Data and Services Discovery projects - Institutional Role in a Data Commons

Title

Data Cooperative for Social Impact and Public Health

Approach

In the scope of this project, Swinburne University has engaged the following organisations on forming data cooperatives: Glen Eira Council, Family Life, Alfred Health, Australian Childhood Foundation, and the Royal Women's Hospital. In addition, Swinburne and ANU have hosted a workshop in Canberra with 26 attendees from different government departments, agencies, not for profit organisations and universities. A similar workshop is also scheduled in Melbourne for 11 Nov.

Furthermore, Swinburne team has worked with researchers at Australian National University and Griffith University on the co-creation of the pilot infrastructure to host and operate the Data CO-OPs. This pilot infrastructure aims to demonstrate how an enabling platform can foster data cooperatives and data-driven research projects using hybrid collections of public and private datasets.

The following are the output of this project:
- A pilot infrastructure (hosted on Nectar) that transforms disconnected social private datasets to linked data packages, augmented with public datasets
- An open linked data packaged (CC-BY) on Social Resilience: https://doi.org/10.6084/m9.figshare.9994418.v1. The future version of this dataset will be available as a linked data package on ADA.
- A peer-reviewed article on Responsible Data Sharing (in progress)
- Presentation: Responsible Data Sharing Framework
- Presentation: Data Cooperative for Social Impact and Wellbeing

FAIR

We started this project by working with the disconnected public record, limited metadata, and unmanaged private datasets. In the course of this pilot project, we have demonstrated how to transform these data elements to a connected linked data package, accompanied by analytics

and GIS mapping using limited by effective cloud-hosted software components. For more information, please refer to the FAIR assessment report for this project.

Collaboration

The Data CO-OP platform has received collaboration from the following organisations:
- Australian National University as the data contributor, collaborators, consumers and providing input into the governance of the platform.
- Swinburne and Griffith researchers as collaborator and participant in governance.
- The city of Glen Eira as data contributor and consumer of linked data packages
- Family Life as a contributor, collaborator and consumer of linked data packages
- Alfred Health, Australian Childhood Foundation, and the Royal Women's Hospital as collaborators in forming the Responsible Data Sharing framework and working toward forming the Data CO-OP.

Sustainability

Swinburne, ANU, Griffith have committed substantial co-funding toward this project, and at the time of writing this report collaborating with UTAS and University of Melbourne eResearch with the view toward long term sustainability of the infrastructure. In addition, we are negotiating with other interested parties to form a broader sustainable community.

The main ingredient that enables the sustainability of this platform is the support of currently funded social science and public health research projects, and close engagement with active researchers. Furthermore, to enable sustaining the data collections and supporting materials generated by this project, we will publish them on ADA[1] and APO[2]. Also to sustain the relationship between the community, we are working on an MoU between contributing organisations, and negotiating a steering committee to lead the community collaboratively.

Learnings

We have identified the following challenges and learned these lessons so far in this project.
- The legal framework to assess data access is often unknown. In most organisations, the legal requirements for extracting information from the enterprise system and terms of use is not well described. In addition, there is no clear data custodian who can authorise access to data and evaluate the risk of data reuse for research purposes. We are currently addressing this issue by co-developing a responsible data sharing framework based on 5 Safes model following the Data Commissioner "*Best Practice Guide to Applying Data Sharing Principles*"[3].

---

[1] Australian Data Archive: https://ada.edu.au

[2] Analysis and Policy Online: Https://apo.org.au

[3] https://www.datacommissioner.gov.au/sites/default/files/2019-08/data-sharing-principles-best-practice-guide-15-mar-2019.pdf

- Working with heterogeneous datasets is a substantial challenge informing data-driven research projects. Data is often structured in different formats across multiple systems, and a centralised data transformation point may not be possible due to data sensitivity concerns. Hence, data needs to be de-identified, aggregated (geospatially or thematically) at each source, and to maintain the data integrity and quality a unified data aggregation model is required across all data sources.
- Managing the output of Data CO-OPs is a key component of the research project. Not all outputs are publicly shareable; however, a proper data management is required for sensitive output beyond the initial analysis. Although this can be managed ad-hoc, a proper data management plan and safety assurance model is required for these outputs.
- Measuring the impact of Data CO-OPs is very much depends on the traceability of the output connections to other scholarly works, this includes the connections between initial funding, researchers, data outputs, link to subsequent projects or grants, and connection to patents and other form of commercialisation. We have identified the need to embed the traceability of research objects in the Data CO-OP lifecycle and we are working on expanding this model beyond the scope of initial pilot project.

Impact

In this project, we are building the foundation for a new generation of data driven research projects where multiple organisations share data toward joint collaborative research. Our work has already led to engagements with state and federal government departments in the area of Mental Health, Social Wellbeing, and Family Violence. In addition, we have create a trusted engagement platform with local governments and not-for-profit organisation to provide data for social good and support initiatives in this space.

The current pilot infrastructure developed as part of this project has demonstrated reusablity of software components, reduction of duplicated efforts in GIS mapping, and increase the efficiency of data analytics across multiple projects. In addition, we are building the social infrastructure to form collaboration between data providers, consumers, researchers and policy makers.

To support research integrity, we have recognised the need to incorporate persistent identifiers including ORCID, RAID, and other elements of Research Graph into Data CO-OP lifecycle, and promote FAIR data output, especially from Open Data CO-OPs including citable accessible public datasets with DOI. In addition, we are working on co-developing the Responsible Data Sharing framework for Social Science and Public Health.

Out work has mainly focused on Social Science and Public Health. However, we are in the process of piloting Data CO-OPs in urban planning, industry 4.0, IoT, and supply chain optimisation.