Journal of Biomedical Informatics 144 (2023) 104437

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

Eos and OMOCL: Towards a seamless integration of openEHR records into the OMOP Common Data Model

Severin Kohler [a], Diego Boscá [b], Florian Kärcher [c], Birger Haarbrandt [d], Manuel Prinz [e], Michael Marschollek [d], Roland Eils [a,c,*]

[a] Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Digital Health Center, Kapelle-Ufer 2, 10117 Berlin, Germany
[b] VeraTech for Health, Avenida del Puerto 237 - Puerta 1, Valencia, Spain
[c] Health Data Science Unit, Heidelberg University Hospital and BioQuant, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany
[d] Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany
[e] Leibniz Information Centre for Science and Technology, Welfengarten 1 B, 30167 Hannover, Germany

ARTICLE INFO

Keywords:
OpenEHR
OMOP
OHDSI
Secondary use
EHR
ETL

ABSTRACT

Background: The reuse of data from electronic health records (EHRs) for research purposes promises to improve the data foundation for clinical trials and may even support to enable them. Nevertheless, EHRs are characterized by both, heterogeneous structure and semantics. To standardize this data for research, the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) standard has recently seen an increase in use. However, the conversion of these EHRs into the OMOP CDM requires complex and resource intensive Extract Transform and Load (ETL) processes. This hampers the reuse of clinical data for research. To solve the issues of heterogeneity of EHRs and the lack of semantic precision on the care site, the openEHR standard has recently seen wider adoption. A standardized process to integrate openEHR records into the CDM potentially lowers the barriers of making EHRs accessible for research. Yet, a comprehensive approach about the integration of openEHR records into the OMOP CDM has not yet been made.
Methods: We analyzed both standards and compared their models to identify possible mappings. Based on this, we defined the necessary processes to transform openEHR records into CDM tables. We also discuss the limitation of openEHR with its unspecific demographics model and propose two possible solutions.
Results: We developed the OMOP Conversion Language (OMOCL) which enabled us to define a declarative openEHR archetype-to-CDM mapping language. Using OMOCL, it was possible to define a set of mappings. As a proof-of-concept, we implemented the Eos tool, which uses the OMOCL-files to successfully automatize the ETL from real-world and sample EHRs into the OMOP CDM.
Discussion: Both Eos and OMOCL provide a way to define generic mappings for an integration of openEHR records into OMOP. Thus, it represents a significant step towards achieving interoperability between the clinical and the research data domains. However, the transformation of openEHR data into the less expressive OMOP CDM leads to a loss of semantics.

1. Introduction

An ever growing number of electronically available health data is captured every day as part of the clinical routine. Its reuse for research is vital in order to improve and enable research itself [1–3]. On the other hand, the lack of this data has major downsides [4] and lead to spurious research findings early into the COVID-19 pandemic [5]. One pivotal cause for this gap between the care and research domain is poor interoperability of health application systems [6,7]. In order to use data from an electronic health record (EHR) for clinical research, information has to be integrated from different sources and systems [8]. To do so, semantic interoperability is needed. Contrary to this, nowadays EHRs are characterized by a heterogeneous structure and semantics of clinical records [7–9]. As a consequence, integration processes required are time and resource intensive [10].

1.1. Clinical Information Models (CIM) and their use

Over the past two decades, health informatics research has put a lot of emphasis on developing methods to harmonize heterogeneous EHRs [7,11]. A part of this process resulted in the definition and

* Corresponding author at: Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Digital Health Center, Kapelle-Ufer 2, 10117 Berlin, Germany.
E-mail addresses: severin.kohler@bih-charite.de (S. Kohler), roland.eils@charite.de (R. Eils).

https://doi.org/10.1016/j.jbi.2023.104437
Received 6 February 2023; Received in revised form 26 June 2023; Accepted 30 June 2023
Available online 12 July 2023

adaptation of Clinical Information Models (CIMs). CIMs are used to describe standardized clinical data elements including the relations among them. These models are then used to exchange computable clinical content. This enables an unambiguous interpretation of the data by each system [11,12] as universally demanded by the FAIR criteria for data management [13], thus making the information 'machine-actionable'. Currently adapted are the HL7 Fast Healthcare Interoperability Resources (FHIR) [14], openEHR [15] and the HL7 Clinical Document Architecture (CDA) [16].

Nevertheless, how much of the EHR and which CIM is used varies between institutions. As an example, in the Nordics the EHR market leaders chose openEHR as their CIM [17], and China decided in 2017 to standardize the national military EHR using openEHR [18]. It has affiliates in seven different languages and key public and industrial partners including the Catalan Health Service, the Norwegian Public Hospitals and Microsoft [19]. On the other hand, FHIR is used to exchange information between institutions of the Medical Informatics Initiative Germany [20–22] and widely used across the United States [23]. The scope of EHR projects is – at best – to enable semantic interoperability on a national level. Clinical research on the other hand often requires the integration of data on an international scale. This is especially true in the times of the global corona virus pandemic. The adaptation of CIM significantly eases the process of integrating EHRs for research, by enabling standardized integration processes. Nevertheless, it does not make this process obsolete, unless every stakeholder of a project agrees in using the same CIM.

1.2. Data models for clinical data reuse

One established approach to solve the above-mentioned issues for the reuse of clinical data, is the use of a Common Data Model (CDM). These models are used to ease the integration of health data from heterogeneous sources and enable their systematic analysis [24,25]. A prominent example of this is the Observational Medical Outcomes Partnership (OMOP) CDM which is maintained by the Observational Health Data Science and Informatics (OHDSI) community [26]. The data from disparate sources is transformed into this OMOP CDM and annotated using terminologies and coding schemes. Researchers can then perform systematic analyses with a set of standardized analytic routines and tools defined for the CDM [25]. Otherwise these large amounts of analytical methods would need to be developed and performed on each proprietary data model and CIM relevant to the research project [27]. In addition, this enables distributed analysis across different systems. Users can process this locally and return the outcomes, without the need of sharing the source data. Nevertheless, if required the data can also be shared easily since it is based on the same model. As of August 2019, the OMOP CDM is already used in over 100 healthcare databases from over 20 countries, capturing more than one billion patient records [25]. With the outbreak of the corona virus disease in 2019, the necessity of shared data collaboration became even more important. One such a collaboration is the National COVID Cohort Collaborative (N3C). Its goal is to integrate EHR data in the United States to enable its secondary use for corona research. A key part of its architecture is the OMOP CDM [28,29].

Another approach to make clinical data accessible for secondary use is the open source clinical data warehouse Informatics for Integrating Biology and the Bedside (i2b2). Instead of defining a common data model, the approach provides an entire software stack to users. Clinical data is populated into this data warehouse and can then be analyzed using i2b2's analytics tooling. Hereby, i2b2 offers a flexible star model of five tables to persist the data [30]. There are two major differences in the data model compared to OMOP. First, terminologies are standardized and harmonized in OMOP. Meanwhile, in i2b2 terminologies or even local codes can be loaded as so called catalogs. These are not standardized or harmonized. Secondly, OMOP has different clinical data types, e.g. measurement, device exposure, drug exposure etc. which are used to categorize data. I2b2 on the other hand has only an observation fact table used to represent all types of information. As a result, i2b2 offers more flexibility than OMOP, but at the cost of semantics and non-standardized terminologies.

1.3. Integrating data for clinical research

To transmit data from source systems into e.g. an OMOP CDM repository, the data has to be Extracted, Transformed and Loaded (ETL). ETL processes are complex and resource intensive [10]. This is especially true for systems that have a heterogeneous data representation like most of the EHRs. For each data model in use an ETL process needs to be defined, this includes mapping the data into the target data model. If source systems use a CIM, this process can be simplified. On the basis of the CIM, a standardized and automatized ETL process can be defined. As an example Haarbrandt et al. [31] defined an automatic ETL tool for integrating openEHR records into the i2b2 data warehouse. Furthermore, publications already covered the ETL process from FHIR into the OMOP CDM [29,32] and an OMOP-on-FHIR [33] ETL tool is accessible. Nevertheless, for the integration of openEHR records into the OMOP CDM there is currently neither a tool, nor a comprehensive investigation about its feasibility. The publication of Li et al. [34] promises that an integration of openEHR and OMOP improves the reuse of clinical data. However, it does not define an approach to do so. A first step in this direction was made by Rinaldi et al. [32]. A single openEHR model for microbiology findings was mapped to FHIR and OMOP. As stated in the paper itself, the data set used is very limited, allowing for no comprehensive review on the integration of openEHR records into the OMOP CDM. Furthermore, no standardized process or ETL tool was defined.

1.4. Objectives

The main goal of this work is to devise a comprehensive approach, theoretically and practically, on integrating openEHR records into the OMOP CDM. The development of such an approach would make openEHR records more easily accessible for secondary use. This improves the interoperability between the clinical and research data domains.

2. Material and methods

2.1. OpenEHR

The openEHR specification defines an interoperable architecture where all health information is represented based on the information model of openEHR. One of the key paradigms of openEHR is the multi-level modeling approach [35]. This requires three levels of models, starting with the Reference Model (RM). The RM defines the basic logical structure of EHR and demographics. This includes simple definitions for how data values are represented or more complex ones like the structure of clinical observations [36].

The second level of models uses these RM classes to specify re-usable clinical content models, the archetypes. These are detailed data models used to describe a maximum set of information about a clinical concept, e.g. blood pressure or heart rate. The data points of these concepts are organized using so called nodes. Each archetype is identified using a unique Archetype ID [36,37].

Archetypes cover a maximum set of information, often containing too much for specific use cases. In order to allow context-specific data sets, templates are used. Templates are the third level of models, they re-use archetypes as building blocks. While doing so, the nodes of an archetype can be limited and specialized in order to fit the specific scope. Each node within an archetype is identified with an archetype node identifier. When embedded in a template, these identifiers are preserved along with the archetype ID and used to construct paths. Paths

Table 1
Mapping from the international archetype openEHR-EHR-EVALUATION.problem_diagnosis.v1 [41] into the CDM CDT Condition_Occurrence [42].

Node name	Node path	CDM Field
Problem/Diagnose name	data[at0001]/items[at0002]	condition_concept_id
Date/time of onset	data[at0001]/items[at0077]	condition_start_date
Date/time of resolution	data[at0001]/items[at0030]	condition_end_date

are used to retrieve data from archetype nodes within templates [36]. Template models can be defined as part of local solutions or even as standardized ones for a country.

