

International Journal of Population Data Science

Journal Website: www.ijpds.org



Using data linkage innovation and collaboration to create a cross-sectoral data repository for Western Australia.

Ferrante, A^{1*}, Boyd, J¹, Eitelhuber, T², Randall, S¹, Brown, A¹, Maller, M¹, Botes, D¹, and Sibma, K³

¹Curtin University

²Western Australian Department of Health

³WA Department of Treasury

Background/rationale

The Western Australian (WA) government and the Centre for Data Linkage (CDL) at Curtin University are creating a large, de-identified researchable database – the Social Investment Data Resource (SIDR) – to support a key government initiative called Target 120 (T120). T120 delivers targeted early interventions to young offenders and their families to reduce the likelihood of re-offending.

Main aim

The SIDR brings together de-identified data from across government to be used for actuarial assessment and social investment analytics to assess long-term costs and benefits of T120 interventions.

Methods

SIDR adopts a distributed linkage model where linkage workload is shared between the Department of Health Data Linkage Branch who curate WA Data Linkage System (WADLS) and the CDL. Design elements of the model included a common spine (embedded into the infrastructure of both groups), methods for leveraging quality from WADLS, and inclusion of family relationships data from the WA Family Connections database. The linkage model uses a combination of traditional and privacy-preserving record linkage (PPRL) methods. PPRL does not require release of personal identifiers; instead, data is irreversibly hashed prior to release for probabilistic linkage.

The resultant SIDR repository has been designed to be securely and strictly managed. Access is by authorised, approved users only.

Results

Use of a distributed linkage model, coupled with traditional and PPRL methods, is an innovative yet pragmatic way of de-

living data linkage services to a large, cross-sectoral research project. PPRL methods enable inclusion of otherwise excluded datasets in the project. Sharing of workload harnesses linkage capacity and capabilities across the state. The SIDR includes health data, education records, justice, child protection, disability and housing data.

Conclusion

SIDR provides a resource for whole-of-government policy development, service evaluation, academic research and social investment analytics for T120 and beyond. The SIDR distributed linkage model has potential for adaptation and use elsewhere.

*Corresponding Author:

Email Address: a.ferrante@curtin.edu.au (A Ferrante)

