

DATA AVAILABILITY AND USE

**Submission to the Productivity Commission
Data Availability and Use Inquiry
from the Population Health Research Network**

29 July 2016



PHRN Population
Health
Research
Network

NCRIS
National Research
Infrastructure for Australia

An Australian Government Initiative

EXECUTIVE SUMMARY

The Population Health Research Network¹ (PHRN) is a national data linkage infrastructure network funded by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) with support from government agencies and partner organisations. The PHRN's primary purpose is to build and facilitate the operation of collaborative, national data linkage infrastructure capable of securely and safely linking data collections from a wide range of sources and providing access to linked data.

With support of the PHRN, Australia now has the capacity to link and provide access to linked data in all jurisdictions. This infrastructure is of international significance. PHRN achievements include:

- Establishment of new data linkage units in Queensland, Victoria, Tasmania and South Australia and an accredited Commonwealth Integrating Authority at the Australian Institute of Health and Welfare;
- New online application and secure data delivery systems which facilitate access to data; and
- Establishment of a remote access data laboratory that enables researchers to access linked datasets in a secure environment from anywhere in Australia.

The PHRN infrastructure supports the linkage of data collections from both the public and private sectors across a range of disciplines including health, education and social services. There are usually few technical barriers to the linkage of data from a variety of sources but a range of legal, policy, resourcing and cultural barriers remain.

The next stage of Australia's data linkage infrastructure development should include:

- Enduring linkage of Commonwealth to Commonwealth data collections e.g. Commonwealth Medical Benefits (MBS) and Pharmaceutical Benefits (PBS) data collections, and Commonwealth to state/territory data collections e.g. hospital to PBS.
- Expansion of the number and type of data collections that are routinely linked at both the Commonwealth and state/territory levels.
- Streamlining of the data custodian and ethical approvals required to access linked data.
- Content data repositories in each jurisdiction to reduce the burden on data custodians and minimise the time to extract data for research and analysis.

In PHRN's experience, this expansion in capability will require coordination, collaboration and concerted effort over a number of years. Significant benefits in consumer outcomes, productivity and economic growth are likely to flow from increased data availability and use resulting from the expansion.

This submission focuses on the following:

- The importance of the availability and use of public and private sector data with an emphasis on the linkage of data for health and human services research.
- The characteristics of high value data collections.
- The barriers to accessing linked data including the legal framework, approvals, governance processes and costs.
- Ways to improve availability and access to linked data.
- The importance of minimising risks to privacy and confidentiality.

¹ The Population Health Research Network. Perth WA (Aus). 2016 [cited 27/06/2016]. Available from <http://www.phrn.org.au>

In order to improve the availability, use and benefits of linked data, the following recommendations are made:

Recommendation 1: An agreed list of high value data collections which should be included in the national linkage system at Commonwealth and state/territory levels should be determined.

Recommendation 2: The recommendation of the Australian Law Reform Commission to amend the Privacy Act to extend the arrangements relating to the collection, use or disclosure of personal information without consent in the area of health and medical research to cover the collection, use or disclosure of personal information without consent in human research more generally should be implemented.

Recommendation 3: Where required, changes to legislation to clarify the collection, use and disclosure of specific data collections for research should be made.

Recommendation 4: A system of national mutual recognition of ethics review of applications for research using linked data should be implemented. It may be possible to adapt the current system for review of clinical trials applications for this purpose.

Recommendation 5: Governments and government departments should implement mechanisms to reduce the number of data custodian approvals required for research projects. Options might include approval at the department or whole of government level rather than at the level of the individual data collections.

1 INTRODUCTION

1.1 Population Health Research Network

The Population Health Research Network² (PHRN) is a national data linkage infrastructure network. PHRN commenced in 2009 and is funded by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS), with support from state and territory government agencies and academic partners. The University of Western Australia is lead agent for the PHRN. The PHRN's primary purpose is to build and support the operation of collaborative, nationwide data linkage infrastructure capable of securely and safely linking data collections from a wide range of sources including within and between jurisdictions and across sectors and providing access to linked data.

