



Australian Research Data Commons

Submission on the Data Availability and Transparency Bill

6 November 2020

The Australian Research Data Commons (ARDC) would like to thank the National Data Commissioner for the opportunity to comment on the Data Availability and Transparency (DAT) draft exposure Bill.

Introduction

The Australian Research Data Commons (ARDC) has made several contributions on this matter since the time of the Productivity Commission's *Inquiry into Data Availability and Use* in 2016. The ARDC remains strongly supportive of the objectives of the proposed data sharing Scheme and thanks the interim National Data Commissioner and their staff for the consultation and work done to date.

The value of this effort to the research sector is reflected by a 2017 OECD survey that established 74% of scientists reported a 'High' or 'Very High' dependence on public sector information for research.¹ A recent analysis by the Institute for Methods Innovation confirmed that public sector data shared effectively with the research sector creates considerable societal value including to the government itself.²

The draft exposure Bill contains many elements essential for achieving improved availability of Commonwealth data, including:

- A National Data Commissioner to provide strong advocacy and oversight,
- Permitted Purposes to enable world class research and policy development,
- Accreditation to regulate and standardise entry of entities into the Scheme, and
- Use of registers to indicate health of the Scheme and facilitate targeted improvements.

The ARDC recommendations are therefore offered in the spirit of realising further benefits through this Scheme by:

- Making Commonwealth data findable by default,
- Further streamlining Data Sharing Agreements,
- Allowing Custodians and Users to use any Accredited Data Service Provider (ADSP), and
- Ensuring the Commissioner sponsors Interoperability Services.

¹ OECD 2019 Enhanced access to and sharing of data (EASD) - Reconciling Risks and Benefits for Data Re-use across Societies.

² <https://ardc.edu.au/resource/investigating-the-link-between-research-data-and-impact/>

Making Commonwealth Data Findable by Default

The Productivity Commission stated that, ‘before data can be used by those other than the original collector, it must first be discoverable’.³ A Key Point in the report was that, ‘the substantive argument for making data more available is that opportunities to use it are largely unknown until the data sources themselves are better known, and until data users have been able to undertake discovery of data’.⁴

It was also noted that, despite numerous recommendations over the years, data catalogues (for finding data) are not commonly used by the public sector and this has been for a range of reasons including, ‘an entrenched culture of risk aversion, reinforced by a range of policy requirements and approval processes, and often perverse incentives...’.⁵

The draft Bill does not currently contain explicit obligations that would ensure a comprehensive register of shareable data. Without global discoverability of shareable data, the Scheme may not result in any actual sharing. At the very least, the Scheme will continue to make finding data time consuming and expensive for Data Users; and for Data Custodians, the number of non-relevant queries about data will likely increase. If this wastage is not addressed by the Scheme, it will dis-incentivise participation by a research community already facing dire economic circumstances.

If the Government is to make a return on its investment in the Scheme, while imposing sizable costs of accreditation, then Commonwealth Data Custodians must be obligated by the Bill to ensure all data they control is findable.

This would significantly strengthen the data supply component of the Bill in line with its objectives, particularly Section 3(a) ‘promote better availability of public sector data’. It should directly address specific findings of the Productivity Commission as agreed by the Government.⁶ The relevant findings are paraphrased below:

- All Commonwealth bodies are to create comprehensive, easy to access registers, that include the metadata and links to the datasets, for data they control. Where datasets are not available for sharing, the register should indicate this and the reasons for it not being shareable. **(Recommendation 6.4)**
- The National Data Commissioner is to offer a public process whereby datasets or combinations of datasets, supported by a public interest business case, can be nominated for sharing. A list of requested datasets, and decisions relating to dataset sharing or otherwise, should be transparent and published publicly. **(Recommendation 6.5)**
- Data Custodians are to ensure that requests to share data they control are dealt with in a timely and efficient manner and are consistent with a risk management approach. Data Custodians are to report annually to the National Data Commissioner on their handling of requests for data sharing. **(Recommendation 6.7)**

To achieve this, the ARDC strongly recommends inclusion in the Bill of the FAIR Data Principles.⁷ The FAIR Data Principles are for data ‘supply’, what the Five Safes Principles are for data ‘use’. The acronym FAIR stands for making

³ Productivity Commission, Inquiry into Data Availability and Use, p.159.

⁴ Productivity Commission, Inquiry into Data Availability and Use, p.2.

⁵ Recommendation 3.5 of the Productivity Commission, Inquiry into Data Availability and Use, p.34.

⁶ <https://www.pc.gov.au/inquiries/completed/data-access/data-availability-use-government-response.pdf>

⁷ <https://ardc.edu.au/resources/working-with-data/fair-data/>

data: Findable, Accessible, Interoperable and Reusable. FAIR does not mean that data is open or that Custodians must share - appropriate protections must still be applied.

ARDC Recommends:

- Data Custodians are to be required by the Bill to make data they control findable, regardless of whether it is subject to a data sharing agreement, by provisioning the data in accordance with the FAIR Data Principles.