OpenEHR systems are built using these templates. They define the structure and constraints of compositions. Compositions are used to capture clinical data in openEHR. Each composition is based on one template. Software solutions implement the RM and are by that enabled to process templates etc. as software objects. In order to exchange or persist templates and compositions, a serialization is required. For that, openEHR defines standardized schemes for the Extensible Markup Language (XML) [38] and the JavaScript Object Notation (JSON) [39].

Besides the use of a common information model, medical terminologies also play an important role for semantic interoperability [40]. OpenEHR allows binding of such terminologies (e.g., SNOMED CT, LOINC) to archetype elements. Apart from this, openEHR also has its own small set of terminologies to provide some value sets for a number of attributes. Users can also define their own set of internal codings if necessary [36].

2.2. Observational medical outcomes partnership (OMOP) common data model (CDM)

The OMOP CDM defines a data model used to implement standardized research repositories for clinical data. Data is transferred into the CDM from the source systems and harmonized using standardized vocabularies. These contain records, so called concepts, that uniquely identify and express clinical information within OMOP. These concepts are described using codes from terminologies and associated descriptions [25]. For each terminology used in OMOP, there is one vocabulary and for each code there is a concept. Nevertheless, the meaning expressed by codes is often duplicated throughout different terminologies. To solve this, OMOP defines a set of concepts as standard for each Domain (e.g. Drug Exposure, Measurement). Mappings from non-standardized concepts to standardized concepts and vice versa are contained as part of the vocabularies. This harmonizes codes from different terminologies and eases analytical processes [25].

These concepts are referenced in the OMOP tables. The CDM consists of a set of tables with fields defined by OHDSI. These tables and rows are maintained by OHDSI as part of CDM versions. Clinical data in OMOP is represented by the Clinical Data Tables (CDT). Here, the concepts are used to unambiguously define which clinical information is represented by an instance. This is done by referencing the concept via a concept identifier specified by OHDSI.

2.3. Mapping the models

In order to define mappings, both data models need to be analyzed and possible mappings identified. The OMOP CDM defines a set of static data tables for different clinical models. Each of these tables has a set of fields. On the other hand, openEHR represents clinical models as archetypes. Archetypes have nodes used to express the different sets of information. As a result, first archetypes and CDM tables have to be mapped. Afterwards, the nodes need to be matched to their corresponding CDM fields. An example of such a mapping is illustrated in Table 1.

OpenEHR and the CDM both support different data types which need to be mapped. Most of these transformations do not require complex processes. An exception is the transformation of openEHR coded text to OMOP concepts. In openEHR, a value coded with a terminology contains the code itself and a terminology ID. In OMOP each code of a terminology is a specific concept identified by a unique ID. CDM fields use these IDs to reference codes. Therefore, the openEHR coded texts have to be transformed into concept, which can then be used to populate the CDM field. This is done by using the code and terminology ID of the coded text to identify the concept in OMOP. An example transformation of such a coded text, contained in the first node path of Table 1, is illustrated in Fig. 1.

Nonetheless, the CDM often requires additional information for the table, that derives from the context of the archetype and not from the data contained in the nodes. As an example, some CDT tables have a type concept field. This field contains the information from which source system type the data originated. In case of records from openEHR, this is an EHR system. As a result, the OMOP concept ID for EHR has to be provided to the transformation in addition to the archetype nodes.

To search such a concept ID and download vocabularies, OHDSI provides the ATHENA tool [43]. This tool was used for the research work described in this paper to determine required concept IDs for the mappings and to download the vocabularies. For the openEHR archetypes, the Clinical Knowledge Manager (CKM) [41] was used to search and download archetypes. The CKM also provides a visualization of the node paths of an archetype which was used to determine the node paths for the mappings.

A more fundamental difference than the type concept mapping is how a patient is represented and identified in both standards. The CDM has a person table that is used to store demographic data like age, gender etc. and a unique identifier. In openEHR, records are linked to an EHR object by referencing its identifier. Therefore, for each person ID there needs to be a related EHR ID. Meanwhile, for demographic data, openEHR does not impose a specific way to represent it. This data can be stored as part of compositions or in an external demographics server. In case of patient data being provided in compositions, this data can be transformed by mapping nodes to fields.

To define and develop the mappings and software of this paper, agile methods were used. Agile methodology is a project management approach that emphasizes flexibility and adaptability to changing requirements. It involves iterative and incremental development, continuous feedback, and close collaboration between the developers and stakeholders. To manage the resulting code and mappings, GitHub was used. GitHub is a platform for collaborative software development, code management, version control, project collaboration, and documentation [44].

3. Results

A software solution, hereinafter referred to as Eos, has been implemented to provide configurable openEHR to CDM transformations. The components are illustrated in Fig. 2. Eos is connected to an openEHR data repository from which it loads the records. The mappings are configured in OMOP Conversion Language (OMOCL) files that are loaded by Eos. Using these configured mappings the records are transformed by the tool into OMOP CDM entities. Here, Eos uses the standardized methods by openEHR and OMOP to communicate with both components. The tool itself is an application server with an Application Programming Interface (API). Using this API, either the process to integrate all EHRs or just a specific set can be triggered. After each composition, Eos populates the transformed entities into the OMOP CDM. Both tools are developed for the latest official version 5.4 of the OMOP CDM and are not compatible with other versions.

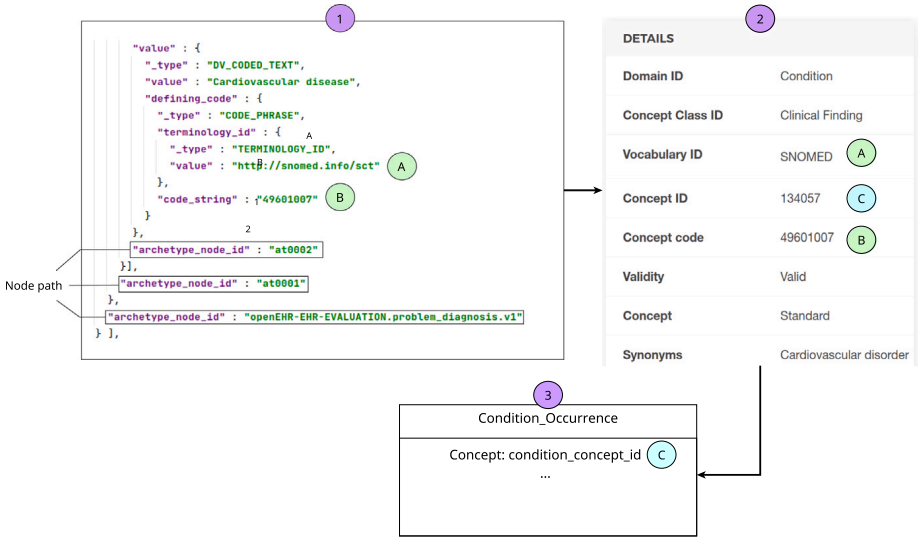

Fig. 1. Transformation of an archetype node into a CDM table field. (1) Snippet of an openEHR composition containing a DV_Coded_Text serialized as JSON. (2) vocabulary entry of the corresponding OMOP concept, displayed using the OHDSI Athena tool [43]. (3) Illustration of the CDM table Condition_Occurrence. (A) The terminology in openEHR that is mapped to the vocabulary ID in OMOP. (B) The code_string that is mapped to the concept code in OMOP. (C) The concept ID that uniquely identifies the concept in OMOP. This value is provided to the CDM table field as a result of the transformation.

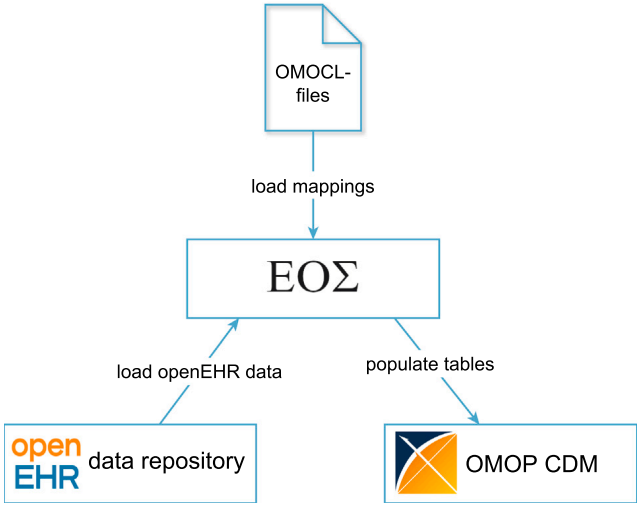

Fig. 2. Eos component diagram.

3.1. OMOP conversion language (OMOCL) and Eos

OMOCL is a Domain Specific Language (DSL) that was implemented to configure mappings from openEHR into OMOP and provide a flexible and easy-to-use solution. The mappings are based on archetypes. This could also be done for templates, yet these are project-specific and therefore less generic. Users would be required to define a configuration for each template. In contrast, archetypes are reused throughout different projects internationally. As a result, the information contained in the archetypes is transformable, no matter which templates they are used in. OMOCL and the mappings done were a collaborative effort with both openEHR experts and the help of certified OMOP professionals. Both the definition of the language and the mappings were done as part of an agile process. Resulting archetype mappings were managed using a git repository, OMCOL is open source and accessible via GitHub.[1] In the repository, mappings are classified by their domain (medical data/personal data) and by RM class. The DSL is based on YAML, which "is a data serialization language designed to be human-friendly and work well with modern programming languages" [45]. This has the advantage that users do not have to learn an entire new language in order to write mappings. An example of the OMOCL configuration of the openEHR-EHR-EVALUATION.problem_diagnosis.v1 archetype is shown in Fig. 3. A railroad diagram of the syntax can be found in Appendix A.

The first line defines the archetype that is to be mapped. Afterwards, a listing of mappings is declared, starting with the *type* that defines which CDM table is populated. In case of the example, it is only one table, the Condition_Occurrence. If multiple tables are mapped, these are defined one after another. Afterwards, the different node/concept to field mappings are declared. Here, keywords are used for multiple transformations. As an example, the *concept_id* key is used to transform not only the Condition_Occurrence field condition_concept_id, but also the condition_source_concept_id and condition_source_value. All of which are always mapped from the same node or concept code. The same keyword syntax is used in OMOCL for all other concept_id mappings with source_concept and/or source_value. The nodes and concepts for the transformation are listed under the key *alternatives*. These are treated as a logical OR.

