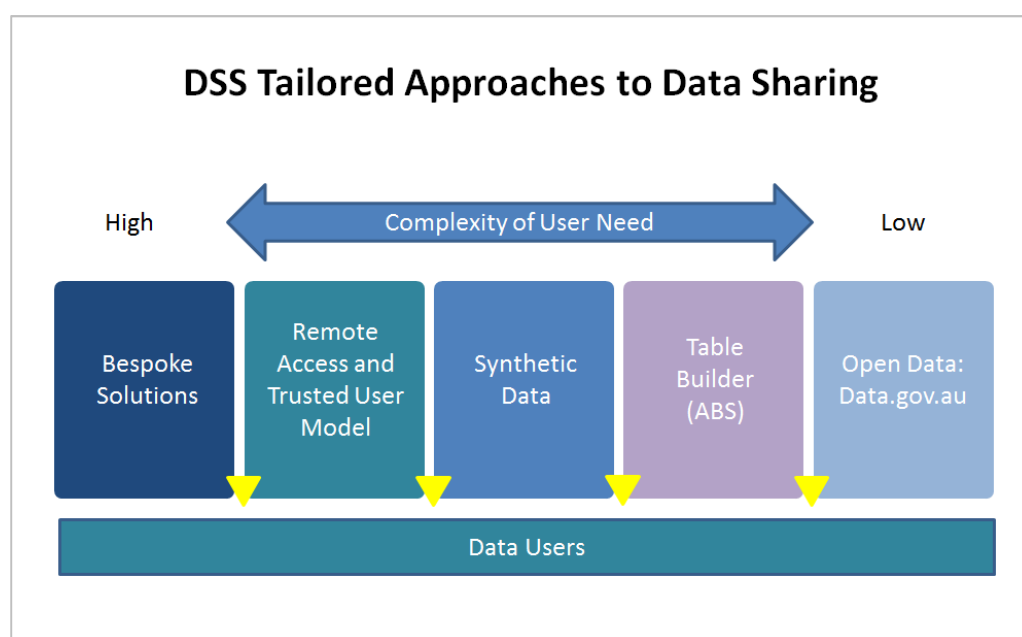
<b>Principle 1</b>	DSS should permit as much public access to data as possible, while protecting the privacy and confidentiality of individuals and organisations
<b>Principle 2</b>	Data access must comply with relevant legislation, interagency agreements and information protocols
<b>Principle 3</b>	Data access must be considered as standard practice
<b>Principle 4</b>	Aggregate or summary data should be easy to access (including making data accessible in formats that promote its reuse, i.e. machine readable)
<b>Principle 5</b>	DSS retains all relevant legal responsibility for social services unit record data at all times
<b>Principle 6</b>	DSS should grant controlled access to unit record data

#### 1.2 Tailored Approaches to Data Sharing

DSS produces data as a result of its administrative responsibilities, policy development activities and interactions with the Australian public. This data is used for research and statistics, service delivery, and compliance. Each type of use is supported differently by legislative arrangements. This submission largely speaks from the perspective of data use for research and statistical purposes.

Building on the *Data Access Policy Principles*, DSS provides tailored approaches to data sharing.

These approaches range from providing data via bespoke solutions to the general provision of public data via the Australian Bureau of Statistics (ABS) Table Builder and de-identified aggregate data via data.gov.au. Bespoke solutions include initiatives such as data linkage projects between agencies and the provision of data to researchers and others under relevant portfolio legislation.



Different types of data provision meet different user needs. The needs of researchers trying to access public data differ from Commonwealth agencies seeking to link datasets. In addition to the approaches taken by DSS to provide access to data, the Department is also undertaking projects that aim to enable the greatest use and re-use of public data by a variety of users.

**Remote Access Project:** DSS has been working with the Australian Institute of Health and Welfare (AIHW) on a proof-of-concept project which improves data access by researchers. The project has achieved its goal of successfully enabling researchers to access selected social services data remotely, subject to relevant legislative requirements.

Access to the data is via a “curated gateway.” Behind the gateway is de-identified information about individuals to which queries are applied and aggregated answers extracted. However, individuals’ records cannot be extracted.

Now successful, the Project will be scaled up and developed under a Trusted User Model.

**Trusted User Model:** The Model (also known as the Trusted Access Model) is being developed to provide a platform for enabling access to relevant administrative data to “trusted” users. Adding a Trusted User Model overlay to existing technology safeguards such as those used with curated gateways adds further security to individual records.

There is currently no one clearly defined Trusted User Model across the Commonwealth. Several organisations are considering this broad concept and developing options with the intent of creating a whole-of-government model.

**Synthetic Data:** DSS is working with Data61 to create a publically accessible synthetic version of a social security dataset. The synthetic dataset will closely reflect the characteristics of individual records in the dataset but without providing individual records relating to real individuals.

