Empowerment of German university hospitals to utilize OHDSI OMOP CDM

Mirko Gruhl[1], Michéle Kümmel,[1] Ines Reinecke[1], Prof. Dr. Dr. Martin Sedlmayr[1]
[1]Center for Medical Informatics, Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany

In order to facilitate the use of the OHDSI tool stack, we created an easy-to-use software bundle. On the basis of Broadsea, databases were generated reproducibly, which contain the necessary vocabularies and adaptations to the German needs. Various available databases enable the general test of the stack using synthetic data. In addition, it is possible to also develop and work with the local data. For this purpose a corresponding ETL process was introduced to import the data into this database. The whole project is hosted on a common GitLab repository and can be downloaded and used within minutes.

Background

The range of different software systems in medical healthcare and diverse ways to store data has led to the demand for a uniform way to represent the clinical information in observational databases and thereby enable distributed analyses for an improved healthcare research. Within the German Medical Informatics Initiative, the MIRACUM[1] consortium has decided to transform their data into the Observational Medical Outcomes Partnership (OMOP) CDM[2]. A major use case of MIRACUM is to build a patient recruitment infrastructure and allow for analyses of patient visit data in a standardized format. To support the uptake of OMOP, we developed a system which includes the needed OHDSI suite of applications, a concise database with the required vocabularies and German extensions as well as an ETL process in one ready-made software stack.

Methods

To provide an OHDSI software stack for our partner sites, we focused on the following three parts:

1. As a starting point, we used the OHDSI Broadsea project. This is a solution which uses OS-level virtualization (Docker) and suits the requirement to share the software among several partner sites in a comfortable way. Thus, we were able to quickly progress in the provision of an integrated software stack.
2. We modified Broadsea so that the database(s) can be provided as Docker volumes. Several databases were generated to select from: for demonstration and testing purposes, the user can bind a database with pre-filled synthetic data. Additionally, a database with the required vocabularies for use of German patient data was provided. For this purpose, a minimal vocabulary set was determined by downloading first the full set of vocabularies that was pruned from unnecessary vocabularies yet extended according to the requirements of German patient data in 4 steps as shown in Figure 1. Based on the work of Maier et al.[3] who already integrated the ICD-10-GM (International Classification of Diseases, German Modification) and the “Operationen- und Prozedurenschlüssel” (operations and procedures) the vocabulary was extended by the current mappings. The enhanced vocabulary set and the database volumes can be (re-)generated via scripts to assure the reproducibility of the process.
3. To fill the OMOP CDM with German patient data, an ETL process was prepared. The source data consisted of four out of six basic modules of the German core data set, which is designed by a cross-consortium team of data scientists and medical doctors (see https://art-decor.org/art-decor/decor-datasets--mide-). Therefore, we could load observational data concerning the person, clinical case, diagnosis and the procedures into the OMOP CDM.

