FINAL REPORT 2024

Which data is best for my research?

Comparing national linked data assets to cross-jurisdictional linked source data

POPULATION HEALTH RESEARCH NETWORK

Mikhalina Dombrovskaya
Erika Hagemann
Felicity Flack

Table of Contents

Contents

1. Executive Summary

1.1. Introduction

The Australian health system's complexity and federated governance have historically created challenges for traditional cross-jurisdictional data linkage, where two or more datasets from different jurisdictions are linked for one-off research projects. Prior to the Population Health Research Network (PHRN) established in 2009, researchers had limited access to such data. Although the situation has improved, obtaining access remains burdensome and time-consuming. Recently, the Australian Institute of Health and Welfare (AIHW) and the Australian Bureau of Statistics (ABS) have developed national linked data assets, which could streamline researchers’ access to cross-jurisdictional data.

1.2. Aims

The aims of the consultation were to:

1. Understand and explain in what circumstances researchers should use certain national linked health data assets versus traditional cross-jurisdictional linked health data.
2. Assist researchers to make informed decisions about which source of cross-jurisdictional linked health data will best meet their needs by providing authoritative, agreed information on access to and use of the linked health data.
3. Develop recommendations and opportunities to better assist researchers with making informed choices with respect to the available data sources for linked data research.

1.3. Scope

The consultation, conducted between February and September 2024, focussed on academic and private sector access to cross-jurisdictional linked health data. Data collections included the National Health Data Hub (NHDH), the Person Level Integrated Data Asset (PLIDA), and equivalent state/territory data collections for hospital admissions and emergency department (ED) presentations. Report findings are for client services teams and those providing guidance to researchers interested in or undertaking linked data research.

1.4. Methodology

Online resources from relevant agencies such as AIHW, ABS, PHRN, and various state/territory data linkage units (DLUs) were reviewed and further discussed with subject matter experts (SMEs). Criteria were developed to assess the differences between national linked data assets and equivalent state/territory datasets linked using traditional cross-jurisdictional linkage. End-user consultation, led by Future Friendly, included end-user (researcher) interviews that focused on gaining clarity, discovering underlying assumptions, and articulating the decisioning process for researchers embarking on a national project using linked data.

1.5. Key Findings

- **Insufficient Online Resources**:
  In-scope websites lack sufficient information for informed decision making regarding the most suitable data sources for cross-jurisdictional linked data projects. There is minimal cross-referencing of alternative data sources on these websites.
- **Variability and Dynamics of Data Sources**:
  Significant differences exist in the type, scope, and range of data available from NHDH, PLIDA, and equivalent state/territory data sources for linked data projects, particularly regarding data collections and year ranges available. Datasets continue to be added to linked national data assets, increasing scope.
- **Decision-Making Criteria for Researchers**:
  Researchers consistently consider available datasets, cohort sizes, costs, and timeframes when making decisions. They prefer the ‘path of least resistance’ when both bespoke and national linked data assets can address their research questions.
- **Quality Preservation**:
  Researchers are unwilling to make trade-offs that compromise the quality of their project outcomes.

1.6. Recommendations and Next Steps

Next steps are to develop resources to facilitate rapid assessment of researchers' project outcomes and dataset needs and help guide decision making.

For researchers:
- targeted website/online content which clearly outlines available datasets and variables
- a ‘first time engaging’ questionnaire to uncover project specific information
- an online digital dataset comparison tool to allow independent, self-directed exploration
- an online dynamic estimation tool to assess project costs and timeframes based on what datasets and parameters are selected

For Client Services Officers (CSOs):
- a standardised discussion guide to help CSOs systematically uncover the overall goals and scope of researchers' projects
- training for CSOs to enhance intentional researcher engagement, identify non-negotiable criteria and document source use cases effectively

These key resources will equip researchers and CSOs to articulate key considerations and trade-off impacts during the initial project consultation phase.

1.7. Conclusion

The consultation highlighted differences in the breadth and depth of data available across a range of data sources for cross-jurisdictional linked data projects. Researchers need support and appropriate resources to reduce complexity, promote clarity and manage expectations to allow informed decision making with respect to their linked data research needs. This support should be provided both online and through direct engagement with client services.

