

International Journal of Population Data Science

Journal Website: www.ijpds.org



Swansea University
Prifysgol Abertawe

Discovering linked data collections through a new national metadata platform

Kate M. Miller^{1,*}, Felicity S. Flack¹, Merran B. Smith¹, Carina Ecremen Marshall¹, and Vicki Bennett²

Submission History

Submitted:	20/10/2024
Accepted:	27/01/2025
Published:	30/04/2025

¹Population Health Research Network, University of Western Australia, Crawley, Western Australia, Australia

²Australian Institute of Health and Welfare, Canberra, Australian Capital Territory, Australia

Abstract

Background

Metadata plays a crucial role in the health research infrastructure ecosystem. Despite the abundance of metadata for data collections in Australia, the vast and diverse data custodian landscape poses challenges for linked data researchers to find relevant information for multiple data collections, often making it an arduous and time-intensive task.

Methods

The project comprised three phases: an initial scoping exercise to understand the current state of metadata and related best practice; a national consultation involving researchers, data linkage staff and data custodians to develop a high-fidelity prototype of a metadata platform; and a final build and implementation phase. The platform underwent several prototyping and testing cycles to refine the digital experience.

Results

Expert interviews confirmed that there is a wealth of metadata available, but it is difficult for researchers to access and evaluate. Consultations with researchers identified opportunities to standardise metadata across collections and provide a centralised platform to enhance the discoverability of data collections for research using linked data. High value platform features included searching, browsing and filtering capabilities, data item list metadata, standardised formats, sample data, and frequently asked questions. The final design and functionality reflected user consultations and data custodian input on feasibility.

Conclusion

The Population Health Research Network developed a metadata platform to enable researchers to evaluate the suitability of Australian data collections for linked data projects more effectively. The platform has standardised the way in which metadata is presented for data collections nationally. Improved metadata quality, readability and accessibility will save time and enhance the quality of applications for linked data.

Keywords

metadata; data linkage, record linkage; administrative data; data collections; health

*Corresponding Author:

Email Address: kate.miller@uwa.edu.au (Kate Miller)



Introduction

Australia, as a federation comprising eight states and territories, has a complex healthcare system delivered by a mix of service providers from a range of organisations, including national and state and territory governments and the non-government sector. This means that data about health service delivery is captured by a wide range of stakeholders [1]. Each of these organisations has a role to play in collecting, collating and making available administrative health data for a range of purposes. However, their ability to do this in a consistent and comparable way varies.

There is a growing need and understanding of the importance of linking many of these data collections for use in research to understand patient pathways through the health system, to inform good public policy decision making regarding health service delivery, and to understand the outcomes of treatments.

The complexity of research projects and the nature of data linkage means that researchers often request the linkage of data collections managed by data custodians across this diverse range of organisations and agencies. Consequently, there can be substantial variations in the quality, completeness, and update frequency of metadata made accessible to researchers.

For researchers to know what data to use and how to interpret these, it is vital that they have a way of discovering what data collections are available for linkage, what data items they contain, and what those data items mean. This can only be achieved if there is good metadata available about the different data collections that can be used for this purpose that is described in a clear and consistent way. While Australia does have some relevant metadata standards published at a national level, these are not necessarily applied or applicable across all data collections and jurisdictions in a federated county such as Australia.

The Findability, Accessibility, Interoperability, and Reuse (FAIR) principles (FAIR Principles - GO FAIR (go-fair.org) emphasise “machine-actionability (i.e. the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data”.

The Population Health Research Network (PHRN) set out to create an online environment that would help deliver on the FAIR principles for researchers seeking to access and use health data collections for data linkage research.

Methods

Phase 1 Scoping exercise and consultation with international metadata providers

The first phase of the project involved a desktop review of the range of metadata methods and tools currently in use for evaluation of data collections, and to scope international best practice and innovation in metadata for linked administrative and population level data.

