PhUSE 2017

Paper RW05

Applying the OMOP Common Data model to Dutch healthcare data: a case study

Rients van Wijngaarden, MSc, PHARMO Institute, Utrecht, The Netherlands

ABSTRACT

The Observational Medical Outcomes Partnership (OMOP) was a public-private partnership aimed at the development of a common data model (CDM) and vocabulary for observational healthcare databases. Once a database is mapped to the OMOP CDM a suite of tools is available to perform quality assessments and various analytical methods. Furthermore, the use of the OMOP CDM over various databases facilitates the execution of large scale research. The PHARMO Data Network (PDN) provides rapid access a rich population based network of databases derived from different primary and secondary healthcare settings in the Netherlands. We have applied the OMOP CDM to the PHARMO diabetes cohort containing detailed healthcare information of approximately 75.000 patients recorded by General Practitioners. The biggest challenge was the mapping of regional coding systems to international standards, but the resulting standardized database enables us to perform specific studies more efficiently and further enables international collaborations.

INTRODUCTION

The PHARMO Institute[1] is an independent research organization dedicated to the study of epidemiology, drug utilization, drug safety, health outcomes, and utilization of healthcare resources and works closely with academic partners and European databases. To facilitate this type of research the patient centric PHARMO Database Network is developed and maintained. The PHARMO Database Network provides rapid access to rich and detailed information of a well-defined population in the Netherlands derived from different primary and secondary healthcare settings. A average of ten years of data is available for more then 4 million residents, roughly 25% of the total population, and through yearly updates the longitudinal follow-up of individuals in the Network increases. Bespoke disease-specific cohorts are derived from the PHARMO Database Network and can be further enriched through linkages with other external healthcare databases.

One of the challenges for PHARMO and other Electronic Health Record (EHR) databases when performing clinical research is that different data standards have been applied and the data is often (partly) coded in national coding systems. A semantic mapping to more generally used international coding systems is needed to enable collaboration with other databases. In addition the underlying structures of collaborating databases are often different and the sharing of methods and algorithms is not always straightforward and can require considerable effort. In order to tackle these challenges a subset of the PHARMO Database Network (PDN) was mapped to the OMOP common data model. This was done within the context of the European Medical Information Framework (EMIF) project[2]; an IMI funded project aimed at providing one platform for data discovery, assessment and (re)use.

The Observational Medical Outcomes Partnership[3] was a public-private partnership aimed at the development of a common data model (CDM) and vocabulary specifically for observational healthcare databases. The partnership ended in 2013, however the Observational Health Data Sciences and Informatics (OHDSI) collaborative[4] has been established as a multi-stakeholder, interdisciplinary collaborative that actively maintains the OMOP common data model and vocabulary. One of the aims of the OHDSI collaborative is the creation of open-source solutions to leverage the value of observational health data through large-scale analytics.

A subset of the PHARMO Database Network was chosen to perform the mapping to the OMOP common data model on. The data collected from the General Practitioner (GP) between 1 January 2006 and 1 February 2017 for a sample of approximately 75.000 type 2 diabetes (T2DM) patients was used giving us a well defined cohort with clear entry criteria. Furthermore diabetes care in the Netherlands is well organized and the majority of T2DM patients are managed within the primary care setting, thus ensuring that a complete profile of the medical history of patients is captured in the cohort.

This paper describes the steps taken to create the mapping, build the Extraction, Transform and Load (ETL) scripts and deploy the tools provided by the OHDSI collaborative as well as the challenges and considerations with each

step. In the first section the syntactic mapping is described followed by the semantic mapping. Next running the ETL process and setting up the tools is described. After that future steps are set out, followed by some concluding remarks.

SYNTACTIC MAPPING

The first step in the process of mapping to the OMOP common data model is the syntactic mapping. In this step the data items available in the source data are mapped to the corresponding data items in the predefined target tables. For example the data items related to medication prescriptions have to be mapped to the *drug exposure* table and data items related to the responsible health care provider have to be mapped to the *provider* table. More information about the OMOP common data model is readily available online.

