



Transformational
Data Collaboration



Data Harmonisation Report

Abridged / Amended presentation used for the
PHRN Strategic Priority Showcase

18 August 2022

Priyanka Pillai, Rachel Canaway, Roger Ward, Christine Chidgey, Katherine
Duszynski, Nicole Pratt, Douglas IR Boyle

What we did

Research project commissioned by the Population Health Research Network (PHRN)
“Opportunities to harmonise jurisdictional data and associated metadata”

Undertaken by The University of Melbourne

- Priyanka Pillai, Rachel Canaway, Roger Ward, Christine Chidgey, Douglas Boyle

University of South Australia

- Katherine Duszynski, Nicole Pratt

Surveyed and interviewed staff from data linkage units at AIHW and all states and territories (Sep-Nov 2021)

- 9 organisations, 24 representatives
 - holding a range of technical/data, client services, business management or leadership positions
- Average interview length 45 minutes (range 29 to 102 minutes)

Interview transcripts were thematically analysed (n=9) and survey data (about infrastructure and collections) were collated (n=9)

Nov 2021 report submitted to PHRN

Project aims

1. Identify which jurisdictions have a health data warehouse with the potential for harmonisation
2. Identify any plans to establish a health data warehouse in jurisdictions that don't have one
3. Describe data and metadata standards used in each jurisdictional health data warehouse
4. Identify the levels of interest in greater harmonisation of data in each health data warehouse
5. Identify what would be needed to achieve greater harmonisation of data:
 - i. in each jurisdictions' warehouse, and/or
 - ii. for making cross-jurisdictional data available to researchers.

Interview question: “From your perspective, what is data harmonization?”

“A lot of our initiatives are striving towards harmonisation even though we don't use that word we have different words for the activity of harmonising.”

Harmonisation was described as an idea or as something concrete – but normally at a local level

- Described **conceptually**
 - assumptions made from harmonised data will be proven correct (“a single truth”, “research ready”)
- Described **operationally**
 - Requires explicit understanding of data: provenance, state (raw vs edited), scope, coverage, caveats, context, measures, comparability, completeness
 - Requires standardisation:
 - input consistency via National Minimum Data from each jurisdiction to a central repository
 - output consistency via Common Data Model approach
- Described from purely **intra-dataset** perspectives
 - validation e.g. correcting single person discrepancies apparent in different datasets
- Described from **intra-jurisdictional** (across datasets within jurisdiction) or **inter-jurisdictional** terms
 - mapping variable lists (and other operational examples above)

Conclusion

- What data harmonisation is, is variably understood or perceived – a common dialogue is needed so that endeavour is not at cross purposes.

Perceived **benefits** of data and metadata harmonisation

- Efficiencies for researchers and for data linkage units (DLUs)
 - DLUs transform / harmonise once and re-used many time = efficiency
 - Quicker access to data
 - Better use of resources
 - Reduced overheads
- Mitigates risks of misinterpreting the data / increased user confidence
- Better quality data, more accurate comparisons, better reporting, bigger picture
- Making data research ready and more useful
- Supports long term benchmarking
- National consistency of data and a complete picture

Perceived **Risks** of Harmonisation

- Loss of data value and context
 - Standardisation leading to loss of nuance and context, lack of appropriate metadata, introduction of errors (inappropriate combining of variables), misinterpretation by users, loss of variables, dataset no longer fit for purpose
- Lack of knowledge among persons undertaking harmonisation
- Relational damage with original data custodians
- Security risk of more data in a single warehouse
- Bureaucracy / limiting local creativity
- Unintended consequences (currently unknown) arising through changing part of the system

Perceived **Barriers** to Harmonisation

Operational and political barriers

- Lack of funding and resources
- Lack of interest / competing priorities
- States / Commonwealth operational divides (red tape)
- Lack of agreement on how

Data and infrastructure barriers

- Lack of consistent data: structure, comparability, transparency, metadata
- Poor data quality at the frontline (free text)
- Lack of infrastructure – available now, appropriate in the future
- Data Linkage logistical barriers
- **Knowledge and trust barriers**
- Lack of know-how
- Lack of data custodian trust (=> resistance)
- Data users ought to work with raw data

Barrier / Risk: Data users “have to understand the pain”