Nodes are referenced using the key *path*, concept codes with *code*. In the example, the key *condition_start_date* contains two node paths. If the node *data[at0001]/items[at0077]* is not present in the composition, the next entry is processed. This entry contains the value *"../context"*, an OMOCL specific node path syntax. All node paths in OMOCL are relative to the archetype entry. Sometimes it is necessary to access the parent of the archetype entry. This can be done using two dots. In the case of the example, the parent nodes context element can contain several standardized fields, specified by the reference model, including a start time. Since the required CDM field is a start_date and the context contains a start time, OMOCL resolves this without requiring additional syntax. In addition, the OMOCL language also includes the following keywords:

- *optional*: sets transformations optional
- *Include*: reference other OMOCL archetype mappings
- *base_path*: iterates openEHR lists
- *multiplicate*: multiplicate values

OMOCL also provides the means to include custom mappings, e.g. programmed ones, if the syntax of OMOCL is not sufficient enough.

[1] https://github.com/SevKohler/OMOCL

```
archetype_id: "openEHR-EHR-EVALUATION.problem_diagnosis.v1"

mappings:
  - type: "ConditionOccurrence"
    concept_id:
      alternatives:
        - path: "data[at0001]/items[at0002]"
    condition_start_date:
      alternatives:
        - path: "data[at0001]/items[at0077]"
        - path: "../context"
    condition_end_date:
      alternatives:
        - path: "data[at0001]/items[at0030]"
```

Fig. 3. Snippet of the OMOCL mapping configuration for the international archetype openEHR-EHR-EVALUATION.problem_diagnosis.v1. The original configuration can be found in Appendix B Fig. B.6.

Another feature of OMOCL is that fields like the type concept, which always contains the same value as explained above, do not need to be configured and can be inferred by the tool processing them.

The Eos tool contains an engine to load these OMOCL files and execute the transformations based on them, as shown in Fig. 2. Eos was developed as part of the same agile processes as OMOCL. The software was implemented based on the requirement of executing OMOCL-files and therefore, the corresponding need of conformance to both openEHR and the CDM. Hereby, Eos and supports the entire grammar of OMOCL. Apart from that, other features were implemented, e.g. automatic transformation of data types and non-standard to standard concepts or tooling support for the programming of custom mappings. Eos is a Java application server, the implementation logic is split into two main components. The first one is responsible for loading the OMOCL-files and interpreting their grammar. The second part is responsible for the execution of the transformation based on the processed OMOCL-files. The tool is available as free-to-use and open-source on GitHub.[2]

3.2. Evaluation

As a proof-of-concept, the openEHR archetypes of different projects including the German Corona Consensus Dataset (GECCO) [46,47] were mapped from openEHR to OMOP. The GECCO data-set is used in the COVID-19 Research Network of University Medicine (NUM) in Germany [48]. In this network, 36 university hospitals forward their GECCO data to a central openEHR platform accessible for researchers. One archetype was not included since it represents imaging data. OMOP requires a specific extension to support this data. These extensions are not part of the standard CDM and are therefore open for future discussion. The resulting mappings are shown in Table 2. In addition, all OMOCL files can be found in Appendix B.

Nearly all of these archetypes from the projects are international ones. The resulting OMOCL configurations cover already 10.5% of the published[3] archetypes contained in the international library. Therefore, these mappings are not only valid for the GECCO data-set, but for all templates using these archetypes. Here, some specific archetypes are more commonly used than others, marking their importance. As an example, the problem_diagnosis archetype is used in 88 templates from the international [41] and German CKM [47] (as of 9.9.2022). Other notable mappings are the laboratory_test_result.v1, medication.v1 and procedure.v1. Each of these clinical models are relevant for most of the healthcare domains. Therefore, the mapping includes some of the most frequently used archetypes in openEHR. The transformed archetypes also cover most of the data required for patient summaries [49].

These configurations were used by the Eos Tool to integrate openEHR sample data into an OMOP CDM. In order to do so, Eos was programmed to process different types of CDT tables using OMOCL or to automatically generate them. As a result, thirteen of the fifteen CDT and Standardized Derived Elements (SDE) tables are supported for the transformation. SDE tables are used to represent eras and episodes of the patient care. These are illustrated in Table 3. All openEHR archetypes were successfully extracted, transformed and loaded into the OMOP CDM using Eos. This included the GECCO test data that was used to test the central openEHR platform containing inpatient patient summary data. Another transformed data-set was real-world inpatient data containing physical examinations of 24 patients. On top of that, sample data was generated from different CKM templates to cover a wider variety of data, including e.g. a template based on the nationwide medication plan standard of Germany. Therefore, the tool provides the means for users to automatically populate openEHR records into the OMOP CDM. Hereby, it supports 80% of the existing CDT and 100% of the SDE tables. Some of the tables are generated automatically by Eos, either by implemented logic or through the use of standardized SQL scripts [42] and the database agnostic SqlRender [50] provided by OHDSI.

SDE tables are used to represent time intervals of the data stored in the CDT tables. An example is the time frame in which a specific drug was given to a person (Drug era). The fields of these tables derive from the data contained in other CDM tables. Therefore, they cannot be mapped using archetypes and rather need to be generated by Eos. In the case of the person table, both methods are provided. If this data is stored in the compositions, OMOCL can be used. If the master patient data is stored in an external server, Eos will still generate persons for each EHR, but these do not contain any data. When a project requires the demographic data from the external server, users can develop their own ETL process to populate the persons generated by Eos

Not all nodes contained in the archetypes could be mapped into the CDM, since the CDM has no representation for them. This is to be expected since the CDM is not designed to cover a full EHR. As an example, the archetype problem_diagnosis contains the following nodes that could not be mapped: clinical description, body site, structured

[2] https://github.com/SevKohler/Eos

[3] The website contains different stages for the maturity of archetypes. The most stable one is the published status.

Table 2
Resulting mapping configurations defined in OMOCL and tested using the Eos tool. The percentage is calculated counting only published archetypes (17 of 169 as of 4.10.2022) released in the international CKM [41].

Archetype ID	CDM tables
EVALUATION.problem_diagnosis.v1	Condition occurrence
EVALUATION.death_summary.v0	Death
EVALUATION.device_summary.v0	Device Exposure
CLUSTER.device.v1	Device Exposure
CLUSTER.dosage.v1	Drug exposure
ACTION.medication.v1	Drug exposure
INSTRUCTION.medication_order.v2	Drug exposure
OBSERVATION.medication_statement.v0	Drug exposure
EVALUATION.tobacco_smoking_summary.v1	Observation
OBSERVATION.pregnancy_status.v0	Observation
ADMIN_ENTRY.person_data.v0[a]	Person
ACTION.procedure.v1	Procedure occurrence
OBSERVATION.blood_pressure.v2	Measurement
OBSERVATION.body_temperature.v2	Measurement
OBSERVATION.body_weight.v2	Measurement
OBSERVATION.body_mass_index.v2	Measurement
OBSERVATION.clinical_frailty_scale.v1	Measurement
OBSERVATION.height.v2	Measurement
CLUSTER.laboratory_test_analyte.v1	Measurement
OBSERVATION.pulse_oximetry.v1	Measurement
OBSERVATION.pulse.v2	Measurement
OBSERVATION.respiration.v2	Measurement
OBSERVATION.laboratory_test_result.v1	Measurement, Specimen, Fact relationship (custom)
CLUSTER.specimen.v1	Specimen
Coverage of archetypes :	**10.05%**

[a]The archetype is not part of the international library, instead it is contained in a German archetype library [47].

Table 3
Supported CDM and SDE tables by the Eos tool. The CDM version 5.4 has fifteen CDT and SDE tables. Currently supported by Eos are thirteen (80%).

CDT & SDE tables	Supported	Type
Person	Yes	OMOCL or generated
Observation period	Yes	Generated
Visit occurrence	Yes	Generated
Visit detail	No	
Condition occurrence	Yes	OMOCL
Drug exposure	Yes	OMOCL
Procedure occurrence	Yes	OMOCL
Device exposure	Yes	OMOCL
Measurement	Yes	OMOCL
Observation	Yes	OMOCL
Death	Yes	OMOCL
Note	No	
Note NLP	No	
Specimen	Yes	OMOCL
Fact relationship	Partly	Custom mapping
Drug era	Yes	Generated
Dose era	Yes	Generated
Condition era	Yes	Generated
Coverage of CDT tables :	**80%**	
Coverage of SDE tables :	**100.00%**	

body site, specific details, course description, severity and diagnostic certainty. Some of these fields can be stored in a non-specific table such as OBSERVATION, like body site, structured body site, cause and severity. The other nodes need to be stored in a completely textual way in the NOTE table. The use of these tables has several limitations. First, these archetype nodes can contain coded or quantified information that would be transformed into text fields when using NOTE, resulting in a loss of information. Secondly, to maintain the relation between them and the resulting diagnosis stored in CONDITION_OCCURRENCE a FACT_RELATIONSHIP needs to be added. A FACT_RELATIONSHIP is used to relate fields in two OMOP tables. Nevertheless, the NOTE, NOTE_NLP and FACT_ RELATIONSHIP are rarely used by the OHDSI community. Even OHDSI's official cohort exploration tool currently lacks support for these tables [51]. Because of these limitations and since the sample data-sets templates rarely contain such nodes, the implementation efforts bear no relation to the value added to the project. Therefore, these tables have not yet been added to OMOCL. If required, the FACT_RELATIONSHIP can be mapped using custom converters. As a proof-of-concept and to faithfully transform one template containing such an archetype node, a custom converter was added for the FACT_RELATIONSHIP mapping of OBSERVATION.laboratory_test_result.v1. Another table that is currently not supported by Eos and OMOCL is the VISIT_DETAIL. VISIT_DETAIL is an optional table, that provides additional information to a VISIT_OCCURRENCE. None of the mapped archetypes required these optional fields, therefore the table was left unsupported. For the

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	2146	19	2165	99%	287	0	287	100%	2433	19	2452	99%
Conformance	683	26	709	96%	104	0	104	100%	787	26	813	97%
Completeness	362	23	385	94%	15	1	16	94%	377	24	401	94%
Total	3191	68	3259	98%	406	1	407	100%	3597	69	3666	**98%**

Fig. 4. DQD output generated by analyzing data transformed by Eos. The plausibility row measures if data is believable, e.g. does a person_id only exist once. The conformance checks if data values adhere to specified standards and formats. The completeness validates for the particular presence of variable like the gender of a patient [25].

Table 4
Overview of archetypes, number of rows imported and the average execution time (in seconds).

Amount of archetypes	CDM rows created	Execution time(AVG)
231737	423756	~ 914 s

same reasons, a small set of table fields is also not supported, like the visit_detail to link an entry to the VISIT_DETAIL table. A listing of these unsupported fields can be found in Appendix C. Clearly, to provide a full coverage of OMOP, these fields and tables need to be added in the future.

