30 July 2018

Department of the Prime Minister and Cabinet
PO Box 6500
Canberra ACT 2600
Australia

Submission to the New Australian Government Data Sharing and Release Legislation

Thank you for the opportunity to make a submission as part of the consultation process for the Data Sharing and Release Bill. We recognise the importance of the New Australian Government Sharing and Release Legislation in the context of Australia's rapidly changing urban environments and the needs of our ever-increasing data landscape. It is here we introduce AURIN; AURIN is an example of a complete data release and delivery service. AURIN's principle role is to provision the underlying research infrastructure through powerful collaboration to support researchers, policy makers, and the private sector develop better urban settlements through the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS).

Currently, AURIN coordinates the distribution of over 4,000 high value datasets from 95 authoritative data sources (public and private sector) and provides tools to access and use data through national infrastructure available from http://www.aurin.org.au/. We believe that AURIN's experience providing a complete data release and delivery service would position it well to provide responses to the questions posed within the issues paper and in the future to operate as an Accredited Data Authority.

The AURIN network of stakeholders is extensive and permits researchers and policy-makers across numerous fields to use AURIN to systematically access relevant data and undertake insightful analysis to answer key science and humanities questions in the fields of population and demographics, economic activity, well-being and quality of life, housing, transport, energy and water consumption and innovative urban design (examples of this work is available from https://aurin.org.au/case-studies/). AURIN supports a network of 70 collaborating organisations with its head office hosted by the University of Melbourne. It is in this context we have prepared this submission.

The following pages address the key questions provided within the issues paper. We are available to clarify or elaborate on any aspects of our submission.

Yours sincerely

Dr Serryn Eagleson
Acting Director, AURIN

Questions about the Key Principles of the Data Sharing and Release Bill

1. **Are the listed factors the correct ones to be taken into account to guide the legislative development?**

AURIN agrees with the listed factors with the following qualifications. First, there are many purposes for research which may not have a demonstrated impact on current society. For example, blue sky research that can be about understanding historical trends and associations.

Much research is conducted by the private sector and used to inform policy and AURIN would like to see priority given to data sharing protocols that encourage and enable researchers to work with industry on projects such as developing and supplying data for the measurement of city performance as required for the National Cities Performance Framework and measurement of the Sustainable Development Goals. Both of these examples are key national projects requiring extensive access to data.

AURIN acknowledges that this would depend on the data being shared, the risks, and the accreditation process.

Second, it is our hope that the factor "[p]romot[ing] better sharing of public sector data" will include the ability to use the data for any purpose that does not infringe on any rights of an individual (i.e. privacy and confidentiality rights).

2. **What else should the Government take into consideration when designing the legislation?**

Release of low-risk data should be fast tracked, whereas data identified as high-risk may need to involve a process, e.g. Ethics Committee review to evaluate.

The legislation should include provisions to de-incentivise government agencies from commercialising data and restricting data use. Data collected by government agencies and statutory authorities should be used for the broadest public use to promote better decision making and innovation for Australians.

In serving the principle of "[e]stablishing institutional arrangements", the legislation should include a broad decision-making framework for Accredited Data Authorities and Trusted User institutions to assist these groups to make data-sharing decisions in a way that promotes innovation and agility while protecting the safety of data.

Finally, the legislation should address how the very likely large numbers of requests for data release should be managed. The administrative requirements for processing applications will require careful design so that the NDC does not become a "bottleneck".

Questions about the Scope of Data Sharing and Release Legislation

3. **Should the scope of the legislation be broader or narrower?**

We acknowledge this as a good first step however state, local government and data generated from federal government grants should also be covered. While the Federal government may not have jurisdiction over the State and territory data sharing processes, there should be a framework

in place to which States can subscribe, in order to create a consistent approach across the country.

Example

Some valuer generals' offices across Australia are actively pursuing models for the commercialisation of property data, however this is at odds with the Open Data Government Initiatives underway nationally, and in Qld, NSW, Victoria, and WA. To address the need for geocoded research ready property data, AURIN previously purchased property price data from Fairfax Australia Property Monitors (APM). The cost of the contract in place is $250,000 per annum. The cost of this data is a major impediment to any research or government activities requiring access to such data. It does not serve the national interest for this exceptionally useful dataset to be restricted in its usage, particularly by publicly funded research organisations, by very large access costs.