This data can be used to test analytic models related to changes to a person’s circumstances when receiving financial support from the government. Researchers will be encouraged to test hypotheses of such explanatory models prior to deciding to make an application to access unit record files in the DSS social security dataset. The information is expected to be provided online without restriction.

**National Centre for Longitudinal Data:** DSS operates the National Centre for Longitudinal Data (NCLD) ([www.dss.gov.au/NCLD](http://www.dss.gov.au/NCLD)) to support the management of critical national data assets. Ongoing support for our national data assets and the review of longitudinal needs is a substantial investment in the Department's future performance data capability, and will contribute to a better understanding of trends, influences and barriers to improving lifetime wellbeing. A major review is being undertaken to determine what Australia's longitudinal data needs are for the future and the best way for meeting these needs. A report on the review findings is expected to be finalised in October 2016.

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

### 1.3 Data Governance

Linking and sharing data will provide a range of economic and social benefits for the Australian public. However, in order to harness the value of this data, it must be more effectively shared both publicly and between Australian Government entities.

The implementation of effective governance structures is key to overcoming agency inaction on data sharing, and gives agencies an opportunity to formally address issues that have become barriers to effectively sharing data.

Internal governance structures are fundamental to participating in more complex data sharing activities, both as part of data integration arrangements between government agencies, and providing data access for the use and re-use of public sector data.

DSS recognises that there are multiple levels of governance required to effectively support data sharing and linkage activities, in line with the range of activities and differing levels of oversight required. These include:

- Internal operational procedures to manage data preparation and collation (e.g. confidentialisation policy)
- Governance processes regulating how data can be used, applying appropriate levels of control for internal and external data use, and
- Governance structures to enable more effective data access systems, where consistent practices across agencies would streamline data access processes.

It would be desirable to have clear and separate rules and arrangements for the different levels of governance as outlined above, providing clear guidance to the distinct groups who undertake the production, regulation, and accreditation of public sector data.

In line with *Guidance on Data Sharing Activities for Australian Government Entities* (PMC, April 2016), a streamlined committee structure has been established to discuss data issues across Commonwealth agencies. DSS is currently a member of the Deputy Secretaries Data Group (DSDG) (reporting to the Secretaries Data Group) and is also on the Data Champions network. Groups such as the DSDG consider cross-agency data issues that will have an impact on the availability and use of data.

DSS has also dedicated resources towards data management and access issues including establishing an internal governance structure. The DSS Data Governance Committee oversees and provides



strategic guidance on data policies, systems, activities and projects, including data related investments.

## 1.4 Data Linkage at DSS

As part of its commitment to supporting the use and reuse of public data, DSS participates in data linkage work that provides benefits to DSS, other Commonwealth agencies and external researchers.

DSS is not an accredited Commonwealth Integrating Authority, and so works closely with accredited Commonwealth agencies (i.e. the ABS, AIHW, and the AIFS) to implement data sharing infrastructure and undertake data sharing and data linkage activities.

**Multi-Agency Data Integration Project (MADIP):** DSS is a partner agency in the MADIP, which is creating an enduring, linked and publicly accessible research data resource to enable agencies and researchers to respond to nationally important cross portfolio policy and service delivery questions.

The Project aims to provide an exemplar for responding to demand for greater access to Commonwealth data for policy and research purposes. It will provide a comprehensive data source for evidence-based policy development across areas of broad social and economic concern.

The initial datasets to be used in the Project are:

- Medicare Enrolments Database (MEDB)
- Medicare Benefits Schedule (MBS)
- Personal Income Tax (PIT)
- Social Security and Related Information (SSRI), and
- 2011 Census of Population and Housing (Census).

The integration of these datasets will provide a foundational research resource to which other datasets could be added.

The Project will initially be used by partner agencies including DSS, the Australian Taxation Office, the Department of Health, the Department of Human Services and the ABS. However, it is envisaged that MADIP will provide access to external researchers in the longer term, possibly via the ABS Data Laboratory and Table Builder, and potentially with remote or online access.

## 1.5 Privacy Protection

Safeguarding the privacy of individuals is very important when sharing data. There are several aspects to consider throughout a data sharing project to preserve privacy and confidentiality, including:

- legislative and policy frameworks
- governance arrangements
- necessary approvals by human research ethics committees
- separation of identifying and analysis variables
- managing risks around dissemination of data
- de-identification or confidentialisation of personal information, and
- data archiving/destruction when/if the data is no longer required.

Legislation and guidelines used when determining whether the Department of Social Services may provide access to data about a person (whether de-identified or not) include:

- *Social Security (Administration) Act 1999*
- *Social Security (Public Interest Certificate Guidelines) (DSS) Determination 2015*
- *A New Tax System (Family Assistance) (Administration) Act 1999*

- *Family Assistance (Public Interest Certificate Guidelines) Determination 2015*
- *Paid Parental Leave Act 2010*
- *Student Assistance Act 1973*
- *Student Assistance (Public Interest Certificate Guidelines) Determination 2015, and*
- *Privacy Act 1988*

The requirements for providing access can vary between Acts and Guidelines. This can add complexity when information is required for research or analysis that spans more than one area.

