



Reference: A1710941

Mr Peter Harris AO
Chairman
Productivity Commission
GPO Box 1428
Canberra City ACT 2601

Dear Mr Harris

I am pleased to provide the NSW Government's Submission to the Productivity Commission's inquiry into data availability and use.

The NSW Government is a leader in data availability and use and strongly supports increasing the availability and use of data across government and between the public and private sectors. Doing so will help drive economic growth, increase innovation and improve the efficiency and effectiveness of public and private sector services. In particular, we support greater data-sharing between state and territory and Commonwealth Governments.

In addition to supporting greater data-sharing the Submission focuses on three areas which are key to unlocking the value and opportunities embedded in data. These are:

- Willingness to share - a mix of cultural, awareness and capability issues
- Ability to share - involving data format, quality, uniformity and technical issues
- Permission to share - covering legislation, regulation, security and governance issues.

NSW will continue to champion data as a key asset to improving productivity and delivering better services. I look forward to assisting you and the Commission as the inquiry progresses.

Should you wish to discuss any of the points raised in our Submission, please contact Jeremy Harris, Director in the NSW Department of Premier and Cabinet

Yours sincerely

Blair Comley PSM
Secretary

NSW Government Submission to the Productivity Commission Inquiry into Data Availability and Use

July 2016

Contents

Introductory comments	3
Public sector data	4
High-value public sector data (Issues Paper p.14)	4
Collection and release of public sector data (Issues Paper p.14)	8
Data linkage (Issues Paper p.14)	14
Private sector data	19
High value private sector data (Issues Paper p.18)	19
Access to private sector data (Issues Paper p.18)	20
Consumers' access to and control over data about them	23
Consumer access to, and control over data (Issues Paper p.21)	23
Resource costs of making data available	25
Resource costs of access (Issues Paper p.22)	25
Trust and privacy	28
Privacy protection (Issues Paper p.26)	28
Data security and Restrictions around the release of particular data	31
Other restrictions (Issues Paper p.27)	31
Data security (Issues Paper p.28)	31
Appendix 1 – sample high-value datasets	33

Introductory comments

Until recently, the ability to access, analyse, use and reuse data has been time-consuming and costly. This has generally resulted in the value of data being limited to the purpose for which it was originally collected. Technological advances have dramatically lowered the barriers to data availability and use. Similarly, data science and predictive analytics are enabling datasets to be used in ways that were not considered when the original data was collected five, ten or fifteen years ago.

These advances are transformational – where once data was an administrative burden, it has the potential to become one of the most valuable resources for improving social wellbeing and driving economic growth. The NSW Government recognises the value of increasing the availability and use of data and is already taking steps to address three key barriers:

1. Unwillingness to share data – which is a mix of cultural, awareness and capability issues
2. Inability to share data – which involves data format, quality, uniformity and technical issues
3. Not being allowed to share data – which crosses legislation, regulation, security and governance issues.

Addressing these issues will help drive innovation and competition for greater client-focused services. It will also drive improvements in the efficiency and effectiveness of both public and private sector service provision.

The NSW Government is a leader in data availability and use and will continue to champion data as a key asset in improving productivity and unlocking opportunity. Since 2013 the NSW Government has implemented several initiatives promoting improved collection, sharing and usability of public sector data while maintaining the rights and privacy of individuals and organisations.

- The 2016 NSW Government Open Data Policy¹ aims to build stronger relationships with the private sector and the research sector to encourage the release of their data to the public and increase their use of public sector data.
- In 2015, NSW established its Data Analytics Centre (DAC)² to facilitate data-sharing across government agencies to inform more efficient, strategic, whole-of-government evidence-based decision-making.
- The NSW Government Information Management Framework³ identifies data standards that enable data to be ‘mapped’ for sharing between systems and agencies or released to the public.
- NSW is also making data available in real time, using technologies such as application programming interfaces (APIs) and web services, increasing its usability and decreasing the costs associated with release.

The responses in this Submission are to selected questions from the Productivity Commission’s Issues Paper released on 18 April 2016.

¹ <http://data.nsw.gov.au/2016-nsw-government-open-data-policy-and-action-plan>

² <https://www.finance.nsw.gov.au/ict/nsw-data-analytics-centre>

³ <https://www.finance.nsw.gov.au/ict/priorities/managing-information-better-services/information-management-framework>

Public sector data

High-value public sector data (Issues Paper p.14)

1. What public sector datasets should be considered high-value data to the: business sector; research sector; academics; or the broader community?
2. What characteristics define high-value datasets?

The NSW Government has defined characteristics for high-value data

Our Open Data Policy, developed in 2013, describes high-value data as data that:

- has the potential to enhance services or service delivery
- will support or demonstrate evidence-based policy-making or research
- will create economic opportunity, generate efficiencies, or reduce costs
- further the core mission of the agency
- increase agency accountability and responsiveness
- increase government transparency
- respond to a need and/or demand identified through public or stakeholder engagement, or supports positive social outcomes.

Whether public or private sector data, key characteristics of high-value datasets are:

- accuracy (e.g. providing a clear and sufficient representation of the activity/interaction)
- validity (e.g. recording and using the data in accordance with agreed requirements, rules and definitions to ensure integrity and consistency)
- reliability (e.g. developing and implementing data collection processes to ensure consistency over time, so that data accurately and reliably reflects any changes in performance)
- relevance (e.g. data should be relevant for the purposes for which it is used and the amount of data collected should be proportionate to the value gained from it)
- currency (e.g. datasets containing current data are more valuable than out-dated data)
- completeness (e.g. data should be complete and should not contain redundant records)
- compliance (e.g. data must comply with regulations on data protection and security) and
- accessibility or usability (e.g. data should be available in a machine-readable format and not require the use of sophisticated software or skills in order to access it).

Data should also be of known quality. Quality is determined by whether or not the data is suitable for its intended use. To make the most of open and shared data, users will need to define what data quality means for their specific aim or objectives. Even if data does not meet the primary user's objectives, it may hold value for a secondary user.

NSW has adopted the Australian Bureau of Statistics (ABS) Data Quality Framework⁴ to describe the dimensions (or characteristics) of data quality. NSW has also developed a standard approach to reporting data quality. A data quality statement helps secondary users understand how the dataset could be used, its comparability to similar datasets, understand limitations in the data and determine whether it is fit for their purpose.

⁴ <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Quality:+The+ABS+Data+Quality+Framework>

The NSW Government releases high-value data through its Open Data Policy

Government and other data custodians cannot always accurately assess the value of data assets – the value is only realised by its use. Many third parties have innovative approaches to using data to generate value that we may not be aware of. That said, the NSW Government holds datasets covering a wide set of issues and policy areas. Datasets are both administrative and performance related. As well as data on service provision, they can include fiscal, extractive industries data, environmental and pollution data, and aggregated and statistical data.

Example datasets considered high-value to both public and private sectors are provided in more detail in Appendix 1 of this Submission, and include the following.

- ABS Census, Centrelink, Medicare and Pharmaceutical Benefits Scheme (MBS and PBS).
- National Assessment Program – Literacy and Numeracy (NAPLAN)⁵ test data for all Year 3, Year 5, Year 7 and Year 9 students – approximately 350,000 students annually since 2008.
- Data relevant to weather, crop, livestock, biosecurity and pest information, water and land lease availability. These data, and their availability, directly contribute to the growth of NSW primary industries, as well as the ability to manage compliance and regulation across the areas of biosecurity, fisheries, food, game licensing and water.
- Geological data relating to mineral and petroleum resources, including the Geological Survey of NSW's (GSNSW)⁶ structured open datasets, used to encourage business sector investment, assist research exercises and support academics studies.
- Local government planning and land-related data such as development applications are of immediate value to local communities, have the potential to yield value for medium- to long-term infrastructure planning, are a measure of the enablement of development and investment in communities and represent information relevant to emergency response services and their predictive prioritisation of incidents.

To assist with data provision the NSW Government is streamlining approval processes to increase the release of public sector data. Further, when re-launching the NSW Open Data Policy in April 2016, the NSW Government also appointed the NSW Information Commissioner as its Open Data Advocate, connecting open data with open government and provisions and protections of the *Government Information (Public Access) Act 2009* (GIPA Act). The Advocate provides information, advice and assistance to agencies and the public on accessing government information.

Longitudinal and spatial datasets are high in value for government agencies

For governments to best understand and drive improvements in social and economic outcomes, they need to be able to measure client outcomes across time, services, sectors and geographies. High-value data of this nature permits governments to identify patterns and trends in outcomes for specific client cohorts and to build a more complete picture of their needs. Longitudinal datasets are particularly valuable to policy-makers in providing evidence of behaviours and trends over time. This type of information assists policy-makers with decision-making and can provide evidence of how to direct investment to interventions that are most likely to improve long-term client outcomes.

For example, the NSW Office of Social Impact Investment (OSII) recently commissioned actuarial modelling and analytics work to understand the government service usage and outcomes of young people leaving out-of-home care (OOHC) over their lifetime. The modelling provided important insight into the major cost drivers and factors influencing individual social and economic outcomes⁷.

⁵ <http://www.nap.edu.au/naplan/naplan.html>

⁶ <http://www.resourcesandenergy.nsw.gov.au/miners-and-explorers/geoscience-information/about/geological-survey-of-nsw>

⁷ NSW Intergenerational Report 2016 Chapter 7 has initial results of a model commissioned from Taylor Fry based on data from NSW human services agencies including Family and Community Services, Justice and Health as well as Finance, Services and Innovation.

