PRODUCTIVITY COMMISSION INQUIRY INTO DATA AVAILABILITY AND USE
AUSTRALIAN RESEARCH COUNCIL SUBMISSION

Introduction

The Australian Research Council (ARC) welcomes the opportunity to provide a submission to the Productivity Commission (the Commission) as part of its inquiry into data availability and use. This submission provides information on the ARC's approach to data in order to inform the Commission's consideration of options for increased availability and improved use of data.

This submission addresses the availability and use of data in two areas:

1. public sector data[1] collected and stored by the ARC, including National Competitive Grants Programme (NCGP) and Excellence in Research for Australia (ERA) administrative and programme data, as well as data on the Australian university research system gathered through ERA

2. research data arising from ARC funded research projects under the NCGP, which are not collected and stored by the ARC.

The ARC is a Commonwealth entity that provides advice to the Government on research matters and administers the NCGP—a significant component of Australia's investment in research and development—and ERA. The ARC's mission is to deliver policy and programmes that advance Australian research and innovation globally and benefit the community. In seeking to achieve its mission, the ARC supports the highest quality fundamental and applied research and research training through national competition across all disciplines.

The ARC provides approximately $800 million in research funding through the NCGP each year. While significant, this represents only around ten per cent of Australian Government investment in research and development. The Commission should therefore consider the issues addressed in this submission within the broader context of public sector data related to funded research activity across government and the research data generated through that activity.

The ARC notes the potential benefits of increased data availability and use identified by the Commission, including efficiency, empowerment of consumers, competition, innovation and accountability of governments.[2] In addition, the knowledge produced through publicly funded research provides further important benefits to researchers and the broader community. Specifically, the ARC considers that open access to research data provides three types of benefits—societal (informing better social, economic and other outcomes), research (contributing to the growth of science and research) and individual (flowing from the use and citation of researchers' data outputs)—and it is committed to maximising these benefits.

Public sector data collected and stored by the ARC

Through its implementation of the NCGP and ERA, the ARC collects and stores a range of administrative and programme data, as well as data on the Australian university research system. As

---

[1] 'Public sector data' defined in Productivity Commission (2016), *Issues Paper – Data Availability and Use*, p. 3.
[2] Productivity Commission (2016), *Issues Paper – Data Availability and Use*, p. 9.

the Commission's issues paper points out, such data can constitute a resource with the potential for use by businesses, researchers, academics and the broader community.[3]

A substantial volume of NCGP data collected by the ARC is made publicly available, including:

- the ARC Grants Dataset (http://www.arc.gov.au/grants-dataset), which provides information about the grants funded by the ARC, including a descriptions of the research, investigators and amount of money provided

- selected data provided through NCGP funding scheme outcomes (http://www.arc.gov.au/outcomes) and funding announcements (http://www.arc.gov.au/rms-funding-announcements-web-page).

ERA, Australia's national research evaluation framework, involves the comprehensive collection of data and information on all research activity in Australian higher education institutions. These data inform detailed measurement of the quality of research produced in Australian universities against national and international benchmarks.

Reports on the three ERA rounds that have been completed, most recently in 2015, are publicly available (http://www.arc.gov.au/era-reports). These reports provide data and information on ERA methodology and implementation, as well as substantive results regarding the state of Australian university research. The ARC plans to produce additional publicly available reports based on analysis of the longitudinal data collected over three ERA rounds, in order to provide further insight into the state of Australian university research, including in areas such as gender, open access and changes in ERA ratings.

The availability and use of these data offer multiple benefits. They provide a mechanism to ensure the ARC remains accountable and transparent in its implementation and delivery of the NCGP and ERA, as well as offering the opportunity for government and the public to monitor the ARC's performance, both in terms of public expenditure and the nature of the research it funds.

By illuminating areas of strength and areas for further development in Australian university research, ERA results provide a robust evidence base for government policymaking in relation to science and research as well as informing the research and investment strategies of higher education institutions.

Public sector data held by the ARC are also valuable for their ability to be used by Australian and international businesses, researchers and academics to help identify partners for collaboration, thus supporting opportunities for innovation and further growth in Australian research and science.