Data sharing is further complicated with varying state and territory legislation and guidelines.

## Part 2 – DSS Vision for Data Sharing

DSS recognises the importance of a trusted, transparent and balanced approach to sharing public data, where privacy and trust are maintained, and is committed to building and maintaining trust in the quality, confidentiality and value of public sector data with government, researchers, business and the public.

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. Increasing the number of users with access to public sector data will provide a range of economic and social benefits.

This provides a strong value proposition for making non-sensitive public sector data open by default, recognising the broader economic and social benefits derived from making non-sensitive public sector data available for use and re-use.

The Department is committed to optimising the use and re-use of public data, to the release of non-sensitive data as open by default, and to collaborating with private and research sectors to extend the value of public data for the benefit of the Australian public.

DSS undertakes a range of activities to support the use and reuse of public data. This includes data sharing arrangements under the DSS *Data Access Policy* that provide a unified and standardised approach to data sharing. Building on this policy, DSS provides a number of tailored approaches to data sharing based on user need, including participating in a wide range of data linkage work that provides benefits to government, researchers and the Australian public.

Australian Government entities are working on initiatives to facilitate more efficient data sharing arrangements, including work currently being undertaken by the MADIP. However, Commonwealth data sharing initiatives are currently undertaken on a largely ad hoc basis between individual agencies. This can be unnecessarily complicated and time consuming, as recognised in PM&C's *Guidance on Data Sharing for Australian Government Entities*.

DSS's Vision is for **the creation of a Data Analytics Hub with a domain for secure access by government and researchers** to streamline data sharing processes to enable greater efficiencies and improve decision making for policy and implementation processes.

This Vision includes the potential for introducing enabling legislation specific to the Data Analytics Hub function. This may have the advantage of providing overriding principle based legislation that is adaptable to technological change, makes explicit capabilities of agencies, and minimises perceived barriers to data linkage while maintaining appropriate privacy protections.

In the short term, government agencies should focus on ways to establish effective data sharing arrangements under existing legislation. However, in the medium term privacy legislation should be reviewed to identify whether privacy laws can be streamlined and modernised to enable data to be better used for policy and research.

A review of privacy legislation could also focus on harmonising privacy legislation across jurisdictions to promote data sharing between Australian governments. In line with DSS's statement on managing commonwealth debt, this should include the review of legislation relevant to transparency of individual debt and the sharing of debt related information.

The DSS Vision also includes the implementation of effective governance structures as critical to overcoming agency inaction on data sharing,

This Vision is aligned and congruent with the *Australian Government Public Data Policy Statement* and the *Guidance on Data Sharing for Australian Government Entities* (PM&C, 2016).



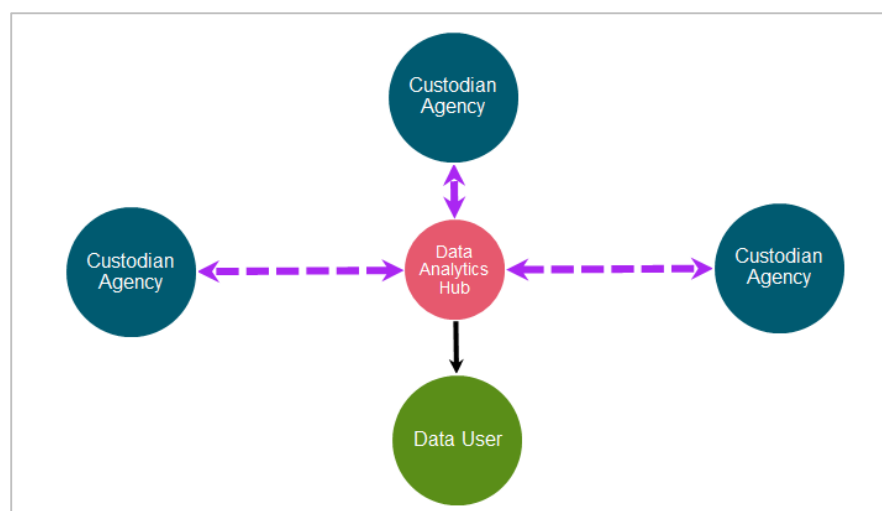
## 2.1 Data Analytics Hub

Under the concept of a Data Analytics Hub, datasets would be supplied by custodian agencies as part of an agile service model that responds to users' needs at a whole-of-government level.

The Hub would streamline data sharing processes to enable greater efficiencies and improve decision making for policy and service delivery. It would play a key role in contributing to the compilation of a national evidence resource that will enable analysts and researchers to create innovative research and statistics through the use of linked and enriched datasets.

