



Productivity Commission Inquiry into Data Availability and Use: Draft Report

Submission by the *ARC Centre of Excellence for Children and Families over the Life Course* (the Life Course Centre)

The *ARC Centre of Excellence for Children and Families over the Life Course* (the Life Course Centre) was established in 2014 as a collaborative research endeavour across university, government and non-government partners to support the needs of children and families experiencing – or at risk of experiencing – deep and persistent disadvantage.

The Centre is funded for seven years (2014-21) to **study the causal mechanisms of deep and persistent disadvantage** in Australia and to **design and test social interventions** that can support Australia's most vulnerable children and families to realise their life potential.

The Life Course Centre is headquartered at **The University of Queensland's Institute for Social Science Research (ISSR)** and has research nodes at the University of Western Australia, the University of Melbourne, and the University of Sydney.

Our **research partners** include four Commonwealth government agencies (the Departments of Social Services, Education and Training, and Employment, and the Australian Institute for Health and Welfare), two state government departments (in Queensland and Victoria), two community organisations (the Brotherhood of St Laurence and Wesley Mission Queensland) and ten international universities.

Executive Summary

The Life Course Centre welcomes the Productivity Commission's Inquiry into the availability and use of public and private sector data.

Despite increasing recognition of the potential of government administrative data as a source of rich and invaluable evidence about our society, Australia lags behind its counterparts in enabling researchers to access and use these resources.

The Centre endorses the Inquiry's overarching recommendation that Australia must transition from a risk averse approach to data access and use to a risk management approach if it is to optimise the use of administrative datasets for the benefit of evidence-based public policy making.

The Centre also broadly supports the recommendations as set out in the Inquiry's Draft Report to support this transition and is keen to work with government at all levels to implement them. In this submission, we suggest some ways to improve some key recommendations to facilitate their implementation nation-wide.



Our Response

In this submission, we:

1. Articulate the value of the Inquiry's findings to social research and evidence-based policy and program development as understood by the Life Course Centre
2. Confirm our support for the key recommendations of the draft report and provide suggestions for enhancement where applicable.

1. The value of administrative and longitudinal data sets

The focus of this Inquiry is central to the work that we undertake at the Life Course Centre to address deep and persistent disadvantage and to the effective and efficient advancement of Australian policy decision-making and development.

One of the primary goals of the Life Course Centre is to unlock the untapped potential of administrative datasets to inform sustainable and effective evidence-based policy and programs to enhance lifetime wellbeing for Australian children and families.

Australia is in an exciting phase where the necessary national longitudinal and administrative data collections have now matured to the point where research into the transmission of social and economic disadvantage within and across generations can be undertaken in this country.

No one dataset can shed light on the complexities of multiple disadvantage in isolation, however, and without advances in Australian social data practices (including data access and integration), we risk falling behind in obtaining the evidence required to improve social and economic outcomes. As the Inquiry has found, impediments exist to researchers accessing both public and private sector data, many of which are founded on Australia's current risk averse approach.

The Life Course Centre supports the Inquiry's finding that Australia needs to transition to a risk management approach to social data if we are to optimise potential for the benefit for individual and communities. We are actively contributing to improvements in social data management and utilisation for our Centre partners, and we are committed to working with government and other organisations to establish the necessary safeguards and processes to ensure Australians are aware of and can have confidence in the way their data is used.

2. Feedback on recommendations

The Life Course Centre provides the following suggestions for the purpose of enhancing the Inquiry's recommendations so that they might better align with social and economic research needs and achieve effective implementation.

Addressing specific impediments to public sector data access

Draft Recommendation 3.2

Publicly funded entities, including the Australian Research Council, should publish up-to-date registers of data holdings, including metadata, that they fund or hold.

Publication of summary descriptions of datasets held by funded researchers but not released, and an explanation of why these datasets are not available, are also essential and would provide far greater transparency about what is being funded by taxpayers but withheld.

We strongly support this recommendation and agree that it should cover publically funded agencies such as the Australian Research Council (ARC), the National Health and Medical Research Council (NHMRC) and the Australian Housing and Urban Research Institute (AHURI).

We do, however, caution against making the datasets funded by these agencies freely accessible to all. Researchers should have the opportunity to hold exclusive access to the data they have collected until at least the end of their period of ethical approval for that research (which may outlast the period of the grant funding). Research is competitive, and researchers should benefit from data collections and research design that were their original idea, before other researchers can access it for their own purposes. A higher order objective, which is consistent with the National Principles of Intellectual Property (IP) Management for Publicly Funded Research (developed by a working party of the Australian Government's Coordinating Committee on Innovation), is to make every reasonable effort to gain benefit for Australia from IP, which may involve protecting the IP, or making it publicly available, appropriate to the circumstances.

