New Australian Government Data Sharing and Release Legislation

Department of the Prime Minister and Cabinet

1 August 2018

Executive Summary

The University of Melbourne welcomes the Australian Government's proposed Data Sharing and Release Legislation, and the opportunity to respond to the Issues Paper for Consultation.

The University also commends the Department of the Prime Minister and Cabinet (the Department) on its approach to what is a program of significant reform. The Department's proposal to release an exposure draft is particularly welcome, giving interested stakeholders an opportunity to thoroughly interrogate the scheme to ensure a robust legislative framework that is fit-for-purpose over time.

Australia's data is a major national resource, offering enormous opportunity for citizens, business and government to make better informed decisions and develop innovative products and services. As a leading research-intensive university, our researchers increasingly draw on large data sets to gain insights and discoveries that can drive innovation, efficiency, productivity, social wellbeing and economic growth for the nation. In many ways, future economic growth and social innovation depends on maintaining, and having access to, a robust research base.

While improved data access and use has the potential to greatly increase Australia's level of productivity and wellbeing, new data sharing arrangements must carefully respond to concerns about data security and privacy. The new data access framework created by the proposed Data Sharing and Release Legislation (DS&R) will need to balance efficient, transparent and timely release of data against careful risk analysis and response to community apprehension around the sharing of personal information.

The University, in welcoming the Issues Paper, notes there are important areas which require further clarity or consultation on implementation and impact. The areas of the Issues Paper on which the University makes detailed comment are:

- Data with relevance to the University;
- The Five-Safes framework and the purpose test; and
- The roles and responsibilities within the system: Accredited Data Authorities (ADA), the National Data Commissioner (NDC), the National Data Advisory Council (NDAC), and trusted users.

The University thanks the Department for its sustained engagement and consultation with our academics and experts, including a recent roundtable. We believe it is important that academics and universities continue to be engaged on policy and practice development related to data creation, transformation, access, and analysis.

As researchers, data generators, data custodians and data users, the University supports ongoing efforts to make data available and secure. The University has well-established ethical and procedural arrangements governing the collection, storage, and use of research data. In line with this expertise, the University proposes to develop a Protocol regarding access, use and storage of government-held health data, ensuring use of data across the university is appropriate. The Protocol, to be developed by experts with the support of the Federal Government, would provide a tested governance model expanding each of the Five-Safes and providing relevant granular detail and an accompanying code of practice. The Protocol, once established, could serve as a model for other organisations and may assist and guide the development of a suitable framework for the DS&R Bill.

List of Recommendations

Recommendation: To build community understanding and support, the Government should prepare an information package that demonstrates how anonymised data can be used for public benefit. Public transport data may be a suitable platform to test this. Commuters may be more willing to have their anonymised data used by planners and service providers, for example, if they can see tangible benefits such as how data can inform and shape public transport routes, timetables, or address congestion.

Recommendation: Suitable control mechanisms should be implemented to ensure that for foreign entities only those with sound data practices and strict ethical codes are granted access to Australian datasets.

Recommendation: Streamline national data linkage and access approval processes, including the linkage of State and Commonwealth data linkage units, subject to comprehensive privacy and security protocols.

Recommendation: Establish incentives for public sector data custodians to conduct a formal process of consultation with researchers to inform future design of public data holdings.

Recommendation: The purpose test should be expanded to encompass public interest more broadly.

Recommendation: Develop a detailed Protocol pertaining to access, use and storage of government-held health data. This Protocol would be based on the Five Safes, taking this one stage further in terms of detailed definitions. The Protocol would be able to be used as a model approach for other like-organisations.

Recommendation: Where possible, the framework for making decisions about releasing data should have checks and balances, decision review components, appeal procedures, and be supported by non-government expert evidence.

Recommendation: The DS&R Bill should include mechanisms to protect against the release of commercially sensitive data. Data shared with Commonwealth agencies due to mandatory reporting or regulatory arrangements will need to be appropriately screened and sanitised to ensure that sensitive or otherwise confidential information, is protected.

