

Transition Database for a harmonized mapping of German patient data to the OMOP CDM – The 'German TDB'

Michéle Kümmel¹, Ines Reinecke¹, Mirko Gruhl¹, Dr. Franziska Bathelt¹, Prof. Dr. Martin Sedlmayr¹

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

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

Target value

4214577

4010105

4123929

4083346

4214577

4014291

4213258

4180080















Background

- Initial situation:
 - Design and implementation of an reference ETL job within MIRACUM (Medical Informatics in Research and Care in University Medicine) [1] using Pentaho Data Integration [2]
 - Including static definition of mapping of German patient data, esp. of four basic modules of the Core Data Set of the Medical Informatics Initiative Germany (MI-I): Person, Visit, Diagnosis, Procedure [3]
- Problems of static definition of mapping:
 - Lack of transparency
 - Difficult to maintain
 - High level of manual effort
 - Lack of consistency

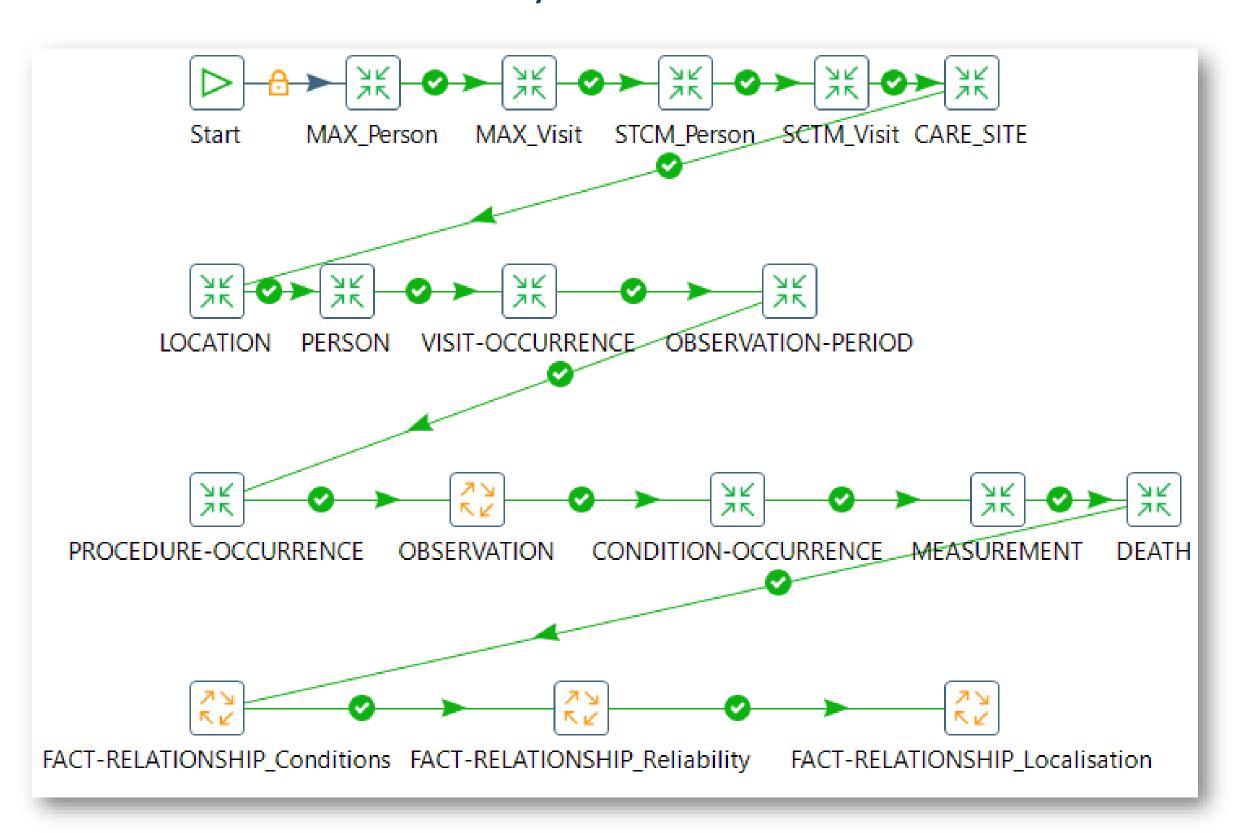


Figure 2: Example of static mapping ("reason of admission") within the reference ETL process

Figure 1: Screenshot of the Reference ETL Process

Results

- Solution: Design and implementation of the Transition Database of German Vocabularies (German TDB) as an extension tool of the reference ETL job
- Advantages:
 - Increased transparency and interoperability
 - Nationwide consistency and comparability of mapping
 - Easy maintaining and versioning of mapping
 - Simple and fast installation without additional software

4	source_value character varying (20)	omop_target_domain_id character varying (20)	omop_target_concept_id character varying (20)	relationship_id integer	valid_start_date	valid_end_date
1	01xx	Observation	4214577	45754864	2019-02-04	2099-12-31
2	02xx	Observation	4010105	45754864	2019-02-04	2099-12-31
3	03xx	Observation	4123929	45754864	2019-02-04	2099-12-31
4	04xx	Observation	4083346	45754864	2019-02-04	2099-12-31
5	05xx	Observation	4214577	45754864	2019-02-04	2099-12-31
6	06xx	Observation	3022007	45754864	2019-02-04	2099-12-31
7	07xx	Observation	4213258	45754864	2019-02-04	2099-12-31
8	08xx	Observation	4180080	45754864	2019-02-04	2099-12-31

Figure 3: Example of mapping ("reason of admission") with the help of the German TDB

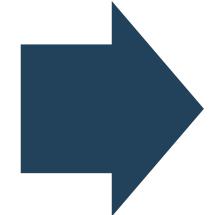
Available on **GitHub** for use and extension:



Methods

Semantic Mapping

- Semantic mapping of German patient data to OMOP Standard Concepts
- Consensus and documentation of individual mappings by an interdisciplinary team of medical doctors, IT specialists and data scientists



Design and implementation of the German TDB

Field values:

01

2 02

3 03

4 04

5 05

6 06

8 08

07

Source value

- Creation of several tables within one database (PostgreSQL) for every kind of source information using the naming convention of OHDSI
- Fill database with SQL statements

Conclusions

- Optimized utilization of the OMOP CDM in Germany by using the German TDB as an extension tool of the reference ETL job
 - Already successful implementation and use in the data integration center of the University Hospital Carl Gustav Carus in Dresden
 - Provision for further MIRACUM partner sites within another release of the reference ETL job
- Possible adoption of the approach regarding other national specifics or different terminologies (esp. Orpha Codes in the context of rare diseases [4] and LOINC) because of high flexibility and fast deployment
 - Extension within the OHDSI Community with further locally specific mappings is conceivable



^[1] Prokosch H-U, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: Medicine: A large data sharing network to enhance translational research and medical care. Methods Inf Med. 018;57(S 01):e82–91.



^[2] Hitachi Vantara, Pentaho Data Integration, Pentaho Documentation. 2018. Available from: https://help.pentaho.com/Documentation/8.2/Products/Data_Integration

^[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)