The Hub could commence as a whole-of-government Commonwealth initiative, and eventually extend to incorporate state and territory datasets in line with relevant legislation.

### 2.1 Data Analytics Hub Model



### *Division of responsibilities*

Success of this vision depends on an efficient division of responsibilities between existing custodian agencies and the entity managing the Data Analytics Hub.

Custodian agencies would retain responsibility for collecting and managing their own data, including managing general data access arrangements (including metadata) for users, implementing quality assurance processes, ensuring data is provided in a format that is machine-readable, and that data is of high quality and complies with agreed open standards.

An infrastructure-light, minimally resourced and nimble coordinating entity would focus on managing appropriate data linkage arrangements and maintaining agreements between agencies, while custodians remain responsible for key data assurance and maintenance functions, leveraging current investments in business intelligence and physical infrastructure.

This function could be managed by an existing Commonwealth Integrating Authority or by a new standalone Commonwealth entity created for this purpose.

### *Creating an independent entity and enabling legislation*

The creation of an independent entity to manage the Data Analytics Hub would address many long-standing issues giving rise to agency inaction on data sharing. Barriers to data sharing, including excessive focus on technological and legislative barriers, could be addressed through creating an entity with responsibility for overseeing data linkage arrangements and maintaining agreements between agencies.

This is of particular relevance in the context of privacy fears hindering the use of data for policy development. It can take several years and multiple MOUs to establish data sharing arrangements between government agencies, whereas the actual process of data linkage can be enacted over a much shorter period. This has been acknowledged in PM&C's *Guidance on Data Sharing for Australian Government Entities* which advises data arrangements should now be formalised through a letter of exchange rather than MOUs or deeds of arrangement.

A central entity with responsibility for overseeing and managing data linkage arrangements within a suitably supportive legislative environment would also support greater efficiencies in data sharing through providing advice to agencies on standard practices for data linkage and data sharing, as well as best practice in establishing governance arrangements for data sharing.

Introducing enabling legislation specific to the Data Analytics Hub function may have the advantage of providing overriding principle-based legislation that makes explicit capabilities of agencies and minimises perceived barriers to data linkage while maintaining appropriate privacy protections.

### ***Investment in infrastructure-light data sharing arrangements***

Consultation with industry experts in data technology and design fields has indicated rapidly diminishing technological barriers to data linkage and establishing large-scale, infrastructure-light integrated data resources from multiple data custodians.

Any investment in this function should be centred upon establishing data sharing mechanisms rather than data storage functions. An infrastructure-light, agile service will also be more adaptable over time in responding to the needs of multiple stakeholder groups.

The Hub will enable greater efficiencies through drawing on existing investments in data asset infrastructure housed in custodian agencies.

For example, the Australian Government has invested in the development of Australia's data capabilities through the creation of the Secure Unified Research Environment (SURE) with funding from the National Collaborative Research Infrastructure Strategy (NCRIS) as part of the Population Health Research Network (PHRN). The PHRN is a collaboration that was set up in 2009 to further develop Australia's data linkage capabilities.

SURE is a computing environment developed to help make best use of our national knowledge base. It is helping to bring researchers together from across Australia to collaborate on large-scale projects tackling major health and social issues such as population ageing, diabetes and mental health.

An infrastructure-light hub will maximise the benefits of investments already made in data infrastructure by the Australian Government, and create greater efficiencies in data sharing through using a 'just-in-time' (JIT) service model that responds to users' needs.

### ***Data-on-demand – a just-in-time service model***

The Data Analytics Hub will enable the provision of 'data-on-demand' using a just-in-time service model. Users of the Hub should be able to request a combination of existing datasets, which a coordinating agency could then request from respective custodian agencies. Enabling a system which can be updated on demand will allow for an iterative interaction that is consistent with the research process.

JIT is an inventory management strategy used in manufacturing to increase efficiency and decrease waste through receiving goods only as they are needed in the production process, which reduces inventory costs.

The advantages of a JIT strategy are applicable to a data sharing scenario. Rather than simply creating infrastructure-heavy resources to support data warehousing, a JIT model will keep costs low by drawing on existing data assets from custodian agencies only as required in real time.

A JIT service model will enable greater efficiencies and security in data sharing through both reducing costs involved in storing data, and being future proofed and adaptable to respond to changes in user need over time.

### ***Domain for secure access and purpose built datasets***

The Data Analytics Hub would be enabled by a domain for secure access by government and researchers. As stated above, custodian agencies would draw on existing data infrastructure to provide datasets to the coordinating entity of the Data Analytics Hub.

Greater efficiencies in data linkage arrangements will be enabled through providing curated datasets purpose built by custodian agencies.

The curated dataset would represent a core research dataset for each agency that is well categorised and supported by a data quality statement and user guide that include metadata definitions. The curated dataset would be a subset of an agency's total public data assets, and would allow custodian agencies to commit time to developing a high utility, linkable dataset that meets the majority of users' needs.