Without the cooperation of State and Territory governments, there will be a patchy and inconsistent data sharing approach in Australia, negating a large part of what AURIN understands this proposed Bill seeks to accomplish.

4. **Are there entities that should be included or excluded from scope? How would this be justified?**

We acknowledge commercial entities have business models that rely on fee-for-service collection of data. However, there are instances when government has commissioned data to be collected and commercial entities have sought to wrap a pricing model around it that restricts use when the data itself serves a very wide national interest. For example, data AURIN currently purchases from the Public Sector Mapping Agency (PSMA). PSMA is a Pty td company with the shares wholly owned and held by commonwealth and state jurisdictions. The PSMA has a deep well of experience in harmonising data across jurisdictions (20 years) and they have been very successful. However, the cost of purchasing foundation spatial datasets such as Road Network, Topography, Features of Interest and Administrative boundaries under a restricted licence is restricting value research and we would like entities such as PSMA to be included, as this data should be in the public domain.

There is a role for the Commonwealth government to manage the state and local government (and statutory authority) collective data arrangements where the benefit of provisioning data free of charge to end uses provides outweighs the collective commercial of pricing data. Working out which datasets fall into the national interest category should be made a high priority. AURIN has much experience in this regard and would be happy to help.

State and Territory authorities should be included in the scope through the creation of a framework to which they could subscribe. The more government agencies across the country that use the same approach as the Federal government, the more consistent data sharing will be in Australia. As many public entities as possible should be included in the scope.

5. **Should any specific categories of data be specifically out of scope? How would this be justified?**

There needs to be a clear set of guidelines to justify release and the safety and proper use of the data once released.

There are a number of factors that may restrict release of government generated data including: ability to de-identify personal data; ability to compromise critical infrastructure, national security or national defence. The capacity for government to run these assessments in a timely way needs to be substantially increased. The research community could well be used to help with this task.

6. **Should exemptions, for example for national security and law enforcement, occur at the organisational level or for specific data categories?**

Datasets created by government agencies should be deemed open by default. Exemptions may exist and there should be a process administered by the NDC which enables Government agencies to apply for datasets to be deemed excluded from the scheme for national security and law enforcement reasons.

A clearer understanding of the consequences of misuse, as well as the Code of Conduct, is required.

7. **Are there instances where existing secrecy provisions should prevail?**

Risks to the integrity of critical built infrastructure and the critical information supply chains upon which they rely is an instance for careful consideration.

Questions about the Data Release Purpose Test

8. **Does AURIN agree with the stated purposes for sharing data?**

The 'research and development' purpose should include research and development beyond government, research institutions and academics. This purpose should include industry as well when clear innovation outcomes are involved. This would be in line with the Department of Industry Innovation and Science's "National Innovation and Science Agenda" report (https://www.industry.gov.au/national-innovation-and-science-agenda-report).

This Bill should allow for free (or cost-recovery pricing) and open access of public data for non-commercial research purposes. The ability to charge commercial entities could be available to government agencies administering the datasets, but access to datasets should not be restricted by licensing. For example, licensees should be allowed to develop derivative work products.

9. **Are there any gaps in the purpose test that would limit the benefits of public sector data use and reuse?**

There is a lot of room for departmental interpretation enabling some government entities to construe the purpose test narrowly and therefore limit data sharing where it serves one government entity's interest over another. The emphasis should shift away from government purpose toward a more general public benefit focus.

The proposed process for data sharing relies on existing processes which are suboptimal. Whilst it is very difficult for the value of a dataset to be known prior to release, or even when it is requested we believe that the research community provides a powerful role and has extensive experience in managing data use and should be built into the release framework.

We acknowledge the difficult trade-off between data quality (including ensuring accuracy and metadata compliance) and just provisioning the data, which could provide a greater benefit than trying to complete metadata and improve accuracy. The reality is that the effort on metadata will increasingly be compromised for expediency. "Easily" sharing data relies on various factors. For

example, one factor we need to know is “what the data is”. This requires accurate and complete metadata for the primary user against a minimum suite of metadata.

The introduction of a ‘spatial’ component to data adds a new and special consideration, particularly in relation to de-identification. Privacy with geographic data relies on specific knowledge to handle consideration of all spatial/temporal/attribute dimensions. Privacy threat models can be complex and communication of "how to process a request", particularly as high value for research, needs careful consideration. Therefore, a requirement of determination as to whether the data can be de-identified effectively should be included to protect privacy when the data consists of spatial coordinates.