Through the support of the PHRN, Australia now has the facilities and capabilities to link and provide access to linked data in all jurisdictions. This infrastructure is of international significance. PHRN achievements include:

- Establishment of new data linkage units in Queensland, Victoria, Tasmania and South Australia;
- Establishment of an accredited Commonwealth Integrating Authority at the Australian Institute of Health and Welfare;
- New online application and secure data delivery systems which facilitate access to data; and
- Establishment of a remote access data laboratory (SURE) that enables researchers to access linked datasets in a secure environment from anywhere in Australia.

The PHRN infrastructure supports the linkage of data collections from both the public and private sectors across a range of disciplines including health, education and social services e.g. hospital admitted patients, cancer registries and the Australian Early Development Census. There are usually few technical barriers to the linkage of data from a variety of sources. However, a range of legal, policy, resourcing and cultural barriers remain.

The PHRN is continuing to improve Australia's data linkage infrastructure, increase access to linked data and expand the use of linked data. The next stage of Australia's data linkage infrastructure development should include:

- Enduring/routine linkage of Commonwealth to Commonwealth data collections e.g. MBS and PBS and Commonwealth to state/territory data collections e.g. hospital to PBS.
- Expansion of the number and type of data collections that are routinely linked at both the Commonwealth and state/territory levels. Ideally there would be a standard set of data collections linked by each state and territory so that researchers could easily conduct national and cross-jurisdictional research.
- Streamlining of data custodian and ethical approvals required to access linked data.
- Content data repositories such as the equivalent of the Custodian Administered Research Extract Server³ in each jurisdiction to reduce the burden on data custodians and minimise the time to extract data for research and analysis.

² The Population Health Research Network. Perth WA (Aus). 2016 [cited 27/06/2016]. Available from <http://www.phrn.org.au>

³ Eitelhuber and Davis Health Information Science and Systems 2014, 2:6 <http://www.hissjournal.com/content/2/1/6> [accessed 25/07/2016]

In PHRN's experience, an expansion in capability of the sort identified above will require coordination, collaboration and concerted effort over a number of years. Key issues are discussed in the body of the submission.

1.2 Submission

The importance and value of access to linked data for monitoring, evaluation and research has been recognised in Australia for some years. The Productivity Commission devoted a whole chapter in its 2012-13 Annual Report⁴ to the use of linked administrative data to achieve better policy outcomes. The Senate Select Committee on Health recently released its report entitled "Big health data: Australia's big potential"⁵ which also highlighted the potential of using linked data to inform health policy and improve health services. However, both of these reports also identified a number of barriers to realising the full potential of linked data for Australia.

This submission has been prepared in response to the Productivity Commission Issues Paper on Data Availability and Use. The PHRN welcomes the opportunity to comment on matters raised in the Issues Paper.

The submission discusses the availability and use of linked data for research purposes. It focuses on the use of personal information in privacy preserving ways for research and evaluation. Broader issues associated with data availability and use in an open access environment are not covered in the submission.

The availability of linked population-based data is important because:

- No single data collection is sufficient to allow an understanding of the complex pathways that result in health or disease and whether Australia's health and social service systems work in optimal ways.
- Australia is a federation and different jurisdictions collect different data. For example, the states/territories collect hospital admissions data and the Commonwealth collects MBS and PBS data. In order to understand individual patient pathways and the health system as a whole these data collections must be linked.
- Collecting data once and using many times for different purposes e.g. service provision and research is more cost effective than collecting data multiple times.

Australia has made significant progress in improving the availability and use of data in recent years. However as explained in the submission, continued improvement is necessary to achieve the benefits in consumer outcomes, productivity and economic growth that flow from increased availability and usage of data.

