Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200458

Specification and Distribution of Vocabularies Among Consortial Partners

Mirko GRUHL [a,1], Ines REINECKE [a] and Martin SEDLMAYR [a]

[a] *Center for Medical Informatics, Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany*

Abstract. Due to the variety of different software systems and disparate observational databases, the need for a uniform data representation rises. Common data models (CDM) support the harmonisation of data. A powerful but compact software setup and a minimum vocabulary set has been composed via Docker to facilitate analysis of data across ten university hospitals. The presented approach also creates the possibility to use a concise database which is easy to deploy.

Keywords. vocabulary, OMOP, Docker, reproducible, data sharing

1. Introduction

To overcome the issues caused by heterogeneous databases, common data models (CDM) support the harmonisation of data. As part of the MIRACUM initiative [1], all data should be transformed into the Observational Medical Outcomes Partnership (OMOP) CDM [2]. This yields a universal format and a corporate representation with corresponding terminologies, vocabularies and coding schemes. The outcome of the implementation is the utilisation and analysis of data across ten university hospitals and the possibility to use the OHDSI suite of applications and exploration tools [3] in one ready-made distribution.

2. Method

First, a minimal vocabulary for a provided synthetic test data was specified by downloading the full set of standardised vocabularies and based on local data and mappings the actually needed concepts where determined. The next step comprised the enhancement of the minimal vocabulary to meet the requirements of the core data. Based on the work of Maier et al. [4] who integrated the ICD-10-GM (International Classification of Diseases, German Modification) and the "Operationen- und Prozedurenschlüssel" (surgery and procedure key) the vocabulary set was cleaned from unnecessary vocabularies that have no relevance in Germany. Finally, a system was planed and realised

[1]Corresponding Author: E-mail: mirko.gruhl@uniklinikum-dresden.de

using Docker. The software bundle includes the OHDSI WebAPI [3] which contains all services that can be called from Atlas [3]. Additionally, the R Methods Library (https://github.com/OHDSI/Broadsea-MethodsLibrary) and Achilles [3] were applied. The databases were generated reproducibly via scripts and shared for the partners as Docker volumes which differ in the setup and size to meet different requirements.

3. Results

All system dependent volumes and images are hosted in a common registry and self-hosted cloud storage, respectively. The developed system uses a Docker compose file which is derived from the OHDSI Broadsea [5] tool set. Thus, the whole setup can be deployed in minutes using simple docker commands. The project is publicly available via GitLab (https://gitlab.miracum.org/gruhlmi/ohdsi-omop-v5).

4. Discussion

The system can be reproducibly deployed. Broadsea was chosen as a basis and customized so that a complete database with all needed content to start work with OMOP is available from one hand. The OMOP CDM was selected because of initial experiences and a big community. The compilation of the provided minimum vocabulary fits the requirements for healthcare research in Germany. Hence, every consortial member can start up an initial database with optimal memory utilisation. The use of the vocabulary is subjected to the license agreement and expects individual responsibility by every end user. Apart from the promising approach there are still challenges regarding the particularities of German classifications and procedures and the mapping to standard vocabularies.

5. Conclusion

The successful realisation of the objectives yields a powerful system which uses OMOP as starting point for federated healthcare research among ten MIRACUM sites. Further work on this topic will include the upgrade of the components and a strategy development to update the vocabularies automatically. The work is part of the project MIRACUM, funded by the German Ministry of Education and Research (FKZ 01ZZ1801L).

References

[1] Prokosch, Hans-Ulrich, et al. "MIRACUM: medical informatics in research and care in University medicine." Methods of information in medicine 57.S 01 (2018): e82-e91.
[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] Hripcsak, George, et al. "Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers." Studies in health technology and informatics 216 (2015): 574.
[4] Maier, C., et al. "Towards implementation of OMOP in a German university hospital consortium." Applied clinical informatics 9.01 (2018): 054-061.
[5] OHDSI Broadsea. [cited 2019 Oct 01]. Available from: https://github.com/OHDSI/Broadsea