

Australian Research Data Commons (ARDC) submission to the Human Rights Commission Consultation on Artificial Intelligence, 21 March 2019

What is the ARDC?

[The Australian Research Data Commons](#) is funded through the Australian Government's [National Collaborative Research Infrastructure Strategy \(NCRIS\)](#). Under the broad recommendations by the Chief Scientist's [2016 National Research Infrastructure Roadmap](#), the ARDC was established in 2018 from the foundation work of the Australian National Data Service project (ANDS), the National Collaboration Tools and Resources project (Nectar), and Research Data Services project (RDS). The ARDC's focus is on research data infrastructure which includes research clouds, Virtual Laboratories, high-value trusted data services, collections, storage, and sector-wide skills and policy development. The ARDC is a transformational sector-wide initiative, working with partners to build a coherent national and collaborative research data commons which delivers a world-leading data advantage, facilitates accelerated innovation, fosters collaboration for borderless research, and enhances researchers' ability to translate their work into benefits.

Considering how artificial intelligence is currently regulated and influenced in Australia:

What existing bodies play an important role in this area?

What are the gaps in the current regulatory system?

Data is an enabler of AI

As noted by the Human Rights Commission White Paper, "Personal data is the "fuel" of AI".¹ However data more broadly is fundamental to the application of all modern AI systems. This is true whether we are considering an AI system underpinning a driverless car or an email spam filter. Therefore regulations governing the access and use of data are fundamental to regulations pertaining to the use of AI systems. As data may be collected and controlled by private, public, and research organisations, under the current regulatory environment, the access and use conditions of data across these different sectors may vary accordingly.² While considerable progress has been made to harmonise the data policies and regulations across different sectors, this is still a work in progress.

Public Sector Data

Accordingly to the Australian Government Public Data Policy Statement released in 2015:

The Australian Government commits to optimise the use and reuse of public data; to release non sensitive data as open by default; and to collaborate with the private and research sectors to extend the value of public data for the benefit of the Australian

¹ Page 9 https://tech.humanrights.gov.au/sites/default/files/2019-02/AHRC_WEF_AI_WhitePaper2019.pdf

² We note that these three types of organisations are not necessarily mutually exclusive. Commercial organisations may also conduct research and generate research data for developing new products or services.

public. Public data includes all data collected by government entities for any purposes including; government administration, research or service delivery. Non-sensitive data is anonymised data that does not identify an individual or breach privacy or security requirements.³

As a part of the Government's Public Data Policy initiative, the Productivity Commission undertook an Inquiry into Data Availability and Use. The full report was released in 2017.⁴ In response to key recommendations of the Productivity Commission, the Government appointed the National Data Commissioner and proposed new Data Sharing and Release legislation in 2019. A key responsibility for the National Data Commissioner is to implement a data sharing and release framework to ensure effective and efficient use and reuse of public and publicly funded research data while maintaining strong security and privacy protections. The National Data Commissioner will work with the national privacy regulator, the Australian Information Commissioner, to ensure that Australia's data sharing and release framework is anchored on strong foundation of privacy and security.

Research Data (research-specific policies)

In addition to the above, the new *Australian Code for the Responsible Conduct of Research* (2018), the *National Statement on Ethical Conduct in Human Research* (2018), and the *Better Practice Guide to Data Management* (expected to be released in 2019) are key research policy documents developed by the Australian Research Council, the National Health and Medical Research Council and Universities Australia. In addition, universities, publicly funded research organisations (e.g. CSIRO) are all relevant stakeholders in the production and management of research data.

In addition to publicly funded research data, commercial organisations may also produce research data in Australia. In these cases, there may be strong commercial reasons to protect these data assets. Intellectual property rights are an important consideration in the access and use of these kinds of data.

Data Infrastructure

A particularly unique feature of the Australian research data landscape is the establishment of key national research infrastructure capabilities through the NCRIS program. NCRIS represents over 30 research infrastructure projects to provide long term access to research data, infrastructures and services to support nationally significant research. For examples, the [Population Health Research Network \(PHRN\)](#) provides key capability to link de-identified population health datasets for research, the [Atlas of Living Australia \(ALA\)](#) is a web based national database of Australian plants and animals, and the [AuScope](#) is an infrastructure system

³ https://www.pmc.gov.au/sites/default/files/publications/aust_govt_public_data_policy_statement_1.pdf

⁴ The report is available at <https://www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf>
For the consultation paper of the Data Sharing and Release Legislation, see <https://www.pmc.gov.au/resource-centre/public-data/issues-paper-data-sharing-release-legislation>

for earth sciences, enabling managed data and software to be used for computer modelling and simulations.

Frameworks for Consideration

Given the close relation between data and AI systems, some of the regulatory frameworks around data could be extended to or re-modeled for AI.

In the area of data linkage, there are a number of frameworks for managing risk. For example, 'privacy preserving data linkage' (a methodology) is commonly used to link multiple health datasets at unit record level while maintaining privacy protection. Some of these frameworks may be useful for consideration in the context of regulating the application of AI technologies to datasets. In Australia, PHRN has considerable expertise in data linkage of deidentified population health datasets.

The Five-Safes framework, developed by the UK Data Service and subsequently adopted, among others, by the Australian Bureau of Statistics and the Office of the National Data Commissioner,⁵ is a framework for managing risks and access to data. The Five-Safes provides flexibility to allow access to higher risk or more sensitive datasets in an appropriate manner. An example is having an openly available metadata record of a dataset but access to the data through application, where the applicant demonstrates the data will be safeguarded and used by an approved authority for a relevant reason. In extreme cases, even a metadata record may be sensitive (e.g. National Security) and here the Five-Safes may well have to be tested against the metadata record as well as the data.

ARDC would welcome the opportunity to discuss this submission further. The contact for this submission is [REDACTED]

⁵ See <http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017>, [Data Sharing Principles Best Practice Guide](#) and [Sharing Data Safely Brochure](#).