Building a Library for Validated OMOP-CDM Queries

Peter Rijnbeek [1], Mees Mosseveld [1], Marcel de Wilde [1], Maxim Moinat[2], Alberto Gil Jimenez[2], Eldar Allakhverdiiev[3], Anthony Molinaro[4]

(1) Erasmus MC, (2) The Hyve, (3) Odysseus Inc, (4) Janssen Research and Development

Introduction

The European Health Data and Evidence Project (EHDEN) aims to build an active and sustainable federated data network across Europe. The choice for the OMOP Common Data Model (OMOP-CDM) and its Standardised Vocabularies for Europe is no-brainer. It is becoming the *de facto* standard for capturing health data for good reasons. First, the large amount of tools and analytical approaches that have been developed by the OHDSI Community, perfectly fit the use cases we will encounter. Second, the large amount of local terminology systems are a big hurdle for federated data analyses in Europe. The harmonization through the Standardised Vocabularies is a necessity to be able to scale up.

In EHDEN a harmonization fund has been made available for data sources to be mapped to the OMOP-CDM. This process will be supported by Small to Medium-sized Enterprises (SMEs) that will be trained and certified in the EHDEN project.

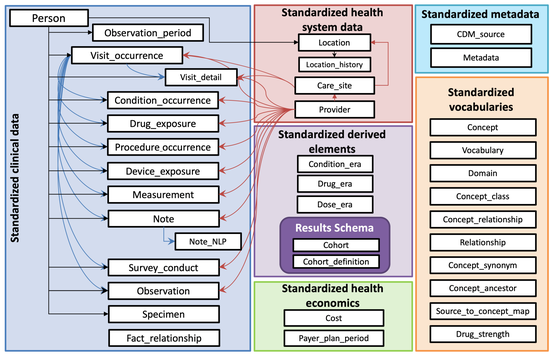


Figure 1. The OMOP Common Data Model

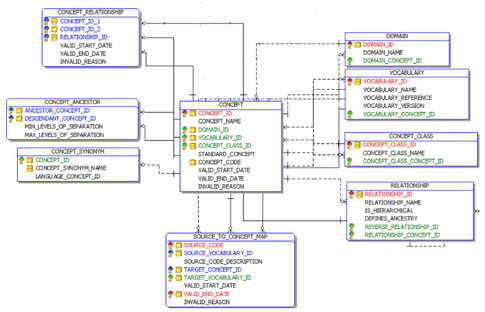


Figure 2. The OMOP Common Vocabulary Model

The OMOP-CDM and Standardized Vocabularies are at the core of the EHDEN ecosystem, and it is absolutely crucial we train all our stakeholders in its use. There are several user-friendly tools, like ATLAS, that automatically executes SQL against the CDM to build cohorts, search the vocabularies, etc. However, much more advanced data manipulations can be done using custom SQL. This does require in-depth knowledge of the CDM and advanced SQL expertise.

OHDSI has build a fantastic set of R-packages (DatabaseConnector, SqlRender) that allow standardized queries that run against the CDM hosted in a large set of Database Management Systems (DBMS), e.g., PostgreSQL, SQL Server, etc.

The EHDEN project is collaborating with OHDSI to develop a Shiny Application that provides access to a library of validated OMOP-CDM queries.

