Creative Commons Australia and the Australasian Open Access Strategy Group Response to the Productivity Commission Draft Report Data Availability and Use

Prepared by Dr Virginia Barbour (AOASG) and Jessica Stevens (CCAU), December 12, 2016.

Summary

Creative Commons Australia ('CCAU') and the Australasian Open Access Strategy Group ('AOASG') welcome the opportunity to comment on the Productivity Commission's Draft Report on Data Availability and Use ('Draft Report'). 'Creative Commons is an international non-profit organisation that provides free licences and tools that copyright owners can use to allow others to share, reuse and remix their material, legally'.[1] CCAU is an affiliate that supports Creative Commons in Australia.[2] The AOASG is a non-profit organisation which aims to advocate, collaborate, raise awareness and lead and build capacity with respect to open access for all the outputs of scholarship in Australia and New Zealand.[3]

CCAU and AOASG support the implementation of policies to increase availability and use of data. We aim to contribute to the discussion regarding consumer rights, specifically, the right to access. As noted by the Commission in their Draft Report, the legal and policy frameworks under which data (both private and public) is collected and shared and accessed in Australia is not as progressive as other parts of the world, for example the European Union's Open Data strategy as a core part of the Digital Single Market.[4] Australia's inaction in this 'global movement' may have a detrimental effect on innovation and research outputs.[5]

The Draft Report proposes a 'fundamental change' to the 'legal and policy frameworks under which public and private sector data is collected'.[6] This proposed fundamental change is timely, sensible and would better align Australia's data practices with those of other international jurisdictions. CCAU and AOASG support the findings and draft recommendations, in particular those contained in Chapter 3, 'Public Sector and Research Data Collection and Access', Chapter 6, 'Making Data Useful, Chapter 8, 'Options for Comprehensive Reform' and Chapter 9, 'A framework'. The recommendations highlight a number of factors, including the importance of public interest with respect to facilitating access to publically funded data and information. There is a clear need to align the Australian legal framework and policies with respect to data availability and use with the best practices and norms of other international jurisdictions. The advantages for Australia to be gained through alignment of legal and policy frameworks for data availability will include the facilitation of sharing of data between jurisdictions.

---

[1] Creative Commons, https://creativecommons.org/; Creative Commons Australia and Organization for Transformative Works Submission to the Australian Government's Online Copyright Infringement Discussion Paper, 5 September 2014, http://eprints.qut.edu.au/78481/1/OnlineCopyrightInfringementCreativeCommonsAustraliaAndOrganizationForTransformativeWorks.pdf.

[2] The views expressed here are those of the Australian affiliate, and are not endorsed by Creative Commons Corporation in the US.

[3] The Australasian Open Access Strategy Group, https://aoasg.org.au/.

[4] European Commission https://ec.europa.eu/digital-single-market/en/open-data

[5] Productivity Commission, Draft report, 'Data Availability and Use', November 2016, 12

[6] Productivity Commission, Draft report, 'Data Availability and Use', November 2016, 12.

Specific Comments on chapters

Overview

There are terms in the report that are not consistently defined such as the use of the term 'open access.[7] We would urge that terms should be specifically defined and used consistently. It is essential to differentiate between 'free access and 'open access'. 'Free access' only denotes there is no cost to the reader. Open access includes the application of an appropriate license such as those endorsed by CCAU and Australian Government's Open Access and Licensing Framework (AusGOAL),[8] secure archiving, as well as free access.

Chapter 3 Public Sector and Research Data Collection and Access

CCAU and the AOASG support the recommendations to implement data registers.[9] We believe that public sector data should be 'open by default'.[10] The implementation of data registers would assist in the aggregation and curation of data. The recommendation that data should be released as a first priority and that the register would provide information with respect to any data sets that are not publically available would provide more transparency in research findings, funding distribution and overall, the implementation of registers for data aggregation would be of significant benefit for Australian consumers. Such registers would increase the visibility and discoverability of research data. This recommendation highlights the importance of metadata in the discoverability process, specifically in the description of datasets.[11] Poor quality metadata is one of the key reasons data is functionally unavailable. There is an urgent need for consistency in the application of standard metadata to datasets. We support the recommendation to implement data registers and that these registers publish up to date lists of data.[12]

