

# Linkage of routinely collected data in practice: the Centre for Health Record Linkage

Katie A Irvine<sup>a,b</sup> and Elizabeth A Moore<sup>a</sup>

<sup>a</sup> Centre for Health Record Linkage, NSW Health, Sydney, Australia

<sup>b</sup> Corresponding author: [katie.irvine@moh.health.nsw.gov.au](mailto:katie.irvine@moh.health.nsw.gov.au)

## Article history

Publication date: September 2015

Citation: Irvine KA, Moore EA. Linkage of routinely collected data in practice: the Centre for Health Record Linkage. Public Health Res Pract. 2015;25(4):e2541548. doi: <http://dx.doi.org/10.17061/phrp2541548>

## Introduction

The Centre for Health Record Linkage (CHeReL) was established collaboratively in New South Wales (NSW) and the Australian Capital Territory (ACT) to expand access to linked health-related data for research, planning and evaluation. Like other international centres, the CHeReL provides a focal entry point for investigators, delivers information and advice, and provides a mechanism for making linked administrative data available (without direct personal identifiers) to approved third parties within relevant regulatory and governance frameworks.

This paper includes information about access rates, with the aim of promoting greater transparency about the use of research assets and the release of administrative data to enable research. It also addresses timelines and planned drivers of improvement because delays in accessing linked administrative data appear to be common internationally (but the literature is largely anecdotal).<sup>1,2</sup> See Box 1 for an overview.

### Box 1. Overview of the Centre for Health Record Linkage (CHeReL)

- 1300 users
- A Master Linkage Key containing continuously updated links within and between 102 million records from 18 core population health-related databases from NSW and the ACT
- 68% of requests for core data are approved by the custodian within 20 days
- Almost 90% of projects linked jurisdictional health system data to existing cohorts, disease registers and administrative collections from other government sectors

## Discussion

Since 2007, more than 1300 investigators have submitted applications for linked data within NSW and the ACT, and more than 98% obtained full approval. Approximately 1020 investigators received data either directly or into secure remote-access facilities. Less than 10% of projects used data exclusively from the NSW Health system or NSW private hospitals. The most common link is of jurisdictional health system data with cohorts, disease registers and administrative collections from other government sectors.

### Time frames

A 2015 survey suggested that 10% of recent CHeReL users perceived overall time frames – including application, data custodian and ethical approval, linkage and data delivery – as a barrier to access. Timelines for access to linked data are highly variable. From 1 July 2012 to 30 June 2014, median time frames for different service types ranged from 3 months to 17 months. For new requests requiring project design and development by the research team, new ethical and data custodian approval, and new linkage or extraction, the median end-to-end time frame was 12 months. Shorter timelines within this range have been driven by several factors, including streamlined information governance and an expansion of the routine linkage system.

### Streamlined information governance

Regulatory approval within 20 days is the target of several British centres. For new research project applications to the CHeReL, the proportion of requests approved by core linkage system custodians within 20 working days increased from 61% in 2012–13 to 68% in 2013–14. Ethical review time frames are reportedly decreasing (unpublished data, NSW Population and Health Services Research Ethics Committee). Initiatives to approve cohesive programs of research within a research group, rather than single projects, have resulted in new project approval and data updates being available within a few months.

### Routine linkage

The CHeReL's core system, the Master Linkage Key (MLK), contains continuously updated links within and between 102 million records from 18 core population health-related databases from NSW and the ACT. The system is expanding rapidly as health identifiers become incorporated into parts of the data linkage process. By sourcing and linking data routinely, under enduring agreements, the MLK provides faster access. From 1 July 2012 to 30 June 2014, the median time for full data access was 6 months faster for MLK extracts than for new, bespoke linkages.

### Further improvements

Within NSW, a 4-month benchmark for 'timely access'<sup>1</sup>, proposed by the Council of Canadian Academies, can be achieved. Key initiatives to minimise variability include:

- NSW Health ethics and governance reform led by the Office for Health and Medical Research
- Expansion of the CHeReL system to include newly emerging clinical information systems, adjusted for technical differences in clinical and population health linkage
- Adoption of optimal operating models used by international centres.

Many data linkage centres hold or access disparate content data that are securely linked and masked, and then provide the data project by project into remote gateways, safe havens or local research environments. In Australia, widespread adoption of a best-practice model<sup>3</sup> has contributed, in part, to uncertain time frames.

The CHeReL is now moving to an internationally accepted model with centralised content data management and delivery, which allows detailed checking of linkage quality, and refinement of linkage algorithms, in a way that is not possible with third-party de-identified linked databanks.<sup>2</sup> It is anticipated that this model will vastly reduce time frames at low ongoing cost, and drive efficiencies in creation of 'analysis ready' datasets. In Western Australia, a change in operating model reduced data delivery times by up to 90%.<sup>4</sup>

## Conclusion

The track record of making linked routinely collected data accessible in NSW and the ACT has been noted in a recent review.<sup>5</sup> Collaborative and sustained development will be necessary as digital transformation drives rapid growth in health-related data. This includes extension of robust access frameworks, which already exist for secondary use of administrative data, to the rapidly evolving and distributed clinical information passing through electronic medical records and other clinical systems.

## Acknowledgements

We gratefully acknowledge the funding and in-kind contributions of the NSW Ministry of Health, the Cancer Institute NSW, ACT Health, and past and present CHeReL member organisations. The CHeReL is a node of the Population Health Research Network, which is an initiative of the Australian Government and conducted as part of the National Collaborative Research Infrastructure Strategy.

## Competing interests

None declared

## Author contributions

KI conceptualised the report. Both authors contributed to the analysis and drafting and both approved the original version and resubmission.

## References

1. The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation. Accessing health and health-related data in Canada. Ottawa: Council of Canadian Academies; 2015 [cited 2015 Aug 11]. Available from: [www.scienceadvice.ca/en/assessments/completed/health-data.aspx](http://www.scienceadvice.ca/en/assessments/completed/health-data.aspx)
2. Wellcome Trust. Enabling data linkage to maximise the value of public health research data: full report. London: Wellcome Trust; 2015 [cited 2015 Aug 11]. Available from: [www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTP056860.htm](http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTP056860.htm)
3. Kelman CW, Bass AJ, Holman CDJ. Research use of linked health data – a best practice protocol. *Aust N Z J Public Health*. 2002;26(3):251–5.
4. Eitelhuber T, Davis G. The custodian administered research extract server: “improving the pipeline” in linked data delivery systems. *Health Inf Sci Syst*. 2014;2:6.
5. Population Health Research Network. 2013 independent panel review: findings and recommendations. Perth: Population Health Research Network; 2014 [cited 2015 Aug 11]. Available from: [www.phrn.org.au/media/80607/phrn-2013-independent-review-findings-and-recommendations-v2-final-report-april-17-2014-2.pdf](http://www.phrn.org.au/media/80607/phrn-2013-independent-review-findings-and-recommendations-v2-final-report-april-17-2014-2.pdf)

Copyright: 

© 2015 Irvine and Moore. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence, which allows others to redistribute, adapt and share this work non-commercially provided they attribute the work and any adapted version of it is distributed under the same Creative Commons licence terms. See: [www.creativecommons.org/licenses/by-nc-sa/4.0/](http://www.creativecommons.org/licenses/by-nc-sa/4.0/)