To enhance understanding and inform the project, a national and international scoping exercise was conducted to identify existing metadata registries. This involved documenting key features of these registries and engaging in interviews with key contacts from international metadata providers. Consultations focused on the development process, exploring aspects such as consultation processes, engagement strategies with end users, the decision-making process for features, and practical value as perceived by end users. We sought insights from evaluations conducted by these providers to discern the attributes most valued by end users and to gather feedback on the functionality of the platforms. The focus extended beyond the initial development to encompass ongoing maintenance, with an emphasis on key learnings that could be incorporated into the next phase of the PHRN metadata project.

Phase 2 Metadata platform consultation and prototyping

The PHRN engaged a strategic design consulting company, Future Friendly, to design, prototype and validate the Metadata Platform.

Expert interviews

Future Friendly conducted a series of initial interviews with experts from medical research institutes, universities, and government agencies to understand what Australian linked data researchers need to effectively explore and evaluate data collections. Interviews centered on the current state of metadata for linked health data collections within Australia, existing pain points in the data linkage process, and key opportunities for the proposed new metadata platform.

A targeted recruitment strategy was used to identify experts in the following areas from across Australia:

- Metadata
- Data custodians of commonly linked data collections
- Client services
- Researchers with experience in working with linked data

Development of a high-fidelity prototype

A high-fidelity prototype was developed to test and validate the attributes for inclusion, key features, functionality, and design of the metadata platform.

The first iteration of the high-fidelity prototype was a modified version of the Health Data Research UK model. Early modifications incorporated learnings from international organisations who had developed metadata registries, as well as additional features to address known pain points in the existing business processes and the needs of the end users identified in the initial expert interviews. The initial selection of metadata attributes was aligned with the Office of the National Data Commissioner Core Metadata Attributes for Data Inventories [2].

User testing sessions

The high-fidelity platform was tested through one-on-one interviews conducted via MS Teams. Researchers were current users of PHRN infrastructure or were referred by members of the Project Control Group or Australian data linkage units. Researchers received a link to the prototype and were requested to share their screens. Interviewers then guided participants to navigate through the platform, encouraging them to vocalise their thoughts in real-time. This approach yielded valuable insights into researchers' preferences and expectations, identified missing metadata attributes, assessed the platform's intuitiveness, and provided constructive feedback on design and features.

The platform was tested with researchers and client services officers from each of the PHRN data linkage units across three rounds of concept testing. The suitability of the Health Data Research UK utility framework for Australian users was also explored within these sessions. Iterations to the prototype were implemented after each week of testing. At the completion of user testing, a workshop was held with data custodians to assess the platform for feasibility.

Metadata acquisition

User testing sessions validated the selection of metadata attributes for inclusion in the platform. A metadata template was created and populated by the PHRN National Office with metadata sourced from relevant websites (e.g. government agencies and data linkage units) and internal documents provided by data linkage units or developed anew. This template was subsequently distributed to client services staff who collaborated with data custodians to revise, finalise, and authorise the metadata for their respective data collections.

Phase 3 Development of the metadata platform

In Phase 3, the centralised discovery platform was completed to improve the exploration and identification of data collections suitable for linkage. The development of the metadata platform from the high-fidelity prototype was conducted using an agile project management approach. This platform features an online data catalogue and accompanying metadata. During this phase, key decisions were made about the technical approach, including the selection of a content management system and hosting arrangements.

Following confirmation of the content, the metadata platform was launched in September 2023 at <https://metadata.phrn.org.au>.

Project control group

A Project Control Group was established to provide guidance and direction for Phases 2 and 3 of the project. The Project Control Group included representatives from each of the PHRN data linkage units, a researcher, a data custodian, a representative from the PHRN National Office and a representative from the Australian Institute of Health and Welfare (AIHW). The Project Control Group ensured that the consultation process was appropriate, informed the

development of the template to standardise the collection of metadata nationally, and provided timely input and feedback on the high-fidelity prototype.

