



HIGH VALUE DATA COLLECTIONS: PRIORITIES FOR DEVELOPMENT OF LINKED DATA RESOURCES IN AUSTRALIA

September 2017

Except for the PHRN logo and content supplied by third parties, this copyright work is licensed under a Creative Commons Attribution 3.0 Australia licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/3.0/au>. In essence, you are free to copy, communicate and adapt the work, as long as you attribute the work to the Population Health Research Network (PHRN) (but not in any way that suggests the PHRN endorses you or your use) and abide by the other licence terms.

NCRIS
National Research
Infrastructure for Australia

An Australian Government Initiative

The Australian Government has provided financial support to the PHRN through the National Collaborative Research Infrastructure Strategy (NCRIS).

CONTENTS

CONTENTS	2
EXECUTIVE SUMMARY	1
BACKGROUND	2
AIMS	3
METHODS	4
RESULTS	5
Survey	5
Desktop audit	9
DISCUSSION	10

EXECUTIVE SUMMARY

This study was funded by the Population Health Research Network (PHRN) as a national strategic initiative to identify data collections that researchers consider to be of high value for routine inclusion in Australian data linkage systems to enable more efficient multi and cross-jurisdictional data linkage.

The study involved an online survey of researchers and others interested in linked data conducted in May-June 2017. There were almost 300 responses to the survey.

Data collections currently linked by some but not all jurisdictions and identified by respondents as their top priority for ongoing inclusion in all linkage systems included:

- Cancer registry
- Births registry
- Ambulance services
- Private hospital admissions
- Mental health

Other data collections identified by respondents as their top priority for ongoing inclusion in the data linkage systems of Australian jurisdictions were, in order:

- Medicare Benefits Scheme (MBS)
- General practice records
- National Death Index (NDI)
- Australian Immunisation Register (AIR)
- Pharmaceutical Benefits Scheme (PBS)

The PHRN will examine what is required to achieve consistent linkage of the prioritised data collections and work with jurisdictions to achieve more consistent linkage. The PHRN is already supporting a project at the Australian Institute of Health and Welfare for enduring linkage of Commonwealth, state and territory data and survey results will inform this work.

The survey also provided an opportunity for respondents to comment on related issues. The majority of comments provided related to concerns about the numerous, complex governance and administrative processes required to obtain approval to access linked data and the time required for receipt of data once approvals were in place. The PHRN is aware of these issues and is working with partners to simplify project approval processes and improve the timelines for delivery of linked data for approved projects. The survey results will assist with this process.

BACKGROUND

In Australia, health and human services data is collected by different government, non-government and private organisations across multiple jurisdictions. All jurisdictions (i.e. Australian, state and territory governments) now have the capacity to link statutory and administrative health and other human services data in privacy preserving ways and make it available for approved research. However there is limited consistency across the jurisdictions in the data included in the data linkage systems.

There are variations across jurisdictions in the years of available linked data and in data quality and/or the variables available for research. Currently, only public hospital admitted patient and emergency department, cancer registry, births and death data are consistently collected and routinely linked in every State/Territory. This lack of consistency for other significant data collections is a source of frustration for researchers undertaking cross or multi-jurisdictional research.

The PHRN is a national network with core funding from the National Collaborative Research Infrastructure Strategy (NCRIS). Its remit includes provision of national leadership in developing the infrastructure for researchers to access linked datasets across Australia for health and human services research purposes. This study was funded by the PHRN as a strategic initiative and led by SA NT DataLink, with support from the PHRN Program Office.

AIMS

The aim of the study was to identify an agreed minimum set of data collections to be consistently included in each of the PHRN members' data linkage infrastructure in order to support projects conducted at national or multi-jurisdictional levels as well as for international collaborations that are valued by researchers and government.

Specific objectives included:

1. Provide a better understanding of the data collections that researchers across Australia consistently regard as being of the highest value to them for cross or multi-jurisdictional research; and
2. Identifying which data collections the PHRN should prioritise to make consistently available based on these findings.

METHODS

An on-line survey was circulated using the list of researchers registered with the PHRN. In addition, each PHRN member organisation in the state, territory and Commonwealth jurisdictions was also asked to circulate the survey to other partner organisations and any researchers or researcher based organisations that they could contact. Other national associations were also asked to circulate the survey invitation. The survey was conducted over four weeks, from 29 May 2017 to 23 June 2017.

The survey consisted of four parts:

Part A: Respondents were asked about their organisation type, jurisdiction, role and experience with data linkage.

