



NATIONAL ARCHIVES OF AUSTRALIA - SUBMISSION

Productivity Commission – Draft Report on Data Availability and Use

Context

As the custodian of high value Australian Government information and data, the National Archives of Australia (Archives) is pleased to respond to the Productivity Commission's draft report of October 2016. The Commission's recommendations will reshape the Australian data landscape and the Archives supports the aim to provide access to more open data for use and re-use.

While the draft report refers to the Archives' role in Chapter 6, it does not acknowledge explicitly the significant role the Archives plays in the long-term management and accessibility of Australian government information, including data. Consistent with the Archives' role in determining, preserving and providing access to all long-term Government information and data, the Archives expects that any entity established to deliver short-term data management will work with the Archives so that enduring data can be preserved and made available for re-use and to support research into Australia's history, democracy and place in the world

The Archives makes this response to ensure that its independent role overseeing Commonwealth information management, identifying the Commonwealth's valuable archival resources and preserving and making those resources publicly available is:

- taken into account by the Inquiry; and
- not affected inadvertently by the Inquiry's final recommendations.

The Archives' response supports the need for reform with regard to data availability and use, while ensuring the nation's most valuable archival resources are not at risk.

The Archives is active in the information and data management space, and highlights that new roles established as a result of the Inquiry's recommendations need to avoid duplication of effort and already established functions.

Recommendations

The Archives recommends that:

1. The Inquiry notes the Archives' legislative role in the selection and long-term preservation and access to Australian government information, records and datasets of enduring national value.
2. The Inquiry recognises the Archives' role as a key institution in a future data framework.
3. National interest data is not put at risk in favour of short-term use, and the Inquiry notes that the Archives has a vested interest in, and obligations for, maintaining archival data assets in perpetuity.
4. The necessity and intent of a Data Sharing and Release Act be carefully examined, especially concerning possible duplication and implications under the *Archives Act 1983*; and the Inquiry notes the commitment to a simpler framework for information access laws, policies and procedures under the Open Government Partnership – National Action Plan.
5. The Inquiry notes the Archives' project on automating the identification of information and data value through machine-learning, and the limitations of a committee approach to identifying Government information and data in today's digital environment.
6. The responsibility entity implementing the proposed data register aligns with Digital Continuity 2020 requirements – so there is no duplication of effort.
7. The Inquiry's final recommendation includes further regulation in system procurement for information and data management specifications (in line with the Archives' Digital Continuity 2020 Policy) to enable interoperability.

Key points

The Archives and the proposed Data Sharing and Release Act and National Data Custodian agency

Under the *Archives Act 1983* (the Act), the Archives determines the status of all Commonwealth records, which in this the digital age, includes all government information and data. It also prescribes the management, retention or destruction at an appropriate time, and preservation and accessibility of the archival resources of the Commonwealth.

The Archives also determines the information management obligations of Australian Government entities in accordance with its powers and responsibilities under the Act. The Archives has legislative responsibility for ensuring government accountability through good information and data

management, which safeguards the national interest and the rights and entitlements of individuals to access government information in the long-term.

The Archives questions the proposed Data Sharing and Release Act, which the report notes would apply ‘across Australia to all digital data’ (page 14), especially around identifying and managing the Commonwealth’s resources, and recommends any new legislation should be carefully considered with a thorough examination of implications under the Archives Act. There is also a commitment in Australia’s Open Government National Action Plan to develop a simpler framework for information access laws, policies and procedures which will need to be taken into account.

A key consideration for the Inquiry is to ensure that any recommendation for a proposed National Data Custodian is not duplicating efforts of the National Archives, and in line with the Belcher Red Tape Review, does not increase the burden on Australian Government agencies by making them engage in separate processes for the identification, management and distribution of data assets that are identified as both a ‘national archive’ and a ‘national interest dataset’. The Archives contends that datasets of national interest are likely to be of archival value and therefore subject to the provisions of the Archives Act.

Identifying value of data assets

Under the Act, the Archives identifies the value of both temporary and archival data assets, and has, over the past 70 years, developed its specialised role in identifying the value of government information. The datasets it identifies and collects signify their value as a national resource.