As noted in the Draft Report, arrangements for sharing and releasing research data in Australia are under review.[13] There are a number of ongoing investigations into open data and open research. As noted in this Draft Report, the Australian Government recently released its Draft Report on Research Infrastructure which noted the importance of discoverability of research data and proposes, in accordance with emerging standards, that research should be, 'Findable, Accessible, Interoperable and Reusable' (F.A.I.R.).[14] Another group noted in the Draft Report, 'The Open Access Working Group',[15] is also investigating the issues pertaining to open access and research data. In addition to these, we also note a group convened by Universities Australia and the Council of Australian University Librarians (of which AOASG is a member) is advocating for an approach to access to research outputs more widely under the F.A.I.R. principles.[16] We believe that the recommendations of the Productivity Commission's Draft Report, to make research data more openly available and to implement registers to aggregate the data, would align closely with the objectives of these other interested groups and ultimately would be beneficial for the Australian research culture. As noted by

---

[7] Productivity Commission, Draft report, 'Data Availability and Use', November 2016, 25.

[8] The Australian Government's Open Access Licensing Framework, http://www.ausgoal.gov.au/.

[9] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, Draft Recommendation 3.1.

[10] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 96.

[11] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 140.

[12] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, Draft Recommendation 3.2.

[13] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 136.

[14] Australian Government, National Research Infrastructure Roadmap, 'National Research Infrastructure Capability' Issues Paper, July 2016, 48 <http://www.chiefscientist.gov.au/wp-content/uploads/20160716-NRIR-Capability-Issues-Paper-16-July-version-proposed-final....pdf>.

[15] Which involves the Department of Education and Training, the Department of Industry, Innovation and Science, the Department of Health, the Australian Research Council, and the National Health and Medical Research Council.

[16] Force 11, The Future of Research Communication and e-Scholarship, 'The FAIR Data Principles', <https://www.force11.org/group/fairgroup/fairprinciples>.

the Commission, increasing access to research data is consistent with recent international academic developments.[17]

There has been significant public policy and investment represented by the Australian National Data Service (ANDS) in the past 6 years.[18] This investment, to support data and its responsible curation and identification as a major element of research infrastructure in Australia has been world-leading. This report complements this infrastructure initiative through the development of stronger policy and legal frameworks.

Whilst the Draft Report mentions the Australian Research Council (ARC) figure for research outputs, we note that this figure should be taken as a very provisional figure, given the current difficulties of collecting such data consistently.[19]

We further agree with the Productivity Commission that the recommendations made in the recent Intellectual Property Arrangements Inquiry, Draft Report, with respect to making publications from publicly-funded research available on an open access basis after one year, should be extended to the 'underlying data'.[20] We would strongly suggest that the one year limit should be an absolute maximum and in general, immediate open access should be the standard.

Chapter 6 Making Data Useful

We support the recommendation for Government agencies to adopt and implement data management standards to support increased data availability.[21] We would further add that this recommendation and its careful implementation is likely to have the most immediate and far reaching effect of the majority of the recommendations.

We note that a concern of the Commission with respect to improving the useability of public data is implementing consistent metadata and metadata standards and the potential upfront costs on initial data custodians.[22] We agree that the curation and aggregation of data comes with associated financial and implementation cost. Data management plans and pre-planning data management strategies are central to minimising the costs associated with data management that have to be borne by custodians. We note that generally, the costs of such management are repaid many times over in the extra value that can be extracted from well curated datasets compared with poorly curated datasets. Data management as part of the lifecycle of research overall is an area which requires more consideration and resourcing, especially within academic institutions.