10. **What further detail could be included in the purpose test?**

To overcome the above issues there needs to be a third-party arbitrator, e.g. the NDC decides if the data will be shared or a more defined test needs to be implemented. There is a need for a defined objective test in both instances.

11. **Should data be shared for other purposes? If so, what are those purposes?**

Yes, for R&D and innovation purposes and to build understanding, see question 8 above.

12. **Should there be scope to share data for broader, system-wide purposes?**

Yes, the NDC should have the authority to deem datasets available to the public without data - sharing agreements for system wide purposes, using CC-by licenses for example, that allow for development of derivative products.

13. **Should the purpose test allow the sharing of data to administer or enforce compliance requirements?**

Yes.

Questions about data safeguards

14. **Is the Five-Safes framework the appropriate mechanism to ensure data is safeguarded?**

The 5 Safes is a good model – and AURIN has adopted its use. Each ADA may have a different interpretation and consequently a different implementation.

15. **Are there any additional safeguards that should be applied?**

Yes. For secure data there will need to be policing/enforcement, including an auditing function.

16. **Are there any instances when the Five-Safes could not be applied?**

Not within the current AURIN context.

17. **Is the Five-Safes appropriate when data is shared and used for the specific purposes in the purpose test above?**

Purpose test should be broader as per the response to question 9, but the Five-Safes framework is still appropriate.

18. **How should the responsibility for managing risks be shared in the framework?**

Via secure platforms that monitor data usage and pass on the legal information regarding any restrictions on use of the data.

Each releasing entity should responsible for their own use of the data, with responsibility passing to end users who agree to abide by any existing conditions in order to access the data provided by an ADA or Trusted User.

19. **How would you envisage Five-Safes principles should be applied over the life-cycle of data to ensure data safeguard are continually met?**

AURIN currently fulfils its data safeguarding obligations by:

a) Managing access
b) Establishing appropriate licensing processes and terms including not licensing inappropriate data (e.g. we will not handle identifiable data)
c) Managing use of data: users should be responsible for proper use of data and its appropriate storage. Our obligation is to communicate this clearly and know what we are responsible for when users fail to act appropriately.
d) Managing compliance: monitoring, auditing and responding
e) Developing clear procedures for handling non-compliant events.

20. **Under what circumstances should trusted users be able to access sensitive data?**

To answer this question, 'sensitive' needs to be clearly defined. Once a dataset is identified as 'sensitive' under the definition, the nature of the sensitive data should be carefully weighed against the potential public benefit, which should also be clearly defined.

Questions about public sector data sharing arrangements

21. **Would this arrangement overcome existing barriers to data sharing and release?**

Existing processes rely on the capabilities of government data teams, which are currently affected by various institutional issues such as number and skills of staff to respond to requests, knowledge of the data (e.g. what boundaries were used in aggregation?) and staff turnover (persistent knowledge issues). AURIN's experience with this is that government data teams are significantly understaffed and take long periods of time to respond to requests. Sometimes these teams do not have assigned and trained data professionals and therefore do not have adequate knowledge to deal with requests. AURIN expects that disseminating custodianship of datasets to ADAs or Trusted Users would help alleviate this issue.

22. **Would streamlined and template agreements improve the process?**

Yes. Currently there are a number of bespoke agreements from individual agencies, an example is the partnership arrangement required to access the Australian Business Register (ABR).