Recommendation: Accreditation for Accredited Data Authorities and trusted users should be for five-year periods.

Recommendation: All participants in the trusted user model should have clear accountabilities and responsibilities. Trusted users should be clearly identified, should agree to well-defined conditions of use, and be employed by organisations that have signed a legal undertaking that sets out appropriate data safeguards.

Recommendation: Building and maintaining a social licence should sit at the centre of the Government's reform package. Through the office of the National Data Commissioner, efforts should be directed towards: showcasing the system's integrity, transparency and accountability mechanisms; and helping people to better understand the potential community-wide benefits of data availability and use.

Recommendation: To support the objective of transparency, the NDC should publish and maintain up-to-date performance reports. The report could include key metrics such as: the leading data sharing custodian; the number of requests received; the number of requests granted; time taken to complete data sharing requests; and information about why certain requests are refused.

Recommendation: The DS&R Bill should introduce a process for dispute resolution. Like the functions carried out by an ombudsman, applicants denied access to data and/or individuals who believe they have been the subject of inappropriate or unauthorised data release should have access to suitable avenues of redress. The design of the review will be important and should ensure a process that is openly accessible, cost-effective, and expedient.

Recommendation: To perform its function as an independent advisory body, the National Data Advisory Council should include community representation in addition to technical experts to ensure that approaches to data sharing and release are consistent with prevailing public attitudes.

Data with relevance to the University

The University is demonstrably committed to finding ways to support data storage and sharing, including investing in platforms to support collaborative access to research data assets of national significance.

Examples of the University's significant data holdings and services include:

- The **Household, Income and Labour Dynamics in Australia (HILDA)** survey. Since 2001, researchers at The Melbourne Institute have run the HILDA survey, a household-based panel study that collects valuable information about economic and personal well-being, labour market dynamics and family life. HILDA follows the lives of more than 17,000 people each year, collecting information on many aspects of life in Australia, providing policy-makers with unique insights.

- The **Australian Urban Research Infrastructure Network (AURIN)**. AURIN is an example of a complete data release and delivery service for researchers and government at a local, state and federal level. AURIN coordinates the collection of data from a variety of data sources (public and private) and provides access to data through the AURIN portal. Currently, AURIN provides access to almost 4,000 high value datasets from 99 organisations. Funded by the National Collaborative Research Infrastructure Scheme (NCRIS), the resource provides researchers and policy-makers with access to data across diverse fields of knowledge such as, population and demographics data, transport, energy and urban design.

- **Twin Research Australia**: Australia's only national twin research centre of excellence, recognised as a world leader in this field with more than 70,000 twins as members. The impact of research is wide-ranging and widespread having contributed to new insights, prevention and treatment for some of the most challenging health problems of our times.

- The **Melbourne Academic Centre for Health (MACH)**: a joint venture between 18 Victorian healthcare providers, research institutes and the University of Melbourne. MACH is focused on the integration of medical research, education, and clinical care. MACH advocates for a strong data network to integrate the findings of health and medical research into patient care and the training of health professionals. MACH supports national programs that will promote rapid translation of discoveries to policy and practice through connecting primary care health data, improving the quality and safety of health systems, and conducting innovative clinical trials.

- The **Medicine in Australia: Balancing Employment and Life (MABEL)**: a national longitudinal study of Australia's doctors is a trusted national resource producing key evidence to help ensure the sustainability of the sector, as well as delivery of better health outcomes and improved access to medical care. Healthcare is the largest sector of the economy at over 10 per cent of GDP and employing 14 per cent of the Australian workforce. It is the largest employer in Australia. MABEL ensures that policy-makers have access to high value data on medical health professionals including their decisions on how many hours to work, where to work in terms of sector, geographic location and specialty, when to stop work temporarily or permanently, and productivity.

A broad scope of data

Data, and, the linking of datasets, forms critical research infrastructure. Access to population data, for example, could provide insight into the changing structure of human populations: chronic disease, fertility, the effects of intergenerational wealth or poverty, or the incidence in mental illness among specific populations such as migrant communities, youth, or senior Australians.

