

Productivity Commission Inquiry into Data Availability and Use

Response by the University of Sydney to the Inquiry's Draft Report

1. General comments

This wide-ranging and detailed report encapsulates the existing issues and deficits in current data policy, and identifies potential improvements that will set the agenda for future Australian research and innovation endeavour, particularly with reference to availability and use of public datasets.

The University of Sydney is supportive of the introduction of new legislation that will facilitate access to, and use of, both public and private datasets in the research environment, and recognises the over-arching importance of maintaining privacy and confidentiality of these data. The public's expectation and confidence that research organisations will have accountable and transparent data handling processes that protect all individuals' confidential data and rights to privacy will continue to inform the University's own policy development in this area.

At the recent public hearing of the Inquiry in Sydney, the Commissioner invited the University to comment specifically on two recommendations in the draft report. This discussion is summarised and continued below.

2. Open access

2.1 Open access of research output

The University recognises the value of new legislation and policy that will support open access to data published by the University's researchers so that research output can be discovered, accessed and used by other researchers, agencies and the wider community (University of Sydney Open Access Policy, 2015). Making research datasets freely available will drive innovation and enquiry across all research disciplines, resulting in enormous scholarly, public and commercial benefit. Thus, it is vital that all research institutions invest in infrastructure and systems that will facilitate open accessibility of research data.

To this end, the University of Sydney is currently sourcing a new public-facing institutional repository that will house digital research outputs (mostly but not limited to data) from all disciplines and will be freely accessible. Where data are confidential or sensitive they will not be made available to the public, but their metadata records will provide information about the dataset and the researchers who have custodianship. This infrastructure will underpin a 'data ecosystem' from which research data can be freely accessed, re-used and re-purposed.

Other research institutions will be planning similar investments in response to increasing expectations from funding bodies and the community to make public-funded research outputs freely available. It is therefore important to recognise the significant investment that is required in providing this technology and infrastructure, and to consider expanding targeted federal funding schemes to assist in the transition to a data-rich open access environment.

2.2 Open access to public datasets

Open access to research datasets will facilitate re-use of data across studies and disciplines, and lead to increased opportunities to link multiple datasets from diverse sources. This raises specific issues concerning public dataset accessibility by researchers:

- 1) legislation around the de-identification and potential re-identification of data needs to be revised to ensure confidentiality is maintained when linking datasets

- containing personal information, while simultaneously removing onerous restrictions on releasing these datasets to researchers;
- 2) investment in data analytical skills training and development in data agency personnel is needed to improve expediency of data linkage requests;
 - 3) mechanisms should be introduced to reduce the costs charged to researchers for public dataset linking; and
 - 4) approval processes from HRECs and other regulatory bodies to access multiple linked public datasets from diverse providers need to be streamlined (e.g. waiting 4 years for linked datasets to be released is obviously not acceptable). The University supports the Productivity Commission's recommendations in this area.

3. Accreditation for researchers

In the draft report, the Commission recommends (in recommendations 9.7 and 9.8) the introduction of a system of 'trusted users' with accreditation overseen by a National Data Custodian. 'Trusted users' would have ongoing access arrangements to confidential datasets approved for the life of a research project (rather than for each dataset used, as is currently the case).

The University is strongly supportive of an accreditation system based on delegated authority, whereby the National Data Custodian (NDC) would effectively accredit a range of 'trusted/accredited organisations', including all universities. Under this model the default position would be that all research personnel of universities and other organisations recognised by the NDC would be "trusted/accredited users". It is important that we specify here that research personnel include academic research staff, research students and research affiliates. When implementing such a scheme, the University would develop accreditation processes agreed by the NDC (complying with any new legislation) and undertake periodical audits as directed by the NDC.

Finally, a question of semantics and public perceptions. Although the term is used overseas, a 'trusted user' or 'trusted researcher' is a loaded term conveying little information, merely implying varying degrees of a personal quality (namely 'trustworthiness'). Indeed, all researchers who currently comply with HREC protocols and the contractual conditions required by data providers are expected to demonstrate their 'trustworthiness' and integrity when handling confidential data. We would suggest that a more appropriate term would be 'accredited user' or 'accredited researcher', which actually informs that some sort of accreditation and training process has taken place.

4. Additional comments

4.1 Deletion of research data

As stated in draft recommendation 5.3, 'the Australian Government should abolish its requirement to destroy linked datasets... at the completion of researchers' data integration projects'. The compulsory deletion of research data at the end of projects using public datasets denies other researchers the opportunity to re-use these valuable resources and is detrimental to future research effort.

However, researchers encounter the same problem with datasets that are generated within the research environment (and not sourced externally from public datasets) as it is a condition of NHMRC funding that data are deleted after a mandatory retention period. Clearly, some funding rules will need to be reviewed when new legislation is introduced in this area. The University believes strongly that retention and re-use of data should be promoted and that research data should be deleted only under exceptional circumstances.