NSW is also building its capacity to use high-value data to track the impact of our investment on breaking the cycle of disadvantage for the most vulnerable people in NSW. The NSW Human Services Outcomes Framework consists of seven domains – home, health, education, economic, social and community, safety and empowerment. The Outcomes Framework is designed to be able to support a more detailed client cohort-level investment approach to measuring the impact of programs and services on client outcomes. This work is only possible through access to a variety of high-value datasets from across government agencies.

Public sector dataset value can be increased by integrating data from different levels of government

When dealing with vulnerable people and complex issues, the NSW Government recognises the benefits of linking data from different agencies and linking data held at the Commonwealth level. There is a need for relevant de-identified data from the Commonwealth and NSW Governments to be shared and linked on a formal, regular basis, rather than in one-off projects. This is particularly the case in policy areas where both jurisdictions are active.

In the area of child protection, both governments hold data on children in out-of-home care (OOHC) and their carers, as well as on education attendance, attainment and service access. Policy and service delivery could also be better informed by integrating data on social determinants of health, or on other potentially co-occurring factors like domestic violence or alcohol consumption.

In the area of health, NSW Health is running simulation trials to test new forms of service delivery including integrated care and telehealth. This work would be greatly enhanced if NSW Health were able to connect high-value Commonwealth data, such as on MBS and PBS claims, public and private hospitals, health insurance claims, disability and aged care and mental health.

Consistently collected spatial and location data has the potential to add significant value

In addition to policy-specific data, spatial datasets have some of the greatest potential for economic impact and could support 'use cases' from improving transport and logistics to helping resolve property disputes. All data 'happens' somewhere and so the spatial component of data associated with the location of services, vehicles, machines, buildings, infrastructure, livestock, weather and people brings added richness to the evaluation of services, identification of needs and new service creation.

Spatial data from Australian state and federal agencies should be released under open source licence. Harmonising the multiple frameworks for recording spatial data under one national standard would add significant public value. Australian spatial data has the potential to create significant value in understanding and driving improvements in social and economic outcomes.

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

Increased availability and use of public sector data improves the efficiency and effectiveness of government programs

The most challenging policy problems faced by the public sector and the broader community are complex and tend to have people's behaviour at their core. If we are able to measure and understand what is effective and what is ineffective in influencing client outcomes we can better direct investment. This would enable us to target interventions to cohorts of people based on their needs and characteristics and reduce the tendency for one-size-fits-all, prescribed public services.

The NSW Social Impact Investment Policy⁸ and transactions have driven improved evaluation and measurement of child and family outcomes. This provides the Government and our service provider partners with a fuller picture of client needs and the services that might benefit them. This enables the provider to adjust and improve their service model to achieve the specified client outcomes.

Another example of efficiency derived from increased availability and use of public sector data is the NSW transport data exchange program⁹. The program makes real-time data on public transport and traffic available in a range of machine-readable, spatially enabled formats. This has provided real benefits to the community by increasing productivity and quality of life, and improving efficient use of public transport infrastructure.

Community benefits are also achievable in community safety and resilience by increasing the availability and use, consistency and compatibility of information for disaster preparedness and emergency response planning. There could also be new economic opportunities in using data in making the best use of land and efficient use of water. Such data could be used by governments and private firms to develop tailored solutions that return value to the broader community.

Community benefits also include increased transparency, confidence and accountability

The benefits the community may derive from increasing the availability and use of public data also include increased confidence, accountability and the ability for the community to use data about itself to better address local needs or market failures. In recognition of this, the NSW Government has released data for the purposes of 'collective impact' initiatives. This enables communities to better understand issues affecting them and where government investment is directed. A good example of this is the Just Reinvest project in Bourke in Far Western NSW¹⁰.

Community benefits also align with the open government principles of transparency, citizen participation, accountability, technology and innovation¹¹. Public access to government data increases the transparency of government, supports accountability and contributes to a greater awareness of government decisions and processes. This increases community trust in government services, improves understanding of government processes, and encourage increased public participation. The NSW Public Service Commission's *State of the NSW Public Sector Report 2015: To the Next Level*¹² identified results of its 2015 Customer Survey showing the need for governments to focus on customer service that promotes integrity and trust, and builds perceptions of accountability.

The public should be involved in determining and contributing to high-value datasets

It is not possible for public agencies to anticipate all the data that could be considered high-value. The NSW Open Data Policy recognises this in requiring agencies to make their data open by default, protected where required, discoverable and usable, primary in nature and timely. However, as being developed by Data NSW¹³, processes could be established whereby potential users of open data can browse data resources, provide feedback on the value of available data, and make requests for access to data that is not yet released. Processes should also enable public input into data collection exercises and methods, to support the wider use and value of publicly-funded data resources.

⁸ <http://www.osii.nsw.gov.au/policy-and-guidance/>

⁹ https://opendata.transport.nsw.gov.au/site/en_us/home.html

¹⁰ <http://www.justreinvest.org.au/justice-reinvestment-in-bourke/>

¹¹ Also see Organisation for Economic Co-operation and Development, 'Modernising Government The Way Forward', (2009), p.29.

¹² <https://www.psc.nsw.gov.au/reports---data/state-of-the-sector/state-of-the-sector-2015>

¹³ <http://data.nsw.gov.au/>

Collection and release of public sector data (Issues Paper p.14)

4. What are the main factors currently stopping government agencies from making their data available?

The NSW Government makes a substantial amount of data available either publicly via sites such as HealthStats NSW¹⁴ or through data release processes to researchers. We estimate that on average 500 million individual-level, linked and de-identified health records are released for analysis yearly. Despite this there are a number of challenges to making data available for use.

Privacy, confidentiality and security of personal information processes are complicated

Government agencies must ensure that data-sharing protects people's rights and privacy. Privacy risks should also be weighed against risks of *not* sharing data and the public interest or benefit of sharing particular information. Government agencies should ensure that their leadership and governance arrangements include a focus on release and an 'access by design' approach to embed information-sharing into agencies' culture and procedures at the design stage of a project.

Data-sharing can be unduly limited by a combination of cultural risk aversion and misunderstandings about using data under privacy and other legislation. The risk of privacy being breached should be addressed through a 'privacy by design' approach¹⁵ and with the appropriate de-identification and application of technological solutions. There is also a lack of understanding of how technological advances can be applied to maintain confidentiality and enable the reuse and accessibility of information. This is explored further in subsequent sections of this Submission.

Governance, control and ownership can lack clarity and hold up release

The lack of standard governance processes is a significant hurdle for streamlined sharing and release. Further, a sense of uncertainty about responsibilities can result in data creators being reluctant to provide open access to agency data. Strong governance is required when releasing or sharing public sector data. At a minimum, this includes:

- Identifying the data to be shared and specifying its retention, storage (and deletion) and publication requirements
- Undertaking security risk and privacy impact assessments to understand the likelihood, consequence and management of a breach
- Having ethics or other relevant approvals in place.

Further, government agencies can have limited control on how data is used once it is in the public domain. The datasets often have a high level of complexity and there is the potential for misunderstanding, misuse and misinterpretation of the information. The use of data quality statements, contextual metadata and development of data dictionaries will help address these concerns.

Data quality is often cited as a reason for not making data available

Agencies may not wish to release datasets which are not considered 'high quality' (e.g. in terms of fitness for use, reliability, validity or completeness). This can be due to a lack of data definitions, fragmented data systems that may be inconsistent, and where data maintenance and quality assurance protocols are lacking. Complex or antiquated systems and infrastructure can also make data extraction complex and error prone.

¹⁴ <http://www.healthstats.nsw.gov.au/>

¹⁵ <https://www.ipc.on.ca/english/Privacy/Introduction-to-PbD/>

Agencies may be unwilling or unable to share real-time raw data that has not been processed because they fear data can be misinterpreted or interpreted in ways that are perceived as unfavourable. It is often considered more appropriate to present data as a 'complete picture' rather than in fragments that may not represent all aspects of a situation, program or broader datasets.

Some data will have good temporal resolution and poor spatial resolution, and others will have good spatial and poor temporal resolution. This means that data analytics projects often require the triangulation of many data sources to get the best of all worlds.

Data collection and secondary uses are not often designed together

Public sector data has generally been collected for an administrative or reporting purpose. Collection may not be well designed or documented and standards may not be consistently applied. As a result, it can be difficult to access and reuse data for other value-adding purposes.

Data reuse can also be difficult, particularly where data platforms (e.g. the Australian National Data Service¹⁶) point to where datasets are found, but do not contain the data themselves. Additionally, the value lost through 'ephemeral data' can limit data availability. Much administrative data is kept in a format that means it is deleted as it is used (i.e. in accounting software a new balance replaces the old). This may require investment in new systems and servers to store additional data.

Costs of sharing data can be prohibitive or are not adequately met

The perceived and real costs of preparing datasets and making them routinely available in a safe and secure manner can be a barrier to making data available. This can be made more difficult when the social and economic benefits to the community are not easily or directly quantifiable.

For example, removing identifiers from data is not only a process of replacing names or dates of birth with other less linked identifiers, but also a process of identifying and addressing where other situations might arise where data could be traced back to an individual or group. This process requires adequate training and expertise and often a significant resource commitment.

Particularly in human services, additional contextual information and data quality statements often need to be documented to avoid misinterpretation and inappropriate use of data. Other policy factors, such as the resource cost of designing caveats or disclaimers for standard collection and release of data, can also create a disincentive to making data public.