Gaining access to data generated by government funded services and infrastructure, like free Wi-Fi or Smart City, should also fall within the scope of any new data sharing arrangements. For instance, with public transport in many Australian cities under strain, and significant infrastructure spending slated by federal and

State governments, there is an opportunity to rethink transport planning. One way to approach this would be to better facilitate data sharing between government agencies and private companies, and to ensure that a project's digital skin – its capacity to generate data – is considered at the time of conception rather than as an afterthought. In the same way that commercial operators offering free Wi-Fi services use data to better tailor product offerings and services, government projects should utilise transport data to inform policy decisions about capacity, future routes, passenger demographics and so on.

Recommendation: To build community understanding and support, the Government should prepare an information package that demonstrates how anonymised data can be used for public benefit. Public transport data may be a suitable platform to test this. Commuters may be more willing to have their anonymised data used by planners and service providers, for example, if they can see tangible benefits such as how data can inform and shape public transport routes, timetables, or addresses congestion.

While there is a public benefit in making broad categories of data available to citizens, business and government, the Department will need to consider how to protect this resource. While data is the key to innovation, protecting data from exploitation must be a priority. For example, foreign governments, agencies or entities may gain considerable advantage by obtaining access to Australia's national datasets.

Recommendation: Suitable control mechanisms should be implemented to ensure that only foreign entities with sound data practices and strict ethical codes are granted access to Australian datasets.

Streamlining data sharing and release

Obtaining data from public sources for research purposes is often difficult, lengthy and requires negotiation with different agencies. The Issues Paper states that the DS&R Bill will provide an efficient pathway for the sharing and release of data based on a modern and flexible framework. This objective relies on the National Data Commissioner's capacity to incentivise agencies to share data where a legislative mandate to do so does not exist. This will involve clear and consistent leadership and significant change to institutional practices.

Thought will also need to be given to the capacity of agencies to deliver data. An increased demand for data will need to be appropriately managed and facilitated to avoid data bottlenecks or monopolies. To maximise public benefit, data will also need to be released swiftly. This is particularly important for researchers attempting to forecast market changes or provide an evidence-based response to pressing policy issues.

Recommendation: Streamline national data linkage and access approval processes, including the linkage of State and Commonwealth data linkage units, subject to comprehensive privacy and security protocols.

Recommendation: Establish incentives for public sector data custodians to conduct a formal process of consultation with researchers to inform future design of public data holdings.

The Five-Safes Framework

The proposed framework seeks to manage risk while also improving data availability. The Government's proposed Five-Safes model and purpose test aims to achieve this. However, further clarity around definitions, mode of operation, and the quality of proposed safeguards is required. The pace of technological development also presents a challenge in that we may not sufficiently understand or appreciate the future risks associated with sharing data publicly.

Case study: In August 2016, Australia's federal Department of Health published medical billing records of about 2.9 million Australians online. These records came from the Medicare Benefits Scheme (MBS) and the Pharmaceutical Benefits Scheme (PBS) containing 1 billion lines of historical health data from the records of around 10 per cent of the population. The records were sanitised, or anonymised, a process intended to protect the identity of individuals. The records were then published on the government's open data website in line with its policy on accessible public data.

In September 2016, one month after publication, researchers at the University of Melbourne informed the Department that the encryption of supplier IDs was easily reversed. The dataset was then taken offline.

The example illustrates the tension between protecting personal information and publishing detailed personal records. This is not to say that data cannot be published safely, but that important technical and procedural issues must first be solved.

The Purpose Test

The Issues Paper canvasses that the following purposes would each constitute sufficient grounds for authorising data sharing under the legislation:

1. informing government policy making;
2. supporting the efficient delivery of government services or government operations;
3. assisting the implementation and assessment of government policy; or
4. research and development with clear and direct public benefits.

These proposed purposes suggest that the Government must be a beneficiary of the activities conducted. Such a formulation of the purpose test could inadvertently limit the circumstances under which data may be released. There is a clear public benefit in connecting datasets, however, the optics of a purpose test focused primarily on benefits to government is unlikely to build community support. It will be important for the government to build a social licence by demonstrating to Australians that their data is safe and secure and that citizens are the ultimate beneficiaries of the use of data.