23. **Do you agree that data sharing agreements should be made public by default?** There should be one default agreement and any approved variances should be recorded by the NDC in a publicly accessible register.
24. **What level of detail should be published?** Metadata about all data should be published in most instances (subject to considerations of national security), by knowing that data exists we can greatly increase our ability to find and publish that data. Full detail should be published except private and confidential information in accordance with other legislation such as the Privacy Act 1988, etc.
25. **What else should a data sharing agreement contain?** **Licensing:** Creative Commons licence terms for non-commercial research purposes. Broad licence for commercial purposes, ability to create and own derivative work. **Price:** Free or cost recovery prices for non-commercial research purposes. Pricing schedule for commercial applications, which should not be used unreasonably to deny access to the datasets. Price should not be a barrier to an ADA. **Penalties:** Penalties for breach of data sharing agreement. **Term:** For non-commercial and research purposes the term should be in perpetuity. For commercial applications yearly terms with automatic rollovers and increase pricing in accordance with Consumer Price Index.
26. **What other transparency mechanisms could be mandated?** Register of data agreements, as per the response to question 23.
Questions about roles and responsibilities within the system
27. **How long should accreditation as an ADA or Trusted user last?** 5 years, we acknowledge that this will depend on the costs to establish and maintain accreditation.
28. **What could the criteria for accreditation be?** Demonstrated knowledge in the following areas should be required to become an ADA: A demonstrated ability to comply with the Act, evidence of ability to secure data effectively thorough knowledge of de-identification processes, and a demonstrated thorough knowledge of data licensing. Sharing of aggregated, geographic data requires understanding of geographic boundaries varying over time. This may require tools to automatically interpret what the data is (e.g. summary stats). Sharing of disaggregated, geographic data requires understanding of privacy and representational issues (noted above) Public data quality sometimes requires significant effort to make the data usable. This includes checking/correcting raw data values to ensure they match data types and adherence to metadata

standards and maintenance. Metadata, once generated, must then be maintained. AURIN has a depth of experience in maintaining metadata.

It appears that the proposed process will push metadata effort from government out to ADAs and would require clear, ongoing communication between the two groups. This is particularly the case for real time data, e.g. IOT and other sensors. AURIN has the capability to communicate with government with demonstrated effective systems in place.

Accreditation levels may vary depending on the type of ADA. Some maybe linking secure data, others have non-sensitive data about the urban environment may have to follow standards for metadata, publishing in machine readable formats.

29. **Should there be review rights for accreditation?**

Yes, but it is reasonable that a fee should be payable for right of review. If the review is upheld, the fee should be refundable.

30. **Should fees be payable to become accredited?**

Fees could be charged on a cost-recovery basis for administration by the NDC. Anything beyond that could prove too great a disincentive.

31. **Is the Australian Government Charging Framework fit for purpose in this context?**

Yes.

Questions about the National Data Commissioner

32. **Are these the right functions for the National Data Commissioner?**

Yes, but they should be expanded to make the NDC an authority which arbitrates data sharing disputes and enforces the DS&R Bill.

33. **What review powers should the National Data Commissioner have?**

Extensive review and arbitration powers for data sharing disputes, data misuse and accreditation issues.

34. **Should the NDC have the power to conduct an investigation into system-wide issues?**

Yes.

35. **What other actions could the NDC be able to take?**

The NDC should also be able to investigate individual cases related to the DS&R Bill.

36. **Are there other ways community values and expectations can be captured and addressed?**

Creation of a complaints system where the public can make reports to the NDC.

37. **What aspects should be taken into consideration when considering consequences for non-compliance with the DS&R Bill?**

The level of severity of the offence.

The extent and degree of privacy or confidentiality breach, this could result in a loss of accreditation and/or financial loss.
38. **Should the consequences differ depending on the type of data involved or the type of misuse, e.g. harsher penalties for intentional misuse?** Yes for the type of misuse, but the Bill should also specify varying degrees and hold higher standards for misuse of personal information. For example, the standard for negligently using personal data should be lower than the standard for treating non-personal data, because the user of the data should know that personal information requires more care to be taken.
39. **Should penalties be strict liabilities?** No, but deterrence should be a factor to consider when penalising.
40. **What would be an appropriate penalty for intentional misuse of data?** Fines at levels that deter the misuse of data. Different levels for individuals to compared to companies.
41. **How would responsibility for misuse of data be shared across the data system?** An investigation should take place by the NDC to investigate where the misuse occurred in the data custody chain. Receivers of data should be able to assume they have received data that meets licensing and/or privacy requirements. Once the NDC discovers where the misuse occurred, the identified entity takes responsibility for the misuse. The NDC may deem entities jointly responsible.
42. **To what extent should there be a complaints mechanism and how should it work?** A complaints mechanism should be enforced for breach of data-sharing agreements, improper sharing and use of data and data access disputes. The NDC should be able to arbitrate these matters.
43. **Should a complaints mechanism provide for complaints by the public?** Yes. Although we recommend this takes the form of a forum to enable people requesting data to communicate, demonstrate value and to make complaints.

Acknowledgements
This submission draws upon input of the following individuals: the AURIN team – Dr Serryn Eagleson, Mr Jason Kreitner, Mr Aaron Magri, Dr Michael Rigby, and Dr Peter Woodgate.