Skills and capabilities in the use of data are not directed to the appropriate areas

The data resources in government agencies are often directed to reporting rather than analysis. As a result, data science and analytics are treated as a second-order priority. The knowledge and mechanisms required to extract, interrogate, manipulate, analyse, communicate and interpret data are not often present in public agencies. When they are present, such as in the research and evaluation units in the NSW government departments of Education, Family and Community Services, Health, Justice and Transport, these units are not necessarily the data custodians and may not always have the readily available access to data that they could do.

Staff with relevant skills in terms of data governance and legal requirements, such as those in records management or privacy and the GIPA Act, may not be involved in the design of processes for data collection and release. These staff also may not have a strong understanding of key technologies such as APIs and web services that can automate the application of policies and processes for ongoing data release.

¹⁶ <http://www.ands.org.au/>

5. How could governments use their own data collections more efficiently and effectively?
6. 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?

The lack of consistent, comprehensive information release frameworks is a major obstacle

At a national level, one of the most important interfaces affecting the flow and value of public data occurs between jurisdictions. The absence of a consistent and nationally harmonised approach limits the capacity of governments to coordinate services, improve planning, and more generally streamline the proactive release of datasets. This can manifest in a number of ways:

- The additional effort needed by agencies or sectors to create one-off arrangements for the flow of information (e.g. current work to establish a national picture on education outcomes)
- Difficulties faced by organisations trying to grapple with national issues but having only partial, fragmented data (e.g. child protection care arrangements and payments)
- The challenges when service provision involves non-government providers which may lie outside or in differing formal jurisdictions of public access and privacy regimes
- The varying approaches between regimes on the extent to which they focus on the agency-public axis alone, neglecting agency-agency sharing opportunities.

Governments should invest in streamlined data collaboration and sharing processes within, across and external to state and federal governments

Governments hold large volumes of data, often in agency silos. This data is of limited value when viewed in isolation, but when combined with other datasets can become a powerful tool for decision-making. Governments should establish streamlined data collaboration, record linkage and sharing processes across agencies and jurisdictions to clarify responsibilities and increase timeliness.

For example, child protection system datasets are high-value datasets, but they provide only part of the picture of factors impacting on children and their families. When linked with health, education, juvenile justice and Commonwealth health and welfare datasets they could provide a 'human services' view of vulnerable children and families.

Commitment to agreed Australian Standards would address many cross-jurisdictional challenges as well as build trust

Agreement on and commitment to national standards and a common lexicon to describe all aspects of data management and handling would significantly enhance the ability for data reuse by government and the broader community. This should incorporate a common test for personally identifiable data.

Such standards would increase ease of use, consumer and user certainty and increased safety and security by helping to ensure consistent applications of legislation and preservation of privacy. Establishing protocols for sharing data between and outside government agencies would also create a greater expectation that data should be shared.

Standards may be principle-based, similar to those of the Organisation for Economic Co-operation and Development (OECD)¹⁷, ABS¹⁸ and Stats Canada¹⁹. This should include standardised classifications, counting rules and definitions.

¹⁷ <http://www.oecd.org/statistics/qualityframework>

¹⁸ <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Quality:+The+ABS+Data+Quality+Framework>

¹⁹ <http://www5.statcan.gc.ca/olc-cel/olc.action?objId=12-539-X&objType=2&lang=en&limit=1>

Having a minimum dataset of standardised data items across all jurisdictions would allow valuable direct comparisons to be made across jurisdictions over time. As an example, the public health sector in Australia has data collection standards that have been established by the Australian Institute of Health and Welfare through its publicly available Metadata Online Registry (METeOR)²⁰.

However, standardisation should not come at the expense of customised data required for specific purposes. It will be important for agencies to maintain autonomy and flexibility in management of data relating to their service. Modern data interrogation and reporting tools mean that both purposes can be served.

NSW's approach (described in our Standard Approach to Information Architecture²¹) requires the application of standards at the 'integration layer', which provides a common point for all systems to map to but allows agencies the freedom to take business-specific approaches to their own internal data management practices.

Governments should invest in analysis and modelling capability

Building analytics capability throughout agencies is key to increasing the level of maturity in the use, analysis and application of data and communication of insight. The best analysis remains of limited value if it cannot be understood and acted upon by frontline staff in making decisions.

Analysis also needs to be rigorous and outcomes-focused to be of greatest benefit. As outlined in the NSW Government Program Evaluation Guidelines 2016²², economic, outcomes, and process evaluations can answer different policy and operational questions to:

- Inform decision-making and promote efficient resource allocation
- Identify who the program works best for and under what circumstances
- Continually improve programs.

Demonstrating the potential of data-sharing can address misconceptions and build support

The creation of the NSW DAC and its accompanying data-sharing legislation provide an avenue to share more data to inform public policy, service design and delivery, and may create increased openness through its enabling legislation and approach. The DAC is also carrying out projects that demonstrate the benefits of cross-agency data-sharing and analytics.

Other NSW projects such as the Human Services Data Hub and the Environmental Data Portal are also demonstrating the value of collaboration on data-sharing and open data initiatives. This is helping to address misconceptions and build support for new ways of working.

7. 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?

Open government legislation supports decisions about the release of data

The GIPA Act establishes a legislative basis for proactive and mandatory release of information. The NSW Open Data Policy aligns with the GIPA Act, establishing a common approach for data release. The objects of the GIPA Act are to maintain and advance a system of responsible and representative democratic government that is open, accountable, fair and effective.

²⁰ <http://meteor.aihw.gov.au/content/index.phtml/itemId/181162>

²¹ <https://www.finance.nsw.gov.au/ict/resources/nsw-government-standard-approach-information-architecture>

²² http://www.dpc.nsw.gov.au/programs_and_services/policy_makers_toolkit/evaluation_in_the_nsw_government

The GIPA Act applies to all NSW government agencies and encourages the proactive release of information by NSW public sector agencies and gives members of the public a legally enforceable right to access government information, ensuring that access is restricted only when there is an overriding public interest against releasing that information.

Decisions about release of data should be based on public use and input

The NSW Government Open Data Policy requires data to be prioritised for release or improvement according to feedback from the public. A 'rolling schedule' for releasing new data and for improving the format and usability of already open data is developed in consultation with agencies. Requests for the release or improvement of data can be made through the Data NSW open data catalogue.

Additional tools and guidance are being developed to support implementation of the Open Data Policy, including a checklist of questions to help agencies identify high-value data or prioritise data for release. Data NSW is being updated to include mechanisms that allow public users to rate open data and provide feedback on the utility of the data.

Other examples of tools used by NSW agencies are as follows.

- NSW Health data is provided to national health agencies and 500 million records per year are released to academics for research purposes under Section 221(1) of the *National Health Reform Act 2011 (Cth)*.
- The Department of Industry works with Geoscience Australia and Government Geoscience Information Committee (GGIC)²³ programs to identify opportunities for circulating GSNSW data to wider audiences.
- The Family and Community Services (FACS) External Research Program provides data to external researchers to support robust analyses that inform policy, service design and local planning.
- The NSW Bureau of Crime Statistics Analysis and Research (BOCSAR)²⁴ provides unit record data to researchers with ethics approval who agree to conditions similar to FACS.

8. 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?

The NSW Government's Data Analytics Centre and data sharing legislation support data access and use within government

As discussed elsewhere, in 2015 the NSW Government established the DAC to facilitate data-sharing across government agencies and to inform more efficient, strategic, whole-of-government evidence-based decision-making. The DAC is responsible for:

- Collecting and analysing cross-cluster or whole-of-government data (including from NSW government agencies, state-owned corporations, and local councils and instrumentalities) in a secure, central and neutral environment
- Building on existing data analytics capability across government by connecting datasets, people, expertise and tools to help realise the potential of the Government's data assets
- Working with agencies to remove barriers to data-sharing and driving connections across government and non-government organisations

²³ <http://www.geoscience.gov.au/about>

²⁴ <http://www.bocsar.nsw.gov.au/>

- Exploring new partnerships and ways to innovate and lead data analytics in NSW
- Promoting access to and sharing of datasets and projects to continuously build the strategic asset of government-held information.

Since late 2015 the DAC has commenced delivery of ten priority projects with 12 to come, working with agencies to share data and develop partnerships with external specialists. Competitions such as hackathons have also been effective as they have encouraged the participation of different communities in using government data.

NSW also established the *Data Sharing (Government Sector) Act 2015* (DSGS Act) to enable the DAC to fulfil its role and address both real and perceived barriers to the sharing of non-personal information. In a manner recognising the protection of privacy, the DSGS Act promotes the management and use of government data as a public resource that supports good policy-making, program management and service planning and delivery. One feature of the DSGS Act is that it requires agencies to share government sector data with the DAC for analytical insight.

The DAC and the DSGS Act could serve as a useful model for other jurisdictions to improve the availability and use of data within government.

Sharing capabilities can build on successful initiatives in particular agencies

In the Department of Education the Education Datahub makes data available to the public²⁵. Also, the Centre for Education Statistics and Evaluation (CESE)²⁶ has been created to provide a 'one-stop shop' for education data and information needs. Other agencies have established portals which will be linked via Data NSW.

In the Health sector, the Centre for Health Record Linkage (CHeReL) provides a collaborative data linkage centre for NSW and the ACT²⁷. CHeReL provides a mechanism for access to linked unit-record data and expands access to linked health and related data for research, planning and evaluation. Additionally, the interactive web-based application HealthStats NSW allows users to access data and tailor reports about the health of the NSW population.

The My School website has increased availability of school-level information but could be improved

My School²⁸ has increased the accessibility of data and the increase in public interest in school-level information in Australia. My School represents a valuable, nationally consistent dataset on schools across all jurisdictions and sectors. For the first time in 2015 results are available for students who have completed the full cycle of year 3, 5, 7 and 9 assessments.