Results

Phase 1 Scoping exercise and consultation with international metadata providers

Australia

Within Australia, the Metadata Online Repository (METEOR), administered by the AIHW, contains metadata standards for the health, aged care, community services, early childhood and housing and homelessness sectors (meteor.aihw.gov.au). METEOR operates as a metadata registry designed to support a disciplined approach to the development, storage and management of metadata, compliant with the international information modelling standard ISO/IEC 11179 released in 2003 [3]. METEOR includes National Minimum Dataset specifications, Dataset Specifications, a National Data Dictionary, and information about metadata standards. It does not, however, contain metadata from subnational government agencies' data collections, thus the need to create an environment that allows a broader range of data to be discovered.

International

The international environmental scan identified Health Data Research UK and Health Data Research Network Canada as having searchable data catalogues for linkable data. Health Data Research UK established the UK Health Data Research Innovation Gateway in 2020 (healthdatagateway.org) as a common entry point or portal for researchers to discover what data collections exist for research use. The portal allows users to search via filters across all data collections, narrowing down the pool of data for a specific purpose. Health Data Research UK also developed a Data Utility Framework which indicates the 'utility' of a data collection for researchers for a particular purpose based on a set of key attributes and classified using predefined criteria into subsequent arbitrary qualitative categories (bronze, silver, gold and platinum) [4].

Health Data Research Network Canada developed the Strategy for Patient-Oriented Research Canadian Data Platform (SPOR CDP) (hdrn.ca/en/dash/inventory/). The SPOR Data Access Support Hub provides a data collection inventory which allows researchers conducting multi-site research to search for a data collection by keywords or via several filters, including region, data-category, and site. A standardised template of 11 metadata attributes is used to describe each of the 573 data collections included in the inventory.

Informal interviews were held with representatives from each of the national and international organisations that had developed online data catalogues. Key learnings from these discussions emphasised the importance of:

- moderating the metadata to ensure quality assurance. Interviewees noted that without proper moderation,

there can be significant discrepancies in the level of detail and the quality of the information provided.

- implementing validation rules to restrict fields for data entry, thereby minimising the need for extensive data cleaning.
- need for developing an agreed standardised template for data collection metadata.
- search and browse functionality, as researchers are accustomed to searching by keywords and filters.

Caution was raised regarding the feasibility of maintaining variable-level metadata and the infrastructure required for its ongoing upkeep.

Current state of metadata discovery

Whilst data custodians are generally responsible for the creation and storage of metadata related to the data collections in their custody, PHRN data linkage units play a key role in providing researchers with access to metadata for linked data projects. The three most common methods of metadata discovery for researchers include visiting the websites of government agencies and/or data linkage units; emailing the data linkage unit or the data custodian to obtain variable lists and data dictionaries that were not available online or contacting a fellow researcher who has previously worked with the data collection.

Phase 2 Metadata platform consultation and prototyping

Over a five-week period Future Friendly conducted 10 expert interviews, 19 testing sessions with end users, and a data custodian (n=5) workshop. End users and experts included researchers, client services staff, data custodians, and metadata specialists from state and commonwealth government agencies, medical research organisations and universities.

The interviews affirmed that metadata is spread across a multitude of different documents and websites, and that there is no unified point of access or simple search mechanism to identify data collections available for linkage. Often the metadata is published in MS Word, MS Excel or PDF format which is not easily searchable. Researchers highlighted the challenge of assessing in advance whether a data collection is appropriate for a specific research project, especially for certain collections that have limited metadata publicly accessible, particularly at a data item level.

The complex data custodian landscape makes it difficult for researchers to compare data collections due to significant differences in metadata quality and the inconsistent frequency of updates and accessibility. These variations depend on the specific data collection and the government agency managing it. The discrepancies are attributed to resource gaps and policy limitations within departments and across government agencies that hold the collections.

Interviews with client services staff highlighted the importance of their role as a conduit between researchers and data custodians. Researchers would often query client services

staff about data collections, despite the information being available online. This unnecessary repetition or duplication stemmed from difficulties in researchers locating the relevant metadata. It was clear that this information provision was resource intensive to support and, in turn, reduced the capacity of staff to provide support for complex matters.