With respect to the 'standardisation and curation of data in the research sector',[23] we would note that although a number of journals do have data sharing policies, they are solely aimed at availability of data associated with specific publications, and not at the wider goal of good data management practices. We note that access to data is just one aspect of such practices. We would strongly advise against any data availability policy or process that directs access to data via publisher sites as an appropriate option. It should be noted that the aims of publishers operate separately from the needs of researchers, institutions or indeed the national interests of a country and the greater good of society globally. Providing access to important datasets at publishers' sites risks replicating the same

---

[17] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 138.
[18] Australian National Data Service http://www.ands.org.au/
[19] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 138.
[20] Productivity Commission, Draft Report, 'Intellectual Property Arrangements' April 2016.
[21] Productivity Commission, Draft Report, 'Intellectual Property Arrangements' April 2016, Draft Recommendation 6.1.
[22] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 243.
[23] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 250.

situation we have now for many research papers, which, despite more than 15 years of global advocacy for public access, remain behind publishers' paywalls.

By contrast, open access repositories at Australian academic and research institutions, and regional and national sites for data storage such as Nectar Cloud,[24] as well as open access discipline-specific repositories (of which many now exist) are appropriate sites. We would urge that all research institutions should be supported to develop robust data management strategies of which the provision of repositories should form one part. Furthermore, given the rapidly evolving landscape of options for data curation and storage, we would re-emphasise the need for ongoing training in this area, and encouragement for institutions to develop training frameworks to support the provision of scaffolded data management training to their research communities.

Chapter 8. Options for Comprehensive Reform

We welcome the overarching views taken in this section.[25] We agree that there is currently 'no shared vision amongst public sector data holders in Australia on how to consistently deliver widespread data sharing and release'.[26] We support the finding and the view that there is now an opportunity to develop this shared vision.

We note the Commission is seeking views with respect to the curation of the data, whether it should occur by the original data custodian or whether '…giving the release authority the ability to curate the data (aggregated model) could provide it with a secondary revenue source to help support and retain its capability'.[27] If the data was available on an open access basis, there would be no reason why the data couldn't be further curated to make it more useful by the release authority.

With respect to research data, we would strongly support the federated model, guided by a well-defined set of policies and standards, including specific standards for metadata. This area of policies and standards in data management is a further area where ANDS has provided important leadership.[28] We further agree that there is a clear need for a designated agency to oversee the policy considerations.[29] Such an agency would be important to ensure accountability for progress and outcomes and further, this designated agency would be a champion with respect to encouraging the necessary cultural change in various sectors.

We would note that there are a number of organisations, with overlapping aims advocating for change towards more openness in research outputs and data. In addition to ANDS, organisations include ourselves; AOASG, which represents a number of Australian Universities, and Creative Commons Australia, as well as Open Data Institute Queensland, (ODIQ').[30] However, we agree that more extensive advocating and championing '…this policy of greater openness' would be beneficial. [31]

With respect to reforms to open up re-use of research data, specifically, the proposed conditions on funding, we would support incentivising institutions and also specific researchers by providing benefits to those that share data.[32] Part of this opening up of data would require the adoption of standards that

[24] Nectar Cloud https://nectar.org.au/research-cloud/
[25] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, Draft Findings 8.1, 8.2 and 8.3.
[26] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, Draft Finding 8.2, 317
[27] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, Draft, 325.
[28] Australian National Data Service http://www.ands.org.au/working-with-data
[29] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 328.
[30] Open Data Institute Queensland, (ODIQ), <http://queensland.theodi.org/>.
[31] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 329.
[32] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 329.

ensure data sharing and reuse was properly tracked, for example by the adoption of standards for citing datasets.

With respect to building on existing journal publication requirements,[33] as noted above, we would not recommend that data sharing in research be led by journal requirements, but instead should be a more comprehensive approach that is driven by data management polices from academic and research institutions. This would also allow an Australian-led approach as opposed to one led from publishers based overseas.

The Productivity Commission asserts that, '[m]aking data available for reuse can be a resource intensive process that requires specific skills and experience. However, the amount of resources required can also be exaggerated'.[34] Our experience is that these costs must not be underestimated. There are substantial costs (especially of time) associated with making even small datasets functionally available in a way that ensures the data are can be properly scrutinised and reused. Especially important is that data are tagged with the right metadata. We believe these costs are outweighed by the benefits that can be derived from well-curated data, but these costs need to be built into projects, ideally at their inception. To do so systematically requires a wholesale approach of policies, standards and incentives and of training for those that generate and curate data. Apart from programs led by ANDS, there are few if any, systematic training programs in place. We believe that this is an area that requires more attention in this Inquiry.

