



techUK

Observational Medical Outcomes Partnership (OMOP) Common Data Model

Roundtable Minutes, Wednesday 1st November 2023

Minutes of the Roundtable on the OMOP Common Data Model, held at techUK on Wednesday, 1st November 2023

Contributors:

Edward Burnⁱ, Alex Eavisⁱⁱ, Jolyon Fairburn-Beechⁱⁱⁱ, Hillary Fanning^{iv}, Suzy Gallier^{iv}, Mustafa Ghafouri^v, Geoff Hall^{vi, x, vii}, Tim Hubbard^{viii, ix, x}, Richard Kavanagh^{xi}, Alex Knight^x, Alexandra Lawrence^{xii}, Ben McAlister^{xiii}, Ian Moody^{xiv}, Jon Payne^{xv}, Marius Posa^{xvi}, Paola Quattroni^x, David Seymour^x and Robert Walker^{xvii}

Introduction

techUK and Health Data Research UK (HDRUK) held a [roundtable discussion](#) on 1st November 2023 to explore the use of the [Observational Medical Outcomes Partnership \(OMOP\) Common Data Model \(CDM\)](#) in Secure Data Environments (SDEs) and elsewhere in the NHS and research. The panel included representatives from industry, NHS organisations and academic research.

The OMOP Common Data Model has been widely adopted to support the generation of Real-World Evidence from routinely collected health data. The use of a CDM permits studies to be conducted across multiple sites, organisations, and countries. Importantly, OMOP has been [adopted](#) as a CDM by the Data for R&D Programme's NHS Research Secure Data Environment Network. The panel discussed the challenges and barriers to the adoption of OMOP, and suggested steps that could enable them to be overcome.

HDR UK Presentation

[\[Presentation\]](#)

- David Seymour outlined HDR UK's overall mission to unite healthcare data across the UK
- The 3 pillars are:
 1. Support health and boost UK science
 2. Provide leadership to fix difficult technical problems
 3. Accelerate and streamline health data
- Members raised questions around if OMOP is being applied to just NHS Secure Data Environments (SDEs) or if it would include OpenSAFELY and the Federated Data Platform (FDP)
- There was general agreement from the group that it's imperative that the implementation of OMOP CDM is consistent. Interoperability across platforms is needed and need to avoid different 'dialects' of OMOP.

Current challenges & opportunities

- It is demonstrated that mapping data to a common data model is needed for international collaborations, sharing codes (not data) between countries and driving federation across countries.
- It can also enable a wider, more heterogeneous population to be used for research, supporting both 'levelling up, diversity in data' and achieving quicker and more impactful research.

- But not everyone understands why mapping data to OMOP is needed, what the use cases are. The UK is in a slightly different position. The historic model has been to link data together from different sources in one place and this can reduce the need for standardisation at source.
- There is a risk of developing different OMOP “dialects” and consistency is needed. A common data model is needed to enable pan-UK (four nations) research use cases.
- There is a role for NHSE in mapping centrally (or at least standardising mapping) of common national core data sets such as HES. NHSE as a data provider not a consumer. The BHF Data Science Centre is already working with these core data sets.
- Scaling up for large research projects is a challenge.
- Many people are developing tools and mappings for free and releasing them “open source”. This area needs professionalisation to be sustainable.

Data Quality (DQ)

- The discussion initially centred around data quality – members saw the potential of OMOP but outlined that data is not always in an accessible state.
- Quality control is important. Mapping to OMOP doesn’t mean that data is research ready – but it does provide a benefit as you find things out when mapping, and it builds knowledge and experience of the data asset. It also standardises ways of looking at and describing DQ.
- But (some) data custodians may not like this... in terms of lifting a stone on the quality of their data.
- There were concerns that many NHS trusts don’t have a systematic record of what data they have, and are working across a heterogeneous range of systems, both presenting a barrier to the implementation of OMOP.
- Deviation of data models across the four nations of the UK, particularly in cancer research, was raised as a barrier.
- The issue of different versions of OMOP being used with the potential of different ‘dialects’ of OMOP developing.