Limits on the availability of public sector data

The public sector data collected and stored by the ARC involve a degree of sensitivity, including privacy concerns related to personal data, and there are some limits on their availability and use.

The ARC collects a large volume of public sector data through NCGP application processes. Applicants provide information in grant proposals to the ARC for a specific purpose—to seek funding to support the conduct of research—and on the understanding of confidentiality, as outlined in ARC

---

[3] Productivity Commission (2016), *Issues Paper – Data Availability and Use*, p. 10.

funding rules.[4] The funding rules also outline circumstances in which information contained in proposals may be disclosed by the ARC, such as for the purposes of assessing or verifying proposals, management of NCGP schemes, research integrity obligations, ministerial advice or to publicise and report offers and awards of funding.

The ARC does not release the full detail of successful research proposals, or the details of people's careers, which may be private. It also does not publicly release any information about unsuccessful proposals (beyond the total number of unsuccessful proposals). Within these limits, and in accordance with its responsibilities under legislation and government policy, the ARC considers that it currently makes public most of the proposal level NCGP data that it is able to.

While the ERA reports contain extensive detail and provide a range of selected data, not all of the underlying ERA data are publicly available. All of the tabulated data from the ERA 2015 national report are available on the ARC website in Excel files (http://www.arc.gov.au/era-reports), but the underlying data submitted by each university are not currently available. To date, institutional submission data for ERA have been provided strictly in confidence, but there is potential for this to change in future rounds.

Following the release of each ERA report, individual institutions are given access to SEER (the System to Evaluate Excellence in Research), which is used by the ARC to manage and implement ERA. This provides institutions with an opportunity to access their submitted data as displayed in SEER, as well as benchmark data. In addition, all benchmark data and all metrics displayed in SEER are provided to each institution in a flat file that is specific to their individual submission.

Where data from ERA are not made publicly available, there are two main reasons. First, staff and output data are not made available due to privacy considerations and the potential for an ERA result to be attributed to an individual (or small group of individuals), and second, citation benchmark data cannot be made publicly available due to licensing arrangements with the ERA citation provider.

Facilitating improved availability and use of public sector data

To maximise the potential benefits from the availability and use of the public sector data it holds, the ARC continually strives to better equip the public and other stakeholders to extract and use those data themselves through enhanced functionality of its systems. The ARC is currently investigating options to make NCGP data more easily accessible, including through possible changes to the operation of the Grants Search Tool and the timing of data releases.

The ERA datasets are considered to be particularly valuable by a variety of government, research, university and business stakeholders. Increasing the ability of these stakeholders to access and use the extremely rich data collected as part of ERA would facilitate independent, tailored analysis of particular issues to meet these stakeholders' needs. Further, the ongoing development of an assessment of engagement and impact in university research as a companion to ERA, an initiative under the National Innovation and Science Agenda, will involve the collection of data that can be used to provide greater insights on a range of issues relating to the translation of research to support innovation.

---

[4] For example, see *Funding Rules for schemes under the Discovery Programme (2015 edition)*, section A11.3, http://www.arc.gov.au/sites/default/files/filedepot/Public/NCGP/DP17/Discovery_Programme_Funding_Rules_2015.docx.

The ARC is considering options to make all ERA 2018 submission data (excluding staff data) publicly available. The same consideration is being given to data submitted for the first full round of the engagement and impact assessment process in 2018. Both privacy and commercial-in-confidence concerns will be taken into account in any process to make these data available to the public.

While the ARC is committed to improving the availability and use of the public sector data that it holds, it must also ensure that, as the Commission states, "greater access to data does not exceed the costs of such greater access, including the potential loss of trust and privacy of individuals".[5] In light of the limits discussed above, it is essential that the ARC appropriately considers the privacy of those who may be identified in NCGP and ERA data. In increasing access, the ARC must also consider appropriate and effective ways to format, describe and present the data, including developing relevant rules and guidance to support accurate extraction and use of the data and minimise the chance of incorrect or misleading use of the data that may have negative consequences for the stakeholders using them.