This approach would allow for efficient and effective data linkage over time, and enable streamlined and efficient processes for updating and maintenance on a regular basis. This approach represents a more efficacious and secure system for data linkage than each custodian agency providing all of their data to a coordinating entity en masse.

### ***Comparison to New Zealand's Integrated Data Infrastructure***

The Data Analytic Hub function would enable a similar data linkage process to New Zealand's Integrated Data Infrastructure (IDI), but with greater focus on providing a model that responds to changing user needs, while maintaining higher infrastructure utility.

Approval to access information in the IDI is granted by the Government Statistician under New Zealand's *Statistics Act 1975*. Research applicants must also have the necessary qualifications and experience to use the integrated data.

Enabling legislation for the Data Analytics Hub should allow for broader use and re-use of linked public datasets while maintaining appropriate privacy protections.

The Data Analytics Hub may build on work currently being undertaken by the Multi-Agency Data Integration Project (MADIP) as outlined above.

### **Part 3 – Managing Commonwealth Debt through Data Sharing**

Sharing of debt-related data within and across Commonwealth agencies would allow for improved efficiency in the management of Commonwealth debt and an improved access for individuals interacting with government agencies.

Commonwealth debt is currently managed on an agency basis, and in some cases on a payment/programme basis, using disparate management practices. For DSS, there are numerous separate pieces of legislation for social services across the three key areas of social security, family assistance and child support which limit the sharing of debt related data.

Legislative limitations in portfolio specific legislation also prevent the effective sharing of debt-related data between agencies. For example, the *Taxation Administration Act 1953* restricts sharing of tax information to specified purposes, and limits the use of Tax File Numbers as a de facto client identifier.

Collectively, these legislative limitations prevent effective debt-related data exchange within and between agencies and limit a coordinated approach to debt management.

#### ***Commonwealth individual debt***

The recovery of individual debt comes at a significant cost to the Australian Government and community. In the social security portfolio alone, there was almost \$3.5 billion of social welfare debt as at 30 June 2016. It is understood a large component of Commonwealth debt is attributable to the taxation obligation of individuals with business related activities, i.e. sole traders and partnerships.

#### ***Review of legislation to enable debt related data sharing***

In line with DSS's position on managing Commonwealth debt, any review of legislation to support the use and re-use of public sector data could also consider legislative limitations to the sharing of debt related information.

As stated above, legislative limitations to sharing debt related data will include portfolio specific legislation in addition to provisions under the *Privacy Act*.

#### ***Tell-us-once approach to debt management***

Sharing information about an individual's debts within and across agencies would provide a platform for a more coherent and streamlined experience for individuals interacting with multiple government agencies. It would support a 'tell us once' approach to interactions with government, providing improved efficiency for individuals interacting with government agencies through reduced compliance and reporting obligations.

This streamlined approach would reduce the administrative and compliance costs of debt collection and increase recovery rates over time.

#### ***Initiatives involving debt related data***

Subject to the passing of enabling legislation, Single Touch Payroll (STP) would streamline and automate tax and superannuation contributions reporting.

As a result of STP, the Australian Taxation Office (ATO) would receive a regular stream of new employer and employee payroll data that would also have significant value to other Commonwealth agencies to undertake their own processes. For example, STP data could be applied to detect undeclared and under-declared employment income in relation to social security payments and for reviewing income estimates for family assistance payments. This would likely have significant

benefits for social security and family assistance payment accuracy and integrity and reduce costs for business and individuals to comply with social security and family assistance legislation.

However, the *Data-matching Program (Assistance and Tax) Act 1990* and existing Commonwealth taxation laws limit and/or prohibit the use of Tax File Numbers for data matching purposes beyond those outlined in these Acts. Without reform of current legislation, the use of third party STP data for social security and family assistance purposes would be limited and the additional benefits for payment integrity and compliance could not be fully realised.

## Part 4 – Data Availability and Use Issues Paper Questions

The Commission's Issues Paper raises questions on matters relevant to the Department. Responses to questions are provided below:

***What public sector datasets should be considered high-value data to the: business sector; research sector; academics; or the broader community?***

DSS supports PM&C's recommendation that a Commonwealth framework for high-value datasets be established (PM&C, *Public Sector Data Management Report*, July 2015). The high-value dataset framework should identify datasets that hold the greatest potential to support economic, social and environmental outcomes, and those regularly requested by stakeholders.

When considering which public sector datasets should be considered high-value to different stakeholder groups, it is important that government responds to needs as identified through public consultation.

