

DATA AVAILABILITY AND USE

Population Health Research Network submission in response to the Productivity Commission Draft Report

December 2016

Dr Merran Smith

Chief Executive, PHRN

Phone: 08 6488 8600

Email: merran.smith@uwa.edu.au



NCRIS
National Research
Infrastructure for Australia

An Australian Government Initiative

RESPONSE

The Productivity Commission's Data Availability and Use Draft Report makes a significant contribution to thinking on the Australian data environment. The Population Health Research Network (PHRN) agrees in principle with the draft recommendations which are bold and far reaching. In particular, the PHRN strongly supports recommendation 5.2 (*The Privacy Act 1988 (Cth) exceptions that allow access to identifiable information for the purposes of health and medical research without seeking individuals' agreement, should be expanded to apply to all research that is determined to be in the public interest*) which was proposed in our initial submission to the Inquiry. The PHRN is pleased that elements of all of the recommendations in our initial submission have been included in the Draft Report.

Overall, we believe the Draft Report understated the significance of the PHRN infrastructure in improving data availability and use in Australia as well as its role in ensuring ethical, safe and secure access to data. The successes of the Western Australian Data Linkage System and Centre for Health Record Linkage in New South Wales were highlighted. However, the strength of the PHRN is its international uniqueness in providing data linkage infrastructure for every jurisdiction in the country and the capability to link data from multiple jurisdictions. In every Australian state and territory, data linkage is now done in an enduring fashion. With PHRN support, the AIHW has commenced a project to link MBS and PBS data to the state/territory linkage maps on an enduring basis so that projects involving linked cross-jurisdictional data will no longer be conducted on a create and destroy basis.

In addition to the national network of data linkage units the PHRN also includes:

- the SURE which was highlighted in the Draft Report as an example of trusted user access,
- an Online Data Application system which streamlines applications for multi-jurisdictional projects
- SUFEX which is secure encrypted data transfer system
- Training for researchers using SURE and for human research ethics committees who assess applications for linked data.

We believe Figure 3.3 in the Draft Report significantly underestimates the volume of requests for linked data across Australia. The figure also does not give any indication of the size and complexity of individual data requests. The PHRN has a mechanism for measuring complexity of requests for linked data and this is being measured over time.

The PHRN has a wealth of skills and expertise and a national infrastructure network that is publically funded. It is not clear how, or if, the Productivity Commission sees this infrastructure contributing to the proposed new framework. For example, the Draft Report proposes the designation/establishment of National Interest Datasets (NIDs). Some of these NIDs would be linked datasets. Is there a role for PHRN data linkage units or is the establishment of another data linkage infrastructure imagined? If so, is the duplication justified?

Recommendation 5.3 - The Australian Government should abolish its requirement to destroy linked datasets and statistical linkage keys at the completion of researchers' data integration projects. Data custodians should use a risk-based approach to determine how to enable ongoing use of linked datasets. The value added to original datasets by researchers should be retained and available to other dataset users.

The PHRN supports this recommendation. However, the Draft Report and this recommendation fails to tease out the difference between an enduring linkage system and a linked dataset. An enduring linkage system e.g. WA Data Linkage System identifies which individuals have records in which data collections. For example a linkage system identifies that a person has both a birth record and a hospital record. This is called a link. The linkage system creates a catalogue of these links, stores them and updates them as new information is added to the system on a regular basis. The catalogue is often called a master linkage map or file. The linkage system only uses identifiers such as name, address, date of birth and sex. The linkage system does not hold any content data. An enduring linkage system enables bespoke linked datasets to be created for analysis quickly and efficiently. All the states and territories currently have enduring linkage systems. To date the Commonwealth has required links to be created for each individual project request and then destroyed. This has been a very resource intensive and inefficient process. The AIHW is currently working on establishing an enduring linkage system. It would be helpful if the Productivity Commission Report could recommend the use of enduring linkage systems by the Australian Government more specifically.

It appears that the use of the term “statistical linkage keys” in Recommendation 5.3 may in fact be referring to the links (or linkage keys) created by Commonwealth Integrating Authorities when they conduct create and destroy linkage. This term has a very specific meaning (a linkage key created using parts of the identifiers) which is probably not intended in the recommendation.

The retention and ongoing use of linked datasets is supported but should be considered a separate activity from the use of enduring linkage systems.

The PHRN is supportive of the designation of high value data collections. In effect Australia already has a system of designating high value data collections i.e. the National Minimum Datasets (<http://www.aihw.gov.au/national-minimum-data-sets/>). The delegation of data custodian responsibilities to Accredited Release Authorities would hopefully reduce the number of approvals required for access. However, in order to ensure that users understand the quality and limitations of the data, detailed and extensive metadata will be necessary. If this is not made available to users there is a high risk of misinterpretation of the data.

The Draft Report clearly encourages as much data as possible to be made open access i.e. publically available. The suggestion appears to be that data pertaining to an individual person, if de-identified, could be made available in the public domain. We would caution against this approach, particularly as re-identification science will continue to be more sophisticated and successful. Individual level data will likely always be most appropriately made available in a trusted user environment.

If implemented these recommendations will transform the data landscape in Australia. They will require substantial funding, resources and effort from all Australian jurisdictions if they are to be successful.

The PHRN would be delighted to discuss options with the Productivity Commission for how Australia's data linkage infrastructure could fit into the proposed framework.