Name:	Michéle Kümmel
Affiliation:	Technische Universität Dresden, Carl Gustav Carus Faculty of Medicine, Institute for Medical Informatics and Biometry
Email:	michele.kuemmel@tu-dresden.de
Presentation type (select one):	Poster

Transition Database for a harmonized mapping of German patient data to the OMOP CDM

Michéle Kümmel[1], Ines Reinecke[1], Mirko Gruhl[1], Franziska Bathelt, Dr.[1], Martin Sedlmayr, Prof. Dr. Dr.[1]

[1]Technische Universität Dresden, Carl Gustav Carus Faculty of Medicine, Institute of Medical Informatics and Biometry

Abstract

To improve medical research and thus increase the quality of health care it is necessary to guarantee distributed analyzes and data interoperability. Therefore, the OMOP CDM was introduced in MIRACUM. In addition to the mapping of German encoding for diagnosis and procedures, there is a need to extract additional observational information and transfer to the OMOP CDM (e.g. reason of admission / discharge, department, reliability of diagnosis). In order to prevent different versions of local mappings, we offer a Transition Database of German Vocabularies as a basis for a harmonized mapping of German patient data to OMOP Standard Concepts.

Introduction

MIRACUM (Medical Informatics in Research and Care in University Medicine)[1] is one of four consortia established within the Medical Informatics Initiative Germany (MI-I). It focuses on digitalization in medicine and therefore on linkage between medical knowledge and data. The use of a CDM (common data model) increases the interoperability and enables distributed analyzes. Maier et al.[2] already demonstrated the successful semantic mapping from ICD-10-GM (International Classification of Diseases, German Modification) to SNOMED for encoding diagnosis as well as the representation of OPS codes ("Operationen- und Prozedurenschlüssel", in English "operations and procedure key") to the OMOP CDM. Based on those results we implemented a reference ETL job to extract and transfer German patient data based on the Core Dataset of the MI-I into the OMOP CDM. To ensure nationwide consistency and comparability of mappings we provide the Transition Database of German Vocabularies (German TDB) as an extension tool of the reference ETL job.

Method

Our approach was divided into two sub-processes: (a) semantic mapping of German patient data to OMOP Standard Concepts and (b) design and implementation of the German TDB.

(a) For the semantic mapping of German patient data we focused on four basic modules of the Core Dataset of the MI-I: Person, Visit, Diagnosis, Procedure[3]. Following the results of Maier et al.[2] we identified additional differences between local source encoding systems and the OMOP Standard Concepts. An all-encompassing mapping of all information (e.g. reason of admission / discharge, department, reliability of diagnosis) form the source data was not yet available. To meet these challenges an interdisciplinary team of medical doctors, IT specialists and data scientists worked on an individual mapping. We identified the OMOP Standard Concepts with the help of browsing through the CONCEPT table, discussed and agreed on the selection.

(b) To use these mappings to load the OMOP CDM we could either implement static definition of mapping within the ETL job, connect to a Meta Data Repository (MDR) or focus on a more flexible yet pragmatic approach. We opted for the latter: A flexible solution by providing a relational database without extra requirements (e.g. installation of a MDR). The German TDB based on PostgreSQL, which is created and filled by SQL statements, is similar to the

SOURCE_TO_CONCEPT_MAP and uses the existing naming convention of OHDSI (Figure 1). The German TDB is available on GitHub for use and extension.

Results

As part of the OHDSI Symposium, we will present the concept of German TDB in detail. The following figures abstracts the database schema (Figure 1) using the example of mapping the reason of admission (Figure 2).

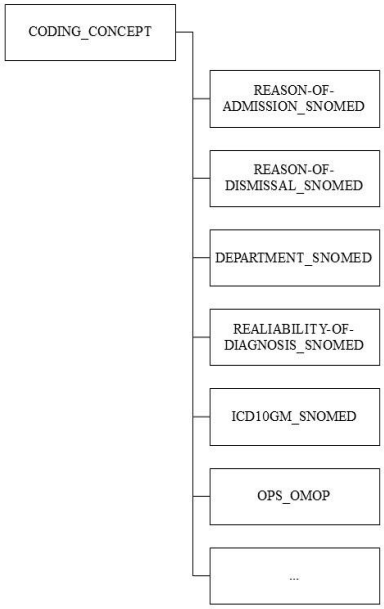

Figure 1. Database schema

TABLE_NAME	COLUMN_NAME	EXAMPLE	
CODING_CONCEPT	coding_concept_id	1	2
CODING_CONCEPT	coding_concept_name	SNOMED	Reason of admission
CODING_CONCEPT	domain_id	*null*	Observation
CODING_CONCEPT	vocabulary_id	*null*	InEK
CODING_CONCEPT	standard_concept	S	*null*
CODING_CONCEPT	valid_start_date	1970-01-01	2019-12-04
CODING_CONCEPT	valid_end_date	2099-12-31	2099-12-31
CODING_CONCEPT	coding_version	*null*	2019

TABLE_NAME	COLUMN_NAME	EXAMPLE	
REASON-OF-ADMISSION_TO_SNOMED	id	1	2
REASON-OF-ADMISSION_TO_SNOMED	coding_concept_id_source	2	2
REASON-OF-ADMISSION_TO_SNOMED	coding_concept_id_target	1	1
REASON-OF-ADMISSION_TO_SNOMED	source_value	01xx	02xx
REASON-OF-ADMISSION_TO_SNOMED	omop_source_domain_id	*null*	*null*
REASON-OF-ADMISSION_TO_SNOMED	omop_source_concept_id	*null*	*null*
REASON-OF-ADMISSION_TO_SNOMED	target_value	*null*	*null*
REASON-OF-ADMISSION_TO_SNOMED	omop_target_domain_id	Observation	Observation
REASON-OF-ADMISSION_TO_SNOMED	omop_target_concept_id	4214577	4010105
REASON-OF-ADMISSION_TO_SNOMED	relationship_id	45754864	45754864
REASON-OF-ADMISSION_TO_SNOMED	valid_start_date	2019-01-01	2019-01-01
REASON-OF-ADMISSION_TO_SNOMED	valid_end_date	2099-12-31	2099-12-31

Figure 2. Example "reason of admission"

Discussion and conclusion

Our approach provides three advantages: (a) increase transparency, (b) increase interoperability and (c) guarantee maintaining and versioning of mappings. In a next step, we will focus on the further development of the German TDB. So, the database will be continuously filled with other terminologies and encoding systems. For example, LOINC (Logical Observation Identifiers Names and Codes) will be added after we compared and harmonized the different mappings of laboratory value of the MIRACUM partner sites. In addition, a particular attention is also paid to expanding the database with terminologies, that are used in the context of rare diseases (e.g. Orpha codes[4]).

Because of the fact, that the German TDB can be used as an extension of the reference ETL job, it optimizes utilization of the OMOP CDM in Germany. Moreover, the flexibility and fast deployment lead to the adoption of the approach regarding other national specifics or different terminologies.

Acknowledgment

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, et al. 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. 2018;57(S 01):e82–91.
2. Maier C, Lang L, Storf H, Vormstein P, Bieber R, Bernarding J, et al. Towards implementation of OMOP in a German university hospital consortium. Appl Clin Inform. 2018;09(01):054–61.
3. Task Force Interoperability of the Medical Informatics Initiative Germany. Definition of the core data set. ART-DECOR. 2020. Available from: https://art-decor.org/art-decor/decor-datasets--mide-
4. Aymé S, Bellet B, Rath A. Rare diseases in ICD11: making rare diseases visible in health information systems through appropriate coding. Orphanet J Rare Dis. 2015;10(1).