Data custodians, especially those overseeing administrative data collections, reported that the provision of data for research is not explicitly outlined in their job descriptions. As a result, there is a lack of dedicated resources allocated to preparing or delivering data to researchers. Furthermore, although custodians are often tasked with developing and maintaining metadata for their collections, the comprehensiveness of this metadata was at times inadequate compared to what researchers required for informed decision-making regarding the relevance or utility of a data collection for a specific research project.

Testing of the Health Data Research UK Data Utility Framework received limited support in user testing, even after modifications were discussed. Researchers noted a preference for data dictionaries, data quality statements, data item level metadata and sample data ('mock data') that would allow them to make their own assessment of the suitability of a particular data collection.

The consultations with researchers highlighted the need for a centralised discovery platform that brings together all the data collections currently available for linkage, across all jurisdictions into a single place. Interviews identified several opportunities for the new metadata platform including:

1. Development of a standardised template to present metadata, allowing for easy interpretation and comparability of data collections
2. Streamlined features to allow researchers to browse and evaluate data collections available for linked data research
3. Provision of frequently asked questions (FAQs) and answers to facilitate the sharing of information.

Phase 3 Development of the metadata platform

Metadata attributes

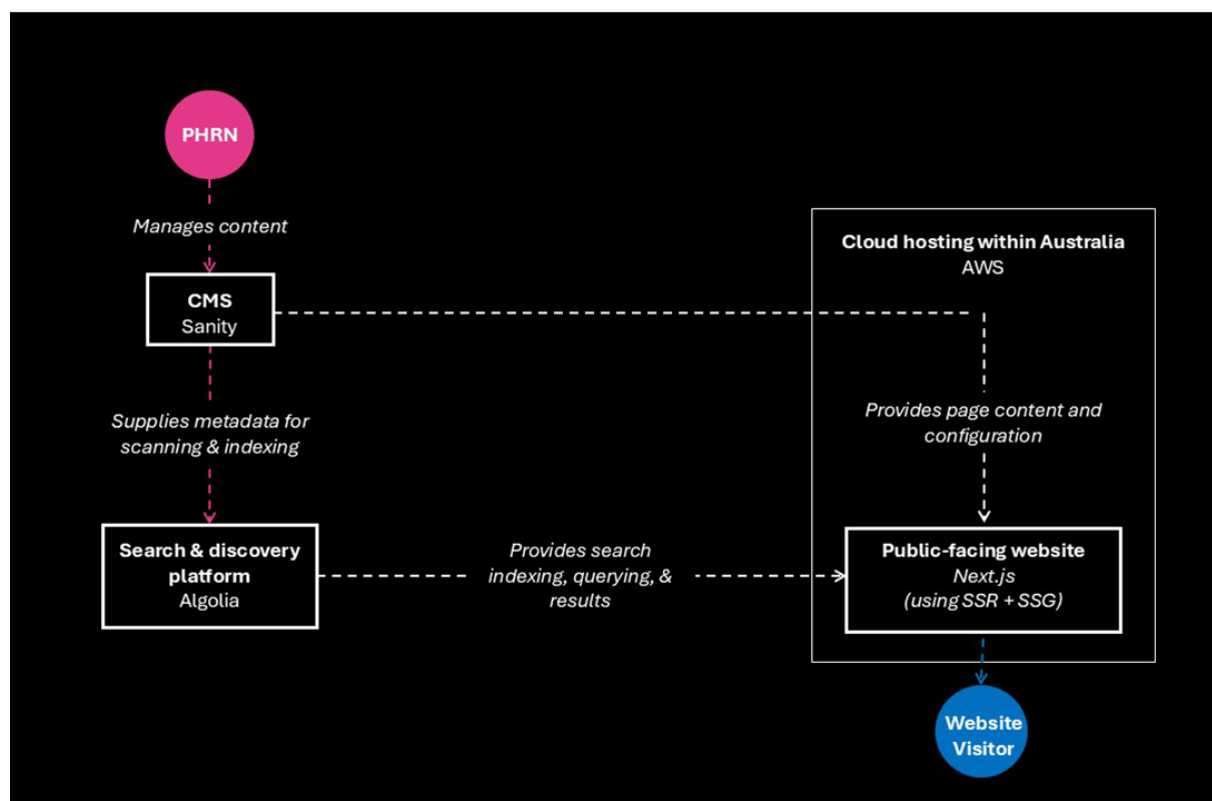
The final standardised template comprised 27 metadata attributes developed through consultations with researchers and tested for feasibility with data custodians (see Appendix A for a list of the attributes). The Project Control Group reviewed the final list, with neither the custodians nor the Project Control Group recommending additions or deletions. To ensure consistency, the Project Control Group suggested aligning attribute names and definitions with the Core Metadata Attributes for Data Inventories [2].

