Mapping Our Future Health Data to OMOP

Magda Meier, Kevin Garwood
Our Future Health, Manchester, United Kingdom.

Background

Our Future Health is a non-profit organisation which aims to recruit 5 million adults to a prospective cohort that will catalyse aetiological and translational research for both common and rare diseases. It is already the world's largest consented cohort study of its kind. Its size, depth and breadth make the cohort data a crucial resource for population health studies. We currently make health care records, genetics, clinic measurements, and questionnaire data available for approved research studies. The health care records will span a variety of health services across the UK's nations. Our Future Health seeks to map its data holdings to OMOP to better support data discoverability and a wider range of studies. It also seeks to leverage tools and data sets that have already been developed by the OMOP community. We are aware that some of our researchers will want to reproduce analyses conducted elsewhere. Supporting these next generation studies will require Our Future Health to invest in interoperability based on common data models. Our Future Health is not just interested in OMOP mappings. We are interested in the stability and scalability of infrastructure that can support those mappings in a production setting as well. Rather than try to build it all ourselves, we are seeking to collaborate with different OMOP groups and to pool collective experience. As part of our early efforts, we have been examining the use of various OMOP tools, libraries and reference data sets that would support whole-cohort processing in a cloud native environment. Currently we are evaluating how tools such as Carrot, White Rabbit, and Rabbit in a Hat would support our data processing characteristics. We are also investigating how OMOP could better support AI-related activities.

Methods

We aim to use the following approach:

1. **Identify exemplar research studies.** We wish to engage researchers whose work with our future OMOP data sets would have the greatest impact on healthcare. This will help us understand the priorities and how we can have the most impact.
2. **Limit the scope of data to map.** We envision that OMOP suits some data modalities better than others. We will initially focus on mapping the participants' routinely collected health data gathered from England, Scotland and Wales based on community input from other cohorts who have had to do similar mappings. We will assess whether OMOP can help reduce the work of harmonising data from the UK's nations that would otherwise be done without the use of a common data model. We will then prioritise other OMOP mappings based on their complexity and stability.
3. **Partition data processing tasks and evaluate suitability of OMOP tools for each.** We will separate concerns related to those of domain experts defining the mappings and those of data engineers implementing them at scale in production. We are specifically interested in the interplay between these two areas of interest.
4. **Prototype OMOP workflows.** Based on our scoping priorities for data, we will develop exemplar data flows that will map participant data sets. We are specifically interested in how community-informed iterations of mapping can be supported.
5. **Assess the performance of the OMOP workflows.** Because our cohort is so large, we will need to assess the computing performance of code used to process the mappings in our data release activities.
6. **Assess maintainability of OMOP workflows.** We are less interested in learning to do one-off OMOP mapping efforts and more in how those would work in regular production activities.

Results

Much of our efforts spent thus far have been about gathering a community of OMOP partners and scoping the data sets that are likely to generate the most return value for our researchers. We have recently examined how our OMOP community partners could work together and how mappings could be managed (see Figure 1).

