

Douglas Boyle ¹, Roger Ward ¹, Daniel Capurro ², Jitendra Jonnagaddala ³, Siaw-Teng Liaw ³, Nicole Pratt ⁴

1. The University of Melbourne and in collaboration with Melbourne Academic Centre for Health (MACH) Advanced Health Translation Centre, Department of General Practice 2. The University of Melbourne, Computing and Information Systems 3. UNSW Sydney, WHO Collaborating Centre on eHealth 4. University of South Australia Quality Use of Medicines and Pharmacy Research Centre, Clinical and Health Sciences. In collaboration with Health Translation SA 4

Introduction

Observational research in Australia has suffered for many years due to an inability to link fragmented data holdings across clinical domains within states and territories in Australia. Additionally, linking data assets across state and commonwealth jurisdictions has been heavily restricted due to data custodian privacy restrictions. Data terminologies, data models, and the quality assurance mechanisms employed vary widely across Australia with no country-wide standard adopted. Such complexity creates significant barriers to the timely generation of high quality evidence on the utilization and safety of health care interventions specific to the Australian setting.

The Formation of the Australian Transformational Data Collaboration (TDC)

The TDC was formed as an initiative under the sponsorship of the Australian Health Research Alliance (AHRA) <https://ahra.org.au/> in October 2019 with the aim of bringing together researchers, government agencies and health care providers to work towards a common goal of breaking down barriers to data use in Australia. The AHRA TDC (www.machaustralia.org/ahra-tdc) is committed to uplifting clinical data for research across Australia and engaging national partners to support a consistent strategy in advancing Health Data Science. The TDC aims to accelerate evidence developments and related improved clinical outcomes and health system sustainability through the use of a common data model to standardize health data holdings across Australia.

Three priority areas have been identified for the TDC 1) enhancing Data Quality, 2) use of the OMOP Common Data Model and 3) Developing a Platform for sharing mappings and key learnings.

Data Translations in Australia

Currently, in Australia the AHRA TDC is supporting the following OMOP CDM conversion initiatives:

- Hospital EHRS
 - Two Cerner hospitals in NSW
- Primary Care Electronic Healthcare Datasets
 - The University of Melbourne, Department of General Practice Patron General Practice data repository (120 practices including a large number in West Melbourne)
 - The Victorian Comprehensive Cancer Centre (VCCC) Data Hub • Administrative Database
 - Australian Government Department of Veterans' Affairs health claims data
 - Australia wide Pharmaceutical Benefits Scheme Data

The AHRA Transformational Data Collaboration is creating tools to help streamline the process of the adoption of the OMOP CDM including an open source shared platform for hosting Australian specific data terminology mappings and an enhanced open source data quality tool.

A Platform for Collective Shared Mappings:

The Australian Digital Health Agency tool Ontoserver technology is being developed to host Australian data terminologies and data mappings as an open resource supporting terminology and mapping convergence in Australia. There is a standard process for creating these "source to standard" maps, but it is time consuming. Rather than individual research groups spending time and resources creating their own mappings we aim to share mappings between groups to speed up the process of mapping data. We have demonstrated this with two primary care data sets and are actively working on hospital data sets. We are developing the mappings in SQL server but ultimately plan to use "Ontoserver" developed by the CSIRO as the platform for sharing¹. Ontoserver is a clinical terminology server based on the Fast Health Interoperability Resources (FHIR) standard. Ontoserver has a REST API which allows sharing and transfer of mappings between systems making our mapping platform extensible. **Ontoserver is also strategically important as it is used by the Australian Digital Health Agency as the National Clinical Terminology Service for managing, developing and distributing national clinical terminologies. This includes being the Australian National Release Centre for SNOMED CT². We also plan to consider the use of Athena to distribute local mappings where appropriate.**

Quality Assessment Tool:

We are developing open source tools to enable the assessment of data quality in data warehouses according to international standards, to allow comparative analyses across repositories and to help researchers comply with research publication reporting standards such as RECORD³. We have developed an open source tool "White Bandicoot" for the examination of the quality of healthcare data prior to conversion to the OMOP CDM. The tool is written in Windows Presentation Foundation (WPF), .NET and based on the Kahn model⁴. In addition, we are allowing researchers to use the tool to annotate the data, to highlight quality issues which may impact on the use of the data. Wherever possible we have designed the tool to complement existing OHDSI tools such as White Rabbit and the OHDSI Data Quality Dashboard.

Figure 1: White Bandicoot in action



Supporting the OHDSI Community in Australia through the OHDSI-Australia Chapter

The formation of the TDC in Australia has led to significant engagement between Australian researchers and led to the creation of the OHDSI-Australia Chapter in 2020. The aims of OHDSI-Australia are to promote use of the OMOP CDM in Australia and to support the collaboration of Australian researchers with the OHDSI international community. OHDSI-Australia has established a series of bimonthly seminars and workshops on the OMOP CDM. Each workshop is recorded and is available here: <https://ohdsi-australia.org/events.html>

Conclusion:

Common data models and data standardisation are tools that permit collaborative research, large-scale data analytics and the sharing of tools, services and methodologies. The TDC will develop resources to support the uplift of Australian data to the OMOP Common Data Model, encourage a local OHDSI community of practice and promote education in the use of common data models.

References

- 1.Mecke-Jimenez, A. Steel, J., Hansen, D. & Lawley, M. Ontoserver: A syndicated terminology server. J. Biomed. Semantics. 2018; 9, 24
- 2.Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, et al. The Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Med. 2015; 12(10)
- 3.Kahn, M. G. et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. eGEMs (Generating Evid. Methods to Improv. patient outcomes). 2016; 4, 18