

6 November 2020

The Australian Academy of Science (the Academy) welcomes the opportunity to comment on the exposure draft of the Data Availability and Transparency Bill 2020 by the Office of the National Data Commissioner (ONDC).

The Academy has a strong interest in information management policy, particularly in the context of research data, and supports the Bill's aims to use data to inform Government policy and programs and for research and development activities. In line with these aims, the Academy also supports the increased role for the research sector in evidence-based policymaking through the safe and responsible sharing of Government data.

The value of data

Although the value of data has become paramount, data itself is not inherently valuable. The value of data lies with the subsequent analyses and actions performed following its initial collection. These actions are usually carried out by highly skilled and trained researchers. For instance, observational earth and climate data collected by the Bureau of Meteorology provide the basis for the development, calibration and evaluation of climate models.¹ Data is only a strategic asset if appropriate mechanisms for its access, sharing, and reuse are in place. These mechanisms must effectively facilitate and not hinder its use. There is a risk that Government data assets would not be used to its full potential, and its value lost if not used.

To realise its full value, data must also be made available in its raw format, with minimal manipulation or conversion and its supporting material (e.g., metadata and linked datasets) also made available. The Productivity Commission report on Data Availability (Recommendation 6.4) has highlighted this issue.²

Data safety and security

The Academy acknowledges that the Bill has been developed based on the Fives Safes framework for effective decision-making for data and risk management, which is useful in highlighting the possible sources of risk in data sharing.³

It is also vital that the Bill promotes and enables best practice protections and thus, is fit-for-purpose. For example, for data at rest or in transit, cryptography is recommended, and differential privacy and active consent should be required for data sharing and release.^{4,5} The USA Census Bureau has fully adopted differential privacy following internal research revealing the limitations and consequent vulnerabilities of systems used for the 2000 and 2010 Censuses.⁶ The Bill can be used to strengthen privacy protections.

Doing so does not prevent data sharing but will increase community trust and the social license to share data. Furthermore, best-practice protections will also support the Bill's aim to enable data to be "as open as possible and as closed as necessary".

Data sharing and access

In the context of research, data is generated, used across and shared between both government and research sectors. The Academy recommends policies be developed that make it easier for data to be shared and to be interoperable across both sectors, i.e. information management policies for both sectors should be aligned.

The Academy also encourages the Office of the National Data Commissioner to ensure that its framework and policies uphold the principles of:

- FAIR: (Findable, Accessible, Interoperable and Reusable): Supporting open data practices to facilitate increased data sharing;⁷ and
- CARE: (Collective Benefit, Authority to Control, Responsibility and Ethics): Supporting the right for Indigenous People to govern the collection, ownership and application of their data.⁸

The 'FAIR Guiding Principles' and 'CARE Principles for Indigenous Data Governance' are complementary and encourage custodians and users of research data to consider both people and purpose.

Government data should also be available in open, non-proprietary standards and formats with a clear definition of licensing and access constraints to promote accessibility, interoperability and reusability.

Finally, given that digital technologies are constantly evolving, the data collected and stored must also be machine-readable. Machine readability will ensure that data assets remain fit-for-purpose and useable in the future, as data becomes more valuable over time.

Data in research

Most of Australia's research activity occurs in universities. In recognising this, universities should have a crucial role in the training and provision of expert advice required for data services including linking, cleaning, quality, ethics and privacy-enhancing technology. This process could be another dimension of the accreditation process but must also acknowledge existing mechanisms in place (e.g. ethics approvals) to prevent duplication and unnecessary administrative burden.

A fundamental pillar of research integrity is the peer review process, which formally takes place before a research paper is published but is an ongoing process conducted by the broader research community. To

enable the ongoing process of peer review, data that underpins research publications (and policy development) must be retained and made accessible. A recent example testifies to the importance of reviewers having access to this data: a potential COVID 19 therapeutic, hydroxychloroquine, was reported to increase the risk of death in hospitalised patients and international trials were halted. It was only when reviewers requested access to the research data, and their request declined, that the data was found to be inadequate and the publication was subsequently retracted and clinical trials resumed.^{9,10} Reviewers' access to research data is particularly important in situations where results could be considered controversial or during investigations of research misconduct. The Academy encourages the ONDC to build this principle and associated access protocols – consistent with the FAIR and CARE principles – into its framework.

The events of 2019 and 2020 have demonstrated that now, more than ever, the need for, and value of data for research. Australia's research capabilities have served Australia well in response to the COVID-19 pandemic: increasing the ease of access to Government data will ensure that researchers can continue to use data in ways that benefit Australians, safely and responsibly.

The Academy supports the Commissioner's intention to develop a consistent and transparent framework for making fair use of the Commonwealth's data. [REDACTED]

References

1. About data and observations. *Bureau of Meteorology* <http://www.bom.gov.au/climate/data-services/about-data-observations.shtml>.
2. Data availability and Use. *Australian Government Productivity Commission* <https://www.pc.gov.au/inquiries/completed/data-access#report> (2016).
3. The five safes. <http://www.fivesafes.org/>.
4. Menezes, A., van Oorschot, P. C. & Vanstone, S. A. *Handbook of applied cryptography*. (CRC Press, 2001).
5. Differential privacy for privacy-preserving data analysis: an introduction to our blog series. *National Institute of Standards and Technology (NIST) - U.S. Department of Commerce* <https://www.nist.gov/blogs/cybersecurity-insights/differential-privacy-privacy-preserving-data-analysis-introduction-our>.
6. Abowd, J. M. The U.S. Census Bureau adopts differential privacy. in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK 2867–2867* (Association for Computing Machinery (ACM), 2018).

doi:10.1145/3219819.3226070.

7. FAIR principles. *GO FAIR* <https://www.go-fair.org/fair-principles/>.
8. CARE principles of Indigenous data governance. *Global Indigenous Data Alliance* <https://www.gida-global.org/care>.
9. Mehra, M. R., Desai, S. S., Ruschitzka, F. & Patel, A. N. RETRACTED:Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *Lancet* **0**, (2020).
10. Mehra, M. R., Ruschitzka, F. & Patel, A. N. Retraction—Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *The Lancet* vol. 395 1820 (2020).