*“When you transform it and you harmonize it (data) and you create that golden record to say: ‘This is the truth’ – you'd better be right, because if it's wrong, you're on the line for it. It's almost better to give the researchers the raw data: ‘Here it is, open, honest, this is the way we collect it. We can give you information about how it's collected and you can make the decisions and caveat that in your research to say how you harmonized it’. I like **the idea of trying to harmonize something to take the pain from researchers**, but **sometimes they have to understand the pain so they can actually understand what they're reporting on and using.**”*

“We're not experts across 30 data sets and 20,000 variables, you just can't be. In certain ways, the researcher that picks up the health data (...for) a cross jurisdictional project may be better placed than us to actually work out those harmonization steps. Therefore, leaving it as raw data for them to put their own coding, own inflection on, and actually harmonize it themselves, may be a better way, at the moment, to do it?”

Perceived **Enablers** of Data Harmonisation

Agreement (Among data custodians & other stakeholders):

- On data structure, processes and education

Motivation and incentives:

- A compelling business case and a National Strategy
- Resourcing and high-level support
- Incentives (at levels of data capture and data linkage units)

Structure and know-how (need):

- Working groups
- Data harmonisation 'champions'
- Utilise the extensive knowledge pool (capture corporate knowledge)
- Greater sharing of code and structures for common data models
- Mandates

Actions Suggested to **Progress** Greater Harmonisation

- Consultation with stakeholders including end-users
- Determine / agree on data and metadata standards
 - Ensure data custodians provide appropriate metadata
- Pilot or exemplar project
- High level support, investment, resources
- Adopt technology to future proof assets and systems
- Raise excitement and knowledge (educate)
 - Including knowledge of data capture context
- Strategy development

Perceptions of Data Harmonisation – start with metadata

“I feel like there is such a big potential to improve the discoverability and usability of data just by working on metadata and before even talking about the data itself, because it seems like most data sets are just really poorly described.”

“There's been a focus on metadata which was disappointing because ... it looked at data set level metadata NOT variable level metadata... the real value in the data is by getting to variable level metadata that could start to actually deal with this harmonization challenge.”

- There is strong support and willingness for national data harmonisation
- Benefits were perceived to outweigh risks
- There was little knowledge about international efforts or common data models

Next steps

- Consultation
- A National Data Harmonisation Steering Committee & Technical group to support NDHS
- A National Data Harmonisation Strategy (NDHS)
- Funding opportunities
- Achieve national harmonised field capture for agreed data fields
- Cross-jurisdictional pilot projects (e.g. metadata standards, Common Data Model)
- Identify use cases to learn from (national or international)

Synopsis of Executive Summary of Report to PHRN

The Population Health Research Network, University of Melbourne and University of South Australia undertook **consultation with entities managing state, territory and Commonwealth health data warehouses**. The main objectives of the study were to:

1. investigate the current state and plans of data warehouses in the jurisdictions,
2. identify barriers, risks, benefits, and enablers associated with further data harmonisation across PHRN and
3. determine the actions needed to achieve greater data harmonisation in the respective data warehouse.

Twenty-four individuals from nine organisations were interviewed. Qualitative data analysis of the pre-interview survey and interview transcripts informed the main findings for this consultation.

Data warehouse capabilities vary across jurisdictions. The majority used Microsoft SQL servers and some used Microsoft Azure Cloud-based access. Staffing levels and resources were mixed across the jurisdictions. Approaches to metadata and data harmonisation vary across the jurisdictions; some align with national standards. Nearly all data warehouses and data linkage units have publicly listed data collections, metadata, and data dictionaries (if available) on their website. Data access and linkage always requires data custodian approval and ethics clearance.

There is a strong willingness for national data harmonisation. It is seen as a pathway to better, more efficient use of data. However, many barriers to data harmonisation are widely acknowledged. A range of recommendations and potential next steps were discussed with a focus on pilot projects in data harmonisation, including:

1. A national strategy in consultation with all jurisdictions and users, implemented by an organisation like PHRN.
2. A Steering Committee to guide a Technical Working Group to investigate the existing standards applied to datasets and improve harmonisation for the other datasets.
3. Learning from successful data harmonisation practices implemented overseas, for example, in the UK.



THE UNIVERSITY OF
MELBOURNE

Thank you

University of Melbourne

Dougie Boyle dboyle@unimelb.edu.au

**Rachel Canaway, Roger Ward, Priyanka Pillai,
Christine Chidgey**

University of South Australia

Katherine Duszynski, Nicole Pratt

PHRN: Merran Smith, Felicity Flack