These issues must be carefully addressed in detail to avoid a loss of trust in the ARC's processes and public investment in them, and to ensure the value of the data is not undermined. Sufficient time and resources will therefore be required for the ARC to increase the availability of the public sector data it holds in a way that maximises the benefits the data offer.

Research data arising from ARC funded research

Research data involve a range of considerations additional to the public and private sector data issues discussed in the Commission's issues paper. In addressing the costs and benefits of options to increase the availability and use of data, the Commission should consider taking into account issues that relate specifically to making research data openly accessible.

The Australian Government Public Data Policy Statement includes a commitment to, "where possible, ensure non-sensitive publicly funded research data is made open for use and reuse".[6] This commitment is consistent with a broad international trend toward open data, which is increasingly supported by both research funders and academic publishers (such as Elsevier,[7] Wiley[8] and others). This trend is typically underpinned by the objectives of maximising the return on public investment in research and, recognising that the value of research data is not a simple monetary argument,[9] increasing scientific and other research benefits for the research community and the public.

The ARC considers that open access to research data provides three types of benefits.

1. *Societal* – publicly-funded research data are a public good; their availability and use support private and public sector innovation and better informed, evidence based policymaking, which contribute to improved social, economic and other outcomes that benefit the public.

2. *Research* – open access to research data contributes to the development and growth of science and research; the availability and use of data reinforce open inquiry and allow the

---

[5] Productivity Commission (2016), *Issues Paper – Data Availability and Use*, p. 21.
[6] Australian Government (2015), *Australian Government Public Data Policy Statement*.
[7] See https://www.elsevier.com/about/open-science/research-data.
[8] See http://olabout.wiley.com/WileyCDA/Section/id-828082.html.
[9] Pryor, G (2012) 'Why manage research data?', in Pryor, G (ed.) *Managing research data*, p. 4, see http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.447.9990&rep=rep1&type=pdf.

research community to analyse, verify, reproduce and build upon findings and methodologies.

3. *Individual* – making research data openly available can provide benefits to individual researchers, including career and reputational benefits that flow from the use and citation of research data outputs.

The ARC is committed to maximising these benefits and, since 2007, has encouraged researchers to deposit data arising from ARC funded research projects in publicly accessible repositories. However, recognising the challenges involved in addressing the rapid growth and significant changes occurring in relation to data, the ARC has not mandated open access to research data.

Since February 2014, the ARC has required researchers to outline how they plan to manage data arising from their research.[10] Effective data management is an important part of ensuring open access to publicly funded research data and helps to outline how data will be collected, formatted, described, stored and shared throughout, and beyond, the project lifecycle.

The ARC's requirement is consistent with the responsibilities outlined in the *Australian Code for the Responsible Conduct of Research* (2007) (the Code), which include the proper management of research data and primary materials by researchers, along with institutional policies addressing data ownership, storage, retention and "appropriate access…by the research community".[11] This includes the need to appropriately consider a range of well-established factors that may limit the availability of research data, including research ethics, privacy, intellectual property and other legal constraints.

Challenges for the availability and use of research data

Open access to research data remains a dynamic and rapidly evolving issue. As the Commission points out, data volumes have been growing exponentially,[12] and in the international research landscape this has led to new models of science and research and the emergence of new fields of study.[13] International and domestic debate reveals the complexities and challenges involved in supporting open data in this environment, as well as the varied approaches to doing so.

Options to increase the availability and use of data arising from publicly funded research should include consideration of the many types of data produced and the different standards and approaches to data management, analysis, archiving, storing, sharing and publication between (and within) disciplines—there is no 'one-size-fits-all' approach.[14]

When available data are used, there is a lack of common standards and consistency in data citation practices and, furthermore, the role of data citations may differ from that of traditional research publication citations, including in terms of the acknowledgement and recognition they offer

---

[10] For additional information see http://www.arc.gov.au/research-data-management.

[11] *Australian Code for the Responsible Conduct of Research* (2007), Section 2.

[12] Productivity Commission (2016), *Issues Paper – Data Availability and Use*, p. 4.