This role has evolved with technology to include datasets, for example the:

- Department of Health’s – Private health Insurance datasets (eg Hospital Casemix Protocol, Dental Data and Risk Equalisation data), including data collected under legislative requirements, used for analysis to support research and evidence based policy development;
- Department of Agriculture’s major biosecurity datasets and associated information about the data fields, for example data dictionaries, indexes, tables, registers and repositories.
- Clean Energy Regulator’s major datasets (eg greenhouse gas emissions) and information about the data fields, for example data dictionaries, indexes, tables and registers.

Under the Act, these datasets will be kept permanently as a national resource, preserved and available for public access, and not merely for compliance, record-keeping or audit as the Productivity Commission’s draft report suggests:

Designating datasets as national interest collections will also signify their value as resources collected in the national interest, not merely (as today) for compliance, record-keeping or audit (Productivity Commission draft report, p19).

The National Archives also holds considerable quantities of paper and other analogue records of national interest which if available digitally can be used in innovative and ground breaking ways for historical, longitudinal and other research. For example, the estimated 600,000 repatriation records of the men and women who served in World War I and returned to Australia present new opportunities for social, medical and other research. The Archives has commenced digitising these records and making them available online to researchers around the world. Current fiscal restraints limit the Archives from continuing the conversion of these valuable records to digital format.

The future of identifying value of data assets

Draft recommendation 2.1 states:

In determining datasets for public release, a central government agency with policy responsibility for data should maintain a system whereby all Australian governments' agencies, researchers and the private sector can, on an ongoing basis, nominate datasets or combinations of datasets for public release...(Productivity Commission draft report, p88).

The Archives supports the Commission's recommendation for a Framework guiding the designation of datasets. The Archives' reiterates its current role in the designation of Government data(sets), and questions the extent to which established expertise and experience has been considered in the proposed Framework. The Commission's approach overlooks provisions such as Principles on Open Public Sector Information, Australian Government Public Data Policy and the Open Government Partnership – National Action Plan, as well as equivalent policies in the Australian States. Strengthening the mandate for open material, while leaving the responsibilities to entities with the appropriate expertise, would provide efficiencies and reduce duplication of effort. The Archives has a wealth of experience in identifying the value of information and data. The Belcher Red Tape Review also recommended that the National Archives be more closely involved in policy development processes and decision-making forums on government information management, including digital transformation-related matters to ensure government information and data is usable for the future and to reduce the administrative burden arising from meeting their responsibilities under the Archives Act 1983. The Archives recommends that the Inquiry recognises the Archives as a key institution in a future data framework.

The draft report suggests that:

For community confidence, consideration should be given to use of a deliberative forum, such as a parliamentary committee, to take community input on and review nominations made, and to make proposals for future designations ...(Productivity Commission draft report, p36)

The Archives supports an approach of identifying and selecting data that builds in community input.

The Archives has practical experience of identifying and selecting data from business systems through its engagement with Australian Government agencies. The Archives considers that some degree of automation is required to identify Government information and data from business systems for multiple purposes, such as sharing, destruction and long-term preservation. An existing

Archives' project is focusing on automated identification of information and data value through machine-learning processes, sentiment analysis, semantic and business analytics, full text indexing and predictive coding.

The Inquiry note the Archives' project on automating the identification of information and data value through machine-learning, and the limitations of a committee approach to identifying Government information and data in today's digital environment.

Through its Digital Continuity 2020 Policy, the Archives issued implementation targets for agencies to identify all information assets and to identify measures to manage those assets appropriately. The Policy prepares Government agencies with the ability to publish data registers in line with the Commission's recommendation 3.1.

The Archives recommends that if the publication of data registers is implemented, the responsible entity aligns with the Digital Continuity 2020 requirements – so there is no duplication of effort, and the Inquiry notes that the Archives' role is long-term preservation and access to datasets of enduring value.

Procurement and design

The Archives agrees with comments in the Commission's draft report (p511) that further regulation in system procurement is required to include information and data management specifications to enable interoperability.