To assure the correctness of the mappings, several quality assessments were made. First of all, a sample set of transformations for each archetype was checked manually by inspecting the resulting CDM tables. In addition, for each archetype mapped, unit tests where developed that ensure the implemented mappings are deterministic in their execution and produce the correct CDM tables and fields. Finally, the official OHDSI Data Quality Dashboard (DQD) [52] was used to validate the data transformed by Eos and OMOCL. An openEHR repository was randomly populated with 6649 compositions from sample test data of the data-sets to provide a diverse sample for the transformation. The results of the DQD are displayed in Fig. 4. The transformed data passes 98% of all checks. The major amount of fails is a result of missing data for demographics and visits, contained in the data-sets. The DQD result file is attached to the resources of the paper.

To evaluate the performance of Eos, an example data set of 116082 compositions from 506 patients was transformed into the CDM. The Table 4 lists the amount of archetypes, the number of CDM rows created and the overall average execution time. A detailed table about the archetypes used and rows created can be found in Appendix D. The performance measurement was conducted using a PC with an Intel Core i7 CPU (4 Cores, 4.7 GHz) and 16 GB of RAM. Five repetitions were made to reliably test the performance of Eos.

The overall execution time took about 15 min (914 s). This means approximately 126 compositions were loaded from the openEHR platform, transformed and integrated into the CDM database each second.

4. Discussion

The Eos tool and a set of OMOCL mappings were successfully implemented and tested by integrating several sample and real-world data-sets. Our results show that a generic standardized ETL process between openEHR and OMOP can be applied, thus opening up a multitude of clinical records for research, which have not been available yet in an automated way. Furthermore, the paper provides a first set of mappings, including an open-source tool to execute them. These mappings cover some of the most used archetypes and lay the foundations for an openEHR to OMOP mapping library. As a result, a fraction of openEHR records is now already convertible to OMOP. This includes the COVID-19 data used by thirty six German university hospitals [48]. Combined with the tooling provided to the community, this lowers the barriers to access openEHR records for secondary use in research.

4.1. OpenEHR as a data source for OMOP

In order to transform openEHR records into CDM tables, archetype mappings were used. This has several advantages. First, an archetype represents the maximum data-set of a clinical model. This makes it suitable for all types of transformations into models that aim for a less exhaustive representation. Users do not edit these archetypes, instead they abstract what they need in templates. Therefore, an archetype mapping is generally valid no matter the template it is used in. This makes the mappings deployable in different locations and contexts without the need for adjustment. Secondly, these archetypes are from an internationally used library and are therefore already valid on an international scale. As a result, the mappings enable a "plug-and-play" solution for openEHR on an international scale.

To further improve this process, clinical modelers could already annotate archetypes with OMOP CDM mappings. These annotations would help to create mappings and may leverage an automatic generation of OMOCL files. Mapping configurations could also be included as part of the archetype libraries. This would make mappings better accessible for use and review, hereby improving quality and making use of already established community modeling workflows. To integrate the openEHR terminologies into OMOP, a vocabulary could be introduced.

However, the OMOP CDM has only a limited set of static tables for observational research. OpenEHR, on the other hand, covers many healthcare domains and has a flexible model that supports semantically rich data. As a result, some of the openEHR data cannot be represented in OMOP. Archetypes are a maximum data-set approach and some of the nodes contained in an archetype have no specific mapping to a CDM table. Depending on the archetype node, non-specific tables like Observation or the completely textual NOTE could be used for that. NOTES represent unstructured information, storing coded or quantified data with them results in a loss of semantics. The resulting tables of a transformation then need to be linked with each other using the FACT_RELATIONSHIP. The NOTE, NOTE_NLP and FACT_ RELATIONSHIP are rarely used by the OHDSI community and are missing tool support [51]. When analyzing the CDM, the researcher is required to know about this and query the tables accordingly, which may not even be possible due to lacking support of the tool. Given the amount and complexity of data, such a distribution of related data leaves room for misinterpretations. On the other hand, openEHR has a standardized field, which can be filtered when defining a cohort for example. Missing health data bears the risk that it can reduce the study power and may even lead to a false conclusion, e.g., in a clinical trial [4]. It is to be expected that new tables and fields will be added in the future to OMOP. Eos and OMOCL are also required to adapt to these changes.

Another limitation is the unspecific way imposed by openEHR for its demographic data. We provide different solutions for this problem. Nevertheless, if demographic data is stored externally, users still have to implement a custom ETL process, if required. Tooling could be provided as part of a future solution to better support this process — for example an endpoint to input person data. On the other hand, keeping the demographic data external also has its advantages in terms of privacy [53] and eases the way data can be provided for research in conformance with privacy standards.

4.2. Eos and OMOCL

The OMOCL managed to represent the mappings in a human and machine readable format. The declarative language provided the means to define all but one mapping without the need of programming and scripting languages. The resulting mappings and language will be free to use and open source. We hope that this will initiate a collective effort to cover all missing archetypes. Ultimately, this would make all openEHR records accessible for secondary use without the need for custom ETL processes. In addition, this free access also enables users to define their own solutions to execute the OMOCL files. OMOCL could also be extended to support other CIM like FHIR. Nevertheless, this is the first version of OMOCL, therefore it is to be expected that the language will require updates and fixes.

With Eos we implemented a tool to execute mappings based on OMOCL files successfully. This was tested using sample and real-world data. The tool offers different API endpoints and configuration parameters, while being open source for the community. In the current state, Eos was only tested within a fixed environment. Using the DQD, the quality of the transformed CDM data was analyzed. The data passed 98% of the DQD tests, which is acceptable. Such a validation also strongly depends on the quality of the input data. Eos and OMOCL rely on correct and complete input data in order to manifest conform CDM tables. The overall performance of the import of data with Eos shows encouraging results, with 126 compositions imported per second. Given the limited hardware resources of our test setup, the approximately 15 min taken to populate the CDM with 423756 rows seems acceptable. To improve its stability and correctness, further testing is required.

The OMOCL mappings currently cover 10.5% of the published archetypes in the CjabreKM. While this number appears to be small, these are some of the most frequently used archetypes that also cover most of the data required for patient summaries. The goal of this work was to define an approach to integrate both standards. Eos and OMOCL enable such an integration. OpenEHR is a community driven standard. With the tooling being provided, we expect that the rest of the archetypes will be mapped as part of a community effort in the near future.

The scope of this study was to make clinical data accessible for secondary use. As a result, a transformation back from OMOP to openEHR was not part of the design and implementation of the tools. Another limitation of Eos and OMOCL is the missing support for the CDT tables: VISIT_DETAIL, NOTE, NOTE_NLP and a declarative method to define FACT_RELATIONSHIP. Nevertheless, all but VISIT_DETAIL are also not supported by OHDSI's official cohort exploration tool [51]. Meanwhile, VISIT_DETAIL is used as an alternative mechanism to define a VISIT_OCCURRENCE. In addition, we are also missing support for six CDT fields which we did not identify as crucial. Therefore, we decided that these fields can be neglected for an initial release. We plan to add support for all CDM fields and tables in the future.

4.3. Comparison to related work

A related work to make openEHR records accessible for secondary use is the approach by Haarbrandt et al. [31], who developed an automated approach to populate an i2b2 date warehouse from openEHR records. The results showed that a transformation of openEHR records into i2b2 is feasible. Similar to OMOP, i2b2 has a less expressive data model than openEHR [31]. Therefore, a transformation always comes with a loss of semantics. When compared to OMOP, i2b2 has no standardized and harmonized terminologies as OMOP does with its vocabularies. Furthermore, i2b2 has no specific representation for measurements, device exposures and drug exposures like OMOP does. As a result, a transformation to OMOP has the advantage of harmonizing terminologies and sustaining more of the initial semantics. Combined, the approaches show that openEHR can be readily transformed into the two major standards currently employed for secondary use of clinical data. OpenEHR is thus transformable into semantically less rich models.

Another approach to make CIM accessible for secondary use are the ones made for integrating FHIR data into the OMOP CDM [29,33]. The OMOP-on-FHIR [33] ETL tool is open source and allows for a bidirectional transformation of FHIR and OMOP data. On the other hand, Lenert et al. [29] successfully transformed FHIR-data into an OMOP CDM database. Every approach made with FHIR has the limitation that mappings need to be altered if different profiles are used. FHIR provides a set of generic resources that are specialized by users for the specific use cases, using so called profiles. As an example, the GECCO data set [46] representation in FHIR contains 66 of such profiles [54]. Each of these profiles can introduce and alter existing fields of the resources [55]. As a result, for each profile, it may be required to reconfigure the mappings accordingly. Therefore, it is not possible to provide a generally valid ETL tool for FHIR, if profiles are used that alter the rudimentary resources. OpenEHR on the other hand has a maximum data set approach using archetypes. Once mapped, these mappings are generally valid.

4.4. OpenEHR for research

In the future, it may be viable to directly use openEHR as a data model for research, and in doing so, enable more data to be represented and therefore make it accessible for secondary use. OpenEHR is especially suited for this since it is FAIR enabling [56] and General Data Protection Regulation (GDPR) [53] compliant by design, has international models, a maximum data-set approach and a native query language. Furthermore, for institutions that already use openEHR, this would be more efficient and resource-saving than integrating it into, e.g., OMOP. Other institutions would be required to write ETL processes in both cases. In case of openEHR, these institutions would also set the foundation for a homogeneous EHR, therewith lowering the barrier for the standardization of the entire institution and closing the gap between the clinical and research domains. In addition, calculations could be shared and executed locally as it is done with OMOP. As an example, the COVID-19 Research Network of University Medicine Germany [48] uses openEHR for its central COVID-19 data platform. The openEHR representation of the GECCO used for this COVID-19 data, contains only international archetypes, except for the archetypes for patient admission, study participation and symptom signs. As a consequence, every calculation made on these archetypes is shareable with other openEHR COVID-19 projects in the very same manner as it is with OMOP, but with no need for initial ETL processes and with more semantically rich data. Nevertheless, openEHR platforms focus on clinical use and therefore more on transactional than analytical purposes. Compared to OMOP, openEHR specifically lacks tooling for research, this needs to be added in the future to make it more viable for global projects.