Draft Recommendation 4.2

All Australian governments entering into contracts with the private sector, which involve the creation of datasets in the course of delivering public services, should assess the strategic significance and public interest value of the data prior to contracting. Where data is assessed to be valuable, governments should retain the right to access or purchase that data in machine readable form and apply any analysis that is within the public interest.

We strongly support this recommendation. For example, this is of particular importance for two sectors: Health and Education.

The lack of accessible data on private hospitals has long hindered health researchers, the development of universal health policies and best practice guidelines as well as benchmarking between the two sectors. Private hospitals in receipt of government funding should be encouraged to undertake and manage the same data collections as public hospitals, using the same metadata standards. This should be done through legislation or by making receipt of government funding contingent on following the same data collection items, processes and standards.

The same lack of accessible data on private schools has also long hindered education and social researchers, as well as the development of universal education policies and best practice guidelines. Private schools receive substantial government funding and should be encouraged to undertake and manage the same data collections as public schools, using the same metadata standards. This should either be done through legislation or by making receipt of government funding contingent on following the same data collection items, processes and standards.

The conundrum of personal data

Draft recommendation 5.1

In conjunction with the Australian Bureau of Statistics and other agencies with data de-identification expertise, the Office of the Australian Information Commissioner should develop and publish practical guidance on best practice de-identification processes.

To increase confidence in data de-identification, the Office of the Australian Information Commissioner should be afforded the power to certify, at its discretion, when entities are using best practice de-identification processes.

One of the primary roles of privacy legislation is to protect individuals from being identified when data or analyses leave the safety of their custodian agency and enter the public domain. Inconsistent definitions of privacy, and differing legislation surrounding it, expose one of the weaknesses of our federation model. Each state and territory, and the Commonwealth, has separately and independently developed their own definition of privacy, and placed their own legal frameworks around this in relation to the administrative data collected under their jurisdiction. These data often relate to individual persons, but also to "things" such as environmental, transactional, or business information. If the Commonwealth plans on creating a system of data integration that spans all administrative sources, including those presided over by state and local governments, it will need to develop a harmonized legal framework that is at least consistent with, or even overrides, existing privacy laws in all Australian jurisdictions.

Draft 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.

INFORMATION REQUEST

The Commission seeks further views on the most practical ways to ensure improvements to linked datasets are available for subsequent dataset uses.

The concept of an enduring composite micro-dataset based around a core spine of key datasets, such as the National Interest Datasets, and updated at least annually via a mostly automated (machine learning) process, would take Australia from administrative data laggard to world leader, with commensurate gains in evidence based decision making and research breakthroughs.

To create large linked datasets for use in a single research project is a time-consuming, arduous process and the Life Course Centre considers the subsequent destruction of these linked datasets, as per the data usage agreements, to be an inefficient use of public funds. The Centre strongly supports ending this obligation in favour of the model described in this recommendation, which will lead to continuous improvement in the linkage quality and linkage rate, and which will support much faster extraction of data following research ethics approval.

Draft Recommendation 5.5

In light of the Australian Government's commitment to open data, additional qualified entities should be accredited to undertake data linkage.

State-based data linkage units should be able to apply for accreditation by the National Data Custodian (Draft Recommendation 9.5) to allow them to link Australian Government data, and the intention of 'open by default' should apply to these exchanges.

The Life Course Centre supports a national standard of linkage hardware, software, methodology and process, which is appropriately funded by the Commonwealth. State Data Linkage Units (DLU) must meet agreed standards and use standardised processes, such that linkage performed in any jurisdiction (state or Commonwealth) would deliver the same quality of data in a similar timeframe. Outputs from any state or Commonwealth DLU should be interchangeable, given the same (hypothetical) data request, which would enable all researchers and policy makers to benchmark their work.

Making data more useful

Draft recommendation 6.1

Government agencies should adopt and implement data management standards to support increased data availability and use as part of their implementation of the Australian Government's Public Data Policy Statement.

These standards should:

- *be published on agency websites*
- *be adopted in consultation with data users and draw on existing standards where feasible*
- *recognise sector-specific differences in data collection and use*
- *support the sharing of data across Australian governments and agencies*
- *enable all digitally collected data and metadata to be available in commonly used machine readable formats (that are relevant to the function or field in which the data was collected or will likely be most commonly used), including where relevant and authorised, for machine to machine interaction.*

Policy documents outlining the standards and how they will be implemented should be available in draft form for consultation by the end of 2017, with standards implemented by the end of 2020.

