Creation of a reusable OMOP transformation workflow for Belgian electronic health record systems

Jared Houghtaling[a], Lore Vermeylen[a], Louise Vandenbroucke[b], Korneel Bernaert[b], Brecht Dekeyser[b], Freija Descamps[a]

[a] edenceHealth NV
[b] AZ Damiaan (Oostende, BE)

Background:

AZ Damiaan is a mid-size regional hospital located in Oostende in Belgium with 513 hospital beds **[1]**. The hospital joined the European Health Data and Evidence Network (EHDEN) in early 2022, and has been working together with edenceHealth to transform electronic health record (EHR), registry, and laboratory information system (LIS) data to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). The structure of the transformations that have been implemented thus far lend themselves toward generalizable tooling for hospitals with comparable medical record structures. More specifically, the team has opted for creating pre-OMOP staging tables that link source tables captured in different vendor systems and serving various purposes within the local data warehouse. In this work, we aim to present the approach taken to transform this rich datasource into OMOP CDM by (1) describing the processes in detail, (2) detailing the inherent challenges and advantages when working with this sort of data, and (3) outlining the next steps in generalizing this approach to apply at other hospitals around Belgium with comparable EHR systems.

Methods:

We performed the majority of the semantic mapping for this harmonization effort using edenceReviewer, a web-application alternative to Usagi that enables collaborative review of mapping suggestions. We first deployed an automated mapping – suggestions generated by a set of algorithms curated by edenceHealth – and uploaded it to the web portal, after which members of the team at AZ Damiaan manually approved or modified a total of 2140 source-to-standard mappings. In instances where community mappings were available – for ICD10CM, ICD10PCS, ATC, GGR and LOINC codes – we created lookup tables using the OMOP standard vocabulary relationships. We did not need to create custom concepts to map the observational data in this project (see **Table 1** for mapping coverage). Note that the semantic mapping is still ongoing; we are prioritizing drugs initially because that domain has the lowest mapping coverage in our dataset, and the 20 most frequent un-mapped drugs represent nearly 1M additional records. The cancer registry also requires additional manual mapping – coverage using automated ICD03 lookups was rather poor, and we have begun looking toward regional and national cancer registries for collaborations in this effort. We intend to share the semantic mappings we have produced thus far to expedite other Belgian harmonization processes and to help validate the mappings themselves.

Table 1. Overview of semantic mapping coverage per OMOP domain

Domain	Source Codes	Source Mapped	Mapped Percent.	Source Records	Target Records	Record Percent.
condition	15,023	13,565	90.29	1,271,831	1,049,359	**82.51**
procedure	10,032	7,769	77.44	4,659,029	4,339,049	**93.13**
device	6	6	100.00	14,378	14,378	**100.00**
drug	8,051	1,048	13.02	8,236,773	6,608,283	**80.23**
observation	2,014	2,014	100.00	1,003,486	1,003,486	**100.00**
measurement	130	127	97.69	3,644,471	3,643,123	**99.96**
visit_occurrence	13	12	92.31	1,876,256	1,875,487	**99.96**
meas-unit	6	6	100.00	65,043	65,043	**100.00**
meas-value	1,303	1,303	100.00	31,503	31,503	**100.00**

The main structure of the Extract-Transform-Load (ETL) process is written in python; transformations are written in embedded SQL using an f-string method (i.e. referencing python-defined objects within raw queries) and the tables are constrained using SQL Alchemy models. The ETL connects to the AZ Damiaan data warehouse, specifically referencing a set of views generated for the purposes of this EHDEN project. The ETL exports those views to comma separated value (CSV) files using the Microsoft bulk copy (bcp) utility, and then loads them into a temporary source schema in the target OMOP database (dedicated SQL Server instance) where they are joined with lookup files and transformed into their associated OMOP tables before being removed. We deployed the ETL as a Docker container on a dedicated Windows Server hosting the other OHDSI applications.

Results:

In total, we transformed approximately 27M unique records (incl. derived OMOP tables) for 182’000 individuals affiliated with AZ Damiaan hospital encounters. We faced several challenges in this harmonization effort: there were multiple formats of patient identifiers across different source systems that we needed to consolidate and subsequently validate, data access within particular source vendor systems was restricted and required workarounds for indirect access, and source concept descriptions for the procedural (via invoicing) records in particular were highly granular. Building on these experiences, we are in the process of compiling an EHR OMOP module that could be deployed at other hospitals around Belgium with similar health record profiles and source vendors. We will construct a set of SQL-Server-based procedures and associated mapping tables that can execute at regular intervals within a hospital data warehouse. We expect this module and its various components to improve the harmonization efficiency and conformance for hospitals among various Belgian-specific consortia in which we are participating.

Conclusion:

The work presented here is a first step toward a general OMOP ETL that could be deployed within Belgian hospitals. While OMOP presence within the country is growing, many hospitals have yet to harmonize their data to the international standard. We hope that by creating a more general harmonization process flow using a multi-stage transformation and tooling that is technology agnostic, we can help promote further adoption of OMOP across Belgium. If successful, we expect the work to expand opportunities for Belgian data owners to participate in cutting-edge, observational health data studies both within Europe and beyond.

References:

[1] *Ziekenhuis Oostende*. Az Damiaan. (2023, April 27). Retrieved April 27, 2023, from https://azdamiaan.be/