



AOASG response to Productivity Commission Issues Paper on Data Availability and Use

Prepared by Dr Virginia Barbour, Executive Director, AOASG, on behalf of the AOASG; July 2016
eo@aoasg.org.au

The Australasian Open Access Strategy Group (AOASG)¹ exists to advocate, collaborate, raise awareness and lead and build capacity for open access in Australia and New Zealand. The AOASG is supported by ten Australian and eight New Zealand Institutions.

General comments

The AOASG welcomes the Productivity Commission inquiry into data availability and use. The inquiry is timely both nationally and globally. We limit our response to the general sections and paragraphs 1 and 4 under the scope of the inquiry.

We especially note and agree that this inquiry should consider domestic and international best practices and the measures adopted internationally to encourage sharing and linking of both public and private data.

Specific comments

Page 3

Definitions

Paragraph entitled: **Open Data**

Open data has most usefully been characterised as having the four characteristics denoted by the acronym: FAIR: Findable, Accessible, Interoperable and Reusable.² This terminology is more useful than the term “open”, which can be interpreted in many different ways. As the developers of the FAIR principles note: “FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.” The move to more open data is part of the drive for more open scholarship generally, that has been highlighted by a number of global initiatives recently including the Amsterdam Call for Action on Open Science, an initiative of the Dutch Government, during their chairing of the EU in Jan-June 2016.³

¹ ‘Australasian Open Access Strategy Group’, *Australasian Open Access Strategy Group* <<https://aoasg.org.au/>> [accessed 18 July 2016].

² ‘Article Metrics - The FAIR Guiding Principles for Scientific Data Management and Stewardship : Scientific Data’ <<http://www.nature.com/articles/sdata201618/metrics>> [accessed 18 July 2016].

³ NL EU 2016, *Amsterdam Call for Action on Open ScienceNL* <<https://wiki.surfnet.nl/display/OSCFA/Amsterdam+Call+for+Action+on+Open+Science>>.

Page 4

Box 1 Paragraph entitled: **Why does data matter?**

Data are also critical in ensuring the reproducibility of the academic literature. Without data to back up published research findings, research is based on trust at best. There are many examples now of researchers being unable to reproduce previously published research findings and where the data behind published papers have been found to be unavailable or uninterpretable. Furthermore, there are many cases where lack of data availability has been linked to fraud in research and publishing.⁴ There is now an increasing global consensus that in order to better ensure the integrity of research and to prevent research fraud and improve its investigation, researchers should be willing to make the data that underpins academic papers available. Such data should be in a format that allows their interrogation, provided that appropriate processes are in place to ensure the protection of sensitive data. The Australian National Data Service (ANDS)⁵ has provided guidance on handling sensitive data, including those from Aboriginal and Torres Strait Islander people.⁶ As data become more open it is essential such guidance remains current.

As well as the principles noted above, there is also increasing infrastructure in place directed at increasing the sharing of these data. National examples, with well-established policies, tools and processes in place include ANDS and its many tools including its portal, Research Data Australia.⁷ International repositories for academic data include long established publicly supported ones such as GenBank⁸ and non-profit and commercially used ones such as Dryad⁹ and Figshare,¹⁰ respectively. Many Australian institutions also have their own data repositories which are linked to Research Data Australia.

Page 8

Box 2 Paragraph entitled: **Insufficient dataset linkage?**

Poor, or absent, linkage between published research and the underlying data is one of the most important reasons leading to poor reproducibility in much of the academic literature, especially in some areas of science and medicine.¹¹ It has also been well established that URLs cited in papers decay very dramatically after publication, having a half-life of at best 4.7 years after publication in one study¹²—which reinforces the need for the development of a secure culture of archiving, not merely linkage to temporary websites for example.

⁴ 'Cases | Committee on Publication Ethics: COPE'

<[http://publicationethics.org/cases/?f\[0\]=im_field_classifications%3A757](http://publicationethics.org/cases/?f[0]=im_field_classifications%3A757)> [accessed 27 July 2016].

⁵ ANDS, 'Australian National Data Service', ANDS <<http://www.ands.org.au/about-us>> [accessed 18 July 2016].

⁶ ANDS, 'Ethics, Consent and Data Sharing', ANDS <<http://www.ands.org.au/guides/ethics-consent-and-data-sharing>> [accessed 27 July 2016].

⁷ 'Research Data Australia' <<https://researchdata.ands.org.au/>> [accessed 18 July 2016].