Part B: Respondents were asked to rank the data collections currently linked by some but not all jurisdictions in order of preference (1 to 5) for consistent linkage across all jurisdictions in Australia

Part C: Respondents were asked to rank in order of preference (at least 5, but up to 10) other data collections they considered should be prioritised for consistent linkage across all jurisdictions.

Part D: Future focus for linked data over the next three years.

Respondents were also able to provide further comment under each part.

A desktop audit to confirm the data collections that are routinely linked by states and territories was also undertaken.

RESULTS

Survey

1. Survey responses

274 respondents fully completed the survey. A further 23 respondents partially completed the survey.

Survey section	No. of responses
Part A	287
Part B	274
Part C	262
Part D	263

2. Respondent information

Part A - Respondents background

Respondents' location

All jurisdictions were represented in the responses. The jurisdictional location of respondents is shown in Chart 1 below.

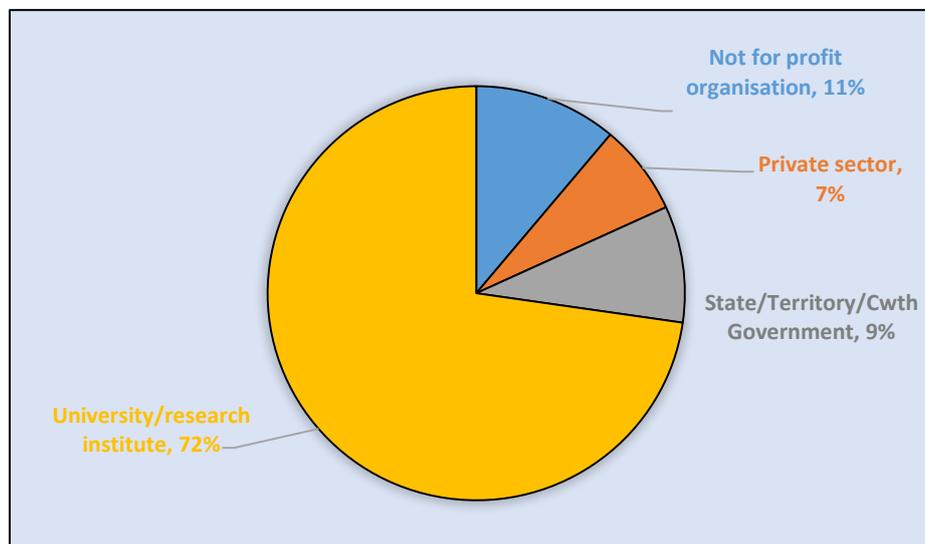
Chart 1 Percentage of respondents by jurisdiction



Organisational sector of respondents

The majority of respondents (72%) were from a university or research institute (see Chart 2 below).

Chart 2: Organisational sector of respondents



Experience in data linkage of respondents

Of the respondents, approximately 59% said they were experienced data linkage users, 23% were new users and 15% had never used data linkage. The remainder said it was not applicable.

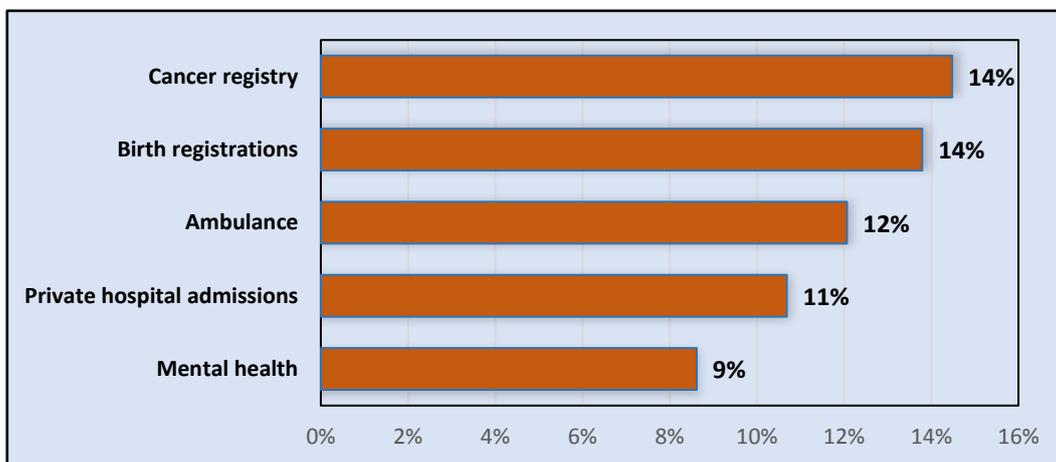
3. Priorities for consistent linkage

Part B - Data collections currently linked by some (but not all) jurisdictions

Chart 3 shows top five first preferences selected by all respondents for the data collections that they consider should be routinely linked by all jurisdictions and available for research using linked data.

