Healthcare Transformation with Informatics and Artificial Intelligence
J. Mantas et al. (Eds.)

doi:10.3233/SHTI230485

Harmonisation of German Health Care Data Using the OMOP Common Data Model - A Practice Report

Nicole HECHTEL[a,1], Johanna APFEL-STARKE[b], Sophia KÖHLER[b], Maikel FRADZIAK[b], Norman SCHÖNFELD[b], Jens STEINMEYER[b] and Steffen OELTZE-JAFRA[a]

[a] *Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig – Institute of Technology and Hannover Medical School, Hannover, Germany*
[b] *MHH Information Technology, Hannover Medical School, Hannover, Germany*

Abstract. Data harmonization is an important step in large-scale data analysis and for generating evidence on real world data in healthcare. With the OMOP common data model, a relevant instrument for data harmonization is available that is being promoted by different networks and communities. At the Hannover Medical School (MHH) in Germany, an Enterprise Clinical Research Data Warehouse (ECRDW) is established and harmonization of that data source is the focus of this work. We present MHH's first implementation of the OMOP common data model on top of the ECRDW data source and demonstrate the challenges concerning the mapping of German healthcare terminologies to a standardized format.

Keywords. Data Harmonization, Data Warehousing, Secondary Data Analysis

1. Introduction

The OMOP common data model is a relevant data model in health care, supported by the growing open science community OHDSI (Observational Health Data Sciences and Informatics) to generate populate-level evidence [1, 2]. Since a few years, the usage of the OMOP common data model in Germany has also been growing. First projects show the possibilities and limitations concerning the mapping of German health care terminologies to OMOP [3, 4]. As the benefit of the OMOP common data model lies in health care data usage explicitly for retrospective research, it is crucial for clinical and epidemiological researchers to access more real world data for their studies in a standardized format.

The Enterprise Clinical Research Data Warehouse (ECRDW) at the Hannover Medical School (MHH) delivers a relational, error-corrected and plausibility-tested data model [5]. In March 2023, the ECRDW contains data from more than 2,4 million patients with approximately 3,2 billion data points. The reuse of these clinical data by expanding their harmonization is realized in different projects [6, 7]. Against that background, the

[1] Corresponding Author: Nicole Hechtel, Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig - Institute of Technology and Hannover Medical School, Hannover Medical School, Karl-Wiechert-Allee 3, 30625 Hannover, Germany; E-mail: nicole.hechtel@plri.de.

MHH started 2022 the first OMOP common data model project funded by EHDEN (European Health Data & Evidence Network [8]) as a data partner. The federated EHDEN network focusses on standardizing health care data sources to the OMOP common data model. The advantage of participating in EHDEN lies in faster, multi-centre and evidence-based studies using the health care data that are already available. Due to the harmonised data model, there is no need to agree on a data model, nor on analysis algorithms in the EHDEN network.

2. Methods

The OMOP common data model is a person-centric model with domain-oriented concepts (a. e. conditions, procedures, measurements) [9]. Standard vocabularies in OMOP are for example SNOMED and LOINC. A key element is the preserved data provenance, which means that original codes are stored within the common data model.

We implemented OMOP common data model for data sets, which include core medical data: demographic data, accounting case data, transaction data, medical risk factors, diagnoses, procedures, laboratory values and data about vital status of patients. The design of the data mapping and the ETL process was performed by using the tools White Rabbit and Rabbit-In-A-Hat [9]. For the mapping, three different procedures are necessary: available standard terminologies, non-standard to standard terminologies and hospital-specific vocabularies to standard vocabularies. The mapping of non-standard to standard terminologies is supported by the OHDSI community and mapping tables are available, even though the tables are not always up-to-date (Tab. 1). The mapping of hospital-specific vocabularies required a manual mapping (Tab. 1), which is performed using the tools Usagi and ATHENA, a web application for standardized vocabularies [9].

Table 1. Mapping tables for the mapping of source data to OMOP common data model at the MHH

Source vocabulary	Description	Status
LOINC (standard vocabulary)	Logical Observation Identifiers Names and Codes (Regenstrief Institute)	Available in ATHENA
ICD10-GM (non-standard vocabulary)	International Classification of Diseases Tenth Revision, German Edition	Available in ATHENA
OPS (non-standard vocabulary)	Operations and Procedures Classification	Available in ATHENA
Laender	Countries of patients origin	Source-to-concept-map
Orgfa	Departments and divisions of the hospital	Source-to-concept-map
Bewegungen	Transaction/transfer of patients	Source-to-concept-map
Risikofaktoren	Medical risk factors, special issues	Source-to-concept-map
Laboreinheiten	Units of laboratory measurements	Source-to-concept-map

For the implementation of the ETL processes, SSIS (SQL Server Integration Services) tools were used. The infrastructure divides between ECRDW as the data source on the one hand (MS SQL based server) and the OMOP infrastructure on the other hand (Linux server based with MS SQL environment).