2. Introduction

2.1. Problem

Navigating the complexities of the Australian federation and its health system has made cross-jurisdictional data linkage a significant challenge. Established in 2009, the Population Health Research Network (PHRN) supports world-class research through ad hoc, bespoke cross-jurisdictional linkage of administrative data. The introduction of new, pre-linked data assets has made providing support more complex. PHRN currently lacks an understanding of how researchers make decisions about different options to access cross-jurisdictional data. This consultation aimed to understand the current state, gain insights, and identify key opportunities to better support researchers and Client Services Officers (CSOs).

2.2. Data sources

2.2.1. National linked data assets (NLDAs)

NLDAs are pre-linked and contain datasets from state, territory, and Commonwealth sources. These assets depend on stakeholders, including government agencies, to provide source data. Researchers can access linked data for various cross-jurisdictional projects, provided they secure the necessary ethics and governance approvals. In some cases, ad hoc linkage to national linked data assets may be possible when a project requires the integration of specific study cohorts or administrative datasets.

2.2.1.1. National Health Data Hub (NHDH)

The NHDH, previously known as the National Integrated Health Services Information (NIHSI), is a key NLDA managed by the Australian Institute for Health Welfare (AIHW). It includes extensive data collections, notably hospital admissions and emergency department (ED) presentations data from state/territory sources, as well as Commonwealth data sources covering immunisations, disability, deaths, pharmaceuticals, aged care and health utility topic areas.

2.2.1.2. Person Level Integrated Data Asset (PLIDA)

The PLIDA is administered by the Australian Bureau of Statistics (ABS), formerly the Multi-Agency Data Integration Project (MADIP), is a secure NLDA that integrates information on health, education, government payments, income, taxation, employment, and population demographics over time. PLIDA offers comprehensive insights into various population groups in Australia, examining how their characteristics interact with service usage and outcomes like health and employment improvements.

2.2.2. State/territory source data collections

Researchers can request state and territory data collections for both single and cross-jurisdictional ad hoc linked data projects. These collections may also be linked to other state/territory and Commonwealth datasets through higher-level agreements between institutions and government organisations, creating assets like the NHDH and PLIDA. Traditional cross-jurisdictional data linkage projects involve bespoke linkages between two or more source data collections from different jurisdictions or at least one state/territory source data collection and one Commonwealth dataset. These linkages are typically project-specific, one-off arrangements; they are not retained for future use, and ethics and governance approvals are also tailored to individual projects.

2.3. Aims

The consultation aimed to:

Clarify how researchers interact with data:

- Understand what criteria are most important to researchers when selecting a data source.
- Understand what trade-offs researchers make when selecting their data.
- Clarify which use cases benefit most from bespoke cross-jurisdictional linkage versus NLDAs.

Enhance support for researchers and CSOs:

- Develop recommendations to better assist researchers in navigating the available data sources for cross-jurisdictional linked data research.

2.4. Scope

The consultation focused on access to cross-jurisdictional linked health data for both academic and private sector researchers.

Data Assets/Collections:

- NHDH
- PLIDA
- State/Territory source data collections (excluding Western Australia and Northern Territory as these are available from the jurisdictional linkage units but not from the NHDH):
  - Hospital admissions
  - Emergency department presentations

Subject matter experts (SMEs):

Data custodians and their delegates, and representatives from ABS, AIHW and participating DLUs were invited for interview (see Appendix A for details).

Researchers (end-users):

Researchers were invited for interview if currently planning or undertaking a data linkage research project, whether using data from one or more jurisdictions, or a NLDA. Participants represented a diverse range of experience levels, from early and mid-career researchers to seasoned experts in data linkage.

3. Our Approach

The consultation started in February 2024, led by the PHRN National Office. The PHRN Chief Executive served as the Project Sponsor, with the Senior Manager of Strategy and Services acting as Project Lead, and the PHRN Coordinator for Client Services managing the consultation processes.

The consultation comprised three phases: **desktop research**, **SME consultation**, and **end-user (researcher) interviews** — each lasting about 2-3 months.