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

Datasets considered to be of high-value of which DSS is custodian include:

- Basic Data Set (BDS) - not publicly available (NPA)
- Bereavement Lump Sum SuperCROSS - NPA
- Bluebook Dataset - NPA
- Cognos cube – Crisis Payments – Grants and Rejections (Post July 2003) - NPA
- DSS – Carer One-Time Payment Dataset - NPA
- DSS – Carers Dataset - NPA
- DSS Essential Medical Equipment Payment Dataset - NPA
- DSS – Income Support for Seniors Dataset - NPA
- Housing Data Set (HDS) - NPA
- Humanitarian Entrants Management System (HEMS) - NPA
- Parental Leave Payments Dataset - NPA
- Settlement Database – Settlement Reporting Facility - NPA
- SuperCross – Special Disability Trust - NPA
- Income Management Summary by Measure inc Basics Card – subset of a dataset that is not publicly available
- Income Management Summary Data
- DSS Payment Demographic Data - subset of the Basic Data Set which is not publicly available
- Settlement Database (SDB) Popular - NPA
- The Longitudinal Study of Australian Children - NPA
- Journeys Home - NPA



- Housing Data Set - NPA
- Unemployment Benefits Data (Bluebook Dataset as above) - NPA
- Household, Income and Labour Dynamics in Australia Survey - NPA
- Longitudinal Study of Indigenous Children (LSIC) - NPA

### ***What characteristics define high-value datasets?***

High-value data has been defined as ‘data that could have a higher value if made available as open data due to its authoritativeness, timeliness accuracy or other traits’ (PM&C, *Public Sector Data Management Report*, July 2015). As outlined above, high-value datasets hold the greatest potential to support economic, social and environmental outcomes, and those regularly requested by stakeholders.

Different approaches have been taken in defining high-value government data internationally. The US Open Government Directive defines high-value government data as any data that can “increase agency accountability and responsiveness; improve public knowledge of the agency and its operations; further the core mission of the agency; create economic opportunity; or respond to need and demand as identified through public consultation.” ([US Open Government Directive, 2009](#))

Definitions of ‘high-value’ data will differ according to the interests of specific stakeholder groups. Individual researchers may place more emphasis on characteristics such as longitudinal capacity, accurate identifiers to enable data linkage, or availability of data on whole populations.

### ***What benefits would the community derive from increasing the availability and use of public sector data?***

There is significant economic value from opening access to public sector data and improving data analytics. A recent study revealed that ready access to government data (or public sector information) in Australia has the potential to generate a value worth up to \$25 billion per year, slightly more than 1.5 per cent of its gross domestic product (GDP, chain volume measures) in 2014 (Lateral Economics, *Open for Business: How Open Data Can Help Achieve the G20 Growth Target*, 2016).

### ***What are the main factors currently stopping government agencies from making their data available?***

There are significant operational challenges in sharing data and establishing linked datasets, including:

- Aligning multiple ethics committees and data custodian approval processes
- The complexities and costs of retaining data over time
- Privacy and legislative constraints, both real and perceived
- Consistency in reporting linkage variables across datasets (i.e. consistency of reported names can impact accuracy of record matching for individuals across collections)
- Data provided by states and territories must be compliant with relevant state and territory privacy legislation and any other legislation relevant to the data involved
- Many commonwealth agencies are also subject to other legislation, policies and procedures relating to how their data may be collected, stored, used an/or disclosed
- Ambiguity in Commonwealth arrangements and associated governance regimes
- A lack of direction and risk mitigation measures for decision makers
- Financial disincentives to preparing data for sharing
- The need to develop secure environments and facilities, and

- Minimal sharing and linkage investments made by departments to develop internal systems for data management and release.

However, these challenges do not present insurmountable obstacles to securely sharing data. Commonwealth initiatives previously referred to already demonstrate diminishing technological barriers to sharing data and establishing large-scale integrated data resources from multiple data custodians.

### ***Perception of privacy legislation as a barrier***

Conservative interpretations of privacy legislation can be a barrier to data sharing. This is recognised in the *National Government Information Sharing Strategy* (2009):

*‘The complexity of privacy laws often results in the default response to requests for information (that might be considered sensitive) as “We cannot share our information because of privacy laws.” This response is often given instead of determining (through the appropriate channels) whether the information can, in fact, be shared.’*

This situation is changing in line with advice provided in the *Australian Government Public Data Policy Statement* as well as PM&C’s *Guidance on Data Sharing for Australian Government Entities*.

The *DSS Data Access Policy* provides clear guidance for agency staff on how to appropriately and effectively make data available.

However, to fully address the perceptions of privacy legislation that give rise to agency inaction on data sharing, government agencies must demonstrate leadership in implementing effective governance and risk management arrangements, and establish a culture of making non-sensitive data open by default in line with the *Australian Government Public Data Policy Statement*.

### ***How could governments use their own data collections more efficiently and effectively?***

In the short term, government agencies should focus on ways to improve data quality and governance, and establish effective data sharing arrangements under existing legislation. However, in the medium term privacy legislation could be reviewed to identify whether privacy laws can be streamlined and modernised to enable data to be better used for policy and research.