Interoperability is central to the Archives' Digital Continuity 2020 Policy, and it expects agencies to work towards meeting minimum metadata standards and requirements for business systems (through the Archives' *Business System Assessment Framework*).

The Archives is working in partnership with other Government agencies on ways to increase semantic interoperability, with the aim to re-use and share data through common metadata and language.

The Archives recommends further regulation in system procurement for information and data management specifications (in line with the Archives' Digital Continuity 2020 Policy) to enable interoperability.

Broadening access to national interest data assets

As the custodian of high value Australian government information and data, the National Archives' experience in the management and provision of access to government information and records is second to none. The Archives is a trusted custodian, preserving and providing online access to records of national significance. The Archives' role in releasing records for public access is largely that envisaged by the Commission for an 'Accredited Release Authority'. The Archives would be happy to share its knowledge and experience in managing and providing access to records of national significance.

The intent of the *Archives Act 1983* is very similar to that of the proposed Data Sharing and Release Act. Under the Archives Act all records subject to the Act, which includes digital records and datasets, are to be publicly released once they reach the open access period unless they contain information that falls within one of the exemption categories in section 33 of the Archives Act. The Act requires the Archives to balance public access rights, openness of Government, privacy, security and national interest. It would appear that the proposed Data Sharing and Release Act duplicates responsibilities currently undertaken by the Archives.

The Archives acknowledges that the Archives Act was drafted in a non-digital world and is undertaking legislative reform to ensure that the Act meets the challenges of the digital age.

The Archives Act does not restrict Commonwealth agencies from making data available as they determine is appropriate and in accordance with other legislative requirements such as the *Freedom of Information Act 1982* and *Privacy Act 1988*. The report notes that the proposed Data Sharing and Release legislation would recognise similar sensitivities.

Inquiry's request for further information - APIs

In response to the Inquiry's request for further information on implementing APIs (p250), lessons learned from the Archives' *Discovering Anzacs* API experience include:

- The importance of a clear business model to ensure authority of source data and clarity for data re-use and citation/attribution.
- A clear understanding of the public need prior to developing APIs and determining access to limit requests for modification of API calls - for example.
- Plan upfront for set up and ongoing costs of maintaining the API based on understanding demand.
- Data reliability and quality is crucial for successful API implementation.
- Consider what additional information the end users will need to understand the context of the data collection and any relationships within the data.

About the National Archives of Australia

The Archives is an Executive Agency in the Attorney-General's portfolio and was established under the *Archives Act 1983*. Its role is to promote the creation, management and preservation of authentic, reliable and usable Commonwealth records and facilitate Australians' access to the archival resources of the Commonwealth.

The definition of 'Commonwealth records' in the Archives Act covers all information in digital and non-digital formats that is created, used or received as part of Government business. As a result all data, including email, social media content, datasets, geospatial information, audio-visual and business-related information in any other format constitutes the records of the Commonwealth.

The Archives is a trusted source of public sector information and is internationally recognised for its leadership in providing access to its archival resources and developing digital information management capability in agencies.

The Archives' [Digital Continuity 2020 Policy](#) centres on digital information governance and complements the Australian Government's digital transformation agenda and the digital economy. Good information governance establishes trust in the quality and integrity of information and data, and is the mechanism for Government agencies to integrate standardised pro-disclosure processes into information life cycles.

The policy is based on three principles, with targets that agencies should reach by 2020, as well as pathways with interim targets. Its governance and interoperability related targets directly impact on public sector data availability, use and re-use by integrating robust digital information management into all technology, systems and processes.

Standards and other guidelines produced by the Archives provide a key resource for agencies in sharing information effectively. Through membership of Standards Australia, the Archives is an active participant in the development, endorsement and use of national and international standards related to information management. These include ISO 16175 *Functional requirements for records in electronic office environments* and AS 5044 *AGLS Metadata Standard to improve the accessibility and discoverability of online resources through search facilities*.

The Archives participates in policy and standards development, including through the Council of Australasian Archives and Records Authorities and its working group Australasian Digital Recordkeeping Initiative.