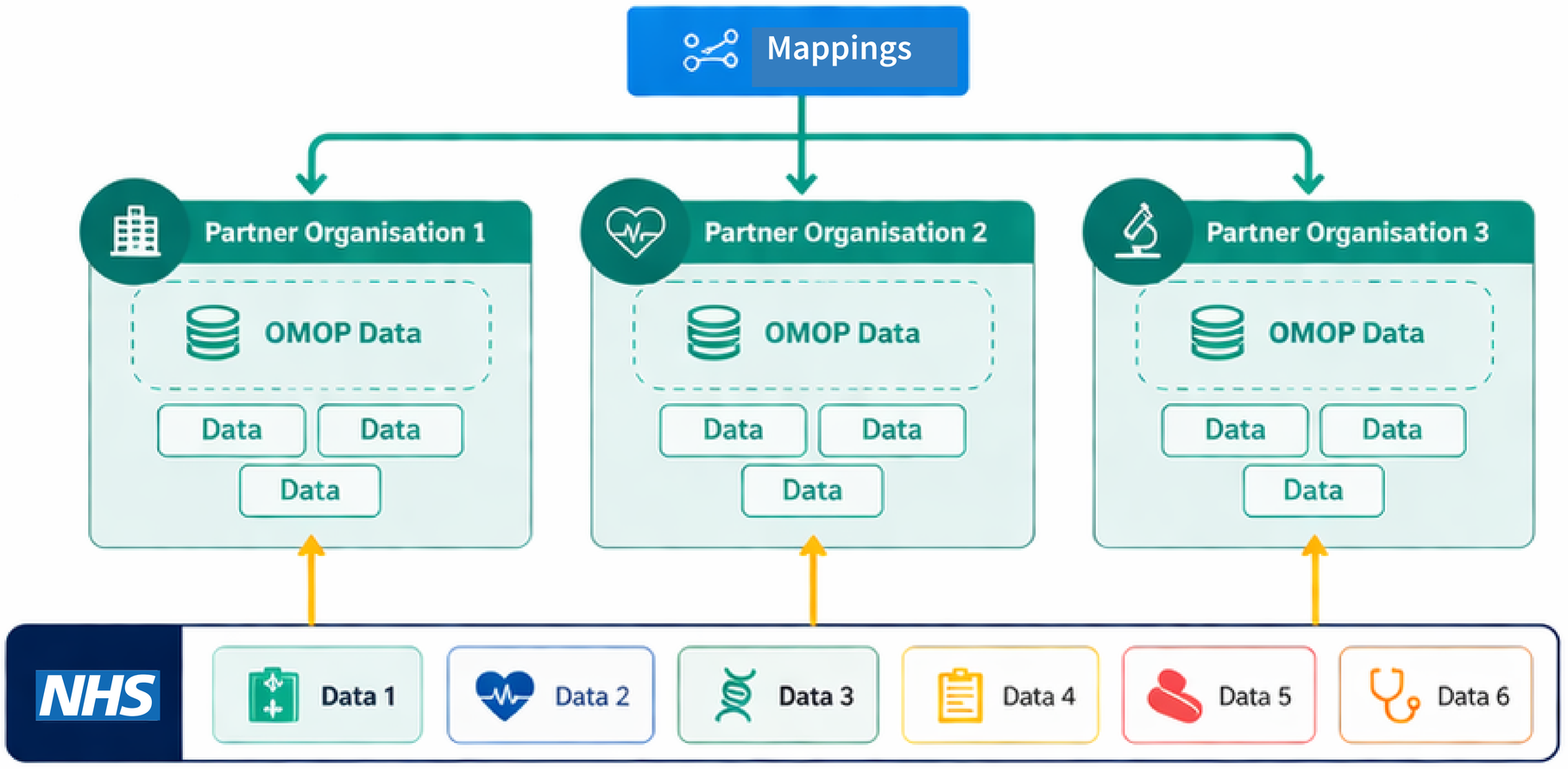

Figure 1. Supporting a community-evolved OMOP mapping effort.

Conclusions

From our work so far, we have concluded:

- Most OMOP mapping projects we have observed so far appear to be one-off efforts that support a specific study rather than provide a long-term, general-purpose service. Most of those projects emphasise the characteristics of particular mappings rather than the software maintenance of those mappings. Maintainability remains an important aspect of a sustainable OMOP effort.
- Groups we have consulted who have significantly invested in OMOP advocate rationalising its coverage by business need rather than by attempting to map entire data assets as a goal in itself.
- Bearing in mind the cost of maintaining mappings in production, we will emphasise the oldest and most stable use cases for which OMOP was originally developed.
- We believe in investing in using a basket of common data models which best suits specific data modalities. OMOP appears well-suited for mapping routinely collected health data and to a lesser extent questionnaires. Other data modalities such as genetic, imaging, and geospatial data may warrant being aligned with more domain-specific common data models.
- Groups we have consulted who use OMOP have an appreciation for how much work is involved when they try to do mappings on their own. This collective realisation has provided the driver for us to engage in our OMOP activities through a cooperative effort featuring shared experience and likely shared mappings amongst our potential partners.

References

Cook MB, Adams N, Adjetey A, Arathimos R, Balabanovic M, Blackwood R, Booth A, Cairns BJ, Connell A, Ellis S, Elsworth B. Cohort Profile: Our Future Health. International Journal of Epidemiology. 2025 Dec;54(6):dyaf171.