Key features

User testing highlighted several key features to be incorporated into the metadata platform, including:

Searching, browsing and filtering: researchers expressed a need to both browse and search for

Figure 1: PHRN metadata platform technology approach



data collections depending on the phase of their research. Searching by keyword was seen as particularly valuable to researchers as it aligned with the way they search the literature. Researchers also requested the availability of filters based on key characteristics such as state/territory, collection type, linkable date range and documentation.

Standardised format: Prior to the development of the metadata platform, there were significant variations in the metadata attributes available for data collections, the format in which it was presented and its availability. The standardised template of agreed attributes in the platform presents metadata in a consistent format no matter which data collection a researcher is viewing. Concept testing indicated that this consistent approach would enable researchers to efficiently compare and assess data collections.

Documentation: The documentation tab provides hyperlinks to data item lists, data dictionaries and data quality statements. It was clear throughout the consultations that providing links to the original source (“as a single source of truth”) was preferred, rather than providing the documents within the platform.

Publications: The publications tab allows researchers to review the journal articles, reports and presentations that have used a particular data collection. This enables researchers to learn about how the data has been used in research previously. The opportunity to connect with researchers who have previously used a particular data collection was deemed invaluable, especially for insights

into any challenges or limitations encountered during their research. Moreover, increasing the discoverability of linked data publications is acknowledged as pivotal for improving future research, reducing duplication, and providing opportunities for better collaboration between researchers nationally and potentially internationally.

Contact details: User testing indicated that researchers wanted ready access to contact details for data custodians as well as client services officers for single and cross jurisdiction project enquiries. Approved contact details were added to this section of the platform.

Technology approach

The platform was built for flexibility, speed, and scale. Figure 1 outlines the technology approach.

Impact of the new PHRN metadata platform

Testing sessions with end users suggested that a new metadata platform would result in more confident researchers, significant time savings for researchers and more time for client services staff to provide guidance to researchers on higher level matters.

Discussion

As the number and breadth of data collections being routinely linked has increased, the discoverability of data collections and associated metadata has emerged as an important consideration for enhancing the efficiency and effectiveness of research. This project identified that there is a wealth

of metadata available in Australia, but it is difficult for researchers to access and evaluate. Data custodians and client services officers find it resource intensive to support individual requests for simple metadata and this reduces their capacity to respond to more complex inquiries. The difficulties in accessing and evaluating metadata may impact research outcomes by contributing to long lead times to develop research projects and write data request applications. In response to these needs, the PHRN built a centralised platform that consolidates metadata for all data collections routinely linked in Australia.

The project's strategic objectives were to enhance the discoverability of data collections suitable for research using linked data and facilitate the standardisation of metadata across diverse collections and government entities. In addition, ensuring effective management of metadata associated with data collections and improving the accessibility of metadata are essential measures to uphold the FAIR Principles [5].

The metadata platform was developed via a three-phase process: national and international scoping exercise; consultations with researchers, data linkage staff and data custodians; and a final implementation phase. The metadata platform underwent a cycle of prototyping and testing to refine the digital experience. The final design, key features and functionality stems directly from user consultations and data custodian engagement. The metadata platform has been designed to work across a variety of platforms, devices, and browsers. Particular attention has been paid to accessibility and inclusivity, with optimisation for searchability, indexability and performance.