There is a range of other information published on My School which serves to improve transparency and accountability at the school-level on a nationally consistent basis. This information is of interest to parents and broader school communities. However, the current information provided and the way in which it is presented does not necessarily address what is needed by parents and school communities to engage more fully with their local schools and their children's performance.

Other jurisdictions locally and internationally provide useful examples for Australia

Further, national examples of improved data access and use are as follows:

²⁵ <http://www.cese.nsw.gov.au/cese-sites/access-education-data>

²⁶ <http://www.cese.nsw.gov.au/>

²⁷ <http://www.cherel.org.au/>

²⁸ <https://www.myschool.edu.au/>

- The National Library of Australia's Trove²⁹ brings together content from libraries, museums, archives, repositories and other research and collecting organisations
- The Atlas of Living Australia³⁰ is a collaborative, national project that aggregates biodiversity data from multiple sources and makes it available and usable online
- The Australian Census Longitudinal Dataset³¹ where a 5% random sample of records has been matched between Censuses using data linkage techniques that protect privacy to create a powerful dataset for many diverse uses
- The National Health Services Directory³² is a joint initiative of Australian governments that provides reliable and consistent information on health services
- National Map³³ provides access to a single platform for Government geospatial and ABS datasets. Users can also overlay their own datasets to provide richer, customised analysis.

In New Zealand, the Data Futures Forum³⁴ has responded to the requirement to ensure effective and comprehensive data governance. It has achieved this through the establishment of an independent data council recognised as guardians of the system, the review of legislation and development of fit-for-purpose legislation.

Data co-ordination bodies such as those in Denmark and Singapore are also considered effective in making data accessible and usable. Singapore Data³⁵ has been operational since 2011 and provides open access to its 70 government agencies. More than 100 apps have been created using data that is simply presented and easily shared in machine-readable form.

The UK Government published in 2013 its second national 'Open Government Action Plan'. One of the most significant operational commitments contained in this plan is the commitment to establish a National Information Infrastructure which will deliver an inventory and publication of government datasets that are "likely to have the broadest and most significant economic and social impact". The plan is supported by the paramount objective of responsible release of information.

Data linkage (Issues Paper p.14)

9. Which datasets, if linked or coordinated across public sector agencies, would be of high value to the community, and how would they be used?

Data on service or support use across agencies or jurisdictions is vital to improving wellbeing of vulnerable groups

Often the value of a dataset is difficult to assess when it is viewed in isolation. The full benefits of data can only be realised through connectivity and linkages with other data, and the insights this linked data provides. Our response to the questions on high-value public datasets (see questions 1 and 2 above) outlines the elements and characteristics of what makes a dataset high in value.

The NSW Government is also seeking to use public and private sector data to help achieve the Premier's and State Priorities³⁶. These 30 priorities are designed to grow the economy, deliver

²⁹ <http://trove.nla.gov.au/>

³⁰ <http://www.ala.org.au/>

³¹ <http://www.abs.gov.au/websitedbs/censushome.nsf/home/acld>

³² <http://www.nhsd.com.au/>

³³ <http://nationalmap.gov.au/>

³⁴ <https://www.nzdatafutures.org.nz/>

³⁵ <http://data.gov.sg>

³⁶ <https://www.nsw.gov.au/making-it-happen>

infrastructure, and improve health, education and other services across NSW. Achieving the targets set under the priorities requires us to specifically consider the following perspectives:

- How will data help inform service delivery
- How data can help individuals and entities choose and access the services and supports they need
- How data can be used to measure the impact of services and supports.

For example, improving outcomes for the most vulnerable members in our community is a joint responsibility across government, non-government organisations, private enterprise and the community. Linkage of public sector datasets would enable better service coordination for people with high social needs. It would help identify client pathways and intervention points to deliver tailored, effective solutions and improve educational, health, justice and other life outcomes for vulnerable people.

This is the policy response that has been adopted by the New Zealand³⁷ and Australian³⁸ Governments' Investment Approaches. The NSW Government is in the early stages of developing investment approach initiatives. These cohort-based reforms put the client at the centre of service design and delivery. To be successful, the investment approach uses data, evidence and monitoring across the service system to build and understand the full view of system cost and efficacy. This requires linked client-level data from both state and federal departments which would enable measurement of client interaction and improvements in wellbeing.

Customer journey mapping can help identify value and opportunities for data linkage

The customer journey is an important indicator of points where data linkage or coordination would provide value. Data on where a customer interacts with service providers, moves between services, from one agency to another, or across jurisdictions, can generate valuable information. These journeys can be mapped to identify points where data is collected or used for a decision. Data linkage or coordination can be used to improve the customer experience of a service but may also help us to understand how a service is performing, how it is being used and points for possible interventions.

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

Complex legislative regimes and competing administrative arrangements present barriers to information-sharing between government entities and with businesses and citizens

With the increase in data linkage activity, barriers arise from an absence of a strategic overarching framework to direct information release and provide confidence that governments will manage information responsibly. Overall, the general approach of governments to linking datasets remains largely passive, relying on interested parties to seek out specific data, the agency to adjudicate on its release and then pass on to the specific requestor. The result of this approach is a relatively low rate of dataset release, not to mention a resource-intensive and inefficient process.

There are several possible reasons for this. Research commissioned by NSW DPC and the NSW interagency Enabling Information Sharing Working Group³⁹ found that:

- Information sharing is perceived as complex and frontline and agency managers are risk averse in sharing information

³⁷ <https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/evaluation/investment-approach/key-findings.html>

³⁸ http://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/BudgetReview201516/Welfare

³⁹ http://www.dpc.nsw.gov.au/__data/assets/pdf_file/0006/171087/Opportunities_for_Information_Sharing.pdf

- There may be a lack of expertise interpreting privacy legislation that leads to information not being released when it could
- There is a need for clear guidance and direction that sets out what information may be shared, the circumstances, authority for information sharing and processes
- Agency technological systems hinder the facilitation of information sharing or easy access to de-identified personal information due to, for example, configuration, formatting and uncontrolled accumulation of data, complex record keeping practices
- Information is held in silos by agencies and not viewed as a resource to be shared.

Additionally, agencies have various unique identifiers that restrict data matching and these identifiers can have internal issues, such as individuals having duplicate identifiers. The inconsistency of data quality across datasets and organisations is also a barrier.

The Commonwealth Government has developed a whole-of-government set of arrangements to support statistical data integration and managed data release. These support the best practice sharing and linkage of data through well-organised, secure data linkage facilities or centres. The Commonwealth could enable more complete and responsive access to Medicare and Pharmaceutical Benefit Scheme (MBS and PBS) data by supporting NSW Health to become a 'Prescribed Authority' under the *Health Insurance Act 1973* and *National Health Act 1953*.

11. How can Australia's government agencies improve their sharing and linking of public sector data?

As discussed at questions 5 and 6, proper and effective implementation of national standards, guidance and legislation would facilitate data-sharing between agencies and jurisdictions. Put simply, organisations that have confidence in their information security measures and those of their partners are more likely to share information and derive the benefits that this can bring. To support this, there is also a need for a shift in the roles and responsibilities for data custodianship and ownership that moves away from public sector data being guarded by individual agencies.

Governments should enter into enduring agreements through a common national approach

Establishing agreements that allow data to be shared and linked on an ongoing rather than project-specific basis would not only enable but would accelerate sharing and linking of public sector data. These should be supported by agreed frameworks and standards for privacy protection, social impact assessment and approval processes to help more agile data analytics within government.

Governments should support development of technologies that reduce linkage costs, increase data handling security and minimise risks to privacy

Several existing technical and institutional barriers could be overcome by increasing the development and adoption of more advanced confidentialisation and anonymisation techniques, such as confidential computing and synthetic data technology being developed by Data61^{40 41}. Sharing and linking of data could also be improved by providing a centralised platform or warehouse for uploading data and online tools for analysing and presenting this compiled data.

Online technology already exists that can provide secure, remote access to analyse data. An example is the Secure Unified Research Environment⁴² (SURE), developed and operated by the Sax Institute with support from Commonwealth and NSW agencies. Cloud adoption can reduce the cost of data storage and handling and offer significantly improved data availability and accessibility.

⁴⁰ <https://n1analytics.com/#technology>

⁴¹ <https://www.nicta.com.au/project/analytix-provable-privacy-guarantees/>

⁴² <https://www.saxinstitute.org.au/our-work/sure/>

Sharing data between states and the Commonwealth represents the greatest opportunity

Improving data-sharing and linkage between Commonwealth agencies and state/territory agencies is one of the greatest challenges and opportunities to generate public value from the better use and access of data. Our federal system creates organisational and political boundaries that become barriers for data-sharing, and using shared data to jointly plan and coordinate services that better meet needs in the community.

A consistent finding of studies on the economic cost of various health and social problems, such as homelessness and obesity, is that the costs impact both state and federal governments. Sharing of data on these groups would be beneficial both for NSW and for the Commonwealth's Investment Approach⁴³ as it would strengthen the evidence base for future interventions. The combination of inter-jurisdictional data could discern common cohorts and further strengthen the underlying investment approach model for both jurisdictions.

Examples of effective data-sharing in NSW could be used as models for others

Established in 2006, CHeReL is a dedicated data linkage unit managed by NSW Health. It supports longitudinal research, planning and evaluation in the health and human services sector, through linkage of multiple identified datasets while protecting the individual privacy. Under the NSW ICT Strategy⁴⁴, CHeReL's core record linkage system is continually expanding and currently contains pointers to over 117 million records relating to more than 11 million people. CHeReL has provided data linkage services to over 1,400 investigators from the government, research and non-profit sectors for hundreds of projects.