⁸ 'GenBank Home' <<http://www.ncbi.nlm.nih.gov/genbank/>> [accessed 18 July 2016].

⁹ 'Dryad' <<http://datadryad.org/pages/organization>> [accessed 18 July 2016].

¹⁰ 'Figshare - Credit for All Your Research' <<https://figshare.com/>> [accessed 18 July 2016].

¹¹ John P. A. Ioannidis, 'Why Most Clinical Research Is Not Useful', *PLOS Med*, 13.6 (2016), e1002049 <<http://dx.doi.org/10.1371/journal.pmed.1002049>>.

¹² P. Habibzadeh, 'Decay of References to Web Sites in Articles Published in General Medical Journals: Mainstream vs Small Journals', *Applied Clinical Informatics*, 4.4 (2013), 455–64 <<http://dx.doi.org/10.4338/ACI-2013-07-RA-0055>>.

Page 9

Paragraph entitled: Benefits of increasing data availability and use

As noted above, one clear benefit of increasing data availability would be to increase the reliability of the published literature. This in turn leads to increased efficiency of research. The issue of lack of data leading to waste in research is mentioned in the campaign by the Reward Alliance, one of whose key recommendations is “Make publicly available the full protocols, analysis plans or sequence of analytical choices, and raw data for all designed and undertaken biomedical research”.¹³

Page 13

Paragraph entitled: More recently the Australian Government...

We very much welcome the Australian Government’s stated commitment to open data. Of particular importance is the requirement for a Creative Commons license,¹⁴ which fulfils the “R” i.e. reusable part of a FAIR framework for data. Guidance will be required to ensure which license is the most appropriate one for specific contexts and recommendations on developing such guidance would be important in ensuring data is optimally re-useable.

Page 14

Question: What benefits would the community derive from increasing the availability and use of public sector data?

See comments above which relate to the reliability of the academic literature and increased efficiency that would accrue through better access to and reliability of data associated with publications. Of note, many academic journals also now recognise the importance of such data sharing including, for example, the PLOS journals.¹⁵

However, currently there are few accepted processes for citing data, though the Research Data Alliance¹⁶ and Force11,¹⁷ two international organisations, both have groups that have worked on citation practices.

One crucial element of improving data accessibility is to ensure that academics who generate the data for others to use are given appropriate credit for it. Systems to reward such behaviour need to be developed and supported by institutions and funders of research.

¹³ ‘Key Recommendations | Research Waste’ <<http://researchwaste.net/about/recommendations/>> [accessed 18 July 2016].

¹⁴ ‘Creative Commons Australia’, *Creative Commons Australia* <<http://creativecommons.org.au/>> [accessed 18 July 2016].

¹⁵ ‘PLOS Data Availability’ <<http://journals.plos.org/plosone/s/data-availability>> [accessed 18 July 2016].

¹⁶ ‘Research Data Alliance Data Citation Working Group’, *RDA*, 2013 <<https://rd-alliance.org/groups/data-citation-wg.html>> [accessed 18 July 2016].

¹⁷ ‘Joint Declaration of Data Citation Principles - FINAL | FORCE11’ <<https://www.force11.org/group/joint-declaration-data-citation-principles-final>> [accessed 18 July 2016].

Page 22

Question: How should the costs associated with making more public sector data widely available be funded?

Making more public sector data available should be ultimately recognised and built in to research costs as an ongoing cost of operation. However, specific time-limited government funding to establish new practices and repositories should be contemplated where need is demonstrated adequately. The ASHER¹⁸ and ANDS initiatives of the Australian government for research data are good examples of targeted, time-limited funding.

Page 22

Question: Is availability of skilled labour an issue in areas such as data science or other data-specific occupations? Is there a role for government in improving the skills base in this area?

There is unquestionably a lack of comfort among many academics in the curation of data associated with their work. There is a need for skills in data management and analysis, especially of complex datasets, to be incorporated into the training of early career researchers. Programmes such as Data Carpentry¹⁹ have been successful in peer to peer training of researchers, though clearly could be scaled up further.

¹⁸ Australian Government Department of Industry and Science, 'Australian Scheme for Higher Education Repositories (ASHER) and the Implementation Assistance Program (IAP)' <<http://www.industry.gov.au/science/ResearchInfrastructure/Pages/ASHERandIAP.aspx>> [accessed 27 July 2016].

¹⁹ 'Data Carpentry', *Data Carpentry* <<http://www.datacarpentry.org/>> [accessed 18 July 2016].