A convenient set of tools has been developed by the OHDSI community to aid in the syntactic mapping. 'White Rabbit'[5] is able to scan the source tables, summarize the content and can even be used to generate fake data while respecting the statistical distribution of data elements in the source. The resulting scan report can be passed on to 'Rabbit In A Hat' which provides a graphic user interface create and annotate the mapping and generate specifications for the extraction, transform and load process.

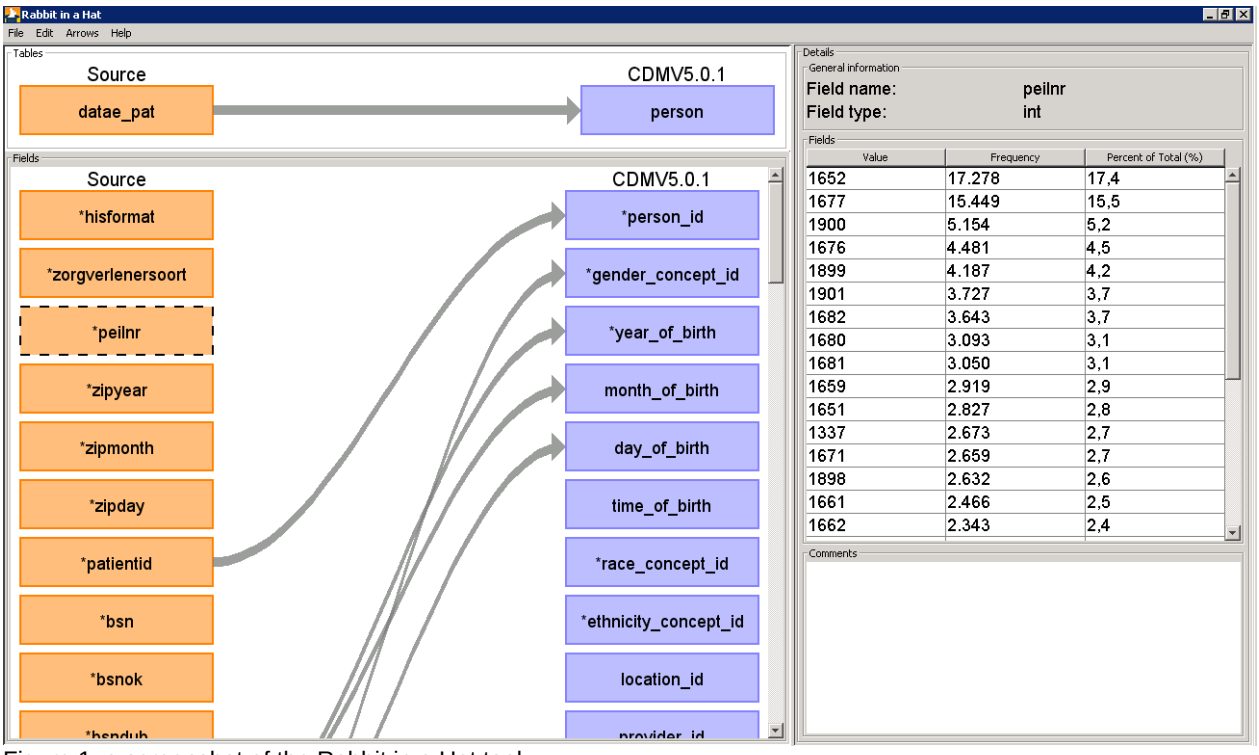

Figure 1: a screenshot of the Rabbit in a Hat tool

In the case of the mapping described in this paper, the syntactic mapping phase and the subsequent generation of the ETL specification was done during two subsequent face-to-face sessions with the respective PHARMO experts, technical staff, OMOP CDM experts and ETL developers. The main challenge was the unfamiliarity with the underlying common data model and the tables, vocabularies and concepts used.

SEMANTIC MAPPING

The second step in the process is the semantic mapping. In this step local coding systems are mapped to the coding systems as used within the OMOP common data model. Tools have been created by the OHDSI community to aid with this mapping; however these are based on the English language and are not directly applicable for Dutch descriptions. The table below describes the semantic mappings that had to be created for this project.