Other NSW agency initiatives further demonstrate the benefits of shared, linked information.

- *NSW Fair Trading's Project Sentinel*: NSW Fair Trading currently has approximately 175 information assets including databases, applications, and high-level groups of records. Project Sentinel is a program to increase the effectiveness of intelligence and analytics functions within the NSW Government.
- *Environmental Data Portal (EDP)*: the NSW Government is developing a whole-of-government environmental data portal, which will make a broad range of environmental data available through a single online source.
- *Biosecurity Business Information System (BIS)*: combines innovative mobile data capture, business intelligence and geospatial capability into a cloud-based 'toolset' for biosecurity managers.
- *Crown Land Management Review System*: uses spatially enabled capability within a structured decision tree to develop a collaborative tool to allocate land on a state significant or local basis. Data is shared with Local Government and Aboriginal Land Councils to help decision-making around the allocation of Crown land.

Enhanced disclosure log requirements could encourage proactive release of information

There is an opportunity to link disclosure logs to the provision of data to the community and provide a one-stop-shop for access to all types of information held by the agency. Jurisdictions' freedom of information and information access regimes commonly include some form of disclosure log requirement. Such logs record the information that government agencies have released in response to a formal application and which may be of interest to other members of the public.

⁴³ http://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/BudgetReview201516/Welfare

⁴⁴ <https://www.finance.nsw.gov.au/ict/>

These are integral to accessing information, supporting government accountability and fostering engagement with the public.

The logs are an efficient measure of ensuring 'self-service' by citizens and avoid the need for more resource intensive mechanisms, such as access applications. Describing and making easily available data sets via disclosure logs would also alert a wider range of stakeholders to the data's existence and availability.

12. What lessons or examples from overseas should be considered?

Overseas models suggest that public confidence in linking and sharing data can be maintained where there is a purposeful, specific and clearly communicated public benefit

The United Kingdom's (UK) approach to promotion of open data by the Information Commissioner and an integrated model offers the opportunity to implement a proactive approach to information release together with promotion of responsible stewardship. The UK Information Commissioner's Office (ICO) is an independent authority established to uphold information rights in the public interest, promote openness by public bodies and data privacy for individuals.

The UK National Health Service (NHS) has been examining how it can harness the power of data and technology to transform health and care services and deliver greater quality and efficiency⁴⁵. In England citizens can access their own linked General Practitioner records online and other digital services in primary care. A key enabler of this initiative was enacting privacy legislation to clearly determine that the owner of a patient's data is the patient⁴⁶.

The New Zealand (NZ) experience with data analytics and the investment approach is an ideal case study to consider in addressing the questions in the Data Linkage section of the Issues Paper. The NZ public sector has used an actuarial approach to measure and inform how interventions are better managing lifetime welfare costs. This has resulted in benefits both for the clients involved and for taxpayers through reducing long-term liabilities⁴⁷.

The data analytics approach has also been extended through the NZ Data Futures Forum⁴⁸ to how NZ businesses, government, researchers, and the public can safely share and use. Statistics NZ has developed an Integrated Data Infrastructure (IDI)⁴⁹ to answer key research, policy, and evaluation questions. The IDI links survey data with administrative data from various government departments to provide an information set that can be used to find new answers and insights to policy challenges.

A further international example for spatial data is the more proactive adoption of World Wide Web Consortium (W3C)⁵⁰ and Open Geospatial Consortium (OGC) standards in live mediums such as web map services (WMS)⁵¹ and web feature services (WFS)⁵². Examining these initiatives will provide insight for greater data interaction within public, private and community-developed interfaces.

⁴⁵ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/384650/NIB_Report.pdf

⁴⁶ <http://techscience.org/a/2015081103/>

⁴⁷ <https://www.msd.govt.nz/about-msd-and-our-work/newsroom/media-releases/2015/reforms-succeed.html>

⁴⁸ <https://www.nzdatafutures.org.nz/why-data-futures-forum>

⁴⁹ http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/researchers-using-idi.aspx#measuring

⁵⁰ <https://www.w3.org/Consortium/>

⁵¹ <http://www.opengeospatial.org/standards/wms>

⁵² <http://www.opengeospatial.org/standards/wfs>

Private sector data

High value private sector data (Issues Paper p.18)

13. What private sector datasets should be considered high-value data to: public policy; researchers and academics; other private sector entities; or the broader community?
14. In each case cited, what characteristics define such datasets?

Private sector datasets that impact or reflect on how people make decisions that affect their social and economic wellbeing are likely to be valuable

It is important to note that it is difficult to assess the value of a data set in the abstract. It may only be once you have used the data to try to answer a research question that it becomes apparent that a data set is valuable or not. For example, social media mentions may have real-time value for police and emergency services and for mental health programs⁵³.

There may also be longer-term private sector data that could be of high-value to public policy-makers in building a more complete picture of client life situation and needs for specific cohorts. In this way, there would be significant value in datasets that could act to complement public sector data and fill in important gaps in understanding how people make decisions that impact social and economic outcomes.

For the NSW Department of Family and Community Services, private sector data may support the development of interventions that address latent demand before it becomes expressed demand. The Aboriginal Housing Office would gain value from Real Estate Institute of NSW and related data to assist in the determination of need and supply of housing for the Aboriginal community.

Some other examples include imagery and location-based information to assist emergency preparedness and response, and financial information for fraud detection and prevention. Primary industries can gain value from private sector data related to precision agriculture, yield monitoring and production. However, this data often uses proprietary systems that cannot be integrated with other data and decision-making solutions. This limits the ability to leverage the power of linked data to make better decisions. System interoperability is key to maximising the value of this data.

Private sector research, monitoring or survey data may have value

Private sector organisations may have greater resources to devote to research and analysis, or focus on measuring or gathering granular data in specific topics, locations or communities. An example might be data from mining or extractive industries which could be significant for environmental assessment.

Global Positioning System data that enables mapping of the movement of particular populations, products or vehicles may have value in understanding peaks, bottlenecks and planning better infrastructure or services. Additionally, any data collected or based on research which was commissioned with public funding should be made publicly available.

Outsourced delivery of government functions means private sector data will be increasingly valuable

Outsourced arrangements for the delivery of government infrastructure, services and functions mean that private sector data will be increasingly valuable in understanding the impact of government policies, programs and to have a good gauge on societal trends. When contracting

⁵³ <http://digitaldog.org.au/programs/classifying-concerning-tweets/>

with the private sector, agencies should recognise what data is valuable prior to entering into an agreement. Sometimes, ownership of intellectual property rights in data is not mentioned in procurement or ICT contracts and agencies may lose the ability to release the information as open data. Datasets from non-government providers in human services such as health, education and housing that complement public sector data could add significant value to government service planning and resourcing.

In the health sector, private hospitals are an important component of national healthcare arrangements. Private hospital data is reported on a voluntary basis via the Private Sector National Hospital Cost Data Collection⁵⁴ cost report, and includes cost weights at a national level. By comparison, public sector hospital data is reported at a much more granular level and across nearly all public hospitals on a mandatory basis.

Notwithstanding provisions around commercial confidence for private hospital operators, there is scope for more detailed and routine public reporting or accessibility to data on private hospital services, for example, information on volumes and average costs of services. This level of information will be important in determining long-term, sustainable hospital funding.

Access to private sector data (Issues Paper p.18)

15. Are there any legislative or other impediments that may be unnecessarily restricting the availability and use of private sector data? Should these impediments be reduced or removed?

Costs for accessing private sector data can restrict realisation of the value it holds

The main impediments to availability and use of data held by the private sector are access costs relating to private sector entities charging for the use of data and intellectual property. Also, commercial sensitivity and competitive advantage is often held up as a basis for not sharing data.

For example, historically, where local governments have commissioned flood studies the intellectual property in those studies has been owned by either the engineering consultant or the local government (or both) and the State has been unable to collate those studies to give the community a complete view of their exposure to flooding or provide them to third parties (such as insurance companies who want to use that information to assess flood risk).

Where private organisations provide access to data, it can be through different types of purchasing arrangements, or via access licensing. A range of different licences may be used by private sector organisations, such as group licences, organisation or individual licences, or licences transferable between individuals or organisations. The type of access cost can have a different effect on whether data is available or not.

Balancing public and private interests can be difficult

Commercial sensitivity and competitive advantage may also contribute to restrictions on access or sharing data. Public and private interests in data sharing may not be aligned nor clear cut. This means it is difficult to say definitively whether restrictions are unnecessary or perceived impediments should be removed.

Different privacy regimes can pose challenges for sharing and access

From a legislative perspective the operation of privacy laws can, in some situations, create an impediment to government and access and use of private sector data. The *Privacy Act 1988 (Cth)* (the Privacy Act) applies to private sector and not-for-profit organisations with an annual turnover of more than \$3 million, as well as private health service providers and some small businesses,

⁵⁴ <https://www.ihpa.gov.au/what-we-do/nhcdc>

while the *Privacy and Personal Information Protection Act 1998 (NSW)* (PIIP Act) applies to NSW government agencies.

Under the PIIP Act, NSW government agencies are generally not permitted to use or disclose personal information without the person's consent where that use or disclosure is not directly related to the purpose for which the information was collected, subject to certain conditions and exemptions. One exemption contained in section 27 of the PIIP Act is where the use or disclosure of the personal information is reasonably necessary for the purpose of research, provided that particular conditions are met.