Other issues to consider include:

- The level of influence the National Data Commissioner will have on how the purpose test should be applied;

- The avenue of appeal where requests are found not to meet the purpose threshold, or where access is granted but disputed by other parties; and

- What happens where data that satisfies one purpose has the potential to disrupt another. For example, where the release of a dataset satisfies the research and development threshold but could also potentially disrupt government services or operations.

Recommendation: The purpose test should be expanded to encompass public interest more broadly.

Our approach to data sharing

The University of Melbourne has well-established policies and processes for the collection, storage, use, re-use, and retention of data and records associated with research projects. The way data is handled by both the researcher and the institution is a key part of the ethical approval process.

In anticipation of the Government's reforms, the University is proposing to develop a Protocol pertaining to access, use and storage of government-held health data that may serve as a helpful guide for other like-organisations to follow, for consideration by the Department.

The draft Protocol to be developed would take the Five-Safes model one step further by creating add-on protocols under each "safe". The Protocol would provide definitional certainty and much needed detail while encouraging data custodians to think adaptively.

The Protocol would expand and provide detail within each of the five safes addressing key issues, such as:

- **Safe data**: How is risk of incidental or secondary findings, or dual use, to be appropriately assessed? What are appropriate mitigation techniques? How is the interpretive potential (as well as pedigree) of data to be addressed within any governance framework?

- **Safe people**: What are the technical skills and qualifications necessary; what ongoing training, assessment, supervision and regulation are good practice?

- **Safe settings**: How will data security, integrity, and confidentiality be assured? What are the appropriate technical, organisational and contractual measures to assure trustworthy governance and safe use of data?

- **Safe outputs**: When can outputs be appropriately described as anonymised? What measures will be introduced to protect commercially sensitive data? How are controls to vary according to use within a safe (internal) setting, publication or dissemination under contract? How is ongoing confidentiality of data to be appropriately assured in different (fluid) contexts? What are appropriate 'reach-through' controls e.g. audit?

- **Safe projects**: How is the appropriateness of a project, given data available, to be assured? How will public benefit be assessed? What mutual recognition of approval processes will streamline, avoid duplication, and constitute trustworthy data governance?

Recommendation: Develop a detailed Protocol pertaining to access, use and storage of government-held health data. This Protocol would be based on the Five Safes, taking this one stage further in terms of detailed definitions. The Protocol would be able to be used as a model approach for other like-organisations.

Recommendation: Where possible, the framework for making decisions about releasing data should have checks and balances, decision review components, appeal procedures, and be supported by non-government expert evidence.

Recommendation: The DS&R Bill should include mechanisms to protect against the release of commercially sensitive data. Data shared with Commonwealth agencies due to mandatory reporting or regulatory arrangements will need to be appropriately screened and sanitised to ensure that sensitive or otherwise confidential information, is protected.

The roles and responsibilities within the system

The University strongly supports the proposed role for accredited bodies and entities within the data system. The most challenging policy problems cut across multiple sectors – having an appropriate release authority with the capacity to navigate diverse data requests and implement best practice procedures is therefore critical. What works well is the diversity in agency responsibilities and opportunity for expert guidance.

Accredited Data Authorities

The University strongly supports the idea of an Accredited Data Authority (ADA) – entities which have strong experience in data curation, collation, linkage, de-identification, sharing and release. This scheme should allow existing organisations, experienced in handling data, to leverage their strengths. A flexible accreditation process that allows institutions to carry out specialist duties may help to secure efficient processing. For example, it is unlikely that many organisations will have the knowledge, expertise or capacity to link, house, clean, and provide access to data. To do so would require significant resourcing and human capital. It may be more efficient therefore to allow organisations to play to their strengths. For instance, while some ADAs will be able to provide the full suite of services required, others should be permitted to participate in their area of expertise. This approach would draw on a wide network of facilitators.