2 BENEFITS OF INCREASING DATA AVAILABILITY AND USE

2.1 Public sector data

Questions on high value public sector data

In Australia, information about an individual's health, education and welfare is recorded throughout their lives as they come in contact with service delivery organisations and

⁴ Productivity Commission 2013, Annual Report 2012-13, Annual Report Series, Productivity Commission, Canberra.

⁵ Senate Select Committee on Health. May 2016. Sixth Interim Report, Big health data: Australia's Big Potential.

agencies, including hospitals (public and private), health departments, schools and other government agencies. The collection of this data is often required under legislation and the information is stored in secure computer databases within the responsible agencies.

The position outlined in the Issues Paper on the benefits from greater availability and use of public sector data is supported. The characteristics of high value datasets and some examples of these datasets are described later in this section, in the context of linked data.

Questions on collection and release of public sector data

There are a range of barriers preventing public sector agencies from making data readily available. These include current policies, processes and practices which limit availability as well as the absence of documented policies, processes and practices to support and enable access. Risk aversion and resource constraints appear to be other contributing factors. In some cases, provisions in legislation are a barrier.

Governments would use their data collections more effectively if they took a whole of agency, if not a whole of government approach to the management of statutory and administrative data. In some cases, amendments to legislation to explicitly provide for the collection, use and/or disclosure of data would assist.

A standardised approach to the use and release of data at agency, jurisdictional and cross-jurisdictional level would be beneficial. While there would be some initial cost in establishing the process(es), it is likely that there would be savings from the application of a more standardised process as well as benefits from a more consistent approach to responding to requests for access to data.

Which datasets, if linked or coordinated across public sector agencies, would be of high value to the community, and how would they be used?

Public sector, health and human services administrative data collected on whole populations e.g. state/territory or national, when linked, is extremely valuable for research. The use of population level data collections in research reduces sample bias and enables research on rare diseases and events.

Through PHRN support, Australia now has data linkage units which service each jurisdiction and also has the capability to conduct linkage between jurisdictions. The core state/territory health data collections (births, deaths, hospital, emergency, cancer registry and perinatal) are routinely linked by these data linkage units. The Australian Institute of Health and Welfare is currently working on establishing enduring/routine linkage of Commonwealth data collections to these linked state/territory data collections. This will enable the high quality state/territory enduring linkage keys to be leveraged to deliver high quality cross-jurisdictional enduring linkage keys.

The characteristics of high value data collections which would deliver value to the community if linked include:

- Longitudinal
- Whole population
- High quality identifiers which can be used for linkage
- Unit record not aggregate data
- Metadata (descriptors of variables) available.

Examples of public data collections that would be considered high value (in addition to those already included in Commonwealth and state/territory linkage systems) include:

- Australian Childhood Immunisation Register
- Medicare Benefits Scheme (linkage in progress)
- Pharmaceutical Benefits Scheme (linkage in progress)
- NAPLAN
- Pathology
- Community mental health
- Non-admitted patient care
- Imaging data collections
- Payments data such as Centrelink payments.

Inclusion of these data collections in Commonwealth and state/territory enduring data linkage systems would significantly expand the research questions that could be answered about causal pathways, health service use and post-market surveillance of pharmaceutical and therapeutics in Australia. Cost effective long term follow up of clinical trials participants would also be enabled.

Recommendation 1: An agreed list of high value data collections which should be included in the national linkage system at Commonwealth and state/territory levels should be determined.

Which rules, regulations or policies create unnecessary or excessive barriers to linking datasets?

The legal framework

The legal framework in which data linkage occurs in Australia is complex and difficult to understand. The linkage of data collections generally requires the disclosure of personal identifiers to the data linkage unit. This attracts privacy protecting principles and provisions in three different bodies of law:

- The authorising legislation e.g. *Health Services Act (2016)* Western Australia.
- The common law duty of confidentiality
- Privacy legislation (Commonwealth, state and territory).