These data collections were confirmed as important when respondents' first, second and third preferences were considered. The proportion of respondents who identified these data collections within their top three preferences was as follows: private hospital admissions (40%), mental health (32%), cancer registry (29%), birth registrations (27%) and ambulance (22%).

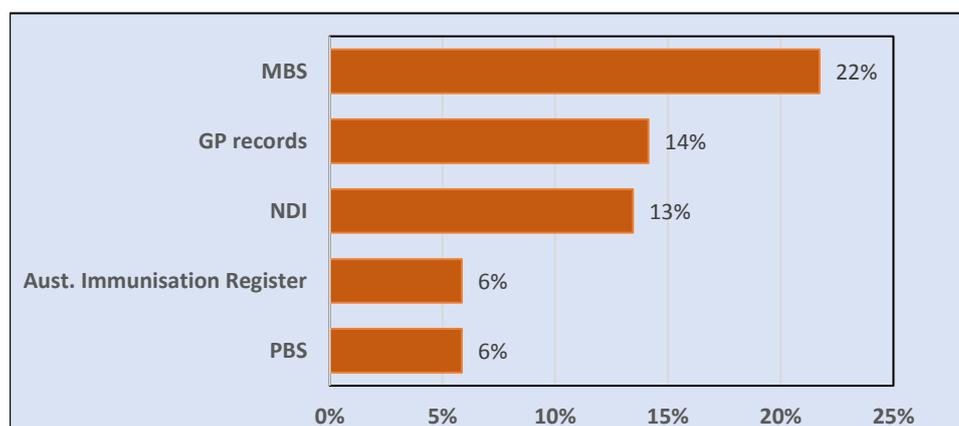
Chart 3 - Top five 1st preferences for consistent linkage for all jurisdictions



Part C - Other data collections not routinely linked by any or linked by only one jurisdiction

The Chart 4 below show the data collections that respondents thought should be considered as a priority for routine linkage by all jurisdictions.

Chart 4: Top five 1st preferences for priority linkage for all jurisdictions



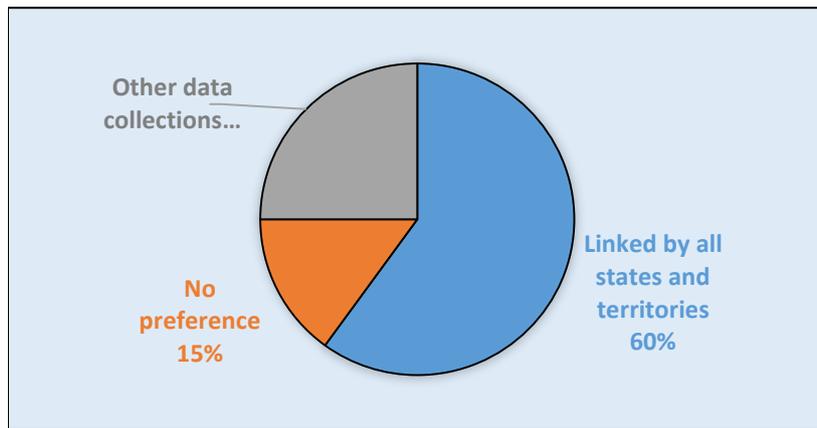
These data collections were confirmed as important when respondents' first, second and third preferences were considered. The proportion of respondents who identified these data collections within their top three preferences was as follows: MBS (55%), PBS (42%), GP records (34%), NDI (29%) and AIR (13%).

4. Future focus for data linkage

Part D – Focus for data linkage activities over the next three years

Preference for the focus of PHRN data linkage activities over the next three years is shown in Chart 5 below. A majority of respondents (60%) supported a focus on routine linkage of the priority data collections as nominated in Part B.