Agencies that do not adopt agreed sector-specific standards would be noted as not fully implementing the Australian Government's Public Data Policy and would be required to work under a nominated Accredited Release Authority (Draft Recommendation 9.6) to improve the quality of their data holdings.

This is a commendable recommendation. Additional Commonwealth funding will be required to support government agencies at both state and Commonwealth levels to smoothly transition to and implement these data management standards in a timely manner. If additional resources are withheld, there is likely to be significant differences in implementation progress between States and between agencies.

Valuing and pricing data

Draft recommendation 7.3

Minimally processed public sector datasets should be made freely available or priced at marginal cost of release.

Where there is a demand and public interest rationale for value-added datasets, agencies should adopt a cost recovery pricing approach. Further, they should experiment with lower prices to gauge the price sensitivity of demand, with a view to sustaining lower prices if demand proves to be reasonably price sensitive.

The Centre would support a model whereby access to data is free as a default, with nominal but escalating fees for requests that incur unreasonable costs. We note that such fees would need to be borne elsewhere, however, and may serve merely to transfer data access costs onto other Commonwealth funding agencies, such as the ARC or NHMRC. Without additional funding support for this activity, a transfer of costs may serve to limit research using that data, or reduce the total pool of available research funding. We would also expect access to be free for government agencies and publicly funded research, given that the creation of this national data infrastructure will have benefits extending far into the future for Australian taxpayers and communities.

Fundamental reform is needed

Draft recommendation 9.3

The Australian Government should provide for broad oversight and complaints handling functions within a reformed framework for individual data access. Key roles should be accorded to the Australian Competition and Consumer Commission (ACCC) the Office of the Australian Information Commissioner (OAIC), and to existing industry ombudsmen.

Any charging regimes, policies or practices introduced to address costs associated with data access, editing or transferability should be transparent and reasonable. The ACCC should be responsible for monitoring and assessing the reasonableness of charges applied. The ACCC, supported by state and territory Fair Trading Offices, should also educate and advise consumers on their new rights in regard to data access and collection.

For specified datasets (such as in banking) the relevant ombudsman scheme would need to be expanded to deal with disputes.

INFORMATION REQUEST

The Commission seeks further views on datasets that are of national interest and that could feasibly be designated as such under the process proposed.

As noted in the Inquiry's Draft Report, it is difficult to identify datasets of national interest without knowledge of the contents and conditions of datasets. Metadata for existing datasets, in addition to information about the quality of the data, will be required to reliably inform discussions about datasets of national interest.

Draft recommendation 9.8

Arrangements for access by trusted users to identifiable data held in the public sector and by publicly funded research bodies should be streamlined and expanded by the Australian Government. The National Data Custodian should be given responsibility to:

- *develop, in consultation with data custodians, a list of pre-approved uses for a dataset, and make decisions on access to data for projects not consistent with the pre-approved uses list*

- *grant, on an approved project-specific basis, trusted user access to personnel from a range of potential entities, including: all Australian Government and state and territory government agencies; all Australian universities; and other entities (be they corporations, not-for-profit organisations or research bodies) that:*
 - *are covered by privacy legislation*
 - *have the necessary governance structures and processes in place to address the risks of inappropriate data use associated with particular datasets, including access to secure computing infrastructure.*

Access would be granted for the life of the specific approved project. Trusted user status for use of identifiable data would cease when the user leaves the approved organisation; a project is completed; or if a breach occurs in that same organisation and/or project.

As set out in our feedback on Draft recommendation 5.1, there is a need for governments to develop a single legal framework that defines privacy and applies to all Australian jurisdictions.

Draft recommendation 9.9

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.

Universities already have well developed policies and practices regarding access of publicly-funded research, and are guided by national frameworks such as the National Principles of Intellectual Property (IP) Management for Publicly Funded Research (developed by a working party of the Australian Government's Coordinating Committee on Innovation). Consistent with our feedback on Recommendation 3.2, the Centre supports researchers having the opportunity to maintain exclusive access rights to the data they have collected, for a reasonable period following data collection, such as until at least the end of their period of ethical approval for that research (which may outlast the period of the grant funding). This position is consistent with a competitive process of innovation, which ensures that creators of intellectual property are able to protect their rights for a reasonable period of time in order to ensure that they are able to realise a return on their investment. By corollary, removing this advantage would serve to discourage academic and private research using publicly-funded data.