In theory, this could also be done with other CIM, but often requires more intense work. As an example, the other major CIM, FHIR, has no large internationally harmonized library of clinical models. Furthermore, FHIR does not aim for a maximum data-set approach. This bears the risk that source systems have different data needs for the same clinical model. In order to harmonize these, either the models have to be standardized or an ETL process has to be established between FHIR systems. In openEHR this can also be the case with archetypes but is less likely due to its internationally shared archetypes. As an example, the openEHR representation of the GECCO uses three non international archetypes. Meanwhile the FHIR representation consists mainly of models that were defined specifically for it or reuses some from other national projects. This shows that FHIR requires additional modeling efforts for the same data set. If data is persisted in FHIR, the standard currently offers no standardized operation to permanently delete all patient data, since the version history of a resource is not removed [57]. Therefore, FHIR data repositories are not per se GDPR compliant and

require the implementation of a permanent delete. As a consequence, the current version of FHIR requires intensive resources compared to openEHR. However, FHIR is a health care data exchange standard [55] and not designed for persistence etc., as stated by the documentation: "In principle, resources are designed for exchange between systems, rather than as a database storage format". [55]. An out of context use also bears the risk of creating a silver bullet thinking [58].

5. Conclusion

We successfully developed a comprehensive approach to integrate openEHR records into the OMOP CDM. With the definition of the OMOCL DSL we were able to represent a first set of archetypes-to-CDM table mappings that cover 10.05% of the internationally published archetypes. This demonstrated that the maximum data-set approach of openEHR provides the possibility for a transformation into less expressive models with generic mappings that are valid on an international scale. The implemented Eos tool used these OMOCL-files to successfully automatize the ETL from real-world and sample EHRs into the OMOP CDM. To the authors' knowledge, this is the first implementation of a generic openEHR to OMOP transformation and thus represents a significant step towards achieving interoperability between the clinical and the research data domains.

However, our approach showed that such a transformation into less expressive data models comes with the loss of information. We suggest that future research projects should export their data into a model that can represent clinical data without information loss. OpenEHR is especially suited for that, providing, apart from its maximum data-set models, a FAIR enabling and GDPR compliant design. Such an adaptation of openEHR would close the gap between the clinical and research domain. Hereby, more semantic-rich data would be made more readily accessible to match the ever-growing complexity of modern and future research.

CRediT authorship contribution statement

Severin Kohler: Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Resources, Writing – original draft, Writing – review & editing, Visualization. **Diego Boscá:** Conceptualization, Methodology, Validation, Investigation, Writing – review & editing. **Florian Kärcher:** Conceptualization, Software. **Birger Haarbrandt:** Conceptualization, Writing – review & editing. **Manuel**

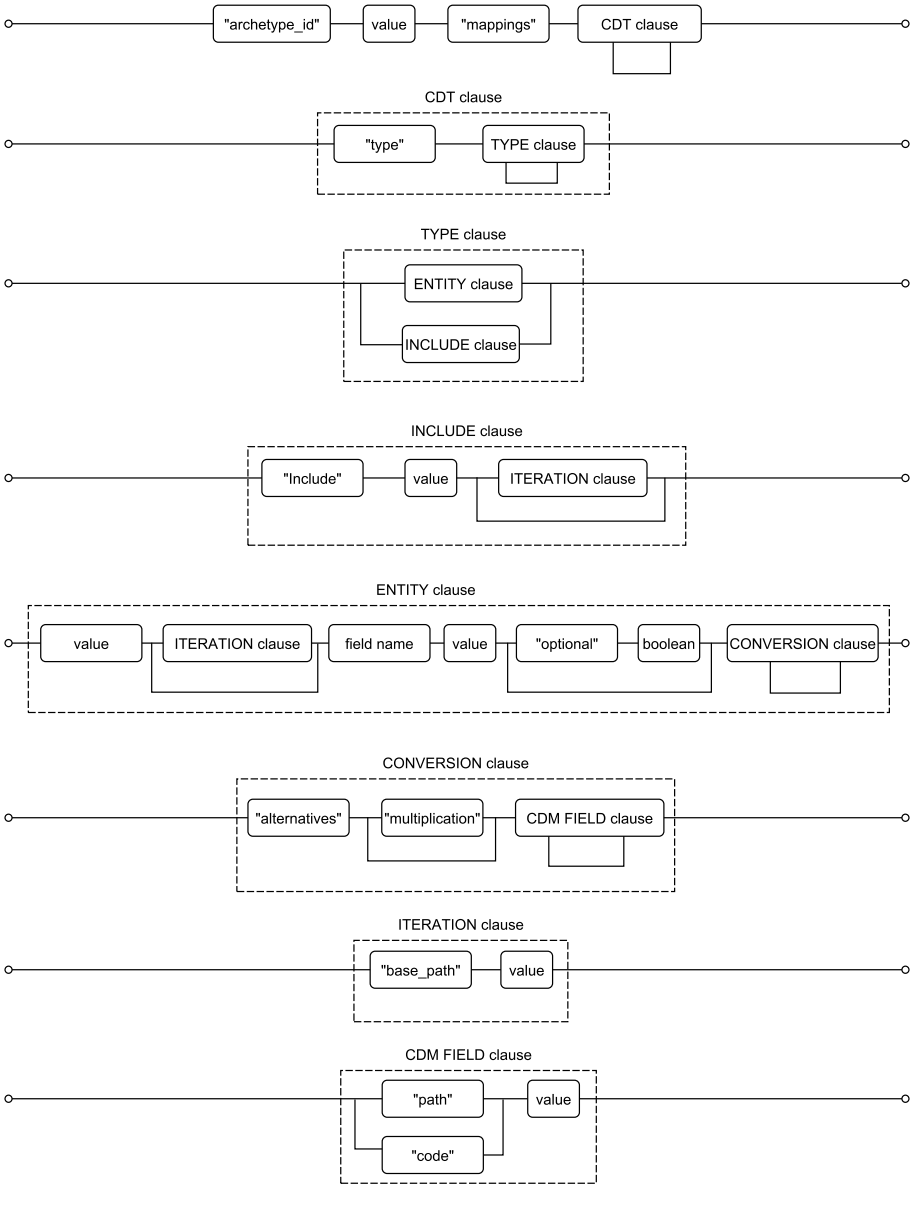

Fig. A.5. OMOCL simplified railroad diagram.

```
archetype_id: "openEHR-EHR-EVALUATION.problem_diagnosis.v1"

mappings:
  - type: "ConditionOccurrence"
    concept_id:
      alternatives:
        - path: "data[at0001]/items[at0002]"
    condition_start_date:
      alternatives:
        - path: "data[at0001]/items[at0077]"
        - path: "../context"
    condition_end_date:
      optional: true
      alternatives:
        - path: "data[at0001]/items[at0030]"
    condition_status_concept_id:
      optional: true
      alternatives:
        - path: "data[at0001]/items[openEHR-EHR-CLUSTER.
            problem_qualifier.v1]/items[at0004]"
        - code: 32893
```

Fig. B.6. OMOCL file for the international problem diagnose archetype.

```
archetype_id: "openEHR-EHR-EVALUATION.death_summary.v0"

mappings:
  - type: "Death"
    death_date:
      alternatives:
        - path: "/data[at0001]/items[at0092]"
```

Fig. B.7. OMOCL file for the international death summary archetype.

```
archetype_id: "openEHR-EHR-EVALUATION.device_summary.v0"

mappings:
  - type: "Include"
    base_path: "/data[at0001]/items[at0022]/items[openEHR-EHR-CLUSTER.
        device.v1]"
    archetype_id: "openEHR-EHR-CLUSTER.device.v1"
```

Fig. B.8. OMOCL file for the international device summary archetype.

Prinz: Software, Validation. **Michael Marschollek:** Writing – review & editing, Data curation. **Roland Eils:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We want to thank Klaus Hendrik Wolf of the Peter L. Reichertz Institute for testing our tool with real world data. We would also like to thank Renaud Subigre of the European External Action Service and Dirk Meyer zu Büschenfelde of the Berlin Institute of Health for discussions of our work.

Appendix A

See Fig. A.5.

Appendix B

See Figs. B.6–B.29.

```
archetype_id: "openEHR-EHR-ACTION.medication.v1"

mappings:
  - type: "DrugExposure"
    concept_id:
      alternatives:
        - path: "/description[at0017]/items[at0020]"
    drug_exposure_start_date:
      alternatives:
        - path: "/items[at0043]"
        - path: "../"
        - path: "."
    drug_exposure_end_date:
      alternatives:
        - path: "/items[at0043]"
        - path: "../"
        - path: "."
    quantity:
      optional: true
      alternatives:
        - multiplication:
            - path: "/description[at0017]/items[openEHR-EHR-CLUSTER.
                dosage.v1]/items[at0144]"
            - path: "/description[at0017]/items[openEHR-EHR-CLUSTER.
                dosage.v1]/items[at0164]"
        - path: "/description[at0017]/items[openEHR-EHR-CLUSTER.dosage.
            v1]/items[at0144]"
```

Fig. B.9. OMOCL file for the international medication archetype.

```
archetype_id: "openEHR-EHR-INSTRUCTION.medication_order.v2"

mappings:
  - type: "DrugExposure"
    base_path: "activities[at0001]"
    concept_id:
      alternatives:
        - path: "description[at0002]/items[at0070]"
    drug_exposure_start_date:
      alternatives:
        - path: "description[at0002]/items[at0113]/items[at0012]"
    drug_exposure_end_date:
      alternatives:
        - path: "description[at0002]/items[at0113]/items[at0013]"
    route_concept_id:
      optional: true
      alternatives:
        - path: "description[at0002]/items[at0091]"
```

Fig. B.10. OMOCL file for the international medication order archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.medication_statement.v0"

mappings:
  - type: "DrugExposure"
    base_path: "data[at0001]/events[at0002]"
    concept_id:
      alternatives:
        - path: "data[at0003]/items[at0006]"
    drug_exposure_start_date:
      alternatives:
        - path: "data[at0003]/items[at0024]"
        - path: "data[at0003]/items[at0019]"
        - path: "."
    drug_exposure_end_date:
      alternatives:
        - path: "data[at0003]/items[at0025]"
        - path: "data[at0003]/items[at0021]"
        - path: "data[at0003]/items[at0026]"
        - path: "data[at0003]/items[at0024]"
        - path: "data[at0003]/items[at0019]"
        - path: "."
    route_concept_id:
      optional: true
      alternatives:
        - path: "data[at0003]/items[at0030]"
```

Fig. B.11. OMOCL file for the international medication statement archetype.