Further Streamlining Data Sharing Agreements

Data sharing agreements need to be streamlined further if the Bill is to deliver its intended objectives of creating 'consistent safeguards' and 'institutional arrangements' for the sharing of data. To achieve this streamlining:

- Custodians need to trust Data Users once they are accredited,
- Custodians do not have full liability and responsibility for the handling of data by others in the Scheme, and
- The new institutional constructs between Users and Custodians, established by the Office of the National Data Commissioner (ONDC), need to be effective.

The remaining transactional friction of agreements could be removed if the Bill were brought further in line with the Productivity Commission findings as agreed by Government.⁸ The relevant findings, paraphrased below, include:

- In assessing trusted user access, the (Data Custodian) should accept existing current approvals of the trusted user's work environment....and grant, on a continuing program-wide basis, data access to trusted users from a range of potential entities. **(Recommendation 6.12)**
- The default position should be that after the applicant and their institution establishes capability to respect the processes and obligations of the...accredited standard, an individual researcher from one of these organisations should be readily approved for access. **(Recommendation 7.3)**

Note especially that, regardless of whether it is for releasable or shareable data:

- Access should be governed by data sharing agreements that embed the 'trusted user' principle.... **(Recommendation 8.3)**

Importantly, the reference to embedding the 'trusted user' principle was not that conformance with the 'Data Sharing Principles' needed to be re-asserted by the Data User in every agreement (as is currently proposed⁹), but that, it is in fact the opposite, that as a principle, *agreements should leverage the existing trust in users*.

Therefore, data sharing agreements should not allow re-visiting the conformance of Data Users with the Data Sharing Principles for each dataset accessed. Given the Scheme establishes accredited users as those trusted to apply the Data Sharing Principles, data sharing agreements should be set at a higher level (e.g. program or organisation rather than project), and should be made in a way that they can be used repeatedly across multiple Data Users and uses.

⁸ <https://www.pc.gov.au/inquiries/completed/data-access/data-availability-use-government-response.pdf>

⁹ Section 18, Items 3 and 7.

The considerable cost of accreditation should result in streamlined sharing for both Data Custodians and Data Users. However, the scheme-wide cost of local granular risk management applied at every data sharing agreement will substantially increase the costs on already financially stressed research institutions while simultaneously undermining the ability of the Scheme to deliver on the objectives of the Bill as stated.¹⁰

ARDC Recommends:

- Remove the clause in Section 18, Item 3 of the Bill requiring ‘outputs’ (as per the ‘outputs principle’) be identified up-front in the data sharing agreement, and
- Remove Section 18, Item 7 requiring accredited users to re-assert conformance with Data Sharing Principles.

Allowing Custodians and Users to use any ADSP of their choice

As a result of submissions received during the Productivity Commission Inquiry it was concluded that Australia needed a data sharing framework that could scale to handle both a large number of datasets as well as cover a great variety of initiatives and sectors.

To achieve this safely, sharing should only be between entities accredited under the Scheme, preferably directly between Data Custodians and Data Users whenever possible. However, if required, Data Custodians and Data Users should be able to choose any ADSP to act on their behalf. This use of an ADSP by a Data Custodian or Data User would result from a separate agreement directly between them and the ADSP. This would allow use of ADSPs in highly flexible configurations including:

- As an agent on behalf of a Custodian to provision data safely to multiple Data Users (or their ADSP), and
- As an agent on behalf of a User to consume data safely from multiple Data Custodians (or their ADSP).

The Bill does not currently require Data Custodians to be accredited as proof of their data handling capacity (as the Scheme requires of Data Users), but it does allow Custodians to mandate use of a specific ADSP in the data sharing agreement.¹¹ And, the Bill additionally defines that sharing involving an ADSP can only occur in two steps - between the Data Custodian and a specific ADSP, then between the specific ADSP and the Data User.¹²

This effectively removes the capacity of Data Users to utilise the services of one or more ADSP of their choosing. More broadly, it specifically prevents utilising a rich array of ADSP to be brought to bear on Commonwealth data.

Ideally, the Scheme should incentivise sharing without the need to specify the intermediaries within each data sharing agreement. This is because mandating the use of a specific intermediary does not have any inherent value for keeping data safe (as long as every entity involved is accredited as per Scheme rules), but will prevent the Scheme from scaling to meet demand in the volume and diversity required. Some example scenarios might include:

- Using data from several portfolios could create a clash of requirements if Custodians use different ADSP.

¹⁰ <https://pmc.gov.au/resource-centre/public-data/guidance-data-sharing-australian-government-entities>

¹¹ Section 18 Item 8.

¹² Section 13(2).

- Researchers pulling data together from multiple jurisdictions, sectors, and/or subject domains will need to use the trusted service appropriate and pragmatic for their program of work.

ARDC Recommends:

- Data Custodians should not be able to mandate use of any particular ADSP by a Data User.
- Data Users should be able to collaborate with any number of ADSP on the datasets shared with them.