The period of accreditation is another important consideration. Testing whether data is fit for release is both time and labour intensive. The duration or period of accreditation for an ADA should therefore reflect this cost. Too short a period of accreditation may work as a disincentive.

Recommendation: Accreditation for ADAs and trusted users should be for five-year periods.

Trusted users

The University supports the idea of creating a community of trusted users. Trusted access would enable approved researchers the right to access sensitive and linked public datasets – a critical research base. In the University's opinion, the trust access model has considerable merit. Creating a community of trusted users will streamline access to aggregate and individual data while maintaining appropriate protection of personal records.

For researchers, having access to sometimes sensitive information can be a critical resource. Regarding the accreditation of trusted users in the research sector, the University recommends that the process for gaining approval be light-touch and streamlined, appropriately acknowledging institutional reputation, existing governance structures and the ethical codes of conduct in place.

Recommendation: All participants in the trusted user model should have clear accountabilities and responsibilities. Trusted users should be clearly identified, should agree to well-defined conditions of use, and be employed by organisations that have signed a legal undertaking that sets out appropriate data safeguards.

National Data Commissioner

The University is broadly supportive of the role of the National Data Commissioner (NDC) as contemplated by the Issues Paper. The responsibilities of the NDC will need to be more clearly defined particularly with respect to its potential role as both an advisory and enforcement body. One important feature of the NDC will be its role in reframing the conversation around data sharing from one that is premised on risk to one that recognises the societal benefits of data sharing.

A key finding by the Productivity Commission Inquiry Report on Data Availability and Use was a lack of trust by both data custodians and users, an obstacle reinforced by the current matrix of legislation based on risk aversion and avoidance.

The Productivity Commission also notes, and the University agrees, that any new arrangement must be supported by initiatives and programs of education aimed at developing community support, building trust, and in so doing, creating the necessary social licence to implement fundamental reform. To this end, the University endorses, and has previously advocated for, the introduction of a National Data Commissioner and National Data Advisory Council. Oversight of institutional and governance arrangements will be necessary to ensure community trust and safe data sharing. Likewise, an advisory panel with cross-sectoral experts must work to advise the Commissioner on safeguarding data, ethical data use and supporting technical best practice.

Recommendation: Building and maintaining a social licence should sit at the centre of the Government's reform package. Through the office of the National Data Commissioner, efforts should be directed towards: showcasing the system's integrity, transparency and accountability mechanisms; and helping people to better understand the potential community-wide benefits of data availability and use.

Recommendation: To support the objective of transparency, the NDC should publish and maintain an up-to-date performance report. The report could include key metrics such as: the leading data sharing custodian; the number of requests received; the number of requests granted; time taken to complete data sharing requests; and information about why certain requests are refused.

Recommendation: The DS&R Bill should introduce a process for dispute resolution. Like the functions carried out by an ombudsman, applicants denied access to data and/or individuals who believe they have been the subject of inappropriate or unauthorised data release should have access to suitable avenues of redress. The design of the review will be important and should ensure a process that is openly accessible, cost-effective, and expedient.

National Data Advisory Council

The University strongly endorses the role of the National Data Advisory Council (NDAC) as a body equipped through cross-sectoral expertise to provide advice on ethical data use, technical best practice and relevant industry developments. The Council can ensure the framework is informed by relevant expert communities drawn from key data practitioners, data experts and government, and is a body capable of evolving and responding to new technologies and capabilities are they are developed.

The Issues Paper contemplates NDAC's role as one that assists the National Data Commissioner to find the right balance between streamlining the sharing and release of data and ensuring the protection of privacy and confidentiality. Part of the role will also involve advising the Commissioner on community attitudes and expectations. In the University's opinion, community representation, in addition to industry and academic expertise should be included. Community representation will help to ensure that the Commissioner's approach is consistent with prevailing and evolving public attitudes.

Recommendation: To perform its function as an independent advisory body, the National Data Advisory Council should include community representation in addition to technical experts to ensure that approaches to data sharing and release are consistent with public attitudes.