Training

- The lack of trained OMOP users and lack of funding for staff training is holding back the use of data assets that have been transformed, and limiting the ‘pull’ for further creation of OMOP data assets.
- NHS analyst workforce: there is a need to train NHS staff. Linking to existing groups such as RAP (Reproducible Analytical Pipeline) NHS group is important.
- OMOP requires epidemiologists to change their ways of working – limited experience of running or taking part in networked studies across multiple data partners using the OMOP CDM and standardised analytical tools.
- Members stated the pressing need to familiarise researchers with the use of OMOP.
- Complementary to training in the use of OMOP is training in cloud use and cloud tooling.
- Industry can offer part of the solution (provision of systems; services; automation, etc.)

Standardisation

- There is a high level/general agreement that OMOP mapping is useful and there is a commitment to standardisation, but getting there is a challenge.
- ‘At source’ OMOPification seems a sensible approach, but heterogeneity of systems and quality of Real-World Data are limitations and it’s not clear what the financial return for hospital trusts is (without a stronger ‘pull’ from potential user base).
- Members felt there was a good opportunity for standardisation across SDEs.
- But variability in approaches, capabilities and skills, resources, and priorities limit adoption across SDEs.
- Adoption of an open-source approach is one way to encourage standardisation through co-creation of resources.

Mapping a Minimum Viable Dataset

- There was consensus that agreeing on a “minimum viable dataset” to implement at the hospital level would be useful.
- Whilst it might be limited in applications, it would lower the barrier to OMOPification and ensure that the data is used. Can then build on that.
- An example of case study is a cancer collaboration led by Leeds Teaching Hospitals to show COVID impact on cancer diagnosis across 6 countries using a simple OMOP dataset. Even data sets with minimal details can enable simple but very broad studies (with large populations, e.g. international).

Primary care data

- Primary care represents a significant opportunity for transformation ‘at source’ due to the small number of system suppliers.
- Suppliers acknowledged that they have a duty to communicate between themselves – TPP, EMIS, etc. should all be talking to each other and standardising data at source
- Is there an opportunity to get a steer from NHS England in using existing contracts to perform OMOP mapping of GP data consistently?
- However, stakeholders need to agree if suppliers are going to standardise data at source. Questions were raised on how to incentivise this.
- There is also a potential opportunity to generate structured secondary care related data from application of Optical Character Recognition (OCR) and Natural Language Processing (NLP) to referral and discharge letters to generate coded data that can be transformed into OMOP.

Fast Healthcare Interoperability Resources (FHIR)

- There was discussion of where FHIR fits into this landscape and if OMOP could be considered a constrained subset of FHIR.
- Relationship between FHIR as implemented for direct care, and reusability of data for research needs further consideration for future proofing
- Members proposed having a minimum set of FHIR mapping as a starting point for all hospitals’ direct care – and then build OMOP maps “on top”. This is a point of potential convergence.

- Members appreciated the UK FHIR CORE implementation guide, which is useful for mapping. Could an OMOP implementation guide be valuable?
- The FHIR terminology server also offers an opportunity to standardise. Scotland is up and running. How does this connect with the Phenotype library, and OpenSAFELY code lists?

How can we move forward?

- The importance of use cases was acknowledged to show what's possible with OMOP and what the impact of using OMOP data is. These are likely to include:
 - UK-wide and international studies.
 - Research into rarer diseases, including cancer as a (set of) rare diseases, where there is a need for scale to get meaningful findings.
 - Speed and reproducibility e.g., pandemic, and speed at which major centres are able to contribute.
- Members raised the idea of the creation of a minimum viable dataset that might be used for different case studies. This should be developed to support priority use cases or research questions that need addressing.
- The work of the NHSE Transformation Directorate can facilitate nationwide OMOP mapping and we need to try and leverage this, but they are not the only players.
- There is a need to identify sustainable sources of funding for the data transformation and early adopter users, now that the European Health Data Evidence Network (EHDEN) project is coming to a close.
- We could explore international funding as well as UK funding but we need to demonstrate the need for it and the usefulness of data transformation.
- The SDE use cases could help demonstrate impact and need for further funding.
- Potential of dedicated funding for OMOP related training.

The role of industry

Industry can:

- Help Trusts and other organisations 'OMOPify' their data warehouses – 'the effort is in building it rather than maintaining it'.
- Provide capacity and expertise, help automate processes, and reduce burden.
- Showcase research that couldn't be done without OMOP.
- There were questions around how to achieve scale and how to lower the technical barrier to access.
- Building skills is key. In Europe, creating a 'fear of missing out' is working – members highlighted that the EMA provides funds to those who want to participate in research studies if they map to OMOP through DARWIN-EU.
- For primary care data, there is a need for the systems providers to agree on OMOP at source.