3. Results

After data mapping and implementation of the ETL processes, following mapping results concerning the mapping of codes and records were achieved (see Tab. 2).

We observe successfully mapped records in the domains *procedure*, *measurement*, *conditions*, *observation*, and *visit_occurence*. Missing results ("NA") are found in the domains *device* and *drug*, where source data are not available in a digital format. We further observe missing results in the domains observation-unit and observation-value, because our observation data do not contain numeric data with units.

Table 2. Results of the mapping concerning codes to standardized vocabularies and the percentage of records.

Domain	#Codes Source	#Codes Mapped	%Codes Mapped	#Records Source	#Records Mapped	%Records Mapped
procedure	10,918	10,918	100.00	3,463,529	3,463,529	100.00
device	0	NA	NA	NA	NA	NA
drug	0	NA	NA	NA	NA	NA
measurement	466	466	100.00	3,262,478	3,262,478	100.00
observation-unit	0	NA	NA	NA	NA	NA
observation-value	0	NA	NA	NA	NA	NA
condition	11,730	11,610	98.98	5,789,740	5,763,302	99.54
observation	198	194	97.98	544,255	541,425	99.48
visit_occurrence	623,862	623,862	100.00	623,862	623,862	100.00
measurement-unit	54	50	92.59	2,266,709	2,095,215	92.43

4. Discussion

We were able to map a large number of patient data. Limitations are clinical data, which are not coded, and especially medication data, which are not available in our case (domain *drug*). Due to a planned conversion of the hospital information system at the MHH, we expect digital available medication data. We are working on improvements in our mapping efforts, but country-specific vocabularies and hospital-specific catalogues will be challenging points in data harmonization.

5. Conclusion

Building on these results we are extending the mapping to intensive care data and data concerning rare diseases (diagnoses codes with orphanet codes). To verify our local structure and to determine the procedure concerning ethical and legal aspects for study requests on OMOP common data model, a proof of concept study is planned. Afterwards we will be able to start first study requests for large-scale data analysis together with other German healthcare institutions, where the OMOP common data model is also implemented.

6. Acknowledgements

Prof. Dr. Christoph Schindler, Head of the Centre for Clinical Studies (ZKS) at the Hannover Medical School, for the initiation of the EHDEN OMOP common data model project at the Hannover Medical School and his support. Martin Preusse (EHDEN-certified SME Kaiser&Preusse) for his support in our project work.

References

[1] Hripcsak G, Ryan PB, Duke JD, Shah NH, Park RW, Huser V, Suchard MA, Schuemie MJ, DeFalco FJ, Perotte A, , Banda JM, Reich CG, Schilling LM, Matheny ME, Meeker D, Pratt N, and Madigan D. Characterizing treatment pathways at scale using the OHDSI network. Proceedings of the National Academy of Sciences of the United States of America. 2016;113:7329–7336. doi:10.1073/pnas.1510502113.

[2] Reinecke I, Zoch M, Reich C, Sedlmayr M and Bathelt F. The usage of OHDSI OMOP–a scoping review. Proceedings of the German Medical Data Sciences (GMDS); 2021. p. 95-95.

[3] Maier C, Lang L, Storf H, Vormstein P., Bieber R., Bernarding J, Herrmann T, Haverkamp C, Horki P, Laufer J, Berger F, Höning G, Fritsch HW, Schüttler J, Ganslandt T, Prokosch HU, Sedlmayr M. Towards Implementation of OMOP in a German University Hospital Consortium. Applied clinical informatics 9.01. 2018;054-061.

[4] Lang L. Mapping eines deutschen, klinischen Datensatzes nach OMOP Common Data Model (Doctoral dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). 2020).

[5] Gerbel S, Laser H, Schönfeld N and Rassmann T. The Hannover Medical School enterprise clinical research data warehouse: 5 years of experience. Data Integration in the Life Sciences: 13th International Conference, DILS 2018, Hannover, Germany. Nov 20-21 2018, Proceedings 13 p. 182-194.

[6] Gerbel S, Laser H, Schoenfeld N, May M, Bahlmann E, Menne J, Schindler C. Real World Data for Clinical Research – Results within the Champion Programme of the EHR4CR (Electronic Health Records for Clinical Research) Project. 64th German Medical Data Sciences (GMDS), Dortmund. 2019 Sep. doi: 10.3205/19gmds172.

[7] Fiebeck J, Laser H, Winther HB, Gerbel S. Leaving No Stone Unturned: Using Machine Learning Based Approaches for Information Extraction from Full Texts of a Research Data Warehouse. Data Integration in the Life Sciences. DILS 2018. Lecture Notes in Computer Science,vol 11371. Springer, doi: 10.1007/978-3-030-06016-9_5.

[8] https://www.ehden.eu/.

[9] Observational Health Data Sciences and Informatics. The Book of OHDSI. 2019. https://ohdsi.github.io/TheBookOfOhdsi/.