In addition, data users are often also bound by confidentiality or data use agreements.

This means that the legality of the linkage of any two data collections must be determined on a case by case basis for each data collection. This results in conservative and variable approaches to the policy and practice of data linkage. Data custodians' primary responsibility is to ensure that they comply with the law when considering requests for access to data for research. They may not always feel confident about decision making in such a complicated legal environment and may act cautiously. In particular there can be reluctance to link data between different sectors and jurisdictions.

Privacy regulation around Australia generally provides a framework for releasing data to data linkage units and researchers without consent, recognising the public interest in research. While this is important, it is also necessary for the relevant authorising legislation to allow the agency to release such information to researchers. In some cases, changes are required to authorising legislation to enable data to be released for research without consent.

The Privacy Act exemption for research using Commonwealth data only applies if the data is used for medical research. The Australian Law Reform Commission in the report "For your information Privacy Law and Practice (ALRC Report 108)" made the following recommendation:

Recommendation 65–2 *The Privacy Act should be amended to extend the arrangements relating to the collection, use or disclosure of personal information without consent in the area of health and medical research to cover the collection, use or disclosure of personal information without consent in human research more generally.*

This recommendation is supported as there appears to be no good reason to restrict research to “medical” research. We now understand the critical interplay of social and educational factors and health and there is no reason to believe that social research is of any less public benefit than medical research.

There are a number of approaches that could overcome the barriers that the legal framework causes. In the short term the provision of guidance materials and training to assist data custodian agencies to process requests for access to data would be worthwhile. In the medium term changes to legislation to clarify the use of specific data collections for research may be required. In the longer term Australia should consider a more uniform national approach.

Recommendation 2: The recommendation of the Australian Law Reform Commission to amend the Privacy Act to extend the arrangements relating to the collection, use or disclosure of personal information without consent in the area of health and medical research to cover the collection, use or disclosure of personal information without consent in human research more generally should be implemented.

Recommendation 3: Where required, changes to legislation to clarify the collection, use and disclosure of specific data collections for research should be made.

Approvals

All research projects involving linked data currently require multiple approvals. Approvals are required from the data custodian of each of the linked data collections included in the research project and at least one human research ethics committee (HREC). In practice more than one HREC often reviews the application and there may also be additional approval requirements in some jurisdictions e.g. the Public Health Act approval in Queensland. For projects involving data from several jurisdictions, 10 or more different approvals may be required. Authorising regimes for linked data projects in developed countries such as the United Kingdom and Canada are less onerous than those in Australia (see comments on lessons from overseas later in this section).

The PHRN has initiated discussions with the NHMRC and others with a view to progressing mutual recognition of ethics review for data linkage projects. There are also opportunities to streamline custodian approvals.

Recommendation 4: A system of national mutual recognition of ethics review of applications for research using linked data should be implemented. The current system for review of clinical trials applications may be able to be adapted for this purpose.

Recommendation 5: Governments and government departments should implement mechanisms to reduce the number of data custodian approvals required for research projects. Options might include approval at the department or whole of government level rather than at the level of the individual data collection.

Governance and institutional arrangements for statistical data integration involving Commonwealth data.

The PHRN supports the principle that linkage of data is best conducted by well-organised, secure data linkage facilities or centres. The High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes are supported. However, the current arrangements where specific datasets are routinely classified as high risk create a barrier to timely and efficient data linkage which is necessary to realise the full benefits of linked data. The limited extent of enduring linkage between key Commonwealth data collections is also a problem.

It appears that the interim accreditation arrangements of the Cross Portfolio Data Integration Oversight Board remain in place and that only Commonwealth agencies can seek accreditation as an Integrating Authority⁶. The consequence of these arrangements are that a very small number of Commonwealth agencies have been accredited for high risk linkage and very few research projects have been reported on the Public Register⁷.