The PHRN metadata platform adopted the Core Metadata Attributes for Data Inventories [2] developed through the Australian Data Champions Network 'Collaboration in Metadata Management and Interoperability'. Use of the Core Metadata Attributes acknowledges that a consistent approach to describing data collections is critical to delivering the infrastructure necessary to enhance data collection discovery and to connect researchers to the data they need more effectively.

A key achievement of this project was the design and implementation of a user-centric national metadata platform tailored for Australia's linked population health data. Further, the adoption of a standardised approach for generating and displaying metadata across all collections and states and territories offers a dual benefit. It allows researchers to compare and evaluate data collections more readily, and also contributes to the enhancement of the overall quality of several data collections' metadata.

Enhancing public accessibility to quality metadata is intended to enable researchers to design better data linkage studies and reduce the burden on client services officers and data custodians for routine enquiries. Dahl et al (2020) noted that even with access to high-quality metadata, working with administrative data remains complex, and requires a thorough comprehension of both the data itself and the methods generating it. Thus, ensuring that client services and data custodians have the capacity to guide researchers on more complex enquiries such as design, feasibility and interpretation of data remains critical [4].

The initial proposal for this project included a rating system for data quality and utility. Due to the limited support from Australian researchers for the Health Data Research UK

Data Utility Framework, [6] even after modifications were discussed, it was removed from scope. It was suggested that the metadata platform provide specific pieces of metadata to support researchers to make judicious assessments of their target data collections, using the lens of their research question, without the need for a quantitative rating system.

The proposed maintenance schedule includes an annual formal update where the PHRN emails templates either to data linkage units (as proxies) or directly to data custodians for verification and updates. Additionally, data linkage units and custodians are encouraged to provide updates to the PHRN as needed throughout the year. Many of the attributes are static with only a few items such as dates, documentation (variable lists and data dictionaries) and publications requiring updating. Since much of the metadata on the platform is not available elsewhere, the maintenance process is currently manual. While this approach is less efficient than automation, it reflects the current limitations of the existing infrastructure. This platform marks the development of metadata standards for data collections, and so manual maintenance offers the benefit of rigorous quality assurance and helps to ensure consistency in the comprehensiveness of metadata and frequency of updates across data collections in this emerging field. Automated programs are however run over the platform to detect broken links to assist with the currency of documentation. Other options for automation will continue to be explored as the infrastructure develops and improves.

Whilst the metadata platform has been deployed, further work is still required to meet the needs of researchers as expressed in the consultation phase of this project. Important first steps include the addition of other frequently linked data collections to increase the visibility of a broader range of data collections that can be readily used for data linkage projects. The inclusion of data item list metadata was a key feature valued by researchers in testing, however further consultation is needed to determine the most sustainable and optimal approach for providing access to this granular level of data. The project identified promising opportunities for integration with existing national registries, such as METEOR, to address this gap. Additionally, there is a strategic consideration to better align the platform with the FAIR principles by introducing persistent identifiers to precisely identify data collections, enhancing discoverability and accessibility [7]. A formal evaluation is scheduled 12 months after the platform's launch to assess its effectiveness in meeting researchers' needs and to inform future improvements and feature development.

Conclusion

The PHRN metadata platform addresses some of the key challenges associated with metadata standardisation and discoverability in the context of Australia's complex data landscape. This platform not only enhances the discoverability of data collections available for linked data research but also promotes standardisation in metadata presentation across jurisdictions. This in turn facilitates easier comparison and evaluation of data collections and contributes to the overall improvement of metadata quality. While the deployment of the metadata platform marks a significant achievement, ongoing efforts are needed to further refine its capabilities, expand its

coverage of data collections, and align it more closely with FAIR principles to ensure continued relevance and effectiveness in supporting population health research endeavours.