[13] Borgman, C (2012), 'The Conundrum of Sharing Research Data', *Journal of the American Society for Information Science and Technology*, 63(6): 1059, http://onlinelibrary.wiley.com/doi/10.1002/asi.22634/epdf.

[14] Lämmerhirt, D (2016), *PASTEUR40A Briefing Paper: Disciplinary differences in opening research data*, p. 1, http://www.pasteur4oa.eu/sites/pasteur4oa/files/resource/Brief_Disciplinary%20differences%20in%20opening%20research%20data%20APS_MP_FINAL1.pdf.

researchers.[15] These unsettled issues can complicate efforts to maximise the benefits of open access to research data described above.

Furthermore, while open data is connected to, and commonly seen as a part of the broader concepts of open access and open science,[16] it involves distinct considerations. Accordingly, in international practice, open data is most commonly (but not always) addressed separately, including in policies, guidelines and requirements with respect to data management and data sharing, and especially in establishing principles on open data. Important international examples include:

- *OECD Principles and Guidelines for Access to Research Data from Public Funding* (2007)[17]

- Research Councils UK (RCUK) *Common Principles on Data Policy* (2011, revised 2015)[18]

- Canadian *Tri-Agency Statement of Principles on Digital Data Management* (2016)[19]

- European Commission *Guidelines on Access to Scientific Publications and Research Data in Horizon 2020* (2013),[20] which are complemented by *Guidelines on Data Management in Horizon 2020* (2013).[21]

Measures to support open data, data management and data sharing have also been implemented by individual research funding agencies in the US (the National Science Foundation and National Institutes of Health), the UK (individual Research Councils under RCUK, Cancer Research UK and the Wellcome Trust) and Canada (Canadian Institute of Health Research and Social Sciences and Humanities Research Council of Canada). Each has varying requirements, such as submission of a full data management plan as part of an application, having a data management plan in place for the duration of a research project, or retaining research datasets after the completion of a grant.

In Australia, the National Health and Medical Research Council (NHMRC) does not currently require data management plans. However, it has released a *Statement on Data Sharing*, which, similar to the ARC's position, encourages data sharing and the provision of access to data.[22] The NHMRC has also released *Principles for Accessing and Using Publicly Funded Data for Health Research*,[23] which focus specifically on health related data. Australia does not currently have a clear set of principles like those developed by the OECD, UK and Canada regarding research data in all disciplines. A clear set of principles for Australia would provide benefits that include reinforcing open, transparent and robust academic enquiry, enabling re-use, and encouraging the combination of datasets from multiple sources. An emphasis on sharing of data, guided by a set of national principles, also has the potential to stimulate new approaches to collection, analysis, validation and management of data.

---

[15] Robinson-Garcia, N, Jiménez-Contreras, E and Torres-Salinas, D (2015), 'Analyzing data citation practices according to the Data Citation Index',
https://www.researchgate.net/publication/271447861_Analyzing_data_citation_practices_using_the_Data_Citation_Index.

[16] Science International (2015), *Open Data in a Big Data World*, p. 8, http://www.icsu.org/science-international/accord/open-data-in-a-big-data-world-long.

[17] http://www.oecd.org/sti/sci-tech/38500813.pdf.

[18] http://www.rcuk.ac.uk/research/datapolicy/.

[19] http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1.

[20] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

[21] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[22] https://www.nhmrc.gov.au/grants-funding/policy/nhmrc-statement-data-sharing.

[23] https://www.nhmrc.gov.au/principles-accessing-and-using-publicly-funded-data-health-research.

<u>Facilitating improved availability and use of research data</u>

The ARC's position is that open access to data arising from publicly funded research should continue to be strongly encouraged. In light of the existing challenges and significant changes in the research data landscape, however, the ARC would not currently propose to mandate open data to increase the availability and use of research data unless significant new funding was made available to the research sector. Further engagement with the sector would be beneficial to first develop a better understanding of the costs that would be involved in enabling compliance with such a mandate.