```
archetype_id: "openEHR-EHR-EVALUATION.tobacco_smoking_summary.v1"

mappings:
  - type: "Observation"
    concept_id:
      alternatives:
        - code: 4041306
    observation_date:
      alternatives:
        - path: "../context"
    value:
      alternatives:
        - path: "/data[at0001]/items[at0089]"
        - path: "/data[at0001]/items[at0043]"
```

Fig. B.12. OMOCL file for the international tobacco smoking summary archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.pregnancy_status.v0"

mappings:
  - type: "Observation"
    concept_id:
      alternatives:
        - code: 42528957
    observation_date:
      alternatives:
        - path: "/data[at0001]/events[at0002]/data[at0003]"
    value:
      alternatives:
        - path: "/data[at0001]/events[at0002]/data[at0003]/items[at0011
            ]"
```

Fig. B.13. OMOCL file for the international pregnancy status archetype.

```
archetype_id: "openEHR-EHR-ADMIN_ENTRY.person_data.v0"

mappings:
  - type: "Person"
    gender_concept:
      alternatives:
        - path: "../content[openEHR-EHR-EVALUATION.gender.v1]/data[
            at0002]/items[at0019]"
        - code: 0
    year_of_birth:
      alternatives:
        - path: "/data[at0001]/items[openEHR-EHR-CLUSTER.
            person_birth_data_iso.v0]/items[at0001]"
        - path: "../content[openEHR-EHR-OBSERVATION.age.v0]/data[at0001
            ]/events[at0002]/data[at0003]/items[at0004]"
        - code: 0
    ethnicity_concept_id:
      alternatives:
        - path: "/data[at0001]/items[openEHR-EHR-CLUSTER.
            ethnischer_hintergrund.v0]/items[at0002]"
        - code: 0
```

Fig. B.14. OMOCL file for the international person data archetype.

```
archetype_id: "openEHR-EHR-ACTION.procedure.v1"

mappings:
  - type: "ProcedureOccurrence"
    concept_id:
      alternatives:
        - path: "description[at0001]/items[at0002]"
    procedure_start_date:
      alternatives:
        - path: "description[at0001]/items[at0066]"
        - path: "."
    procedure_end_date:
      optional: true
      alternatives:
        - path: "description[at0001]/items[at0060]"
```

Fig. B.15. OMOCL file for the international procedure archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.blood_pressure.v2"

mappings:
  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 4152194
    measurement_date:
      alternatives:
        - path: "data[at0001]/events[at0006]"
    value:
      alternatives: &valueAlternatives1
        - path: "data[at0001]/events[at0006]/data[at0003]/items[at0004]"
    unit:
      alternatives:
        - code: 8876
    range_low:
      optional: true
      alternatives: *valueAlternatives1
    range_high:
      optional: true
      alternatives: *valueAlternatives1
    operator_concept_id:
      optional: true
      alternatives: *valueAlternatives1

  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 4154790
    measurement_date:
      alternatives:
        - path: "data[at0001]/events[at0006]"
    value:
      alternatives: &valueAlternatives2
        - path: "data[at0001]/events[at0006]/data[at0003]/items[at0005]"
    unit:
      alternatives:
        - code: 8876
    range_low:
      optional: true
      alternatives: *valueAlternatives2
    range_high:
      optional: true
      alternatives: *valueAlternatives2
    operator_concept_id:
      optional: true
      alternatives: *valueAlternatives2
```

Fig. B.16. OMOCL file for the international blood pressure archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.body_temperature.v2"

mappings:
  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 3020891
    measurement_date:
      alternatives:
        - path: "/data[at0002]/events[at0003]"
        - path: "."
    value:
      alternatives: &valueAlternatives
        - path: "/data[at0002]/events[at0003]/data[at0001]/items[at0004
            ]"
    unit:
      alternatives: *valueAlternatives
    range_low:
      optional: true
      alternatives: *valueAlternatives
    range_high:
      optional: true
      alternatives: *valueAlternatives
    operator_concept_id:
      optional: true
      alternatives: *valueAlternatives
```

Fig. B.17. OMOCL file for the international body temperature archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.body_weight.v2"

mappings:
  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 3025315
    measurement_date:
      alternatives:
        - path: "/data[at0002]/events[at0003]"
        - path: "."
    value:
      alternatives: &valueAlternatives
        - path: "/data[at0002]/events[at0003]/data[at0001]/items[at0004
            ]"
    unit:
      alternatives: *valueAlternatives
    range_low:
      optional: true
      alternatives: *valueAlternatives
    range_high:
      optional: true
      alternatives: *valueAlternatives
    operator_concept_id:
      optional: true
      alternatives: *valueAlternatives
```

Fig. B.18. OMOCL file for the international body weight archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.body_mass_index.v2"

mappings:
  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 4245997
    measurement_date:
      alternatives:
        - path: "/data[at0001]/events[at0002]"
        - path: "."
    value:
      alternatives: &valueAlternatives
        - path: "/data[at0001]/events[at0002]/data[at0003]/items[at0004
            ]"
    unit:
      alternatives: *valueAlternatives
    range_low:
      optional: true
      alternatives: *valueAlternatives
    range_high:
      optional: true
      alternatives: *valueAlternatives
    operator_concept_id:
      optional: true
      alternatives: *valueAlternatives
```

Fig. B.19. OMOCL file for the international body mass index archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.clinical_frailty_scale.v1"

mappings:
  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 40483383
    measurement_date:
      alternatives:
        - path: "../context"
    value:
      alternatives: &valueAlternatives
        - path: "/data[at0001]/events[at0002]/data[at0003]/items[at0004
            ]"
    unit:
      optional: true
      alternatives: *valueAlternatives
    range_low:
      optional: true
      alternatives: *valueAlternatives
    range_high:
      optional: true
      alternatives: *valueAlternatives
    operator_concept_id:
      optional: true
      alternatives: *valueAlternatives
```

Fig. B.20. OMOCL file for the international clinical frailty scale archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.height.v2"

mappings:
  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 3036277
    measurement_date:
      alternatives:
        - path: "/data[at0001]/events[at0003]"
        - path: "."
    value:
      alternatives: &valueAlternatives
        - path: "/data[at0001]/events[at0002]/data[at0003]/items[at0004
            ]"
    unit:
      alternatives: *valueAlternatives
    range_low:
      optional: true
      alternatives: *valueAlternatives
    range_high:
      optional: true
      alternatives: *valueAlternatives
    operator_concept_id:
      optional: true
      alternatives: *valueAlternatives
```

Fig. B.21. OMOCL file for the international height archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.pulse.v2"

mappings:
  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 4224504
    measurement_date:
      alternatives:
        - path: "/data[at0002]/events[at0003]"
        - path: "."
        - path: "../"
    value:
      alternatives: &valueAlternatives1
        - path: "/data[at0002]/events[at0003]/data[at0001]/items[at0004
            ]"
    unit:
      optional: true
      alternatives: *valueAlternatives1
    range_low:
      optional: true
      alternatives: *valueAlternatives1
    range_high:
      optional: true
      alternatives: *valueAlternatives1
    operator_concept_id:
      optional: true
      alternatives: *valueAlternatives1
```

Fig. B.22. OMOCL file for the international pulse archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.pulse_oximetry.v1"

mappings:
  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 4196147
    measurement_date:
      alternatives:
        - path: "/data[at0001]/events[at0002]"
        - path: "."
    value:
      alternatives: &valueAlternatives1
        - path: "/data[at0001]/events[at0002]/data[at0003]/items[at0006
            ]"
    unit:
      optional: true
      alternatives: *valueAlternatives1

  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 40760909
    measurement_date:
      alternatives:
        - path: "/data[at0001]/events[at0002]"
        - path: "."
    value:
      alternatives: &valueAlternatives2
        - path: "/data[at0001]/events[at0002]/data[at0003]/items[at0044
            ]"
    unit:
      optional: true
      alternatives: *valueAlternatives2

  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 36305056
    measurement_date:
      alternatives:
        - path: "/data[at0001]/events[at0002]"
        - path: "."
    value:
      alternatives: &valueAlternatives3
        - path: "/data[at0001]/events[at0002]/data[at0003]/items[at0045
            ]"
    unit:
      optional: true
      alternatives: *valueAlternatives3

  - type: "Measurement"
    concept_id:
      alternatives:
        - code: 3007930
    measurement_date:
      alternatives:
        - path: "/data[at0001]/events[at0002]"
        - path: "."
    value:
      alternatives: &valueAlternatives4
        - path: "/data[at0001]/events[at0002]/data[at0003]/items[at0046
            ]"
    unit:
      optional: true
      alternatives: *valueAlternatives4
```

Fig. B.23. OMOCL file for the international pulse oximetry archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.respiration.v2"

mappings:
  - type: "Measurement"
    base_path: "/data[at0001]/events[at0002]"
    concept_id:
      alternatives:
        - code: 4154772
    measurement_date:
      alternatives:
        - path: "."
        - path: "../../"
    value:
      alternatives: &valueAlternatives
        - path: "data[at0003]/items[at0004]"
    unit:
      alternatives:
        - code: 8541
    range_low:
      optional: true
      alternatives: *valueAlternatives
    range_high:
      optional: true
      alternatives: *valueAlternatives
    operator_concept_id:
      optional: true
      alternatives: *valueAlternatives
```

Fig. B.24. OMOCL file for the international respiration archetype.

```
archetype_id: "openEHR-EHR-OBSERVATION.laboratory_test_result.v1"

mappings:
  - type: "Include"
    base_path: "/data[at0001]/events[at0002]/data[at0003]/items[openEHR-
        EHR-CLUSTER.laboratory_test_analyte.v1]"
    archetype_id: "openEHR-EHR-CLUSTER.laboratory_test_analyte.v1"

  - type: "Include"
    base_path: "/data[at0001]/events[at0002]/data[at0003]/items[openEHR-
        EHR-CLUSTER.specimen.v1]"
    archetype_id: "openEHR-EHR-CLUSTER.specimen.v1"

  - type: "Include"
    base_path: "/data[at0001]/events[at0002]/data[at0003]/items[openEHR-
        EHR-CLUSTER.laboratory_test_panel.v0]/items[openEHR-EHR-CLUSTER
        .laboratory_test_analyte.v1]"
    archetype_id: "openEHR-EHR-CLUSTER.laboratory_test_analyte.v1"

  - type: "CustomMapping"
    name: "FactRelationshipCustomConverter"
```

Fig. B.25. OMOCL file for the international laboratory test result archetype.