Next Steps

1. **Researcher/analyst skills and capabilities.** The group agreed that increasing familiarity with the use of OMOP across UK researchers and NHS analysts will be important. Currently, there are only a small

number of experts in the use of OMOP, and very few research centres specialise in this. Improvements in capacity, training and skills development are important to drive use of OMOP in research (and hence the pull for further data transformation).

- 2. Use cases and priority research.** It was agreed that NHS England's investment in development of SDEs can be used as an opportunity to drive OMOP mapping as well as the use of OMOP data. Understanding the priority research questions that can be addressed through the adoption of the OMOP common data model would demonstrate the impact of OMOP transformation. This will help make the case for more sustainable solutions and long-term funding mechanisms. Funders and decision makers need to be part of the conversations, as there is also a **lack of funding** available for development of open-source analytical tooling that can be deployed for rapid and reproducible analysis of data that conforms to OMOP CDM. The work of OHDSI UK could help both use cases and researcher skills and capabilities by bringing researchers together to achieve that (using examples of Hackathons/Studyathons).
- 3. Minimum viable dataset.** There is a case for agreeing a minimum dataset to propose for OMOP mapping in order to answer key priority research questions and also the feasibility use case ("cohort discovery"). Community work could facilitate agreement. It will be important to not reinvent the wheel and keep bringing communities together and sharing knowledge. Closer work with the NHS Reproducible Analytical Pipeline group might help.
This can be a focus of OHDSI UK or Alliance Special Interest Group work, and can include input from the Driver programmes and HDR UK's wider networks.
- 4. National data collections and GP Data.** Work with EMIS/TPP is key to achieve standardisation across GP data systems. GP data are one of the key priorities for access, availability and usability. It seems clear that systems providers will play a key role. HDR UK could facilitate the conversation between NHS and systems providers. For national (secondary care) collections such as Hospital Episodes Statistics (HES), National Disease Registration Service (NDRS) data, need to engage with relevant teams in NHSE to influence adoption of OMOP and support transformation to be done once centrally and then shared back to local data custodians.

About tech UK

techUK is the technology trade association that brings together people, companies and organisations to realise the positive outcomes of what digital technology can achieve.

With around 1000 members (the majority of which are SMEs) across the UK, techUK creates a network for innovation and collaboration across business, government and stakeholders to provide a better future for people, society, the economy and the planet.

By providing expertise and insight, we support members, partners and stakeholders as they prepare the UK for what comes next in a constantly changing world.

About Health Data Research UK

HDR UK is the national institute for health data that includes England, Wales, Scotland and Northern Ireland. Its mission is to unite the UK's health data to enable discoveries that improve people's lives. It is a registered charity with core funding provided by UK Research and Innovation, the Department of Health and Social Care

in England and equivalents in Northern Ireland, Wales and Scotland, and leading medical research charities.
www.hdruk.org

The UK Health Data Research Alliance is an independent alliance of leading healthcare and research organisations united to establish best practice for the ethical use of UK health data for research at scale. Convened by Health Data Research UK, the UK Alliance develops and co-ordinates the adoption of tools, techniques, conventions, technologies, and designs that help researchers to answer some of the most difficult questions and address the most important health challenges faced in the UK. The UK Alliance is part of the UK-wide capability for the responsible use of health data for research and innovation, now funded through HDR UK core funding, having been initially funded by Innovate UK as part of the Industrial Strategy Challenge Fund (ISCF). www.ukhealthdata.org

ⁱ Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford

ⁱⁱ EMIS

ⁱⁱⁱ GlaxoSmithKline

^{iv} University Hospitals Birmingham NHS Foundation Trust

^v CloudKubed

^{vi} Leeds Teaching Hospitals NHS Trust

^{vii} University of Leeds

^{viii} Genomics England

^{ix} King's College London

^x Health Data Research UK

^{xi} Kainos

^{xii} techUK

^{xiii} Oracle

^{xiv} Clinical Architecture

^{xv} InterSystems

^{xvi} Palantir

^{xvii} techUK