The Commission asserts that there is significant room for improvement in sharing of research in the research sector.[35] The Draft Report notes that, 'Researchers often complain about the lack of openness of public sector data, but their sector remains vastly behind the public sector in terms of openness and availability of data — the pot calling the kettle black, as it were'.[36] We agree that there is room for improvement within the research sector, however, we note that there a number of reasons this progress is restricted. First, there is a noted absence of incentives for researchers with respect to data sharing. Currently there are strong disincentives to share, which is predominantly a result of the research publishing system. Second, there is limited expertise with respect to data management and limited training available. Third, there is limited infrastructure, systems and processes to support researchers to make research data openly available.

Chapter 9 A Framework for Australia's Data Future

CCAU and the AOASG supports the implementation of National Interest Datasets (NIDs) with a default position of immediate release of these datasets unless classified as sensitive. Further, we agree that sensitive data that is able to be de-identified should be done so and publically released within a minimal period of time. Whilst there are some instances in which it may not be appropriate to release datasets, we argue that such situations should be restricted to a very limited set of situations. We further agree that community participation and input with respect to NIDs is central to ensuring transparency in the decision making process. We note that the scope of what may be classified as a NID is unclear.

We support the draft recommendation to establish an Office of the National Data Custodian.[37] A centralised body which oversees the data management policy within Australia will provide stability and certainty with respect to the use of datasets within Australia. We support the draft recommendation to

---

[33] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 330.
[34] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 330.
[35] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 331.
[36] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, 331.
[37] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, Draft Recommendation 9.5.

implement accredited release authorities (ARAs).[38] We believe that ARAs will provide an important safeguard to ensure that the datasets and registers are maintained and up to date. We support the use of ARAs to promote trust and transparency in organisations.

We believe that it is important for unreleased data sets to be accessible for specific purposes such as research undertaken by Universities. As such, we support the recommendation for the National Data Custodian to provide accreditation to trusted users to make certain uses of the datasets.[39] Use of these unreleased datasets would enable research to progress and increase transparency in findings. Whilst we support this recommendation, ultimately our position is that datasets should only remain unreleased in specific circumstances where the data is unable to be de-identified or the data is sensitive.

As noted above, we support initiatives for both researchers and institutions to reward the sharing of data. We support the draft recommendation to prioritise public research funding on the basis of institutions making their research data available.[40] Openly available data promotes good research practices and aligns with objectives of open data such as shareable and useable data. We would note that institutions will need time to implement the necessary change to comply with this recommendation.

CCAU and the AOASG advocate that the default position with respect to data should be released on an open access basis so that data is easily accessible to the public. We thus support the draft recommendation that 'all non-sensitive public sector data should be released'.[41] We believe that the introduction of an Act which promotes the release of data, especially data in the public interest and data that is publically funded is in the public interest.[42] Further, legislation which improves the rights to access data by individuals and institutions would be of significant benefit to the Australian community.

Conclusion

Overall, CCAU and the AOASG support the draft recommendations and findings contained in the Draft Report. The recommendations highlight the necessity for a fundamental change in Australian legal policy and framework with respect to the management, release and availability of data. The Draft Report acknowledges the steps required to address Australia's current deficiency in the area of data management and the importance of better aligning Australia's practices with those of other jurisdictions. We would, however, note that there is a specific and urgent need to address the lack of training for those that generate and curate datasets.

---

[38] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, Draft Recommendation, 9.6
[39] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, Draft Recommendation, 9.7 and 9.8.
[40] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, Draft Recommendation, 9.9.
[41] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, Draft Recommendation, 9.10.
[42] Productivity Commission, Draft Report, 'Data Availability and Use', November 2016, Draft Recommendation, 9.11.