16 December 2016

The Productivity Commission
<online submission>

Go8 submission to the Productivity Commission Draft Report on Data Availability and Use

Introduction

The Group of Eight (Go8) welcomes the opportunity to comment on the Productivity Commission (PC) Draft Report on Data Availability and Use.

The Go8 institutions may make individual and more detailed responses to the report and therefore the Go8 response is high-level. This submission focusses on the use of data for research, though the Go8 acknowledges the diverse and many other uses of externally available data by its members, and conversely of its data by others.

The Go8 represents Australia's leading research-intensive universities, accounting for more than two-thirds of Australian university research activity, spending some $6 billion each year on research.

Collectively the Go8 generates a significant proportion of the sector's research data, much of which is made available to others through collaboration, partnerships and publication.

Our researchers correspondingly benefit from the use of public and private data in a range of fields such as health, social sciences, agriculture, environment and climate, financial and economic, and they engage in numerous international research collaborations where Australian data intersects with internationally available datasets.

The Go8 promotes the availability and use of research data in the way it handles and manages the datasets it generates. The Go8 also has a significant role in contributing to making research data available on a national level, through its members' leadership of a range of thematic or disciplinary initiatives that generate, link and share research data – such as the Population Health Research Network (University of Western Australia), the Terrestrial Ecosystem Research Network (University of Queensland) and the social sciences Australian Data Archive (Australian National University).

Underpinning national data or research infrastructure is vital to ensuring research data is more available, re-usable and linked. Collectively the Go8 leads a number of key such infrastructures in close collaboration with the research sector – the data sharing and reuse focused Australian National Data Service (Monash University), the high performance computing National Computational Initiative (Australian National University), the data manipulation National Research Collaboration Tools and Research (University of Melbourne), and the data storage enhancing Research Data Services (University of Queensland).

Key points:

1. The Go8 commends the depth and breadth of the PC's inquiry into data availability and use, and the draft report as a broad ranging, significant and widely impacting study of the use and potential benefits of making public and private data more available.

   a. The Go8 is committed to making research data as widely available where possible – in keeping with the policies and codes under which our researchers and institutions operate, such as the Australian Code for the Responsible Code of Conduct and the policies of the two major funding bodies: the Australian Research Council [1] and the National Health and Medical Research Council[2].

   b. Our approach recognises broader international movements such as that by the OECD[3] and the G8[4] to open up access to publicly funded research data. In part this commitment reflects that the Go8 must embrace open access research practices as a basis for collaborative and successful research partnerships.

2. The Go8 recognises and supports the clear intent to raise the discoverability and accessibility of public and private datasets as a means of driving opportunities from additional yet untapped productive uses of the data. In view of the unknown potential of many of the datasets, the Go8 notes that, to the degree possible and without restricting desirable access to the data, measures should accompany the release of these datasets to ensure that inappropriate, unwanted or criminal use of the data is minimised.

   a. The Go8 recommends that the Office of the Australian Information Commissioner work closely with the university research sector to develop guidance pertaining to recommendation 5.2 in the draft report around enabling access to identifiable information without seeking individuals consent where research is determined to be in the public interest. There are significant ramifications from extending the exceptions that apply to health and medical research to other types of information, in weighing up public interest against the potential ramifications on individuals should their data become known.

---

[1] The ARC strongly encourages the depositing of data arising from a Project in an appropriate publicly accessible subject and/or institutional repository (http://www.arc.gov.au/funding-rules)

[2] In addition to requiring metadata from any publications arising from NHMRC supported research projects to be deposited into an open access institutional repository immediately upon publication, the NHMRC also encourages researchers to share and deposit research data arising from NHMRC supported research projects through an open access database (https://www.nhmrc.gov.au/book/nhmrc-funding-rules-2016/NHMRC-funding-rules-2016)

[3] OECD 2007 Principles and Guidelines for Access to Research Data from Public Funding (http://www.oecd.org/sti/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm)

[4] On 12 June 2013, G8 Science Ministers committed in a Statement to openness in scientific research data and support for a set of principles for open scientific research data (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206801/G8_Science_Meeting_Statement_12_June_2013.pdf)