Conflict of interests

The authors Miller, KM; Flack, FS and Smith MB are all employed by the University of Western Australia and their salaries are funded by the Population Health Research Network through the Australian Government National Collaborative Research Infrastructure Strategy.

Funding

This project was funded by the Australian Government through the National Collaborative Research Infrastructure Strategy.

The funder of the study had no role in the design or development of the metadata platform or writing of the report.

Acknowledgements

The PHRN acknowledge Future Friendly who conducted the consultations and collaborated with the PHRN to develop, plan, and build the metadata platform. We acknowledge the Project Control Group members who provided valuable support and guidance to the project. The PHRN National Office would also like to thank the data users and experts who participated in expert interviews and concept testing.

Data availability statement

Not applicable.

Ethics statement

This project involved the development and user testing of a metadata software platform to inform its functionality and usability. The testing process was conducted as part of routine product development and quality assurance, rather than as a formal research study aimed at generating new knowledge. Participants were engaged to provide feedback

on the software's features and user experience, with their informed consent. No personal, sensitive, or identifiable data were collected, and testing posed no foreseeable risks beyond those encountered in everyday software use. As this activity falls under software development and usability testing rather than human research as defined by the National Statement on Ethical Conduct in Human Research (NHMRC, 2023), formal ethics review was not required.

References

1. Australian Institute of Health and Welfare. Health System Overview Canberra: Australian Institute of Health and Welfare; 2022 [Available from: <https://www.aihw.gov.au/reports/australias-health/health-system-overview>.
2. Office of the National Commissioner. Metadata Attributes Guide. Canberra: ONDC; 2023.
3. Standardization IOF. ISO/IEC 11179-3:2023 Information technology Metadata registries (MDR). Data management and interchange. Switzerland: International Organization for Standardization; 2003.
4. Dahl LT, Katz A, McGrail K, Diverty B, Ethier J-F, Gavin F, et al. The SPOR-Canadian Data Platform: a national initiative to facilitate data rich multi-jurisdictional research. *Int J Popul Data Sci*. 2020 Nov 9;5(1):1374. <https://doi.org/10.23889/ijpds.v5i1.1374>
5. Stausberg J, Harkener S. Metadata of Registries: Results from an Initiative in Health Services Research. *Stud Health Technol Inform*. 2021 May 27;281:18-22. <https://doi.org/10.3233/SHTI210112>
6. Gordon B, Barrett J, Fennessy C, Cake C, Milward A, Irwin C, et al. Development of a data utility framework to support effective health data curation. *BMJ Health Care Inform*. 2021;28(1):e100303. <https://doi.org/10.1136/bmjhci-2020-100303>
7. Australian Research Data Commons. Australian National Persistent Identifier (PID) Strategy 2024. Victoria: ARDC Ltd.; 2024. <https://doi.org/10.5281/zenodo.10656275>



