UNSW Contact:  Associate Professor Grainne Moran, Pro-Vice-Chancellor (Research Infrastructure)
Email: g.moran@unsw.edu.au

UNSW Australia welcomes the opportunity to provide a response to the Productivity Commission Draft Report on Data Availability and Use. The report is very comprehensive and will greatly assist in informing planning for both generators and end users of data, including for research purposes.

UNSW agrees with the views expressed in the Draft Report, in particular with the notion that fundamental and systemic changes are needed to the way Australian governments, business and individuals handle data.

Specifically, UNSW endorses the following aspects of the proposed new framework:

• the proposed risk-based approach to determine the accessibility of data;
• the measures to improve the use of linked datasets (Draft Recommendation 5.3 and 5.5); and
• the call for an independent review of the pricing of public sector data for the publically funded research sector (Draft Recommendation 7.2).

On the other hand, UNSW suggests amending the following recommendations as outlined further in the attachment, to ensure that there are no unintended consequences for research:

• providing trusted users access to higher risk data on a project-specific basis;
• prioritising research funding on the basis of institutions making their researchers’ data widely available; and
• publication of registries of data by publicly funded entities.

UNSW would be pleased to provide any clarification or further information relating to this submission.

Yours sincerely,

Associate Professor Grainne Moran
UNSW Response to the Productivity Commission Draft Report

Data Availability and Use

Publication of registries of data by publicly funded entities

Draft recommendation 3.2 states that publicly funded entities, including the ARC, should publish up-to-date registries of data holdings, including metadata, that they fund or hold.

UNSW suggests that requiring the ARC (and other individual entities or agencies) to set up registries may not be the most effective or cost-efficient way to achieve the goal of data availability. Agencies should require the data generated by researchers they fund to be registered through recognised national, international or discipline-specific registries; and that the data itself be made available either openly or to trusted parties subject to the various restrictions identified in the preceding section of the Report.

UNSW also suggests that a significant transition period will be required in order to achieve full implementation, given the need to develop the requisite tools and infrastructure.

Ensuring improvements to linked datasets are available for subsequent uses

Draft recommendation 5.3 proposes abolishing the requirement to destroy linked datasets and statistical linkage keys at the completion of researchers’ data integration projects. The Commission is seeking further views on the most practical ways to ensure improvements to linked datasets are available for subsequent dataset uses. UNSW suggests that an alternative to requiring researchers to supply their enhanced research-ready datasets would be to require that they provide and document the code that they used to create these from the source datasets that they were supplied with. The code could then be made available to other users. This avoids a requirement to transfer and curate multiple different research datasets based on the same source datasets.

Prioritising funding for academic institutions

Draft Recommendation 5.4 states that to streamline approval processes for data access, the Australian Government should prioritise funding to academic institutions that implement mutual recognition of approvals issued by accredited human research ethics committees. UNSW is in favour of mutual recognition being the norm and suggests that to achieve the desired outcomes, implementation of mutual recognition by and with ethics committees operated by government agencies needs to be prioritised and resourced.

Draft recommendation 9.9 specifies that public research funding should be prioritised on the basis of progress made by research institutions in making their researchers’ data widely available to other trusted researchers on conclusion of research projects. This recommendation as worded has the potential to actually disadvantage some research groups that focus on secondary use of public sector data, but who cannot make these data more widely available by virtue of the agreements that govern their use. Perhaps there should also be prioritisation of public research funds based on progress in making effective use of public sector data for policy-engaged research?

Workforce shortages

The full Draft Report (Section 6.5) includes discussion of the acute shortage of data skills, but there are no matching findings or recommendations. UNSW would suggest that the final report include a recommendation in this area, for example prioritising postgraduate programs in data science and data analytics for graduate Commonwealth Supported Places (CSP).
Proposed new model for data access for trusted users

Draft recommendations 9.7 and 9.8 both recommend accreditation of trusted users. UNSW would suggest that completion of formal, assessed, training for individual researchers should be part of accreditation. The CITI researcher training consortium (https://www.citiprogram.org) offers one such model.

Draft recommendation 9.8 specifies that approval for data access to trusted users is given on a "project-specific" basis. Research using large-scale data is often programmatic in nature and extends over years, so it may be preferable to consider approving "programs" of work, with annual reporting to the Accredited Release Authority of the component projects that have been completed.

The description of the new model on pages 21 and 22 of the Overview Draft Report outlines that access would occur in a specified secure computing environment with "output from the dataset reviewed by an automated process…to ensure confidentiality requirements have been satisfied". UNSW suggests that such automated review may not be feasible (or at least not universally feasible) given the huge variety of output types and formats produced by researchers, and the difficulty of evaluating context and risk for such varied outputs. We would instead suggest that a complete audit trail of all file movements is kept (including copies of the files themselves), allowing for "spot check" audits, and that responsibility for ensuring that confidentiality requirements are met, like other responsibilities for the responsible conduct of research, rests with the trusted user.