Data Cooperative for Social Impact and Public Health

Presenter: Amir Aryani

Amir Aryani[1] , **Jane Farmer[1], Timoleon Sellis[1], Josephine Barraket[1], Luis Salvador-Carulla[2], Nasser Bagheri[2], Tom Gedeon[2], Tom Verhelst[3] , Scott Baum[3], Sharyn Rundle-Thiele[3], Richard Sinnott[4], Bruce Tranter**

1- Swinburne University of Technology, Victoria, Australia
2- Australian National University, Australian Capital Territory, Australia
3- Griffith University, Queensland, Australia
4- University of Melbourne, Victoria, Australia
5- University of Tasmania, Tasmania, Australia

ABSTRACT

The Data Cooperative (Co-Op) Platform for Social Impact aims to advance data-driven research in the social sciences by making access to linked datasets more effective and affordable. The main focus of this infrastructure is supporting research and practice in public health, community wellbeing and social policy. This platform will form the foundation for cross-sectoral data-driven collaboration between universities, government and the not-for-profit community sector organisations in Australia and internationally to improve community health, wellbeing and social outcomes for disadvantaged and vulnerable groups. This project takes an innovative approach to the difficult problem of data aggregation and integration assisting the community to realise the benefits of the diverse range of data types and data sources now available, both existing and dynamically generated to drive new research insights.

BACKGROUND

Globally, there is growing recognition of the benefits of investing in data platforms and work is underway to deliver supporting ethics, governance and security infrastructures that enable organisations to come together in data collaboratives to produce collective intelligence [4]. The Australian government has recognised the need for data platforms that assemble comprehensive data to provide insights on complex, interconnected social, health and community issues [2]. This requires a major shift toward better, smarter and ethical data use [3]. A transition to incorporating data as part of evidence-based policymaking is already underway in many sectors and industries around the world; however, Australia is significantly behind in leveraging government and community sector data for social and public policy research [3]. The costs of more open data access and use will be substantially outweighed by the numerous opportunities presented by better collection, integration and visualisation of research, government, public service and other community data that is currently unavailable or unusable

Research into critical societal challenges requires the integration and analysis of data from across sectors and collaborating partners. Engaging in partnerships involving researchers, funding bodies, practitioners and more. Complex data processing techniques remain completely out of reach, financially and technically, for community services and humanitarian organisations [1]. The Data Co-Op Platform enables, for the first time in Australia, community sector organisations to securely and ethically apply their data to address crucial societal research questions.

OBJECTIVE

The objectives of the Data Co-Op Platform are to:
- Develop a secure, safe and ethical data platform for cooperative data projects using community sector datasets within the boundaries of ethics, data governance and security.
- Provide the technology for automated data linkages across open data sources from census, social data repositories, public health national collections, public policy records, and other government data.
- Provide the technology for automated geospatial mapping to enable place-based data analytics.
- Create 'cleaned', linked data packages that enable more productive data-driven research in public health, community wellbeing and social policy.
- Build a platform of structures and governance that facilitates organisations coming together in cooperative data projects, thus making it easier and more straightforward for researchers to collaborate with policymakers and community service organisations.
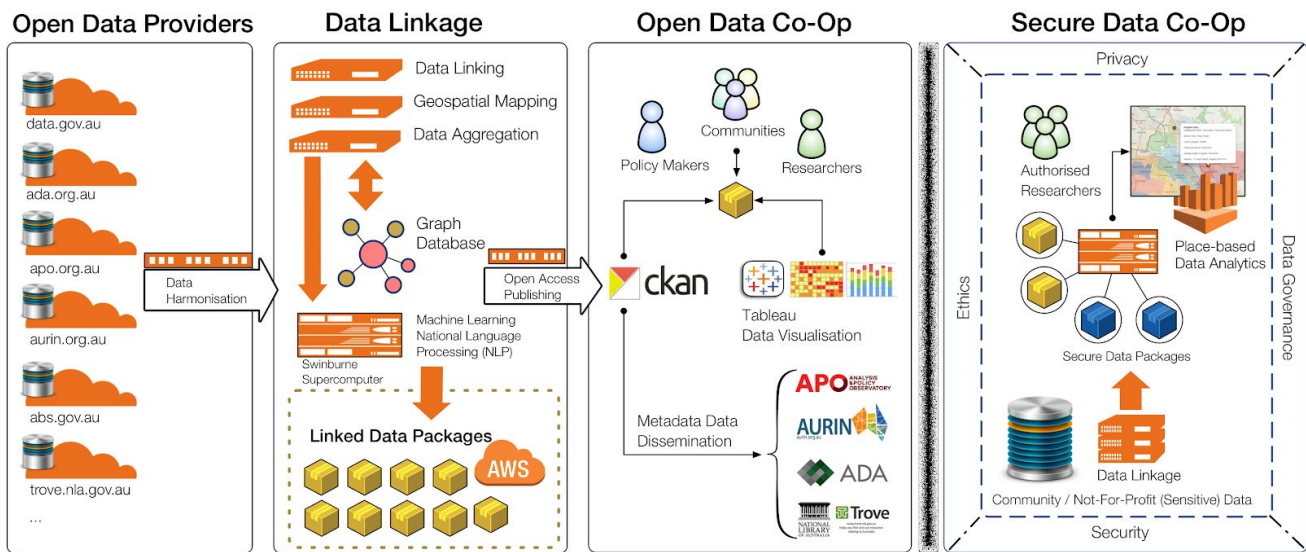
FIGURE 1: MODEL OF THE DATA CO-OP PLATFORM

The proposed solution consists of four main components: Open Data Providers, Data Linkage, Open Data Co-Op and Secure Data Co-Op (Figure 1). The key innovation of the Data Co-Op Platform is transforming currently unconnected open datasets into harmonised, connected and geospatially enabled linked data packages. These data packages will be aggregated from multiple underlying data sources, based on a connected graph of topics and subject taxonomies. The platform will distribute the resulting Open Linked Data Packages, developed by aggregating and linking open datasets, through national data repositories including Research Data Australia, ADA, APO, AURIN, and Trove. To optimise the discoverability of these data packages, the platform will leverage the DataCite collaboration to disseminate the metadata records across global data services, including Google Data Search, Harvard Dataverse Network, and Elsevier Data Search.

In this presentation, we will describe the blueprint for these components. In addition, we aim to discuss the following questions:

1. What legal frameworks and data management practices are needed to link aggregated government data (e.g. Census) to sensitive public health datasets (e.g. hospital records) for research while protecting the privacy and confidentiality?
2. What technical capabilities are required to link heterogeneous health and social data at the national level?
3. How ARDC and international initiatives (e.g European Open Science Cloud) can collaborate to link public health data across international boundaries?
4. How such a platform can leverage current investments in this space?

REFERENCES

1. Carlisle, K. and Farmer, J., 2018, Participatory approaches in planning and implementation of new healthcare services in northern Australia. International Journal of Health Planning & Management, 33,704-22
2. Infrastructure Australia, 2017, Trends to 2040, ISBN 978-1-925401-81-3
3. PC, 2017, Data Availability and Use, Productivity Commission, Canberra
4. Saunders, T. and Mulgan, G. 2017. Governing with Collective Intelligence. Nesta.

BIOGRAPHY OF THE PRESENTER

A/Professor Amir Aryani is the Head of the Swinburne Social Data Analytics (SoDA) Lab. The Lab applies contemporary and emerging co-op data analytics techniques to provide insight into health and social problems. Amir is the founder of Research Graph Technology and on the board of Research Graph Foundation. He has experience with large-scale and cross-institution projects in Australia and Europe.