As the Commission points out, increasing the availability of data requires resources, including skilled staff, to ensure sufficient quality of data for release and consistency with evolving standards in storage and security.[24] Furthermore, in the context of research data, it is also essential to account for the variation between disciplinary approaches if effective infrastructure and mechanisms for curation and storage of research data and metadata are to be established, and common or shared data standards and approaches that facilitate data use are to be developed.

To support continued movement toward open data without imposing a mandatory requirement, the ARC suggests that the Commission may wish to consider including in its recommendations the options of developing principles on open access to research data and establishing appropriate research data repositories.

1. Principles on open access to research data

The development of national principles—carefully constructed to be flexible enough to address the range of complex and dynamic issues related to research data—would signal a national commitment to improve the availability and use of research data, help to provide greater clarity at a high level on priorities and issues in open data, and reinforce the momentum towards open data in Australia's research culture.

As outlined above, there are useful international examples that Australia could draw from in developing such principles. In particular, the ARC notes that the *OECD Principles and Guidelines for Access to Research Data from Public Funding* also provide guidance on the management of data and primary materials. As an OECD member, Australia is expected (not legally bound) to implement these principles and guidelines.

The ARC notes that relevant work is currently underway at the whole-of-government level, in line with the *Boosting the Commercial Returns from Research* agenda for action[25] and the *National Innovation and Science Agenda* initiative for "Promoting innovation through publishing and sharing public data",[26] as implemented through the *Australian Government Public Data Policy Statement*, to consider a national approach to open access to the outputs of publicly funded research. The ARC is actively contributing to this work and is well placed to ensure that its policies and practices are consistent with any national approach that is developed.

2. Research data repositories

As noted above, appropriate infrastructure and mechanisms are required to improve the availability and use of research data. With an increasing focus on data intensive research, there are growing

---

[24] Productivity Commission (2016), *Issues Paper – Data Availability and Use*, p. 21.

[25] https://www.education.gov.au/boosting-commercial-returns-research

[26] http://www.innovation.gov.au/system/files/case-study/Factsheet%2026%20-%20Promoting%20innovation%20through%20publishing%20and%20sharing%20public%20data.pdf. Also see http://www.innovation.gov.au/page/data-sharing-innovation

arguments that research funders and national infrastructure providers have an obligation to support sustainable research data repositories.[27] Substantial work has been undertaken in this regard, for example in the UK, with government funding provided to national research data capabilities such as the UK Data Service,[28] Natural Environmental Research Council data centres[29] and Archaeology Data Service.[30] The ARC suggests that the Commission could consider including in its recommendations the option of establishing and supporting national research data repositories in Australia.

Such repositories should be tailored to address the needs of different disciplines (and support use across disciplines). They should also be sustainable and allow for ongoing developments in data and data management. It is important to acknowledge that sufficient time will be required in establishing repositories to address these challenges and complexities and ensure that the repositories effectively facilitate data availability and use.

If maximising the societal, research and individual benefits that flow from increased availability and use of publicly funded research data is a priority for government, it will need to consider and identify appropriate funding sources to meet the costs of establishing and supporting the repositories.

A report commissioned by the Australian National Data Service in 2014 suggests that while the costs of research data curation and sharing are material, they are small compared to the potential benefits.[31] Extrapolating from studies of data centres in the UK and scaling the findings to the Australian context, the report estimated that the value of efficiency savings, reinvestment of those savings into research and the additional use and reuse of data would overwhelmingly outweigh the cost of repositories.

The ARC would welcome continued engagement with the Commission regarding the development of options to improve the availability and use of public sector data collected and stored by the ARC and publicly funded research data, particularly data arising from ARC funded research.

---

[27] Science International (2015), *Open Data in a Big Data World*, p. 14, http://www.icsu.org/science-international/accord/open-data-in-a-big-data-world-long.

[28] https://www.ukdataservice.ac.uk.

[29] http://www.nerc.ac.uk/research/sites/data.

[30] http://archaeologydataservice.ac.uk.

[31] Houghton, J and Gruen, N (2014), *Open Research Data: Report to the Australian National Data Service (ANDS)*, http://www.ands.org.au/working-with-data/articulating-the-value-of-open-data/open-research-data-report.