The extensive network of dedicated data linkage units across Australia that are part of the PHRN and established with Commonwealth funding are not being used for State to Commonwealth data linkage. These data linkage units, some of which have been operating for more than ten years, have a strong track record of productivity, community acceptance and security. Accreditation of these state based data linkage units as Integrating Authorities would provide a system with much greater capacity, flexibility and competitiveness in effectively servicing the growing number of requests from the research and government sector for access to linked data. It should be noted that the Sixth interim report of the Senate Select Committee on Health includes a recommendation supportive of this approach.

How can Australia's government agencies improve their sharing and linking of public sector data? What lessons or examples from overseas should be considered?

The Senate Select Committee on Health's report, "Big health data: Australia's big potential"⁸ made a range of recommendations on how government agencies could improve their sharing and linking of data all of which are generally supported. In addition, the PHRN suggests the following activities which would also result in significant improvements:

- Enduring/routine linkage, NOT one-off/create and destroy linkage.
- Streamline data cleaning and extraction e.g. CARES⁹, SAIL¹⁰, Manitoba Centre for Health Policy¹¹. These approaches can include streamlined data custodian approvals.
- Clarify and make transparent approval criteria and processes for data custodians.¹²
- Partner with academic institutions to access analytic expertise.
- Partner with academic institutions to train analysts e.g. NSW Biostatistics Training Program.¹³

⁶ <https://statistical-data-integration.govspace.gov.au/topics/accreditation/> [accessed 26/07/2016]

⁷ <http://www.nss.gov.au/nss/home.NSF/disProjectsByAuthority?OpenView&Start=1&Count=1000&Collapse=1> [accessed 26/07/2016]

⁸ Senate Select Committee on Health. May 2016. Sixth Interim Report, Big health data: Australia's Big Potential.

⁹ Eitelhuber and Davis Health Information Science and Systems 2014, 2:6 <http://www.hissjournal.com/content/2/1/6> [accessed 25/07/2016]

¹⁰ <http://www.saildatabank.com/> [accessed 25/07/2016]

¹¹ http://umanitoba.ca/faculties/health_sciences/medicine/units/community_health_sciences/departmental_units/mchp/ [accessed 25/07/2016]

¹² Carolyn Adams and Judy Allen, 'Government Databases and Public Health Research: Facilitating Access in the Public Interest' (2014) 21 *Journal of Law and Medicine* 957;

¹³ NSW Biostatistics Training Program. Sydney, NSW (Aus). 2016 [cited 27/06/2016]. Available at <http://www.health.nsw.gov.au/training/botp/pages/default.aspx>

2.2 Private sector data

What private sector datasets should be considered high-value data to public policy; researchers and academics; other private sector entities; or the broader community? In each case cited, what characteristics define such datasets?

Private sector data is essential to establish a complete understanding of all aspects of the lives of Australians. Over 30% of expenditure on health and education in Australia comes from non-government sources. Linkage of private with public sector data enables a comprehensive picture to be drawn.

The characteristics of private sector data that could be of high value for data linkage is the same as described above for public sector data. In the health and education sectors in particular a significant proportion of services are provided by the private sector. Therefore, it is essential to have the data from private providers included and linked if a comprehensive understanding of the health, education and human service sectors and outcomes is to be obtained.

Some of the highest value private data collections that should be incorporated into the national data linkage system include:

- Private hospital data
- General practice data
- Pathology
- Radiology
- Dental
- Health insurance data
- Health and well-being data from other sources e.g. wearable devices.

Are there any legislative or other impediments that maybe unnecessarily restricting the availability and use of private sector data? Should these impediments be reduced or removed?

The Privacy Act exemption for research using private sector data only applies if the research is relevant to public health or public safety. This is similar to the issue explained above in Section 2.1. Recommendation 2 also applies to the use of private sector data.

What principles, protocols or legislative requirements could manage the concerns of private sector data owners about increasing the availability of their data?