A review of privacy legislation could also focus on harmonising privacy legislation across jurisdictions to promote data sharing between Australian governments. In line with DSS’s position on managing Commonwealth debt, this may include the review of legislation relevant to transparency of individual debt and the sharing of debt related information.

Governments could use their data collections more efficiently and effectively through implementing the proposed Data Analytics Hub function. This initiative would streamline data sharing processes to enable greater efficiencies and improve decision making for policy and service delivery.

The Hub would also address a number of the issues raised by the Commission’s Issues Paper, including improving data sharing between agencies through improved data linkage processes and the use of aligned standards for data sharing. The solution would maximise the potential of public data for informing evidence based policy.

As outlined above, the MADIP is a practical example of how government can use its data more efficiently and effectively. The Data Analytics Hub may solve residual public data access issues.

Data collections can be used more efficiently and effectively through aligning with the *Australian Government Public Data Policy Statement* and *Guidance on Data Sharing for Australian Government Entities* provided by PM&C in March 2016.

***Should the collection, sharing and release of public sector data be standardised? What would be the benefits and costs of standardising? What would standards that are ‘fit for purpose’ look like?***

Whole-of-government data standards and principles should be published drawing on existing (local and international) guidelines that outline how collection, sharing and release of public sector data should be standardised. Clear data standards should outline technical specifications for data and address electronic interchange of data.

For reference, the Australian Governments Open Access and Licensing Framework (AusGOAL) has published practices that aim to protect confidential and private information, third-party copyrights, and intellectual property, while managing the risks involved in providing non-sensitive public data as open by default. The Digital Transformation Office has also published design guides on improving government services through machine-readable and accessible data.

***What criteria and decision-making tools do government agencies use to decide which public sector data to make publicly available and how much processing to undertake before it is released?***

In deciding which public sector data to make publicly available and the processing required before its release, DSS uses the *Data Access Policy* and its overarching policy principles as outlined above.

The *Data Access Policy* sets out the guiding principles to increase access to DSS data and standardise processes relating to data access. It outlines the principles for access to all types of data for which DSS is custodian and criteria and decision making tools for DSS decisions on which public data to make publicly available.

The policy’s unit level record principles also provide guidance on how much processing should take place before data is released, including that aggregate or summary data should be easy to access and accessible in formats that promote its reuse (machine readable). It also outlines data suppression and perturbation practices to confidentialised data where necessary.

The policy supports the access of unit level record data in a controlled environment through secure infrastructure, and provides that only unit record data essential to meet the purpose of a request be accessed by a user. Where DSS grants controlled access to unit level record data, DSS must assess, and if necessary prioritise, each unit record data request on its own merit with focus on potential benefits accruing to government, and supporting the development of evidence based social policy.

DSS adheres to the *Commonwealth Arrangements for Data Integration*, which outline the arrangements for the integration of Commonwealth data for statistical and research purposes.

DSS aligns with the *Privacy Act 1988*, the *Australian Privacy Principles* (APPs), and advice from the Information Commissioner including the Privacy Management Framework.

DSS also aligns with relevant departmental specific legislation as outlined above, in addition to aligning more broadly with the *Australian Government Public Data Policy Statement*.

***Criteria for data linkage via the Data Analytics Hub***

In relation to data linkage, government agencies should use more selective criteria as part of sharing information via the Data Analytics Hub. Agencies could provide a curated dataset (representing a subset of an agency’s total public data assets) as part of data linkage arrangements. The curated dataset would be well categorised and supported by a data quality statement and user guide that include metadata definitions.

This approach would allow custodian agencies to commit time to developing a high utility, linkable dataset that meets the majority of users’ needs. This approach represents a more effective and

streamlined system for data linkage than each custodian agency providing all of their data to a coordinating entity en masse.

***What specific government initiatives (whether Australian Government, state, territory or local government, or overseas jurisdictions) have been particularly effective in improving data access and use?***

Government initiatives that have been effective in improving data access and use include:

- The National Centre for Longitudinal Data (as outlined above)
- The MADIP (as outlined above)
- Data.gov – an online central catalogue for accessing public data. data.gov.au provides an easy way to find, access and reuse public datasets from Government. It provides hosting for tabular, spatial and relational data with hosted Application Programming Interfaces (APIs) and the option for agencies to link data and services hosted by other government sources
- The National Map – a website for map-based access to spatial data from Australian government agencies
- Population Health Research Network - The PHRN is a collaboration that was set up in 2009 to further develop Australia's data linkage capabilities, and
- Commonwealth Data Integration Arrangements.

***Which rules, regulations or policies create unnecessary or excessive barriers to linking datasets?***

DSS recognises the importance of a trusted, transparent and balanced approach to sharing public data, where privacy and trust are maintained. DSS is committed to building and maintaining trust in the quality, confidentiality and value of public sector data with government, researchers, business and the public.

