Automation of the OMOP conversion process using an end-to-end software tool

Sonia Araujo, PhD[1], Christian Reich, MD PhD[2], Katie Mannering[3], Tony Aldridge, BSc[3]

[1]IQVIA, London, UK; [2]IQVIA, Cambridge, MA, USA; [3]IQVIA, Brighton, UK

Abstract

Observational data such as that in EHRs and claims databases can significantly contribute to academic and commercial research. As the COVID-19 pandemic has proved, observational research can help health providers make better decisions, and ultimately lead to better outcomes for patients.

However, observational data in its native format presents challenges, such as: disparate data structures, country- or organization-specific medical vocabularies, privacy concerns, etc. Converting observational data to the OMOP format results in harmonization of data, reducing or eliminating some of those challenges, and enabling systematic and collaborative research.

Whilst having data in OMOP format brings enormous benefits and opportunities for research, the path to getting there is often hard, unreliable and time-consuming, especially when an organization is new to conversion processes.

The IQVIA OMOP Converter is a software tool that semi-automates the conversion of data to the OMOP format. It lowers the threshold of joining the OHDSI community, and lowers the cost of maintaining high-quality OMOPed datasets.

Introduction

The OHDSI community has produced a series of scripts and tools that help at specific points of the conversion process. However, those tools don't have a consistent or transparent user experience, can be complex to use, are not connected to each other and have no common technology. This leads to conversion scripts being non-reusable and non-scalable, and conversion projects being slow and expensive. This is an issue particularly for those OHDSI members who lack technical/SQL expertise in their teams.

The IQVIA OMOP Converter is addressing these challenges. It takes users from importing their data asset to completing field mappings and re-using mappings for data refreshes; from running conversions to analyzing results and exporting the OMOPed data. All on the same interface, requiring little technical knowledge.

In this talk, we will be presenting version 1 of the software. We will share the overall software workflow and walk through a few scenarios, such as: profiling source data, creating complex-logic field mappings without writing any SQL, and viewing progress for a conversion run as it's happening.

High-level Functionality

The IQVIA OMOP Converter aims to emulate the workflow of a traditional conversion project (Table 1). One notable difference is that the ETL specification document is no longer an input for developer work, but an iterative output of the mappings that are completed interactively on screen.

Table 1. Components and high-level functionality of the IQVIA OMOP Converter

Component	High-level functionality
Pre-processing	Importing data assets. Reviewing/confirming inferred data schema. Profiling data. Sampling data.
Mapping	Mapping source to target tables (many to one) & source to target fields (many to many). Defining target field logic. Creating lookups, staging tables, filters and joins (for those more SQL inclined). Importing source-to-concept-map files and managing vocabularies. Viewing ETL spec doc.
Conversion	Running a conversion for full or sample datasets. Viewing progress of a conversion run. Stopping a conversion run early if errors reported. Exporting OMOPed data.
Dashboards	Viewing conversion run results, overall and per CDM table. Viewing all runs' summary & statuses, and all mappings' summary & statuses. Exporting charts for offline review and further analysis.

The visual interface guides the user through the different actions, with guard rails present to keep errors down (and catch them early). Additionally, the ability to incrementally map and run a conversion, table-by-table, allows users to check their assumptions, catch errors early and try things out quickly.

Mappings can be copied across different OMOP CDM versions and different data asset releases, ensuring users don't have to start from scratch every time. Thus helping good maintenance of OMOPed datasets.

We support OMOP CDM v5.2 and v5.3.1. Support for v6 and CDM extensions (e.g., Oncology) will follow once those are more stable and OHDSI tools like Atlas support them.

The software's tech stack includes Spark (framework), Scala (backend) and React (frontend), making the software resilient and able to handle well the large amounts of data characteristic of real-world data assets, along with a responsive interface. The software can be deployed on-premise or private cloud, according to company infrastructure and security policy; or on IQVIA's SaaS cloud environment.

Conclusion

The OHDSI Community has a constant stream of network studies, where several members collaborate on specific research questions. The quality of the data used is vital to ensuring reliable study results. For that, source data needs to be reliably and accurately converted to OMOP. However, conversions are often hard, unreliable and time-consuming, especially when an organization is new to the process.

The IQVIA OMOP Converter presents an option for organizations who are OMOP-knowledgeable but lack technical expertise or funds to contract vendors, to complete conversions themselves in a way that is transparent, reusable, reliable and cost-effective. Hence, it lowers the threshold of joining the OHDSI community, and lowers the cost of maintaining high-quality OMOPed datasets. So together, we have a better network.