The current privacy and security arrangements for data linkage through the PHRN are appropriate and no additional requirements should be necessary for private sector data. For example, private hospital data is already routinely linked in many jurisdictions.

Jurisdictional data linkage units should be used to encourage private organisations to link and share their data collections in a way that is safe, secure and allows the organisation to remain in control of what their data is used for.

What lessons or examples from overseas should be considered?

See examples above in Public Data.

2.3 Consumer access to and control over data

Consumers generally recognise the benefits of health research and support the use of data for this purpose. Consumers should be actively engaged in discussions about use of public and private sector data and involved in data governance processes as appropriate.

3 MANAGING THE COSTS

Questions on resource costs of access

Consistent approaches to setting costs/charging for access to public sector data should be considered. These should take into account the cost of providing access to data as well as the benefits to government and the wider community from making data available. Some funds are generally available from research grants to support data access. However, these funds are generally insufficient to cover current full cost-recovery.

Some countries such as the United Kingdom have invested heavily in infrastructure for data linkage and management. Examples include the FARR Institute¹⁴¹⁵ and the Administrative Data Research Network initiatives¹⁶. Australia has also made important investments in data linkage infrastructure through the NCRIS programs and related co-investments. These have helped to limit growth in access costs/charges.

Questions on privacy protection

It is important to achieve a balance between the right of people to control their personal information with the competing public interest in research. The PHRN's approach to linking data and providing access to linked data for research minimises the risks to individual privacy whilst enabling valuable research.

The PHRN uses the separation principle first described in 2002¹⁷. In this approach the data used for linking is restricted to the specific data items that are needed to ensure accurate data matching e.g. name, date of birth, sex and address. Specialised units (data linkage units) use these linkage variables to create linkage identifiers. The staff in the linkage units do not have access to the content information in each record e.g. diagnosis and treatment information. The content information remains with the original data collection. The data custodians of each collection are the only people who have access to the complete data collection. For specific approved research projects, the data custodians extract the required records for the project, remove the identifying information and replace with a project specific linkage key created by the data linkage unit. The researchers can then access the dataset created specifically for their project without identifying information.

The PHRN also includes a number of other facilities that assist in minimising risks to privacy. A secure, encrypted file transfer system (SUFEX¹⁸) enables data to be securely transferred

¹⁴ <http://www.mrc.ac.uk/research/initiatives/health-and-biomedical-informatics/initiatives-in-informatics-research/> [accessed 29/07/2016]

¹⁵ <http://www.mrc.ac.uk/news/browse/20-million-for-new-health-informatics-research-institute/> [accessed 29/07/2016]

¹⁶ Economic and Social Research Council. Annual Report and Accounts 2013-14. <http://www.esrc.ac.uk/news-events-and-publications/publications/corporate-publications/annual-report-and-accounts/>[accessed 29/07/2016]

¹⁷ Kelman C, Bass J, & Holman C. Research use of linked health data – a best practice protocol. ANZ J Public Health. 2002. 26:251-255.

¹⁸ SUFEX. Perth WA (Aus). 2016 [cited 27/06/2016]. Available from <http://www.phrn.org.au/for-researchers/services-for-researchers/sufex/>

between data custodians and data linkage units or data custodians and researchers. The Secure Unified Research Environment¹⁹ (SURE) provides researchers access to linked data through virtual workspaces. The data cannot be copied, downloaded or transmitted by email or other means. Researchers can take their analyses from SURE but not the original data.

The PHRN is a good example of how personal, sensitive data can be made available for research at a unit record level in a way that has minimal risks for individual privacy.

Questions on data security

Structured processes are required to provide for the reporting and investigation of alleged breaches. Proportionate penalties should be implemented for substantiated breaches. Public reporting of breach statistics would assist in risk management and support continuous improvement in data security.

¹⁹ SURE. Sydney, NSW (Aus). 2016 [cited 27/06/2016]. Available from <https://www.saxinstitute.org.au/our-work/sure/>