However, there is particular complexity where data linkage involves multiple jurisdictions – data linkage between different jurisdictions must be compliant with relevant state and territory privacy legislation which is not always consistent across jurisdictions and does not exist in the case of Western Australia and South Australia.

A review of privacy legislation could also focus on harmonising privacy legislation across jurisdictions to promote data sharing between Australian governments.

Many Commonwealth agencies are also subject to other legislation, policies and procedures relating to how their data may be collected, stored, used an/or disclosed. For example, in addition to complying with its obligations under the *Privacy Act*, DSS is also subject to requirements under a range of legislation and guidelines specific to the Department. This is discussed in more detail in the section on privacy protection above.

The requirements for providing access can vary between Acts and Guidelines. This can add unnecessary complexity when information is required for research or analysis that spans more than one area.

Government agencies would benefit from administrative arrangements that clearly and effectively outline best practice for data linkage arrangements. The ambiguity of current arrangements creates an unnecessary barrier to linking datasets.

In relation to any review of privacy legislation, DSS suggests that the Information Commissioner review the recommendations made as part of the Australian Law Reform Commission's 2008 inquiry into the *Privacy Act*.

As outlined above, legislative limitations also prevent the effective sharing of debt-related data within and between agencies, which limits a coordinated approach to debt management. In line with DSS's position on managing Commonwealth debt, any review of legislation to support the use and re-use of public sector data could also consider legislative limitations to the sharing of debt related information.

***How can Australia's government agencies improve their sharing and linking of public sector data? What lessons or examples from overseas should be considered?***

Improved sharing and linking of public sector data will be facilitated through the implementation of a shared commonwealth Data Analytics Hub as outlined above. The Analytics Hub would be effectively supported by a Trusted User Model, whole-of-government data standards and principles, consistent and transparent user charging for data, and a legislative environment that supports the use of data while maintaining privacy.

As mentioned above, the introduction of enabling legislation specific to the Data Analytics Hub function may have the advantage of providing overriding principle based legislation that makes explicit the responsibilities of agencies and data users respectively, and minimises perceived barriers to data linkage while maintaining appropriate privacy protections.

***How should the costs associated with making more public sector data widely available be funded?***

Where Australian Government agencies (non-corporate Commonwealth entities) publish data, it should be provided at no cost to the user unless there are substantial additional costs associated with providing the data. Public sector data has strong public good characteristics, including being non-rivalrous (i.e. use by one individual does not reduce availability to others), and non-excludable (i.e. once it is made available its use is cannot be easily limited).

Public sector data is a strategic national resource that holds considerable value for growing the economy, improving service delivery and transforming policy outcomes. Increasing the number of users with access to public sector data will provide a range of economic and social benefits, in addition to providing economies of scale for government agencies in collecting data.

This provides a strong rationale for providing public sector data at no cost to users where possible, recognising the broader economic and social benefits derived from making non-sensitive public sector data available for use and re-use.

In addition, costs involved in data linkage initiatives could be offset by efficiency and productivity gains with the implementation of the Data Analytics Hub.

Where there are substantial costs associated with providing data, a 'user pays' model is appropriate. For example, specialist researchers requiring access to highly detailed data in a secure environment will involve costs beyond the scope of normal expenditure to make non-sensitive public data available for use and re-use.

***To what extent are data-related resources in agencies being directed towards dealing with data management and access issues versus data analysis and use?***

DSS has dedicated resources towards data management and access issues including establishing an internal governance structure. The DSS Data Governance Committee oversees and provides strategic guidance on data policies, systems, activities and projects, including data related investments.

***How could coordination across the different jurisdictions in regard to privacy protection and legislation be improved?***

A review of privacy legislation could also focus on harmonising privacy legislation across jurisdictions to promote data sharing between Australian governments.

***How effective are existing approaches to confidentialisation and data security in facilitating data sharing while protecting privacy?***

DSS currently uses a blanket suppression method to ensure privacy and security principles are adhered to while sharing data. This has proven over time to be a very effective way to maintain individual privacy.

Currently, the Department is investigating the use of several methods of rounding and perturbation to support the release of data in a machine readable format. We would envisage this to have the same success as blanket suppression while allowing end users more usability of flat data sets.

***How can individuals' and businesses' confidence and trust in the way data is used be maintained and enhanced?***

DSS recognises the importance of a trusted, transparent and balanced approach to sharing public data, where privacy and trust are maintained.

Following recommendations by New Zealand's Data Futures Forum, the Data Futures Partnership was established as a cross-sector group of influential citizens to work on issues involved in creating a framework for a trusted data-use environment at a national level.

The public facing working group is actively coordinating with citizens, consumers, businesses, non-governmental organisations, and government to drive greater trusted data sharing and use.

Australia may benefit from a similar public facing body that engages with the issues involved around the social licence for data use and reuse.