OMOP CDM and Vocabulary
Helpful Bookmarks

https://ohdsi.github.io/CommonDataModel/

https://athena.ohdsi.org/

https://ohdsi.github.io/TheBookOfOhdsi/
Why a Common Data Model
Why a Common Data Model

https://dash.ohdsi.org/research
The OMOP CDM is a system of tables, vocabularies, and conventions that allow observational health data to be standardized. It is this standard approach that facilitates rapid innovation in the areas of open-source development, methods research, and evidence generation.

https://ohdsi.github.io/CommonDataModel/index.html
The OMOP CDM is a system of tables, vocabularies, and conventions that allow observational health data to be standardized.

It is this standard approach that facilitates rapid innovation in the areas of open-source development, methods research, and evidence generation.

https://ohdsi.github.io/CommonDataModel/index.html
The OMOP CDM is a person-centric model
A typical patient journey within a healthcare system

- Visit
- Encounter
- Lab Tests
- Treatment
- Discharge
A typical patient journey within a healthcare system into data

https://ohdsi.github.io/CommonDataModel/index.html
A typical patient journey within a healthcare system into data

[Diagram showing a patient journey with labels for different stages: Visit, Encounter, Lab Tests, Treatment, Discharge. Each stage has associated data types like person, visit_occurrence, visit_detail, location, care_site, measurement, specimen, provider, condition_occurrence, death, observation, note, drug_exposure, procedure_occurrence, device_exposure, cost.]
The OMOP CDM is a system of tables, vocabularies, and conventions that allow observational health data to be standardized. It is this standard approach that facilitates rapid innovation in the areas of open-source development, methods research, and evidence generation.
Conventions

General conventions of the model
General conventions of schemas
General conventions of data tables
General naming conventions of fields
General conventions of domains

Technical conventions
Table-specific conventions
Source data-specific conventions

https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html#data-model-conventions
Technical Conventions

Fields

Variable names across all tables follow one convention:

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>_SOURCE_VALUE</td>
<td>Verbatim information from the source data, typically used in ETL to map to CONCEPT_ID, and not to be used by any standard analytics. For example, CONDITION_SOURCE_VALUE = '787.02' was the ICD-9 code captured as a diagnosis from the administrative claim.</td>
</tr>
<tr>
<td>_ID</td>
<td>Unique identifiers for key entities, which can serve as foreign keys to establish relationships across entities. For example, PERSON_ID uniquely identifies each individual. VISIT_OCCURRENCE_ID uniquely identifies a PERSON encounter at a point of care.</td>
</tr>
<tr>
<td>_CONCEPT_ID</td>
<td>Foreign key into the Standardized Vocabularies (i.e. the standard_concept attribute for the corresponding term is true), which serves as the primary basis for all standardized analytics. For example, CONDITION_CONCEPT_ID = 31967 contains the reference value for the SNOMED concept of ‘Nausea’</td>
</tr>
<tr>
<td>_SOURCE_CONCEPT_ID</td>
<td>Foreign key into the Standardized Vocabularies representing the concept and terminology used in the source data, when applicable. For example, CONDITION_SOURCE_CONCEPT_ID = 45431665 denotes the concept of ‘Nausea’ in the Read terminology; the analogous CONDITION_CONCEPT_ID might be 31967, since SNOMED-CT is the Standardized Vocabulary for most clinical diagnoses and findings.</td>
</tr>
<tr>
<td>_TYPE_CONCEPT_ID</td>
<td>Delineates the origin of the source information, standardized within the Standardized Vocabularies. For example, DRUG_TYPE_CONCEPT_ID can allow analysts to discriminate between ‘Pharmacy dispensing’ and ‘Prescription written’</td>
</tr>
</tbody>
</table>

http://ohdsi.github.io/CommonDataModel/dataModelConventions.html
Table-specific Conventions

PERSON

Table Description

This table serves as the central identity management for all Persons in the database. It contains records that uniquely identify each person or patient, and some demographic information.

User Guide

All records in this table are independent Persons.