3.1. Desktop Research

To lay the groundwork for deeper discussions with Subject Matter Experts (SMEs), the PHRN Coordinator, Client Services, conducted thorough desktop research which involved:

- **Collecting Data**: Sourcing information on in scope data sources by exploring AIHW, ABS, PHRN, and various government websites (see Appendix B).
- **Assessing Similarities and Differences**: Developing criteria to evaluate the distinctions and commonalities between data sources (see Appendix B).
- **Spotting Gaps**: Highlighting areas where information was unclear or lacking.
- **Follow-Up Identification**: Identifying questions for SMEs (e.g. data custodians) to confirm the accuracy of publicly available information and address knowledge gaps.

3.2. SME Consultation

In this phase, the PHRN reached out to stakeholders to nominate liaisons for interviews. SMEs received a one-page project summary and a Phase 1 findings template related to their datasets. Before the interviews, they were encouraged to review and add any relevant information.

Each SME participated in a one-hour online interview with the PHRN Coordinator, Client Services. SMEs clarified details and gave insights where previous data fell short. Finalised criteria were used as a starting point for Phase 3, end-user interviews.

The PHRN Coordinator, Client Services

- **Consolidated Findings**: Incorporating insights from SMEs on their data sources.

3.3. End-user Researcher Interviews

Future Friendly (FF), an independent design and research firm, led the end-user consultation. Their goal was to uncover researcher needs and decision-making processes through interviews and testing sessions. Working closely with PHRN and SMEs, FF adopted a co-design approach to create high-fidelity interactive prototypes for testing concepts and understanding what researchers value most in their data source decisions.

FF conducted two rounds of 1:1 qualitative testing to gain first-hand insights, guided by interview scripts aligned with defined learning goals.

Round 1 focused on:

- Researchers' familiarity with PHRN
- Data-specific information needed for decision-making
- The impact of operational criteria on researchers' choices

Round 2 aimed to explore:

- How project type and outcomes influence data decisions
- Researchers' trade-offs in data selection
- Relevant use cases for different data pathways

4. Our Learnings

Researcher consultation led by FF sought to: uncover what is most important to researchers when making a data decision; understand what trade-offs researchers make when selecting their data; and clarify which research questions benefit most from bespoke cross-jurisdictional linkage or national linked data assets.

- Accessing cross-jurisdictional data is inherently complex, regardless of the source.
- Researchers make decisions through a consistent stratum by understanding:
    - the available datasets and their variables
    - the available cohort size within these datasets (through temporal and geographic scope)
    - the accompanying cost and timeframes
- There are notable differences in the type, scope, and range of data available from NHDH, PLIDA and state/territory data sources for linked data projects.
    - particularly data collections and year range for the NLDAs in comparison to the state/territory data sources. State/territory data sources of hospital admissions and emergency department presentations are more current

        and extensive (year range available) and comprehensive (variables available) than what is included in the NLDAs.

- There is insufficient information publicly available for researchers to make an informed decision on which data source/s may best serve their cross-jurisdictional linked data project.
- There is limited information publicly available for researchers conducting cross-jurisdictional linked data projects
    - to easily quantify costs
    - pertaining to governance, application, and approval processes
- Based on SME responses, timeframes for projects accessed through NHDH and PLIDA would appear generally shorter compared to state/territory counterpart datasets using traditional cross-jurisdictional linkage.
- Where both bespoke cross-jurisdictional data linkage and national linked data assets can appropriately answer a research question, researchers will prioritise the 'path of least resistance'.
- Data custodian and ethics approval remain a requirement for accessing hospital admissions and ED presentations data, whether directly from states/territories, or from the NHDH. While data custodian approval is required for PLIDA data access, requirement for ethics approval is assessed on a case-by-case basis.
- Researchers will not make trade-offs which negate the quality of their project outcome.

5. Insights

- To support decision making, there needs to be a layer of abstraction for the complexity of accessing cross-jurisdictional data.
- Researchers require trusted mentors and guides.
- When engaging with researchers, DLUs and Client Services Officers should
    - answer their foundational questions and look to meet their non-negotiable criteria first and foremost
    - provide the right kind of information (relevant and accurate), to enable educated trade-offs whilst maintaining overall project integrity

6. Recommendations and Next Steps

Next steps are to create resources to facilitate the rapid assessment of researchers' project outcomes and dataset needs. This can include an online self-assessment tool for researchers and semi-structured, guided discussion with Client Services Officers.