---

3. The Go8 supports recommendations that will enhance the ability of researchers to undertake their work as a result of greater access to datasets that might not be otherwise known or available to them, to the degree that those recommendations can be workable, and effectively implemented. These include the following:

   a. 2.1 for a system to enable nomination by researchers and others of datasets for public release;
   b. 3.1 requiring Australian Government data registers;
   c. 3.2 on publication of registers of data holdings by publicly funded entities including the Australian Research Council;
   d. 5.5 to support additional qualified entities to be accredited to undertake data linkage;
   e. 6.1 for government agencies to adopt and implement data management standards to support increased data availability and use;
   f. 9.4 to enable nomination and designation of public and private datasets as National Interest Datasets (NIDs) and public or restricted access to these NIDs depending on their nature;
   g. 9.6 to accredit Australian and state/territory government agencies as release authorities by the National Data Custodian;
   h. 9.8 to streamline and expand arrangements for access by trusted users to identifiable data held in the public sector and publicly funded research bodies;
   i. 9.10 release of all non-sensitive public sector data

4. The Go8 recommends that the Australian Government and other key funders of research take careful note of – and undertake steps to quantify and address – the resourcing implications that accrue to universities and researchers in the event that those recommendations requiring additional outlay are adopted.

   The Go8 further recommends that the principle underpinning recommendation 7.4 to provide funding to government agencies to make datasets available should also extend to other publicly funded entities such as universities and research funding bodies to implement the additional costs accruing as a result of making datasets known and/or available. However, this view does not diminish the Go8's recognition of the importance of these recommendations in promoting desirable outcomes for researchers around data availability and use, nor in driving greater access to data collected, generated, or re-purposed into new forms by researchers.

   Relevant recommendations are:
   a. 3.2 on publication of registers of data holdings by publicly funded entities including the Australian Research Council.
   b. 7.2 on an independent review to further ascertain the pricing of public sector datasets to the research community for public interest purposes.
   c. 9.3 recommending public research funding should be prioritised on the basis of progress made by research institutions in making their researchers' data widely available
   d. 9.4 where the nomination of research data sets under the custodianship of universities as National Interest Datasets may have implications for their release and availability.
   e. 9.7 on accreditation of trusted users including all Australian universities by the National Data Custodian, a process whose costs to universities is yet unknown.

5. The Go8 supports in principle recommendations around public research funding being contingent on researchers' data being widely available, and on implementing streamlined practices to facilitate access.

   a. **In practice, the costs accruing to universities to extend their practices above and beyond what is currently required under funding and other provisions would be significant** – including for curation and infrastructure costs. For example, it has been estimated that Australian national research disciplinary data repository costs alone may be in the order of $130 million to $200 million per annum[5].

      This can be balanced against the potential untapped annualised benefits of research data of around $1.4 billion to $4.9 billion[6]. These figures alone may demonstrate a clear argument for making the data available, but also that a percentage of the value that would accrue from the use of the data could be well employed in offsetting the costs of making the data available.

      Relevant recommendations are:

      i. 5.4 recommending funding be prioritised to academic institutions that implement mutual recognition of approvals issued by accredited human research ethics committees
      ii. 9.3 recommending public research funding should be prioritised on the basis of progress made by research institutions in making their researchers' data widely available
      iii. 9.9 to prioritise public research funding on the basis of progress made by research institutions in making their researchers data widely available to other trusted users on conclusion of research projects.

   b. The Go8 also notes that not all researchers' data can and should be made widely available, and would point to the impracticability of doing so where data is of such a significant scale as to make wide access exorbitantly expensive as in the case of data from large astronomical facilities such as the Australian Square Kilometre Array Pathfinder, or where the data is of an ephemeral nature relevant only to the particular research objective at the time of analysis and can be recreated in future.

6. The Go8 recommends that deeper analysis and investigation occur into the implication of the draft report's recommendations on the types and incidence of skills that will required to implement the scale and breadth of activity promoted – and the role of university and other education providers in building such skills in Australia. While the draft report acknowledges that data skills needed are in short supply, it does not attempt to quantify or significantly qualify what these are. Yet it is clear that funding alone will not result in optimal outcomes, and that a concerted strategy is needed to ensure that at a minimum the requisite data scientists and specialists are created and encouraged to thrive.

---

[5] Houghton, J and Gruen, N 2014 *Open Research Data: Report to the Australian National Data Service*, p. v (http://www.ands.org.au/__data/assets/pdf_file/0019/393022/open-research-data-report.pdf)
[6] Ibid, p.v