ETL Conventions

All Persons in a database needs one record in this table, unless they fail data quality requirements specified in the ETL. Persons with no Events should have a record nonetheless. If more than one data source contributes Events to the database, Persons must be reconciled, if possible, across the sources to create one single record per Person. The content of the BIRTH_DATETIME must be equivalent to the content of BIRTH_DAY, BIRTH_MONTH and BIRTH_YEAR.

<table>
<thead>
<tr>
<th>CDM Field</th>
<th>User Guide</th>
<th>ETL Conventions</th>
<th>Datatype</th>
<th>Required</th>
<th>Primary Key</th>
<th>Foreign Key</th>
<th>FK Table</th>
<th>FK Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>person_id</td>
<td>It is assumed that every person with a different unique identifier is in fact a different person and should be treated independently.</td>
<td>Any person linkage that needs to occur to uniquely identify Persons ought to be done prior to writing this table. This identifier can be the original id from the source data provided if it is an integer, otherwise it can be an autogenereated number.</td>
<td>integer</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

https://ohdsi.github.io/CommonDataModel/cdm54.html#PERSON
Source data-specific Conventions

Observation Period Considerations for EHR Data

By Melanie Philofsky and the EHR Working Group

The EHR WG convened on July 24, August 7, and August 21, 2020 to discuss the creation of an Observation Period from EHR data. The current and future conventions are not prescriptive enough and leave room for various ways of interpretation. The goals of our discussions were to increase the standardization for the implementation of the OBSERVATION_PERIOD table by providing some general guidelines for determining the start, end, and gaps in Observation Periods. The suggestions we came up with are only "suggestions" at this point. More research should be done to understand how these choices might impact evidence generated using these data. All of these decisions should be tempered by local understanding of patients in the EHR you are ETLing.

- Note - These suggestions are not intended for HMO EHR sites since HMO EHR Observation Periods more closely resemble claims data Observation Periods.

Observation Period Start Date

- Generally an Observation Period does NOT begin before birth, however, it might begin before birth IF the pregnant mother receives care recorded in your EHR. The child's record is then split from the mother's record at birth but may retain care given during pregnancy. For these children in your dataset, the field observation_period_start_date should be the birth date minus 9 months.

- An Observation Period does NOT begin before the implementation of the EHR at your site. Any records prior to implementation are probably "history of" record types and not a complete EHR record of clinical events.

- Special consideration should be given to migration from previous EHR, implementation at different sites within your healthcare system, implementation of different modules, etc.

Observation Period end date

Set the observation_period_end_date as the first date from the following:

- Date of death + 60 days
  - This is a CDM convention to allow events after death (autopsy, final notes, etc).

- Last clinical event + 60 days
  - The assumption is that person will return to the same health provider if an adverse reaction/complication/unresolved condition occurs.

- Date of the data pull from the system

https://ohdsi.github.io/CommonDataModel/ehrObsPeriods.html
The OMOP CDM is a system of tables, vocabularies, and conventions that allow observational health data to be standardized. It is this standard approach that facilitates rapid innovation in the areas of open-source development, methods research, and evidence generation.

https://ohdsi.github.io/CommonDataModel/index.html
Comparison of common data models

Balancing trade-offs in data management vs. analysis complexity

- Common protocol
  + Common structure
  + Common conventions
  + Common vocabularies

For 1 study:
- Common protocol
  + Common structure
- Cohort identification
  - Clinical characterization
  - Population-level effect estimation
  - Patient-level prediction

For N studies:
- Common protocol
  + Common structure
  + Common conventions
- Cohort identification
  - Clinical characterization
  - Population-level effect estimation
  - Patient-level prediction

Complexity for data management
(source data → input format for analysis)

Complexity for analyst
(input format for analysis → final analysis results)
Vocabularies

https://ohdsi.github.io/CommonDataModel/index.html
OMOP Standardized Vocabularies