Clinical concept	Local coding system	(Inter)national	Mapped to
Drug prescriptions	KNMP	National	RxNorm
Conditions	ICPC	International	SNOMED
Observations	WCIA	National	SNOMED
Measurements	WCIA	National	LOINC

Table 1: semantic mappings created

This step was the most time intensive step and not all source concepts could be adequately mapped to the target coding systems. The order in which source concepts were mapped was based on the frequency i.e. mapping the most frequent source concepts first. Although a 100% mapping of source to target concepts was not feasible for all four created semantic mappings at least 80% of the cumulative count of occurrences have been mapped.

RUNNING ETL AND SETTING UP TOOLS

The result of the syntactic and semantic mapping steps was a set of SQL scripts and mapping tables designed to perform the entire process from setting up the target tables to performing the ETL. In our case 4 iterations were needed to iron out all the wrinkles. One of the main challenges, apart from technical intricacies, was deciding where to handle data issues; in the source tables or in the ETL. For example some observations had a start date before the birth date of the patient in the source data. In the end we chose to solve these issues as much as possible at the source data to keep the ETL process from becoming too complex.

After the ETL process was successfully completed the next step was to run the 'Achilles'[6] tool for data characterization. 'Achilles' consists of an R package and summarizes the data loaded into the OMOP common data model and creates several standard reports presenting patient characteristics, drug exposures, conditions etc. In a second step the resulting .JSON output can be visualized in a web application. Screenshots of the Achilles Heel dashboard and the Drug exposure report have been added in figure 2 and figure 3 respectively.