Similar to the PIIP Act, the Privacy Act generally prevents private sector organisation from using or disclosing personal information for a purpose other than the purpose for which it was collected without consent, subject to some exemptions. However, the Privacy Act does not include an exemption in equivalent terms to section 27 of the PIIP Act. The lack of such an exemption may limit private sector organisations' capacity to share information with government agencies.

In the NSW education sector, independent and Catholic school data is considered private and therefore unable to be linked (without permission) for use in education policy development, teaching/learning research and program evaluation. In order to ensure consistency in the evaluation of the education system as a whole, data between different sectors would need to be linked. The inconsistency in privacy legislation whereby independent and Catholic school sectors are not subject to the NSW legislation represents an impediment to the availability and use of data.

16. What principles, protocols or legislative requirements could manage the concerns of private sector data owners about increasing the availability of their data?

Protocols may need to be applied to manage privacy risks

Private sector data owners will need assurance that commercial sensitivity is protected to an appropriate degree, balancing shareholder interests with broader public interests. Agreement that a private organisation's data not be made available for use by other private organisation may help to manage concerns.

Another mechanism would be to obtain consumer consent to share information collected about them by the private sector with public or other private third parties. This would be particularly useful in situations where government commissions services from non-government providers, and for the purposes of cross-agency service planning.

Legislatively, it is worth examining the value to be gained from harmonising the different privacy regimes and seeking international leading practice guidance. These initiatives would also benefit from clear direction from Information and Privacy Commissioners on ways to manage data.

17. What role can governments usefully play in promoting the wider availability of private datasets that have the potential to deliver substantial spill-over benefits?

Governments should play a role in facilitating better connectedness and interoperability between public and private datasets

In recognition of consumers using a mix of public and private services, governments must play a role in facilitating better connectedness and interoperability between public and private datasets. Having only the public picture, or only the private picture, will not suffice. This is particularly relevant in areas of shared public and private activity, such as health, education, and transport.

Opportunities exist for government to encourage private data holders to make data publicly available through promotion of partnerships with government agencies, or incentive-based payments for providers.

Outsourced arrangements for the delivery of government infrastructure, activities and functions, mean that private sector data will be increasingly valuable in understanding the impact of government policies, programs and to have a good gauge on societal trends. The DAC has already embarked on a number of predictive analytics projects addressing difficult policy issues. These projects are relying heavily on existing government datasets. As these projects progress, the value of certain types of private sector datasets may become more apparent.

Governments can also take a lead by publishing private sector data submitted for regulatory purposes and making it discoverable through publicly searchable portals. Governments could also make open data portals and similar platforms available for private sector data publishing.

18. Who should have the ownership rights to data that is generated by individuals but collected by businesses? For which data does unclear ownership inhibit its availability and use?

Clearer, well-understood definitions of data ownership and custodianship are needed

Ownership and access conditions may be established by business through service agreements, but should align with best practice information principles. Individuals must be given a choice about whether to provide data to that business, and understand for what purpose it will be used.

Ownership and intellectual property rights can be unclear in data or data-related products created by a private organisation procured by agencies. Agencies are encouraged to include data ownership and intellectual property provisions in procurement contracts but there is no standard approach. The NSW Government has adopted the AusGOAL approach to Creative Commons licensing⁵⁵ for open data. There is a variety of licenses that could be used to enable the release of data for a range of reuse scenarios depending on ownership and public or private interests.

⁵⁵ <http://www.ausgoal.gov.au/creative-commons>

Consumers' access to and control over data about them

Consumer access to, and control over data (Issues Paper p.21)

19. What impediments currently restrict consumers' access to and use of public and private sector data about themselves? Is there scope to streamline individuals' access to such data and, if there is, how should this be achieved?

The NSW Government is working to promote public rights to open data

The NSW Minister for Innovation and Better Regulation is taking a number of actions to promote consumer rights. In April 2016 the Minister announced the NSW Information Commissioner's new role as Open Data Advocate. Further, the 2016 NSW Open Data Action Plan identifies the development of a Charter of Rights for Open Data Users. The Charter is the consumer companion to the responsibilities of government agencies set out under each principle of the policy.

The Open Data Policy Action Plan also outlines the implementation of a 'dMarketplace' initiative. The 'dMarketplace' is a single data discovery platform and interface that will be available for private sector and research sector data, similar to the functionality of iStore or TripAdvisor. This will be developed and integrated with Data NSW.

The 'dMarketplace' will help companies, individuals and governments discover data that has been published. Users will be able to search for data products similar to the way a user can search for apps in the iStore, by interest and relevance rather than by agency area or by data type.

There is an opportunity for greater public sector capability to support rights of access

Individuals already have a right under the PPIP Act to access their personal information held by government agencies and to request amendments or alterations to that information to ensure that it is accurate. There is a need to develop agency capability to design privacy management frameworks and consent mechanisms to better support consumer controls. This is referred to as 'privacy by design'⁵⁶. Privacy by design is an internationally used, leading practice to protecting privacy by embedding it into the design specifications of technologies, business practices, and physical infrastructures.

Technology is rapidly advancing in this space. World-wide web creator Tim Berners-Lee is involved in developing a platform for managing privacy preferences based on existing web architecture^{57 58}. The project's goals⁵⁹ are to:

- Enable personal choice and informed consent
- Establish commitment from publisher about use of data
- Provide technology to negotiate these two
- Provide hooks to "enforcement" mechanisms which promote confidence

Cloud Insurance⁶⁰ is an Australian company working to bring simple ways to enable individuals to control, audit and effectively self-manage privacy and data permission rights, particularly

⁵⁶ <https://www.ipc.on.ca/english/privacy/introduction-to-pbd/>

⁵⁷ http://www.csail.mit.edu/solid_mastercard_gift

⁵⁸ <https://www.w3.org/Privacy/>

⁵⁹ <http://crosscloud.org/>

⁶⁰ <http://www.cloudinsurance.com.au/>

mechanisms to manage opt-in or opt-out scenarios. These kinds of developments could significantly change the existing balance of power in relation to the collection and use of personal data, irrespective of any currently existing barriers.

20. What datasets, including datasets of aggregated data on consumer outcomes at the product or provider level, would provide high value to consumers in helping them make informed decisions? What criteria should be used to identify such datasets? What, if any, barriers are impeding consumers' access to, and use of, such data?

Datasets that address information asymmetry in consumer markets would provide value

Datasets that address information asymmetry in consumer markets would provide significant value to consumers in helping them make informed decisions. This has been partly achieved by the advent of price comparison websites allowing consumers to compare offerings for services such as phone and internet or insurance and investment. However, these do not offer information on consumer outcomes, perhaps because it may not be commercially advantageous to do so.

Other examples include the introduction of the National Disability Insurance Scheme (NDIS) and the creation of a client-directed market for disability services. As the NDIS rolls out and moves from block funding to individualised funding consumers will seek and generate information about the quality of supports they receive from service providers. This would inform consumer decision-making about how they spend their NDIS funds (see, for example, Hireup⁶¹ and Clickability⁶²).

From a service delivery and commissioning perspective, opening up datasets provides a tremendous opportunity for government agencies to change the way they engage with the private sector, research and academic organisations. Customer satisfaction data is an important data source for the NSW Government and its collection aligns with the Premier's stated priorities.

⁶¹ <https://www.hireup.com.au/>

⁶² <https://clickability.com.au/about/>

Resource costs of making data available

Resource costs of access (Issues Paper p.22)

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

NSW's Open Data Policy requires data to be made available for free where appropriate

Charging may be appropriate where there is a commercial benefit from the use of the information, or where there are costs involved in making information available beyond the level of information than can be justified on public benefit grounds⁶³. Under the NSW Open Data Policy access to open data is free where appropriate. Cost recovery can be undertaken for specialised data services and fees for these services must be published.

Greater use of technologies such as APIs and web services, once established, can minimise the costs and administration associated with both privacy and ongoing release. Under the GIPA Act agencies may impose an application fee and a processing charge when providing government information in response to an 'access application'.

Agency investment and procurement strategies should include costs associated with making public sector data available

Funding of public sector data availability is an extremely complex issue as it depends on numerous variables such as type and quantity of data, as well as privacy requirements, degree of public interest and what the data is to be used for. Upfront investment in technology that makes data sharing easier and less arduous could reduce the long-term resource costs of making more public sector data available for use.

Agencies must build into investment and procurement strategies the recurrent costs associated with making public sector data reasonably available. As a result, a level of additional resourcing is likely necessary to make fundamental and meaningful changes to existing data availability and usage practices. Before undertaking amendments to existing datasets the benefits of making more data available to the public (and the specific purpose for doing so) needs to be clear to ensure investments are warranted.

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

Disproportionate effort is put into collection, cleaning and basic reporting of data compared with analysis and use

Data, including its linkage and access, is just one part of the story. Well curated and 'narrated' data, that is presenting and explaining the data in a way the public can understand, provides immense public value. Currently, much effort is directed at collection, cleaning and basic reporting of data compared with analysis and use. This is exacerbated when data is prepared on a one-off basis that requires it be destroyed after use.

For example, BOCSAR has an open data policy and as a consequence is required to spend a considerable amount of resources on data management and access. Staff are devoted to in-house data analysis and separate staff are devoted to receiving data requests from other agencies and disseminating data, and other staff compile output. The resources involved are considerable.