Ensuring the Commissioner Sponsors Interoperability Services

The Scheme describes a complex interaction of users, custodians and datasets regulated by accreditation and agreements. The ability for these interactions to happen at scale, with appropriate speed and at reduced marginal costs across multiple organisations and sectors assumes a level of automation.

Increased automation requires a sophisticated facilitating layer enabled by contemporary information and communications infrastructure and services. Some commonly accepted elements include scheme-wide registers, catalogues, identifiers, authentication and authorisation.

This facilitating layer extends well beyond technology to include interoperable business processes and common legal frameworks such as licences and terms of use. The Data Interoperability Maturity Model¹³ from the National Archives of Australia offers a comprehensive approach beyond technology that the Commissioner could apply to the Scheme.

As an example of business process interoperability, it is imperative the Scheme recognises existing mature research sector mandated procedures for ensuring safe projects (e.g. Human Research Ethics Committees¹⁴) or safe data and people (e.g. The Code for Responsible Conduct of Research¹⁵ and the National Statement on Ethical Conduct in Human Research¹⁶). The Commissioner will need to clearly define equivalent external business processes to influence Custodian behaviors in a way that ensures efficient and 'consistent' recognition of these existing trust systems.

To improve interoperability, the Commissioner needs an explicit operational function under Section 41 of the Bill, including an expectation that the Commissioner will provide Interoperability Services for the Scheme. Of note, these services need not be run directly by the Commissioner, but instead can be operated by bodies sponsored by the Commissioner. This ensures perceived objectivity of the Commissioner and better enables the Commissioner to mediate when required service levels are being negotiated or when disagreements arise during operations.

Scope for inclusion of these Interoperability Services can already be seen in the DAT Bill 2020 Exposure Draft Consultation paper, which includes a graphic titled 'Figure 1 - Components of the data sharing scheme established by the Bill'.¹⁷ At the bottom of this graphic is a panel titled 'Operational and Administrative Systems: The National Data Commissioner will establish systems and processes to implement (and presumably 'operate') the scheme'.

¹³ <https://www.naa.gov.au/information-management/data-interoperability-maturity-model>

¹⁴ <https://www.nhmrc.gov.au/research-policy/ethics/human-research-ethics-committees>

¹⁵ <https://www.nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2018>

¹⁶ <https://www.ahmrc.org.au/publication/national-statement-on-ethical-conduct-in-human-research/>

¹⁷ p. 13

Mandating use of Interoperability Services by entities whenever they are involved in data sharing should be a defining characteristic of the Scheme, contributing to Objective 3(b) of the Bill 'consistent safeguards for sharing'.

Lastly, the Scheme should incentivise proliferation of accredited services that offer more advanced capabilities such as transformation, data linkage, analytics and privacy enhancing technologies just to name a few. These services could be used at the discretion of any participating entity based on their application of Scheme principles and project needs. This would better meet Objective 3(d) of the Bill 'confidence in the use' (of data).

ARDC Recommends:

- The Bill includes an 'operational' function of the Commissioner and obligates them to provide Interoperability Services that ensure consistent safeguards in the Scheme when sharing data.
- The Bill allows the Commissioner to sponsor bodies to provide Interoperability Services on their behalf.
- The Bill makes mandatory the use of Interoperability Services by any entity involved in the sharing of data.
- The Commissioner matures interoperability of the Scheme as a whole (not just each participating entity) over time and in line with the Data Interoperability Maturity Model from the Australian National Archives.

Conclusion

The Data Availability and Transparency Bill represents an important opportunity for the Government to be a modernising force for society, utilising the latest information sharing approaches in a way that benefits citizens.

ARDC is excited by the opportunity the intent of the Bill represents. It is an opportunity to bring even more of Australia's intellectual capacity to bear on highly complex research and policy challenges facing Australia and the world. As already indicated above, better availability of Commonwealth data is essential to achieving this outcome.

The draft Bill includes many strong elements, legislating important data institutions and infrastructure critical for the better use of Commonwealth data, but there remain critical gaps. And these gaps will be further exacerbated by the current economic crisis faced by universities and research institutions. This has, and will for some time, make them hypervigilant about ensuring every scarce resource spent on accessing data will provide a worthwhile return.

Ideally, the Bill should reflect a clearer vision of the key lines of effort required to build a national Scheme including by whom and how they relate to each other. As such, the Bill should result in:

- Data Custodians applying the **FAIR Data Principles** to ensure data is findable and provisioned appropriately.
- Data Users applying the **Five Safes Principles** to ensure data is accessed and used safely.
- The Commissioner provides **Interoperability Services** that facilitate and safeguard all data sharing.
- Data Service Provider numbers and value offered increase over time, incentivising entity participation.

If these changes recommended by the ARDC are adopted, the Bill will be an important piece in transforming data capabilities nationally so as to support leading edge research and policy innovation.

A graphic summarising ARDC recommendations is provided below:

ARDC Recommendations for the DAT Bill