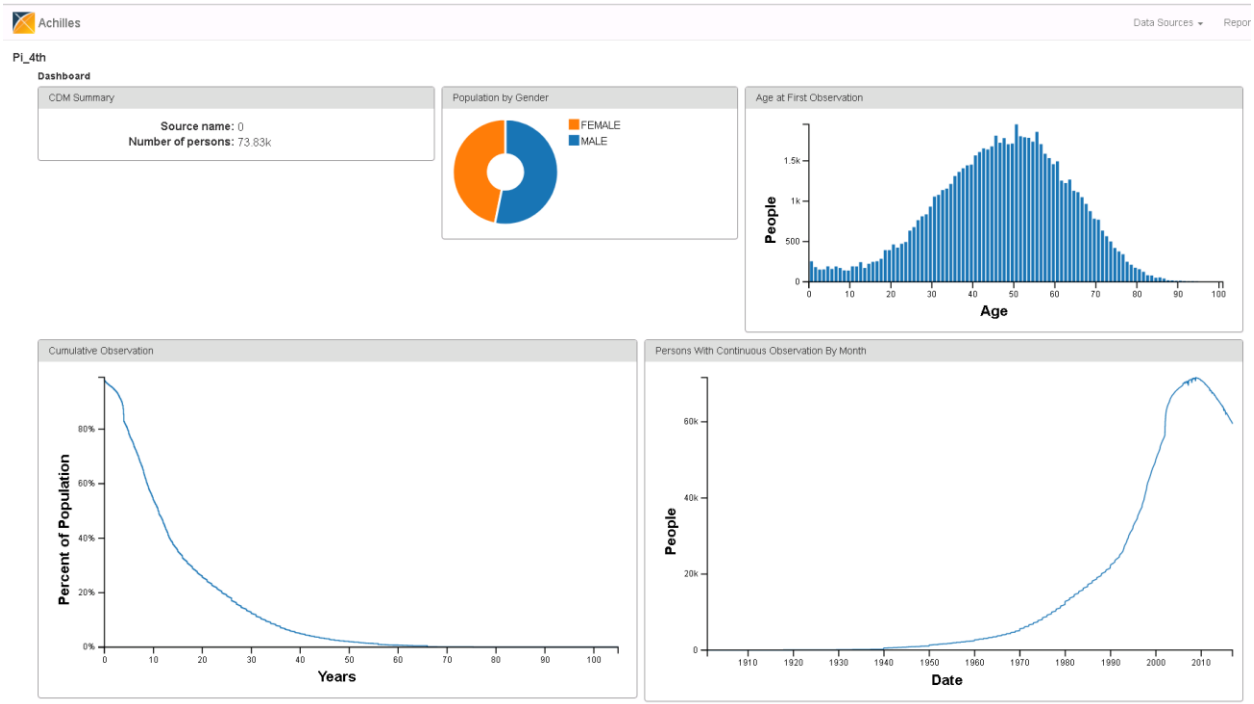

Figure 2: A screenshot of the Achilles Web dashboard

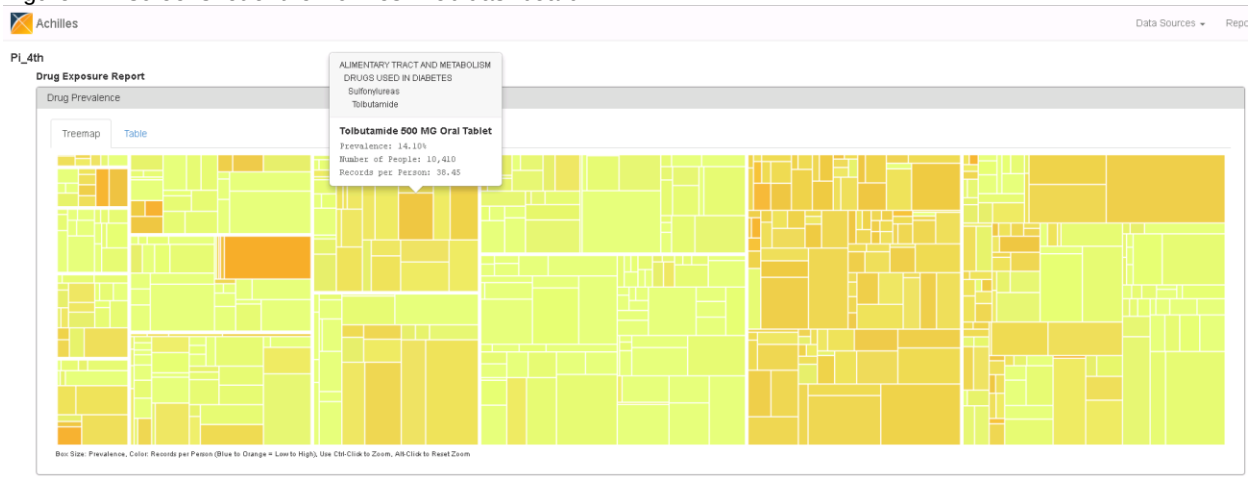

Figure 3: A screenshot of the Achilles Web drug exposure report

Finally we installed the 'Atlas'[7] tool. 'Atlas' has been created to bring different individual components together in one user interface and aid researchers to perform scientific analysis on the OMOP data. It allows the user to create cohorts based on entry and exit criteria, enables the user to browse the various vocabularies and also includes the functionality of the 'Achilles' tool to visualize the underlying characteristics of a cohort. The deployment of the 'Atlas' tools is not very straightforward due to many software dependencies.