6.1. Enhance Website and Researcher Questionnaire:

- Update website content to clearly outline available datasets and their variables. Implement a 'first-time engaging with PHRN' questionnaire to gather specific project information that aids decision-making.

- Introduce a digital dataset comparison tool that allows researchers to compare options independently and at their convenience.
- Provide dynamic estimations of project costs and timeframes based on selected datasets and data sources, highlighting the impact of data modifications.

6.2. Standardise Discussion Guides for Client Services Officers:

- Develop a standardised discussion guide to help Client Services Officers systematically uncover the overall goals and scope of researchers' projects.
- Train Client Services Officers to enhance intentional engagement with researchers, enabling them to identify non-negotiable criteria and document source use cases effectively.
- Equip Client Services Officers to articulate key considerations and trade-off impacts during consultations with researchers.

7. Conclusion

The consultation highlighted differences in the breadth and depth of data available across a range of select data sources for cross-jurisdictional linked data projects. Researchers need a trusted mentor and guide to reduce the complexity for them, and provide resources to enable informed decision making. Specifically, this trusted mentor role needs to:

- ask questions in the order that researchers have them
- provide clear and easy to understand use-cases for when each data source might be most valuable
- be clear and upfront about expectations and processes so that researchers can make their own informed decision

This support should be provided both online and through direct engagement with client services.

8. Acknowledgements

PHRN is a national research infrastructure capability with core funding from the Australian Government’s National Collaborative Research Infrastructure Strategy. The University of Western Australia is lead agent for the PHRN. The PHRN National Office would like to thank the end users and subject matter experts who participated in review of findings, interviews, and concept testing. The National Office would also like to thank the PHRN Board and Participant Council for their review and support.

Appendix A

Subject Matter Experts

Data asset/collection	Organisation	SME/s
NHDH (NIHSI)	AIHW	Director, NHDH (data custodian); NHDH Unit team members
PLIDA	ABS	Program Manager, Data Integration Services; Director Data Linkage Centre, Data Integration Services
Hospital admitted data ED presentations	ACT	Director, Epidemiology, ACT Health
	CHeReL (NSW)	Principal Data Integration Manager; Manager Client Services; Data Linkage Team members
	CVDL (Vic)	DLU Client Services Officers; DLU Content Team Manager
	DLQ (Qld)	DLU Client Services Officer; Manager, Data Access, Governance and Partnerships
	TDLU (Tas)	DLU Manager; DLU Operations Manager
	SA NT DataLink (SA)	DLU Client Services Officers; Associate Director SA NT DataLink

Appendix B

The websites listed below were accessed to collect data on the NHDH (formerly NIHSI); PLIDA (formerly MADIP); and the hospital admissions and ED presentations data collections from Queensland, New South Wales (NSW), Australian Capital Territory (ACT), Victoria, South Australia (SA), and Tasmania.

- AIHW NIHSI website
- AIHW NHDH website
- ABS PLIDA website
- ABS website, The Australian Statistician Speeches
- PHRN Metadata Platform
- NSW Centre for Health Record Linkage (CHeReL) website
- Data Linkage Queensland (DLQ) website
- Centre for Victorian Data Linkage (CVDL) website
- SA NT DataLink website
- Tasmanian Data Linkage Unit (TDLU) website

Criteria on which different types of data assets/sources were compared were grouped into themes including:

- Data availability – including dataset(s), jurisdictions in scope, year range, update frequency, data granularity, and variables
- Underlying infrastructure
- Services available – including ad hoc linkage, family linkage, control/comparison group selection, and use of privacy preserving record linkage
- Metadata – including data dictionary links
- Governance, application and approval processes – including ethics requirements
- Eligibility and conditions – including researcher eligibility, access and storage requirements, outputs and dissemination
- Costs – including application, data linkage, extraction, access, and storage
- Timeframes – including for approvals, data provision and output review
- Future – including additions and changes in coming years