```
archetype_id: "openEHR-EHR-CLUSTER.device.v1"

mappings:
  - type: "DeviceExposure"
    concept_id:
      alternatives:
        - path: "/items[at0001]"
        - path: "/items[at0003]"
    device_exposure_start_date:
      alternatives:
        - path: "../items[at0022]/items[at0008]"
        - path: "../"
        - path: "../../"
        - path: "../../../../../context"
    device_exposure_end_date:
      optional: true
      alternatives:
        - path: "../items[at0022]/items[at0009]"
        - path: "../"
        - path: "../../"
    unique_device_id:
      optional: true
      alternatives:
        - path: "/items[at0021]"
    production_id:
      optional: true
      alternatives:
        - path: "/items[at0020]"
```

Fig. B.26. OMOCL file for the international device archetype.

```
archetype_id: "openEHR-EHR-CLUSTER.dosage.v1"

mappings:
  - type: "DrugExposure"
    concept_id:
      alternatives:
        - path: "../items[at0020]"
    drug_exposure_start_date:
      alternatives:
        - path: "../items[at0043]"
        - path: "../../"
    drug_exposure_end_date:
      alternatives:
        - path: "../items[at0043]"
        - path: "../../"
    quantity:
      optional: true
      alternatives:
        - multiplication:
            - path: "/items[at0144]"
            - path: "/items[at0164]"
        - path: "/items[at0144]"
```

Fig. B.27. OMOCL file for the international dosage archetype.

Appendix C

See Table C.5.

Appendix D

See Table D.6.

Appendix E. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jbi.2023.104437.

```
archetype_id: "openEHR-EHR-CLUSTER.laboratory_test_analyte.v1"

mappings:
  - type: "Measurement"
    concept_id:
      alternatives:
        - path: "/items[at0024]"
    measurement_date:
      alternatives:
        - path: "/items[at0025]"
        - path: "../../"
        - path: "../../../"
    value:
      alternatives: &valueAlternatives
        - path: "/items[at0001]"
    unit:
      optional: true
      alternatives: *valueAlternatives
    range_low:
      optional: true
      alternatives: *valueAlternatives
    range_high:
      optional: true
      alternatives: *valueAlternatives
    operator_concept_id:
      optional: true
      alternatives: *valueAlternatives
```

Fig. B.28. OMOCL file for the international laboratory test analyte archetype.

```
archetype_id: "openEHR-EHR-CLUSTER.specimen.v1"