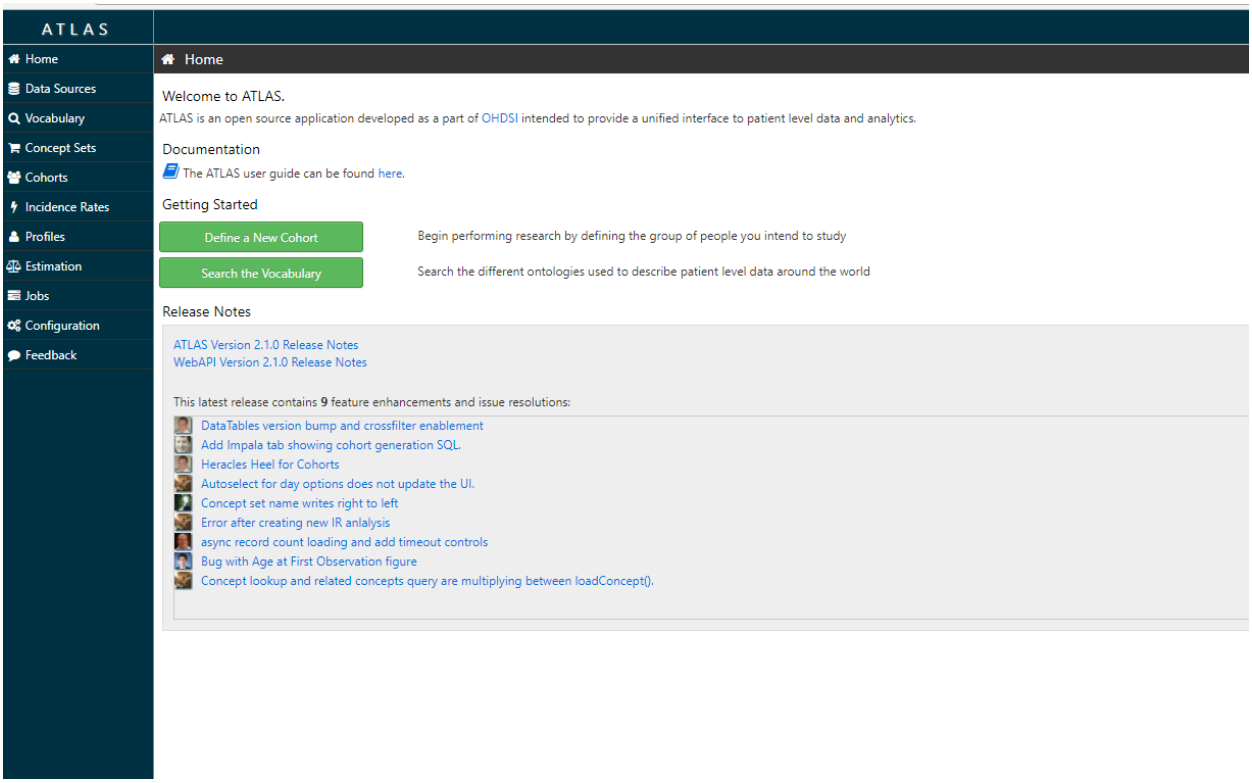

Figure 4: A screenshot of the Atlas tool

FUTURE STEPS

Currently we have covered the basics of mapping EHR data to the OMOP common data model and deploying the basic toolset created by the OHDSI collaborative. We see the following future steps:

- Besides the basic toolset, the OHDSI collaborative has developed several R packages enabling more in depth analyses on the data. For example, a package has been created for patient level prediction[8] and we would like to run the analyses methods on the subset of PHARMO diabetes type 2 patients.
- We would like to explore the possibility of giving external researchers remote access to the mapped data either by enabling them to run queries on the data remotely or by running predefined queries locally and sharing the results.
- The current mapping was focused on data collected from General Practitioners. We would like to expand the current mapping with data from pharmacies, clinical laboratories and hospitals.
- Beside the current diabetes type 2 cohort we would like to map other disease specific cohorts.
- The mappings of local coding systems to the coding systems used in the OMOP common data model need to be fine-tuned.
- We would like to further collaborate with other EHR databases mapped to the OMOP common data model in international multi-database studies.

CONCLUSION

All in all getting from the source data to a fully functional OHDSI environment was quite a journey. However,we feel that this is just a starting point. There is still a lot of work to be done to both improve and expand the current mapping. The OHDSI collaborative is a very active community with new tools and packages being released regularly. Mapping to the OMOP common data model and adhering to its design principles further aids us in conducting and collaborating in studies in a standardized and transparent way and is a viable and valuable addition to our research capabilities

REFERENCES

1 www.pharmo.nl
2 www.emif.eu
3 www.omop.org
4 www.ohdsi.org
5 www.ohdsi.org/analytic-tools/whiterabbit-for-etl-design
6 www.ohdsi.org/analytic-tools/achilles-for-data-characterization
7 www.ohdsi.org/web/atlas
8 www.github.com/OHDSI/PatientLevelPrediction

ACKNOWLEDGMENTS

We would like to thank Michel van Speybroeck and Lars Halvorsen of Janssen Pharmaceutical and Maxim Moinat and Ilona Pinter of The Hyve for helping us with mapping the data, building the ETL and setting up the OHDSI environment.

CONTACT INFORMATION

Rients van Wijngaarden
PHARMO Institute
Van Deventerlaan 30-40
3528 AE Utrecht
Work Phone: +31 30 7440 819
Fax: +31 30 7440 801
Email: r.van.wijngaarden@pharmo.nl
Web: http://www.pharmo.nl