⁶³ A fuller discussion of Pricing Principles for Provision of Government Information can be found at: http://www.applieconomics.com.au/pubs/papers/pa01_services.htm

On the other hand, in the health sector, the NSW Government is committed to efficient and effective data management and access. The Business Information Program (first investment in the Health Information Exchange and now upgrading to the Enterprise Data Warehouse) is an example of a continued investment in data management and access within the health system.

The streamlined portal of HealthStats NSW is a demonstration of this commitment, as is the availability of data through the Bureau of Health Information. NSW Health has also recently released the NSW Health Analytics Framework⁶⁴, focussing on transforming health through data and insights, in alignment with the eHealth Strategy for NSW Health setting out a 10-year program of innovation, investment and implementation.

23. What pricing principles should be applied to different datasets? What role should price signals play in the provision of public sector data?

Using market supply and demand pricing dynamics could be considered for data provision

Moving beyond cost recovery approaches, governments could consider applying market supply and demand pricing dynamics to the provision of public sector data as in any market for a good or service while ensuring that the right to access one's personal information remains free. Examples already exist in real-time public transport data. However, pricing decisions should ensure that data is open and free where appropriate.

Other examples of current practice include BOCSAR, which charges data recipients to compile data if it spends over two hours on the request. This is important for significant requests as it both deters flippant requests and helps focus the recipient to consider what they really seek. Clients not from the NSW Government are charged \$220 an hour for programming time. Most requests are completed in less than two hours or are for government. Only 5-10% of clients are charged and this is an average of \$660.

Another option is to licence data with free use for non-commercial purposes, or data used to enable a business process (such as a licence check or identity check) but to charge for use in commercial services that involve resale or value added data.

Net benefits could be established that support the agencies allocating resources to making data public

Rather than adopt a purely financial pricing basis to datasets a broader net benefit approach could be taken. For example, most projects using human services data would have a public benefit (e.g. improved education and health, leading to greater employment opportunities). Similarly, the benefits gained through datasets that enable more effective services flow to governments in the form of future avoided costs and efficiency dividends.

A further example is the Geological Survey of NSW (GSNSW) which is being partially funded through the New Frontier Exploration Initiative, which sources funding through mineral and petroleum title rental fees. This funding is used to resource projects to help stimulate mineral investment in frontier areas that are under-explored across the state.

This model of direct funding from government revenue obtained through administrative processes could be applied more broadly (essentially an indirect "user pays" system). GSNSW does not apply a pricing structure to its data and typically only administrative production costs are covered. Industry demand and the ability to support meaningful exploration opportunities drives development, which would mean more funding is required in these areas.

⁶⁴ http://www.ehealth.nsw.gov.au/publications/nsw_health_analytics_framework

24. Is availability of skilled labour an issue in areas such as data science or other data-specific occupations?

The ability to use data effectively relies on the combined skills of statisticians, computer scientists and content experts to curate, extract and interpret insights

The ability to use data effectively relies on skills not commonly held by single individuals to curate, extract, interpret and act on insights from the masses of data available. There is a perception that the public sector will find it increasingly difficult to attract high-level data scientists due to strong competitive demand for these skills from other employers such as the private IT industry, multinational technology firms and universities. This is a critical risk for success in this area, as the public sector is more likely to use short-term data science skills on a contract basis, posing challenges for knowledge management and business continuity with staff turnover.

To improve data skills the NSW Government is introducing a program to engage final year university students in developing responses to real data issues and a pilot placement project in relevant schools to develop a recruitment pipeline. National effort to build science, technology, engineering and mathematics capabilities provides a further channel for capability in applying data.

Some agency challenges are domain specific. For example, it may be challenging to find skilled labour in data science and other data-specific occupations that can appropriately apply their skillset to the health care industry. Acquiring suitable people with the required skills and knowledge to comprehend the process and deliver geoscientific data is a known issue, particularly for organisations such as GSNSW who are based in regional areas.

In education, building skills in data literacy increases the effectiveness of teachers and school leaders for decision-making. Effective principals have been found to systematically use assessment data to constantly plan, coordinate and evaluate teaching and the curriculum. At best, there is forgone opportunity if data is not used strategically to advance educational outcomes. At worst, misinterpretation and misapplication of evidence could lead to program design or funding decisions with poor outcomes.

Leading data science organisations in Australia could be leveraged for public benefit

Alongside increasing demand, there are also organisations like the 1,100 data-scientist strong Data61, as well as private data science firms such as Data Republic⁶⁵ and Quantum⁶⁶. Data61 in particular has an objective to partner with governments for a data-driven future and has important experience helping to address public policy challenges through data.

Whether it is availability of staff within an organisation with the skills to achieve desired results, or costs associated with preparing data for public release, issues will arise in the expertise available for performing specific data tasks. This is particularly the case in specialist areas such as de-identification and anonymisation of data, and in machine learning and predictive modelling techniques.

⁶⁵ <https://www.dataarepublic.com.au/>

⁶⁶ <https://www.quantium.com/>

Trust and privacy

Privacy protection (Issues Paper p.26)

25. What types of data and data applications (public sector and private sector) pose the greatest concerns for privacy protection?

The risk that an individual's identity can be determined through data sharing and use is one of the greatest concerns surrounding privacy

The most valuable datasets often contain sensitive information about individuals. While there are numerous ways to de-identify or anonymise the information contained in single datasets, when linked with other datasets there is an increased risk of (inadvertently) identifying individuals.

While data remains within government and private sector organisations it is subject to governance mechanisms which generally minimise the risk of re-identification and improper use to acceptable levels. The NSW Government Digital Information Security Policy's risk management approach sets out minimum requirements to ensure all information is appropriately protected and handled.

When data is opened up publicly governance and proper use are difficult to enforce. It is also difficult to know what other datasets are in the public domain, what information these contain and how they may be applied. It is important to note that while control over data is reduced when it is made open, there are a range of techniques that can assist in managing the risk of re-identification.

The NSW Government recognises that there are some datasets that should remain confidential. For example, under the GIPA Act information is not made available to the public if there is an overriding public interest against disclosure. For certain classes of information, such as cabinet information, there is a 'conclusive presumption of overriding public interest against disclosure' and the information will not be released. Where there is no such conclusive presumption, information will not be released if the public interest considerations against disclosure outweigh the public interest considerations in favour of disclosure.

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

Enhanced data and cyber security would increase public confidence in how we use data

A better understanding and use of de-identification and anonymisation techniques, combined with getting the 'benefit and risk' balance right, is critical for community confidence at all levels. As discussed in our response to question 11 above, there are a range of techniques used to de-identify or anonymise data and to reduce or remove the risk of re-identification.

Confidence can also be increased by public and private sector organisations being more transparent with the types of data that is used and the purposes for which it is used. This would enable individuals and businesses to better consider the costs and benefits arising from different use cases. Building on this, individuals and businesses need to be confident that should there be a breach in the agreed usage protocols that enforcement will be taken.

In its response to the Commonwealth Government's cyber security review, the NSW Government highlighted the need for improved sharing of information about security incidents with the states, and for support to the business community to improve cyber security capabilities. The recent establishment of a node of the industry-led Cyber Security Growth Centre in Sydney is an opportunity to develop new approaches and support capability development.

27. What weight should be given to privacy protection relative to the benefits of greater data availability and use, particularly given the rate of change in the capabilities of technology?
28. How effective are existing approaches to confidentialisation and data security in facilitating data sharing while protecting privacy?

More extensive public consultation and communication may be required

There is a need for greater public consultation, public dialogue, education and research into public opinion on data privacy. This is the best way to understand the main concerns of businesses and individuals and how much weight should be given to privacy protection, and to build trust.

Additionally, greater openness and discussion around the advantages and disadvantages of data-sharing and data analytics, and how privacy protections work will help people make more informed choices about their privacy and help government ascertain public interest.

Advances in technology and data analytics should be leveraged to protect data

It is critical that public confidence of governments' collection and use of data is maintained. Privacy protection will continue to grow in importance given the move towards digitisation of services and products. Currently, this must be balanced against the great but often unquantifiable benefits to society of increased availability and use of data. The question remains one of balancing cost and benefit, which includes both financial and social aspects.

Achieving this balance should become easier as technology increasingly enables greater levels of data anonymisation. The public and private sectors should increase support for the development and use of technology and sophisticated data analytics techniques such as those of Data61⁶⁷ and the Australian Census Longitudinal Dataset⁶⁸.

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

Harmonising legislation and clarifying rules would enable data-sharing and protect privacy

The impact of privacy legislation on the disclosure of information has been discussed throughout this Submission. The complexity that results from the intersection of Commonwealth and states' privacy legislation poses a challenge in terms of promoting greater sharing of government data. This is because the complexity of the intersecting legislative regimes may promote a risk-averse culture within government agencies that discourages disclosure of government information. This is particularly the case in instances where adherence to both Commonwealth and NSW Privacy legislation is required, which is discussed above in our response to question 15.

The resolution of complex regulatory issues creates benefit as data-sharing projects won't be delayed due to uncertainty for government agencies and concern about compliance risk. Likewise an overly cautious approach to managing all-inclusive legislative requirements has resulted in a plethora of specific exemptions which collectively contribute to an arcane framework of soft law solutions.

For example, under the PPIP Act there is an exemption that allows government agencies to share personal information for research purposes in some circumstances. However, different, more restrictive circumstances apply in the equivalent exemption contained in the Privacy Act. Similarly, the NSW legislation requires there to be a 'serious and imminent' threat to safety for service delivery purposes. At the federal level the requirement has been reduced to serious.