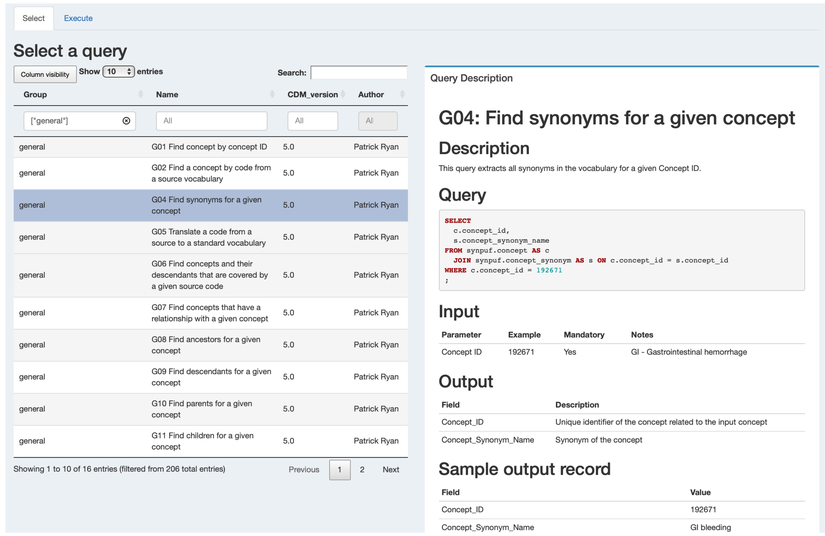
The Design Principles

1. **Adding new queries should be possible without additional programming.**
   Each query is added as a Markdown file following a predefined structure (Meta Data, Description, Query, Input and Output variables, and example output). These files can simply be added in the queries folder and will be automatically added in the Shiny Application.

2. **The tool should be able to connect to a data source and execute the query.**
   We added functionality to load a database configuration file and the user can then select a query, modify it, and execute it against the database. The query can be downloaded for external use as well.

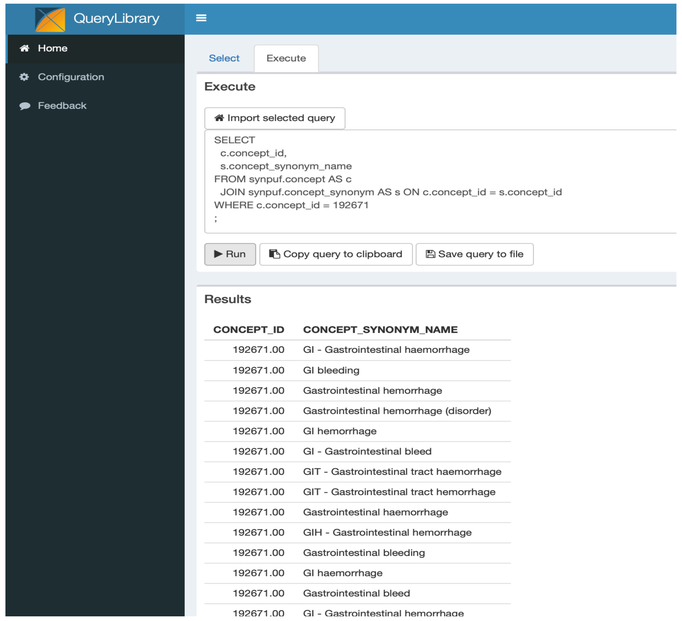
3. **Queries should be tested against multiple database management systems using an automated approach.**
   We added functionality that allows the developers to test all queries against a set of databases. The result is an Excel file that contains an overview of the duration and potential errors.

The QueryLibrary R-Package



The R-package contains a Shiny Application in which the user can search for a query. After selection, the query automatically renders to the Target Dialect as specified in the Server Configuration.

Currently, more than 200 queries are available in the library. These queries are targeted at all the data domains, such as drugs, conditions. Furthermore, multiple more general queries are added, for example how to traverse through the vocabulary tables.



The selected query can be imported under the Execute tab and can be further modified if needed, for example to use another concept_id.

This query can then be executed against the CDM and the results will appear below the query.

The user can copy the query to the Clipboard or can save it to a text file for future use.

Current Development Status

EHDEN is in the process of testing the first release and will make the tool available to the OHDSI Community in April 2019

Conclusion

As EHDEN, we are committed to enabling the adoption of the OMOP-CDM in Europe. Training our stakeholders is a prerequisite for its success and therefore we believe that simple tools like the Querybuilder R-package are valuable. We will maintain the library and its quality during the EHDEN project. Moreover, we invite the OHDSI community to suggest new queries that are of value to everyone that is using the CDM in their evidence generation pipeline.

P.RIJNBEEK@ERASMUSMC.NL

www.ehden.eu @IMI_EHDEN IMI_EHDEN github.com/EHDEN

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.