Appendix

Appendix A: PHRN metadata platform metadata attribute list

Attribute No.	Attribute name	(i) Description as provided by the Office of the National Data Commissioner.
Att_1	Title	The most common useful name by which the data asset is known by your agency and by your largest perceived audience. The title should be relatively unique and can be created by following a naming convention used by your agency.
Att_2	Equivalent datasets	Availability of a comparable data asset in other jurisdictions. For cross jurisdictional projects, this indicates which jurisdictions have a comparable data asset.
Date range		
Att_3	Temporal coverage from	The start period for which this asset is applicable. Temporal coverage refers to the time period that a data asset covers and may be broader than a single collection. E.g. a data service or data portal. This field is related to attribute Temporal coverage to.
Att_4	Temporal coverage to	The end period for which this asset is applicable. The asset may not have an end date if it is being continually added to, in which case, a value is not required. E.g. a data service or data portal. This field is related to attribute Temporal coverage from.
Att_5	Linkage temporal range	The time period for which this asset is linked. Linkage temporal range refers to the time period that data within the data asset has been linked and may be less than the temporal coverage of a dataset. The data may not have an end date if it is being continually linked, in which case, a value is not required.
Last updated		
Att_6	Update frequency	The frequency at which new, revised, or updated versions of this asset are made available. For assets regularly released, one data asset record will represent a series; separate records will not be required per update. Agencies will determine when a new record is required for an asset, based on changes in methodology, collection and related policies.
Att_7	Linkage update frequency	The frequency at which, new, revised, or updated versions of this asset are linked/added to the master linkage map.
Att_8	Metadata updated	The most recent date the metadata was either created, changed, updated or modified. This refers to the metadata of a data asset, not a date pertaining to the data within a dataset.
Summary		
Att_9	Description	A descriptive statement of the asset. Easily read and understood information about the data asset; the purpose is to enable users to find, categorise and evaluate the fitness of a data asset to their needs. The Description attribute is also searchable and is typically several sentences long. It is useful to consider what keywords your potential audience may use to search for the asset. This field could be supplemented by attributes Keyword and Purpose.
Att_10	Scope	Limitations of the data asset or the linkage of the data asset.
Att_11	Limitations	
Provenance		
Att_12	Purpose	A descriptive summary of the intentions with which the asset was developed and proposed to be used. This field supplements the attribute Description.
Att_13	Location	The scope of the geographic area or location that the data asset covers. Location represents the geographic scope of the entire asset (e.g. “Australia”) and is not intended to represent location values contained within the asset, for example street, suburb or region which could be captured within Keyword, Description or Purpose.

Continued

Appendix A: Continued

Attribute No.	Attribute name	(i) Description as provided by the Office of the National Data Commissioner.
Att_14	Legal authority	All legal mandates under which the asset was collected, created, received, used or disclosed. Legal mandates could include: Memorandum of Understanding; Legislation; Machinery of Government; Government policies or acts; etc. It could include the authority, e.g. (Australian Government) Federal Register of Legislation or Data Availability and Transparency Act 2022. Where multiple legal mandates exist, separated their URLs with a comma ','. This information may be sourced through the agency's legal department.
Att_15	Access Rights	Specifies access to the data asset. Access will be based on agency privacy, security, or other policy approaches that apply to this data asset. Access can be: <i>Open</i> – Data is publicly accessible online (registration may be required); <i>Conditional</i> – Data is publicly accessible subject to certain conditions. For example: a fee applies; or the data is only accessible at a specific physical location; <i>Restricted</i> – Data access is limited. For example: during an embargo period; to a particular group of users; or where formal permission is granted
Formats and Standards		
Att_16	Language	The language used within the asset. This refers to the language used within the asset – e.g. "English". The default value may be set to "English". Some agencies may have assets containing languages other than English, in which case the Australian Standard Classification of Languages (ASCL) can be used.
Att_17	Format	Format(s) in which the dataset is available to researchers e.g., xlsx, csv, application, text, etc.
Att_18	Coding	Relevant terminologies / ontologies / controlled vocabularies, such as ICD codes, data dictionary national codes or SNOMED CT International, that are being used by the dataset.
Contact information		
Att_19	Point of Contact – Single Jurisdiction	Position; Agency; Email
Att_20	Point of Contact – Cross Jurisdiction	Position; Agency; Email
Approvals and Governance		
Att_21	Approvals required for access	List of approvals the researcher must acquire to obtain access to the data. For example, data custodian, Data Linkage Unit, HREC, and any other bodies. Researchers may wish to publicly present research findings or publish articles in journals or other publishing forms. Please list any approvals that researcher must obtain prior to publishing.
Att_22	Approvals required for publication	
Other details		
Att_23	Data Linkage Unit	The custodian(s) of the asset. The custodian is the agency who has the control of the asset and has the authority for sharing and disclosure. The custodian may not be the publisher (see Publisher attribute).
Att_24	Data Custodian	
Documentation		
Att_25*	Variable list	
Att_26*	Data Dictionary	
Att_27*	Data Quality Statement	
Att_28*	Other Documentation	