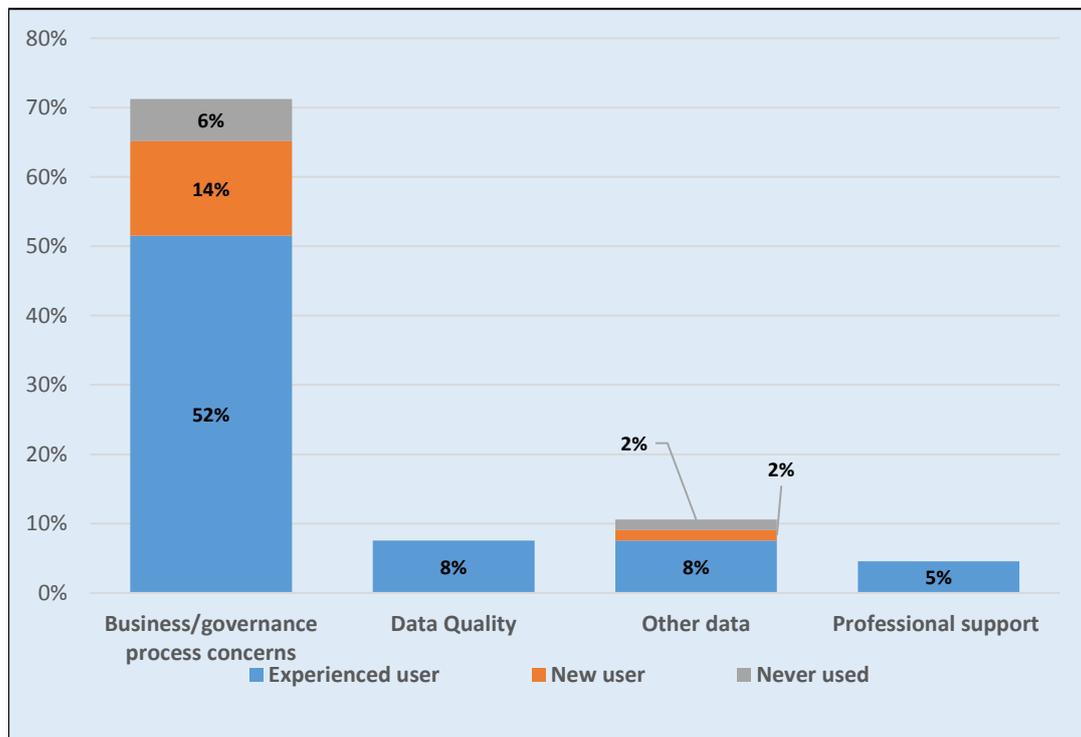
Chart 5: Preferred focus of data linkage activities



5. General Comments

Chart 6 shows the particular concerns identified by the researchers, as a percent according to the researcher's experience using data linkage.

Chart 6: Business/governance process, data quality and other concerns by user status