mappings:
  - type: "Specimen"
    concept_id:
      alternatives:
        - path: "/items[at0029]"
        - path: "/items[at0098]"
    specimen_date:
      alternatives:
      - path: "/items[at0034]"
    quantity:
      optional: true
      alternatives:
        - path: "/items[at0099]"
    specimen_source_id:
      optional: true
      alternatives:
        - path: "/items[at0001]"
    anatomic_site_concept:
      optional: true
      alternatives:
        - path: "/items[at0013]/items[openEHR-EHR-CLUSTER.
            anatomical_location.v1]/items[at0001]"
```

Fig. B.29. OMOCL file for the international specimen archetype.

Table C.5
Currently unsupported CDM version 5.4. fields by OMOCL and Eos.

CDT tables	Unsupported fields
Condition occurrence	provider_id , visit_detail_id
Death	–
Device Exposure	provider_id, visit_detail_id
Drug exposure	provider_id, visit_detail_id
Person	provider_id, care_site_id, location_id
Procedure occurrence	provider_id, visit_detail_id
Measurement	provider_id, visit_detail_id, measurement_event_id, meas_event_field_concept_id
Observation	provider_id, visit_detail_id, observation_event_id, obs_event_field_concept_id
Specimen	–

Table D.6
Amount and type of archetypes and CDM entities transformed for the performance tests. Clusters archetypes were not counted.

Archetype ID	Amount
EVALUATION.problem_diagnosis.v1	39733
ACTION.medication.v1	20791
INSTRUCTION.medication_order.v2	19097
OBSERVATION.medication_statement.v0	14151
EVALUATION.tobacco_smoking_summary.v1	16219
OBSERVATION.pregnancy_status.v0	10575
OBSERVATION.blood_pressure.v2	6543
OBSERVATION.body_weight.v2	3843
OBSERVATION.body_mass_index.v2	3557
OBSERVATION.clinical_frailty_scale.v1	8069
OBSERVATION.height.v2	15557
OBSERVATION.pulse_oximetry.v1	29749
OBSERVATION.pulse.v2	5357
OBSERVATION.respiration.v2	2351
OBSERVATION.laboratory_test_result.v1	36145
CDM table	**Amount of rows**
Visit occurrence	115876
Condition occurrence	19867
Drug exposure	28961
Measurement	133815
Observation	4937
Specimen	1787
Fact relationship	3574

References

[1] M.R. Cowie, J.I. Blomster, L.H. Curtis, S. Duclaux, I. Ford, F. Fritz, S. Goldman, S. Janmohamed, J. Kreuzer, M. Leenay, A. Michel, S. Ong, J.P. Pell, M.R. Southworth, W.G. Stough, M. Thoenes, F. Zannad, A. Zalewski, Electronic health records to facilitate clinical research, Clin. Res. Cardiol. 106 (1) (2017) 1–9, http://dx.doi.org/10.1007/s00392-016-1025-6.

[2] C. Safran, M. Bloomrosen, W.E. Hammond, S. Labkoff, S. Markel-Fox, P.C. Tang, D.E. Detmer, Toward a national framework for the secondary use of health data: An American medical informatics association white paper, J. Am. Med. Inform. Assoc JAMIA 14 (1) (2007) 1–9, http://dx.doi.org/10.1197/jamia.M2273.

[3] W.R. Hersh, Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance, Am. J. Manag. Care 13 (6) (2007) 277–278.

[4] C. Heneghan, B. Goldacre, K.R. Mahtani, Why clinical trial outcomes fail to translate into benefits for patients, Trials 18 (1) (2017) 122, http://dx.doi.org/10.1186/s13063-017-1870-2.

[5] C. Piller, Disgraced COVID-19 studies are still routinely cited, Science 371 (6527) (2021) 331–332, http://dx.doi.org/10.1126/science.371.6527.331.

[6] K. Dentler, A. ten Teije, N. de Keizer, R. Cornet, Barriers to the reuse of routinely recorded clinical data: a field report, Stud. Health Technol. Inform. 192 (2013) 313–317.

[7] E. Commission, D.-G.f.t.I. Society, Media, M. Virtanen, B. Ustun, J. Rodrigues, G. Surjan, A. Rector, K. Stroetmann, P. Lewalle, P. Zanstra, D. Kalra, Semantic Interoperability for Better Health and Safer Healthcare: Deployment and Research Roadmap for Europe, Publications Office, 2013, http://dx.doi.org/10.2759/38514,

[8] P. Coorevits, M. Sundgren, G.O. Klein, A. Bahr, B. Claerhout, C. Daniel, M. Dugas, D. Dupont, A. Schmidt, P. Singleton, G. De Moor, D. Kalra, Electronic health records: new opportunities for clinical research, J. Internal Med. 274 (6) (2013) 547–560, http://dx.doi.org/10.1111/joim.12119, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/joim.12119.

[9] H.U. Prokosch, T. Ganslandt, Perspectives for medical informatics, Methods Inf. Med. 48 (1) (2009) 38–44, http://dx.doi.org/10.3414/ME9132, Publisher: Schattauer GmbH.

[10] R. Kimball, The Data Warehouse Toolkit: the Definitive Guide To Dimensional Modeling, third ed., Wiley, Indianapolis, IN, 2013.

[11] W. Goossen, A. Goossen-Baremans, M.v.d. Zel, Detailed clinical models: A review, Healthc. Inform. Res. 16 (4) (2010) 201–214, http://dx.doi.org/10.4258/hir.2010.16.4.201, Publisher: Korean Society of Medical Informatics.

[12] C.G. Parker, R.A. Rocha, J.R. Campbell, S.W. Tu, S.M. Huff, Detailed clinical models for sharable, executable guidelines, in: MEDINFO 2004, 2004, pp. 145–148, http://dx.doi.org/10.3233/978-1-60750-949-3-145, Publisher: IOS Press.

[13] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR guiding principles for scientific data management and stewardship, Sci. Data 3 (1) (2016) 160018, http://dx.doi.org/10.1038/sdata.2016.18, URL https://www.nature.com/articles/sdata201618, Number: 1 Publisher: Nature Publishing Group.

[14] D. Bender, K. Sartipi, HL7 FHIR: An agile and RESTful approach to healthcare information exchange, in: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, 2013, pp. 326–331, http://dx.doi.org/10.1109/CBMS.2013.6627810.

[15] D. Kalra, T. Beale, S. Heard, The openEHR foundation, Stud. Health Technol. Inform. 115 (2005) 153–173.

[16] R.H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F.M. Behlen, P.V. Biron, A. Shabo (Shvo), HL7 clinical document architecture, release 2, J. Am. Med. Inform. Assoc. JAMIA 13 (1) (2006) 30–39, http://dx.doi.org/10.1197/jamia.M1888.

[17] H. Pohjonen, Chapter 20 - Norway, Sweden, and Finland as forerunners in open ecosystems and openEHR, in: E. Hovenga, H. Grain (Eds.), Roadmap To Successful Digital Health Ecosystems, Academic Press, 2022, pp. 457–471, http://dx.doi.org/10.1016/B978-0-12-823413-6.00011-2.

[18] L. Marco-Ruiz, B. Haarbrandt, S. Liosland Bakke, D. Zhao, B. Fabjan, I. McNicoll, Towards open electronic health records and data reuse networks: a multinational perspective on current and future OpenEHR implementations, in: MedInfo 2019, MedInfo, 2019, Available from: https://openehr.atlassian.net/wiki/spaces/resources/pages/365363201/Towards+Open+Electronic+Health+Records+and+Data+Reuse+Networks+a+Multinational+Perspective+on+Current+and+Future+OpenEHR+Implementations. (Last Accessed 9 January 2023).

[19] openEHR Website, 2023, Available from: https://www.openehr.org/. (Last Accessed 20 April 2023).

[20] B. Haarbrandt, B. Schreiweis, S. Rey, U. Sax, S. Scheithauer, O. Rienhoff, P. Knaup-Gregori, U. Bavendiek, C. Dieterich, B. Brors, I. Kraus, C.M. Thoms, D. Jäger, V. Ellenrieder, B. Bergh, R. Yahyapour, R. Eils, H. Consortium, M. Marschollek, HiGHmed – an open platform approach to enhance care and research across institutional boundaries, Methods Inf. Med. 57 (2018) e66–e81, http://dx.doi.org/10.3414/ME18-02-0002, Publisher: Schattauer GmbH.

[21] F. Prasser, O. Kohlbacher, U. Mansmann, B. Bauer, K.A. Kuhn, Data integration for future medicine (DIFUTURE), Methods Inf. Med. 57 (S 1) (2018) e57–e65, http://dx.doi.org/10.3414/ME17-02-0022, Publisher: Schattauer GmbH.

[22] H.-U. Prokosch, T. Acker, J. Bernarding, H. Binder, M. Boeker, M. Boerries, P. Daumke, T. Ganslandt, J. Hesser, G. Höning, M. Neumaier, K. Marquardt, H. Renz, H.-J. Rothkötter, C. Schade-Brittinger, P. Schmücker, J. Schüttler, M. Sedlmayr, H. Serve, K. Sohrabi, H. Storf, MIRACUM: Medical informatics in research and care in university medicine, Methods Inf. Med. 57 (S 1) (2018) e82–e91, http://dx.doi.org/10.3414/ME17-02-0025, Publisher: Schattauer GmbH.

[23] M.L. Braunstein, Health Informatics on FHIR: How HL7's API Is Transforming Healthcare, second ed., Springer, http://dx.doi.org/10.1007/978-3-030-91563-6.

[24] J.M. Overhage, P.B. Ryan, C.G. Reich, A.G. Hartzema, P.E. Stang, Validation of a common data model for active safety surveillance research, J. Am. Med. Inf. Assoc. 19 (1) (2012) 54–60, http://dx.doi.org/10.1136/amiajnl-2011-000376.

[25] OHDSI, The Book of OHDSI, 2023, Available from: https://ohdsi.github.io/TheBookOfOhdsi/. (Last Accessed 9 January 2023).

[26] G. Hripcsak, J.D. Duke, N.H. Shah, C.G. Reich, V. Huser, M.J. Schuemie, M.A. Suchard, R.W. Park, I.C.K. Wong, P.R. Rijnbeek, J. van der Lei, N. Pratt, G.N. Norén, Y.-C. Li, P.E. Stang, D. Madigan, P.B. Ryan, Observational health data sciences and informatics (OHDSI): Opportunities for observational researchers, Stud.Health Technol. Inform. 216 (2015) 574–578.

[27] P.E. Stang, P.B. Ryan, J.A. Racoosin, J.M. Overhage, A.G. Hartzema, C. Reich, E. Welebob, T. Scarnecchia, J. Woodcock, Advancing the science for active surveillance: Rationale and design for the observational medical outcomes partnership, Ann. Internal Med. 153 (9) (2010) 600–606, http://dx.doi.org/10.7326/0003-4819-153-9-201011020-00010, Publisher: American College of Physicians.

[28] M.A. Haendel, C.G. Chute, T.D. Bennett, D.A. Eichmann, J. Guinney, W.A. Kibbe, P.R.O. Payne, E.R. Pfaff, P.N. Robinson, J.H. Saltz, H. Spratt, C. Suver, J. Wilbanks, A.B. Wilcox, A.E. Williams, C. Wu, C. Blacketer, R.L. Bradford, J.J. Cimino, M. Clark, E.W. Colmenares, P.A. Francis, D. Gabriel, A. Graves, R. Hemadri, S.S. Hong, G. Hripscak, D. Jiao, J.G. Klann, K. Kostka, A.M. Lee, H.P. Lehmann, L. Lingrey, R.T. Miller, M. Morris, S.N. Murphy, K. Natarajan, M.B. Palchuk, U. Sheikh, H. Solbrig, S. Visweswaran, A. Walden, K.M. Walters, G.M. Weber, X.T. Zhang, R.L. Zhu, B. Amor, A.T. Girvin, A. Manna, N. Qureshi, M.G. Kurilla, S.G. Michael, L.M. Portilla, J.L. Rutter, C.P. Austin, K.R. Gersing, The national COVID cohort collaborative (N3C): Rationale, design, infrastructure, and deployment, J. Am. Med. Inform. Assoc. JAMIA 28 (3) (2020) 427–443, http://dx.doi.org/10.1093/jamia/ocaa196, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7454687/.

[29] L.A. Lenert, A.V. Ilatovskiy, J. Agnew, P. Rudisill, J. Jacobs, D. Weatherston, K.R. Deans Jr., Automated production of research data marts from a canonical fast healthcare interoperability resource data repository: applications to COVID-19 research, J. Am. Med. Inform. Assoc. JAMIA 28 (8) (2021) 1605–1611, http://dx.doi.org/10.1093/jamia/ocab108.

[30] S.N. Murphy, G. Weber, M. Mendis, V. Gainer, H.C. Chueh, S. Churchill, I. Kohane, Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), J. Am. Med. Inform. Assoc. JAMIA 17 (2) (2010) 124–130, http://dx.doi.org/10.1136/jamia.2009.000893, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3000779/.

[31] B. Haarbrandt, E. Tute, M. Marschollek, Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository, J. Biomed. Inform. 63 (2016) 277–294, http://dx.doi.org/10.1016/j.jbi.2016.08.007.

[32] E. Rinaldi, S. Thun, From OpenEHR to FHIR and OMOP data model for microbiology findings, Public Health Inform. (2021) 402–406, http://dx.doi.org/10.3233/SHTI210189, Publisher: IOS Press.

[33] T. Pope, M. Choi, R. Star, E. Soto, S. Crumpton, M. Riley, OMOP on FHIR, 2023, URL https://omoponfhir.org, Available from: https://omoponfhir.org(Last Accessed 9 January 2023).

[34] B. Li, R. Tsui, How to improve the reuse of clinical data– openEHR and OMOP CDM, J. Phys. Conf. Ser. 1624 (3) (2020) 032041, http://dx.doi.org/10.1088/1742-6596/1624/3/032041, Publisher: IOP Publishing.

[35] T. Beale, Archetypes: Constraint-based domain models for future-proof information systems, in: In Proceedings of the 11thOOPSLA Workshop on Behavioural Semantics, Seattle, WA, USA, 4–8 November, 2002.

[36] openEHR foundation, openEHR - reference model (RM) component - latest, 2023, Available from: https://specifications.openehr.org/releases/RM/latest. (Last Accessed 9 January 2023).

[37] openEHR foundation,archetype technology overview, 2023, Available from: https://specifications.openehr.org/releases/AM/latest/Overview.html. (Last Accessed 9 January 2023).

[38] openEHR foundation, openEHR - XML schemas (ITS-XML) component - 2.0.0, 2023, Available from: https://specifications.openehr.org/releases/ITS-XML/Release-2.0.0. (Last Accessed 9 January 2023).

[39] Openehr foundation, openEHR - JSON schemas (ITS-JSON) component - latest, 2023, Available from: https://specifications.openehr.org/releases/ITS-JSON/latest. (Last Accessed 9 January 2023).

[40] C.N. Mead, et al., Data interchange standards in healthcare it-computable semantic interoperability: Now possible but still difficult. do we really need a better mousetrap? J. Healthc. Inform. Manag. 20 (1) (2006) 71, Publisher: Citeseer.

[41] Ocean informatics, OpenEHR Clinical knowledge manager, 2023, Available from: https://ckm.openehr.org/ckm/. (Last Accessed 9 January 2023).

[42] OHDSI, OMOP common data model, 2023, Available from: https://ohdsi.github.io/CommonDataModel/index.html. (Last Accessed 9 January 2023).

[43] OHDSI, Athena, 2023, Available from: https://athena.ohdsi.org. (Last Accessed 9 January 2023).

[44] GitHub, 2023, Available from: https://github.com. (Last Accessed 21 June 2023).

[45] O. Ben-Kiki, C. Evans, YAML Ain't Markup Language (YAML™) Version 1.2, 84.

[46] J. Sass, A. Bartschke, M. Lehne, A. Essenwanger, E. Rinaldi, S. Rudolph, K.U. Heitmann, J.J. Vehreschild, C. von Kalle, S. Thun, The German Corona Consensus dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond, BMC Med. Inform. Decis. Making 20 (1) (2020) 341, http://dx.doi.org/10.1186/s12911-020-01374-w.

[47] Ocean informatics, highmed clinical knowledge manager, 2023, Available from: https://ckm.highmed.org. (Last Accessed 9 January 2023).

[48] Netzwerk Universitätsmedizin, Home | Netzwerk Universitätsmedizin, 2023, Available from: https://www.netzwerk-universitaetsmedizin.de. (Last Accessed 9 January 2023).

[49] EN ISO 27269:2022 - Health informatics - International patient summary (ISO 27269:2021), ITeh Standards URL https://standards.iteh.ai/catalog/standards/cen/6e4cf23d-3aa7-4171-a08a-b3612242609e/en-iso-27269-2022.

[50] OHDSI, SqlRender, Observational Health Data Sciences and Informatics, 2022, Available from: https://github.com/OHDSI/SqlRender. (Last Accessed 9 January 2023).

[51] OHDSI, Atlas, 2023, Available from: https://atlas.ohdsi.org#/home. (Last Accessed 9 January 2023).

[52] C. Blacketer, F.J. Defalco, P.B. Ryan, P.R. Rijnbeek, Increasing trust in real-world evidence through evaluation of observational data quality, J. Am. Medi. Inform. Assoc. 28 (10) (2021) 2251–2257, http://dx.doi.org/10.1093/jamia/ocab132.

[53] M. Sousa, D. Ferreira, C. Santos-Pereira, G. Bacelar, S. Frade, O. Pestana, R. Cruz-Correia, openEHR based systems and the general data protection Regulation (GDPR), Stud. Health Technol. Inform. 247 (2018) 91–95.

[54] Forschungsnetz Covid-19, 2023, Available from: https://simplifier.net/ForschungsnetzCovid-19/~introduction. (Last Accessed 9 January 2023).

[55] Fast healthcare interoperability resources (FHIR) specification, 2023, Available from: http://hl7.org/fhir/index.html(Last Accessed 9 January 2023).

[56] F. Frexia, C. Mascia, L. Lianas, G. Delussu, A. Sulis, V. Meloni, M. Del Rio, G. Zanetti, openEHR is FAIR-enabling by design, in: Public Health and Informatics, IOS Press, 2021, pp. 113–117.

[57] M. Anywar, B. Schreiweis, rn, H. Ulrich, Strategies and Recommendation for Data Loading of FHIR-Based Data Marts with Focus on GDPR Compliance, in: Digital Professionalism in Health and Care: Developing the Workforce, Building the Future, 2022, pp. 127–131, http://dx.doi.org/10.3233/SHTI220921, URL https://ebooks.iospress.nl/doi/10.3233/SHTI220921, Publisher: IOS Press.

[58] F.P. Brroks, No silver bullet: Essence and accident of software engineering, Computer 20 (4) (1987).