⁶⁷ <http://www.csiro.au/en/Research/D61/Areas/Big-data-and-analytics/Privacy-Confidential-data>

⁶⁸ <http://www.abs.gov.au/websitedbs/censushome.nsf/home/aclid>

An agreement on a set of common Privacy Principles and the development and adoption of Privacy Codes of Practice for use that are well understood will help encourage government agencies to open up and share data. Efforts have been made at both federal and state levels, however, adoption of a more proactively open data policy which asks “if not, why not?” framed within the scope of existing legislation would go a long way to providing the necessary clarity.

30. What lessons from overseas jurisdictions can Australia learn from regarding the use of individuals’ and businesses’ data, particularly in regard to protecting privacy and commercially sensitive or commercially valuable information?

Australia has an opportunity to adopt the good work done by organisations internationally

Organisations such as the Open Data Institute UK⁶⁹, the Sunshine Foundation USA⁷⁰, and Privacy by Design Canada⁷¹ provide principles for proactive disclosure which also seek to maximise individual privacy. The Flemish system for online information collection^{72 73} should also be considered for applicability to the Australian context. This system is said to have safeguards and rules for access to protect privacy that comply with European Union directives on data protection.

⁶⁹ <https://theodi.org/about>

⁷⁰ For more information including “proactive disclosure” principles see <http://sunlightfoundation.com/opendataguidelines/>

⁷¹ <https://www.privacybydesign.ca/>

⁷² <http://governance-flanders.be/e-government>

⁷³ <http://www8.hp.com/h20195/v2/GetPDF.aspx/4AA5-6171ENW.pdf>

Data security and Restrictions around the release of particular data

Other restrictions (Issues Paper p.27)

31. Is there need for a more uniform treatment of commercial-in-confidence data held by the Australian Government and state and territory governments?

More uniform treatment of commercial-in-confidence data would increase confidence in dealing with the government

Commercial-in-confidence classifies information that, if disclosed, may result in damage to a party's commercial interests or intellectual property. The NSW Government enters into a wide range of contracts with external organisations who want their information protected.

As part of greater uniformity, it will be important to be clear about what types of data or uses of data, including where confidentialised, should be permitted by government agencies. This is particularly the case for data that is highly valuable for government service delivery. There would also be value in national consistency as common data is kept across jurisdictional boundaries.

As discussed above, clear guidelines outlining protocols for the de-identification of data for use in analytics systems should be developed. This will ensure agencies have clear advice regarding methods to securely disclose data while ensuring privacy is protected.

Data security (Issues Paper p.28)

32. Are security measures for public sector data too prescriptive? Do they need to be more flexible to adapt to changing circumstances and technologies?

Citizens must be able to trust that their data is secure in digital services

Government services are increasingly delivered through digital channels. Customers need to be confident that the digital services they use are stable, secure and resilient to cyber-attacks. The NSW Government Digital Information Security Policy takes a risk management approach and stipulates a set of minimum requirements to ensure that all information is appropriately protected and handled by government agencies.

To assist in the detection and management of security incidents, the NSW Government has adopted an Information Security Event Reporting Protocol. This protocol establishes common practice to assist the reporting of information security events in the NSW Government. The protocol uses the expertise and experience within the Government's Information Security Community of Practice.

Security needs to be flexible to facilitate increased data use and changes in technology

Current security measures can be prescriptive but have been generally appropriate for the culture and environment that has existed until recently. As public sentiment shifts to a greater acceptance of digital service delivery and provision of personal information in general, data security measures should become more principle-based than prescriptive to maintain flexibility and create further opportunities for data-sharing, linking and analysis.

The dynamic nature of technology will require ongoing changes to data management processes and practices, governance arrangements and policies. However, strong security measures are essential for maintaining the privacy and confidentiality of the data and the trust of those that provide and use the information.

33. How should the risks and consequences of public sector and private sector data breaches be assessed and managed? Is data breach notification an appropriate and sufficient response?

Collaboration and awareness will be increasingly important for managing breach risks

Data breach notification may be an appropriate response for some breaches but may not be sufficient to prevent or address the most severe consequences of data and privacy breaches. In order for a data breach notification and any subsequent investigation to be an appropriate and sufficient response to data breaches, there should also be suitable consequences flowing from the notification and investigation including penalties for responsible parties and remedies for individuals who have suffered damage.

Other considerations may also be relevant for considering what is appropriate and sufficient, such as the extent of data and systems compromise, individuals/bodies affected, secondary sources and other datasets compromised and implications for ongoing sharing arrangements. It may also be noted that a deterrent structure built around fines may have limited impact if the compromise is perpetrated extra-nationally.

NSW notes developments at the Commonwealth level, including the introduction of the *Health Legislation Amendment (eHealth) Act 2015 (Cth)* and the release of an exposure draft of the *Privacy Amendment (Notification of Serious Data Breaches) Bill 2015*.

Appendix 1 – sample high-value datasets

Some examples of datasets that would be of high value if available for linkage are as follows.

- Socio-economic Indexes for Areas (SEIFA) scores at ‘meshblock’ level, held by the Commonwealth Government – access would enable the inclusion of socio-economic deprivation in, for example, health risk stratification models and service responses.
- Public and private hospital data linked with data on Medical and Pharmaceutical Benefits Schemes (MBS and PBS) claims, health insurance claims, disability, mental health, residential and community aged care, and the health workforce would enable analysis and policy development to improve health outcomes and patient experience as well as reduce duplication of services.
- ABS Census, Centrelink and Health data, if linked with Coroners’ and emergency incident data would allow for assessments of true cost of emergency incidents and provide a greater understanding of community risks and hazards for particular suburbs and communities and allow for better targeting of prevention, preparedness and response effort.
- ABS Essential Statistical Assets (ESA) for Australia identifies data which are critical to decision-making in a complex and sometimes fragmented information environment across Australia. The ESA aims to support more efficient use of government resources by identifying areas of duplication and underutilisation. Agencies can ensure that the critically important information (statistics/data) are of sufficient quality and identify critical information gaps, so that these can be addressed.
- Linkage of specific offender information across partner agencies, although not accessible to the community, could be of benefit to community safety and services agencies – relevant data may concern child protection, reoffending, violent offenders, mental health, disability, violent extremism, cybercrime, and disaster management.
- Geo-spatial mapping, emergency planning and preparedness, and police and emergency telecommunications data held by several public and private organisations are critical to law enforcement and emergency response.
- Datasets relevant to weather, crop, livestock, biosecurity and pest information, water and land lease availability directly contribute to the growth of NSW primary industries, as well as the ability to manage compliance and regulation across the areas of biosecurity, fisheries, food, game licensing and water (e.g. NARCLiM – NSW/ACT Regional Climate Modelling⁷⁴).
- Datasets relating to mineral and petroleum resources geological data, including the Geological Survey of NSW’s (GSNSW)⁷⁵ structured open datasets, are used to encourage business sector investment, assist research exercises and support academics studies.
- High-value public sector data of relevance to primary industries, land and water include:
 - Katmandoo⁷⁶ - biosciences data management system. While created for plant breeding and evaluation, the system is applicable for other data derived from designed experimental research within the biosciences.
 - National Variety Trials⁷⁷ - access to independent results on the performance of recently released grain and field crop varieties.

⁷⁴ <http://www.cccr.unsw.edu.au/sites/default/files/NARCLiM/index.html>

⁷⁵ <http://www.resourcesandenergy.nsw.gov.au/miners-and-explorers/geoscience-information/about/geological-survey-of-nsw>

⁷⁶ <http://www.katmandoo.org/Help/Index.htm>

⁷⁷ <https://grdc.com.au/Research-and-Development/National-Variety-Trials>

- RESDAT - integrated management of information on forest research and management for individual studies.
- National Assessment Program – Literacy and Numeracy (NAPLAN)⁷⁸ test data for all Year 3, Year 5, Year 7 and Year 9 students – around 350,000 students annually since 2008.
- Higher School Certificate (HSC) and Record of School Achievement (ROSA) participation and performance data for students in Years 10, 11 and 12 – approximately 240,000 students annually since 1991 (and non-digital data for prior years).
- Accreditation information for all NSW teachers accredited since 2004 and from 2018 for all NSW school and early childhood teachers - approximately 120,000 teachers.
- Registration data for NSW schools and school systems, and for home-schooled students.

An example of high-value data linkage from NSW Family and Community Services (FACS) is the Pathways of Care Longitudinal Study (POCLS)⁷⁹. POCLS aims to provide a strong evidence base to inform policy and practice in the OOHC service system, improve decision-making about how best to support children who have experienced abuse and neglect and thus improve their outcomes. The study provides unique, high-quality information about the wellbeing trajectories of children in OOHC, with immediate relevance to policy and practice decision-making. Linked data including on education, health and outcomes justice will be linked to the survey data and will provide critical evidence on outcomes that will underpin the Outcomes Framework and the OOHC Quality Assurance Framework. FACS is also exploring the idea of establishing a longitudinal administrative linked dataset aimed at being able to address a broad range of research requests and to supply data for reporting against the Human Services Outcomes Framework.

A good example of an initiative that uses big data, modelling and re-analysis to get the most out of datasets is the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS)⁸⁰. The ACEMS brings together a critical mass of Australia's best researchers in mathematics, statistics and machine learning. This unique group works with partner researchers to engage in research programs that combine new methods for the analysis of data with theoretical, methodological and computational foundations provided by advanced mathematical and statistical modelling to develop new insights.

⁷⁸ <http://www.nap.edu.au/naplan/naplan.html>

⁷⁹ <http://www.community.nsw.gov.au/research-centre/pathways-of-care-longitudinal-study>

⁸⁰ <http://acems.org.au/>