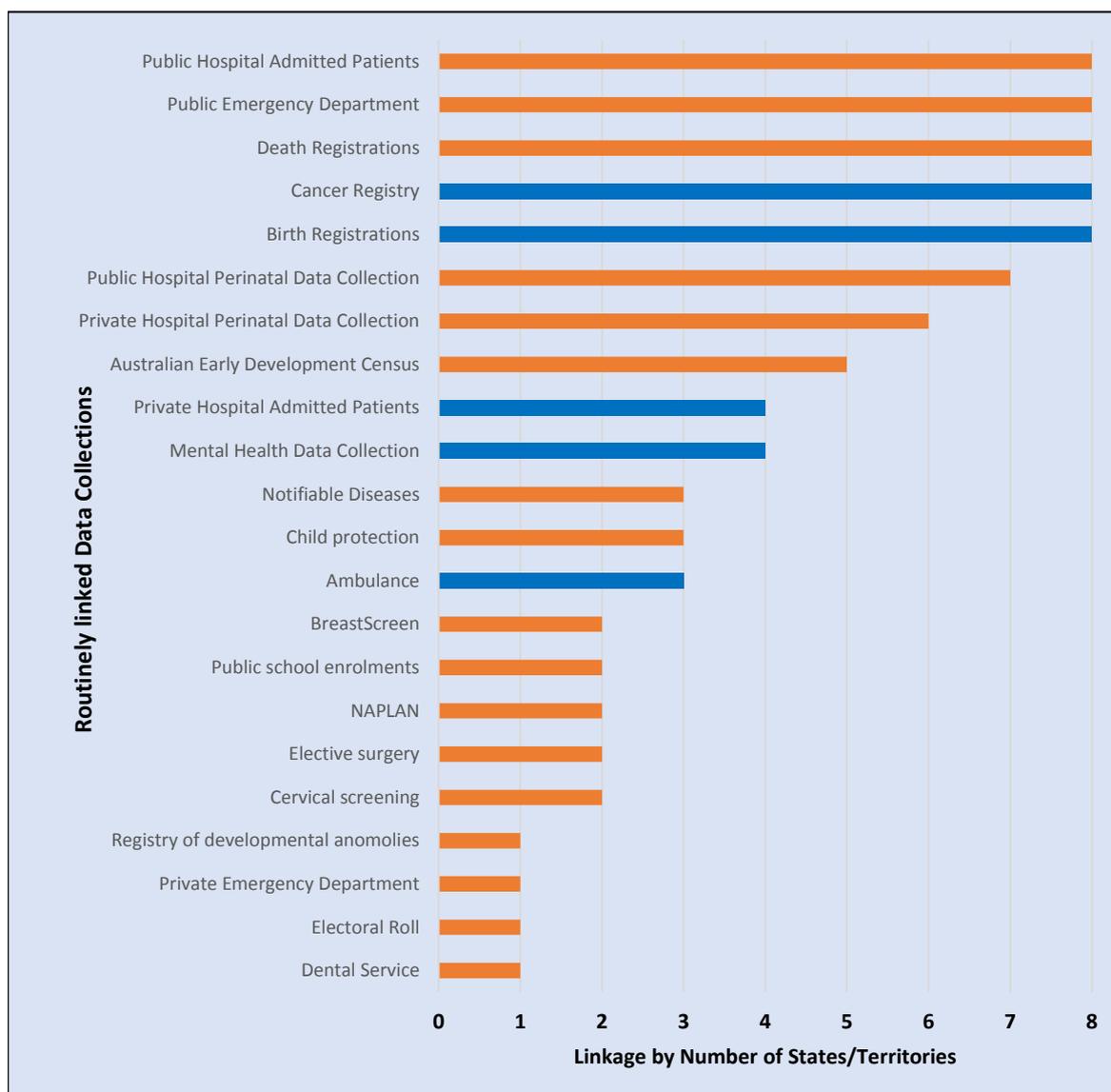
Desktop audit

6. Data Collections Routinely Linked by States/Territories

A desktop audit of data collections routinely linked by the PHRN state and territory data linkage units was conducted (Chart 7). The results were compared to the priorities identified by respondents in the survey.

The audit showed that state/territory cancer registries and birth registrations nominated as a priority for linkage by the respondents are already linked by all state/territory data linkage units. However, the remaining data collections identified in the top five priorities (ambulance, private hospital admissions and mental health - shown in blue) are not yet linked by all units.

Chart 7: Data routinely linked by states/ territories (as at June 2017)



DISCUSSION

There was strong support from the respondents for consistent linkage of data collections already linked in some jurisdictions to be linked across all jurisdictions with a focus on:

1. Cancer registry
2. Birth registrations
3. Ambulance
4. Private hospital admissions
5. Mental health

This result indicates that many respondents were unaware that cancer registries and birth registrations are already linked by all state/territory data linkage units. The PHRN will look at opportunities to better promote the extent of the linked data resources already available.

With respect to the other prioritised data collections, the PHRN will examine what is required to achieve consistent linkage of ambulance, private hospital admissions and mental health data by all states and territories and work with jurisdictions as appropriate to achieve more consistent linkage.

Researchers identified a number of other data collections that they considered to be of high value and which they considered should also be made more readily available. These in ranked order were:

1. Medical Benefits Scheme (MBS)
2. General practice collections
3. National Death Index (NDI)
4. Australian Immunisation Register (AIR)
5. Pharmaceutical Benefits Scheme (PBS)

The Australian Institute of Health and Welfare (AIHW) is currently able to link the MBS, NDI, AIR and PBS data collections. The PHRN has funded a project for AIHW to collaborate with states and territories to establish enduring national master linkage keys for routine linkage of state/territory data collections with these Commonwealth data collections. While the processes involved are complex, good progress is being made.

Results from the survey highlight the importance of linking general practice data with other jurisdictional data. Linkage of general practice data is challenging as Australia does not currently have large population level general practice data collections that can be linked with other data. However some jurisdictions have made progress with linkage of general practice data.

While not related to a specific question in the survey, researchers expressed very strong concerns about the complex and numerous governance and administrative processes related to ethics, application and data approvals as well the time required for receipt of data once approvals were in place. The PHRN is working to address this issue at national and

jurisdictional levels. For example, the PHRN is working within the existing limitations and federated data governance and approval processes to improve application and approval processes, particularly for cross and multi-jurisdictional projects. The survey results are being used to support work at the national level to minimise duplication of ethics review.

The PHRN will also use the researchers' comments to support its arguments for the need for change, bringing to the attention of the Australian Government the findings in this area as being strongly supportive of the findings and recommendations of a number of recent inquiries¹²³.

¹ Senate Select Committee on Health. *Big health data: Australia's big potential* (Sixth interim report), Parliament of Australia, Canberra, 4 May 2016. See http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Health/Health/Sixth_Interim_Report

² Productivity Commission. *National Education Evidence Base*, Melbourne, 24 May 2017. See <http://www.pc.gov.au/inquiries/completed/education-evidence/report>.

³ Productivity Commission. *Data Availability and Use*, Melbourne, 8 May 2017. See <http://www.pc.gov.au/inquiries/completed/data-access/report>