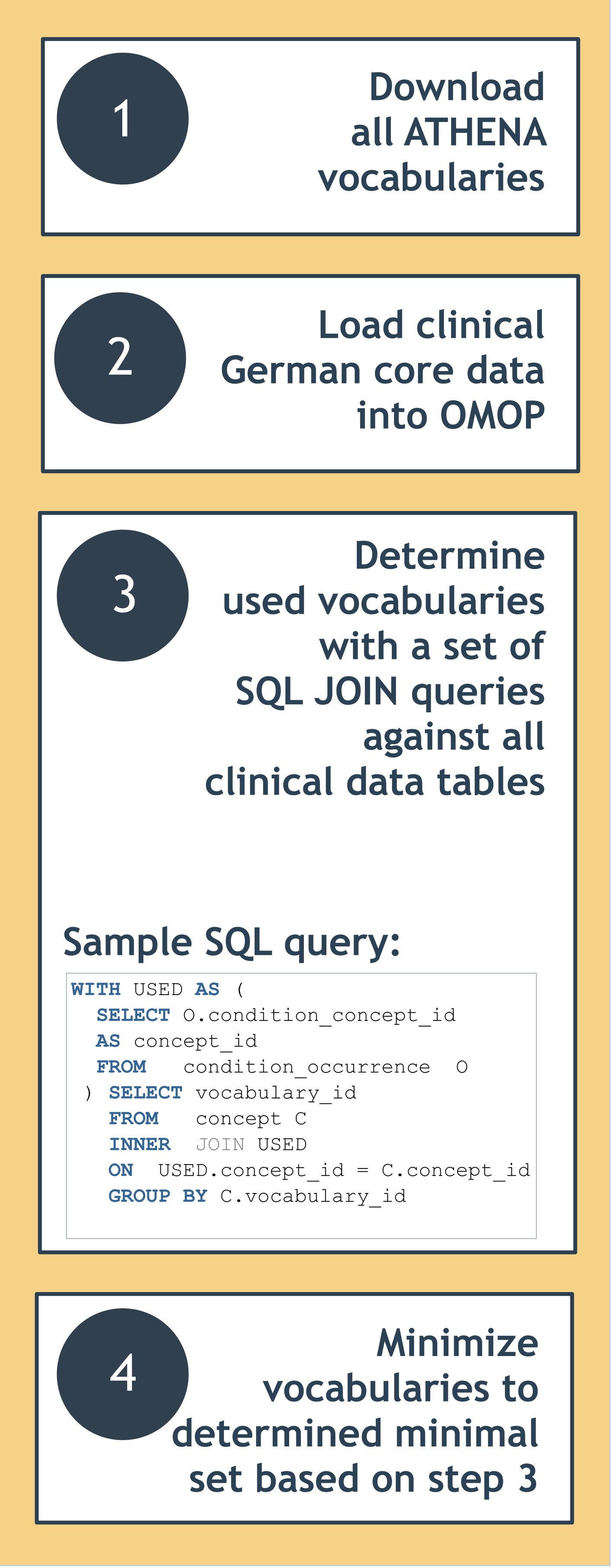

Figure 1: Determine needed Vocabularies

Results

The designed system consists of a number of docker images manged through docker compose as shown in Figure 2 and comprises the OMOP CDM including the vocabularies, the WebAPI, Atlas and sample data and can be used by each partner site with little to no technical knowledge of the involved software and database. The system dependent docker images are hosted via a common registry. The created volumes are downloadable from a self-hosted cloud storage. Thus, the project is deployable within minutes. Due to the simplicity and the flexibility to use synthetic test data besides the German patient core data, the partner sites started to use the OHDSI tools as well as the OMOP CDM for basic analyses and proof of concept.

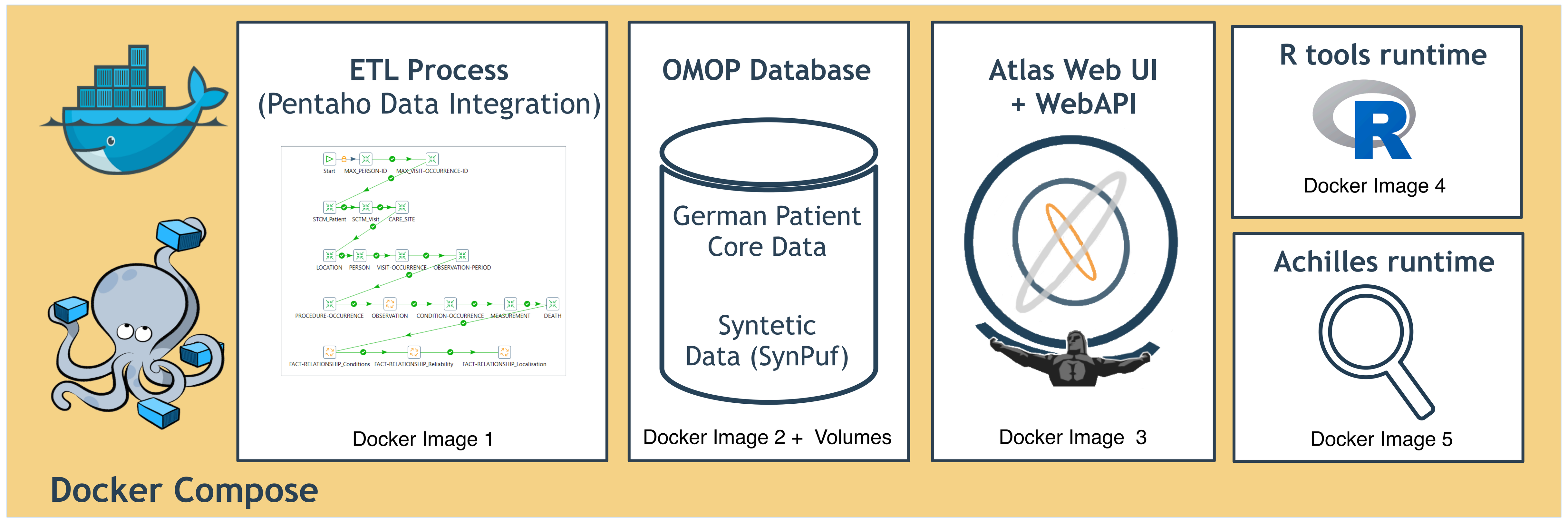

Figure 2: Docker Compose Environment Overview

Conclusions

The realized software stack provides a high degree of flexibility and a maintainable workflow for the generation of the basic database. The OMOP CDM was chosen because of some fundamental expertise and the promising big community. Nonetheless, the realization of the project showed some challenges especially in relation to the requirements for healthcare research in Germany. While the Apache Shiro framework was integrated into the OHDSI stack, we did not find a convenient way to use it in the Docker stack, yet. Another open issue is on securing the access to the generated databases, because the credentials are the same for all deployed databases and hence must be changed manually at the sites. Although our stack still requires further work, it allowed the MIRACUM sites to deploy an OHDSI tool stack within minutes locally for demonstration and development purposes. The scripts will be provided to the OHDSI GitHub repositories.

The thriving implementation of the basic software stack and the usage of the vocabulary set yields a potent system for federated healthcare research among ten MIRACUM partner sites. Following work on this project will include an upgrade of the components such as the WebAPI and Atlas. Moreover, the software stack must be reflected upon security aspects like authentication of the user interface.

This work is part of the project MIRACUM, funded by the German Ministry of Education and Research (FKZ 01ZZ1801A/L).

References : 1. Prokosch H-U, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, u. a. MIRACUM: Medical Informatics in Research and Care in University Medicine: A Large Data Sharing Network to Enhance Translational Research and Medical Care. Methods Inf Med. Juli 2018;57(S 01):e82-91.
2. Overhage, J. Marc, et al. ”Validation of a common data model for active safety surveillance research.” Journal of the American Medical Informatics Association 19.1 (2011): 54-60.
3. Maier C, Lang L, Storf H, Vormstein P, Bieber R, Bernarding J, u. a. Towards Implementation of OMOP in a German University Hospital Consortium. Appl Clin Inform. 2018;09(01):054-61.