All content: concepts in `concept`

Direct relationships between concepts in `concept_relationship`

Multi-step hierarchical relationships pre-processed into `concept_ancestor`
<table>
<thead>
<tr>
<th>Concept</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
<td>313217</td>
</tr>
<tr>
<td>Concept Name</td>
<td>Atrial fibrillation</td>
</tr>
<tr>
<td>Domain ID</td>
<td>Condition</td>
</tr>
<tr>
<td>Vocabulary ID</td>
<td>SNOMED</td>
</tr>
<tr>
<td>Concept Class ID</td>
<td>Clinical Finding</td>
</tr>
<tr>
<td>Standard Concept</td>
<td>S</td>
</tr>
<tr>
<td>Concept Code</td>
<td>49436004</td>
</tr>
<tr>
<td>Valid Start Date</td>
<td>01-Jan-2002</td>
</tr>
<tr>
<td>Valid End Date</td>
<td>31-Dec-2099</td>
</tr>
<tr>
<td>Invalid Reason</td>
<td>Code in vocabulary (SNOMED)</td>
</tr>
</tbody>
</table>

- **Unique identifier in OHDSI**: 313217
- **English description**: Atrial fibrillation
- **Domain**: Condition
- **Vocabulary**: SNOMED
- **Class in vocabulary (SNOMED)**: Clinical Finding
- **Standard/Non-standard/Classification**: S
- **Code in vocabulary (SNOMED)**: 49436004
- **Valid during time interval**: 01-Jan-2002 to 31-Dec-2099
- **Invalid Reason**: Code in vocabulary (SNOMED)
Mapping to OMOP Standardized Vocabularies

SOURCE_CODE
XYZ
i.e. ICPC-1 Dutch codes, ICD9, etc.

STANDARD_CONCEPT_ID
123456789
i.e. SNOMED for conditions and RxNorm, RxNorm Extension for drugs

• What is standardized:
  – TABLE_CONCEPT_ID: standard concept the source code maps to, used for analysis
  – TABLE_SOURCE_CONCEPT_ID: concept representation of the source code, helps maintain tie to raw data
  – TABLE_SOURCE_VALUE: original source code as given in the source table, helps to review data quality

• Ways to get a source code to standard code:
  – OMOP Vocabulary (concept_relationship)
  – USAGI
Mapping to OMOP Standardized Vocabularies

- If your source data’s codes are in the OMOP vocabularies, you can use it to translate to an OMOP standard
  - For example: ICD9 $\rightarrow$ SNOMED or NDC $\rightarrow$ RxNorm
OMOP Standardized Vocabularies In a Nutshell

• What it is:
  – **Standardized structure** to house existing vocabularies used in the public domain
  – **Compiled standards** from disparate public and private sources and some OMOP-grown concepts

• What it’s not
  – **Static dataset**: the vocabulary updates regularly to keep up with the continual evolution of the sources
  – **Finished product**: vocabulary maintenance and improvement is ongoing activity that requires community participation and support
Demo: ATHENA

- [https://athena.ohdsi.org/](https://athena.ohdsi.org/)
Exercises

Find standard concept IDs for the following conditions:

• Asthma
• Plague
• Ingrown toenail

Find standard concept IDs for the following drug ingredients:

• Metformin
• Tolazamide
• Telmisartan
Exercises

Find standard concept IDs for the following conditions:

• Asthma
  - 317009
• Plague
  - 434271
• Ingrown toenail
  - 4065236, 4290993

Find standard concept IDs for the following drug ingredients:

• Metformin
  - 1503297
• Tolazamide
  - 1502809
• Telmisartan
  - 1317640
Exercises

• What is the standard concept ID for the ICD10 code E11.9?
  – What domain does E11.9 belong to?
• What is the standard concept ID for the ICD10 code Z02.1?
  – What domain does Z02.1 belong to?
• What ICD10 codes are mapped to the concept ID 443767?
• What is the standard concept ID for the ICD10 code X67.0?
Exercises

• What is the standard concept ID for the ICD10 code E11.9?  
  – What domain does E11.9 belong to? **Source domain = OMOP domain**

• What is the standard concept ID for the ICD10 code Z02.1?  
  – What domain does Z02.1 belong to? **Source domain ≠ OMOP domain**

• What ICD10 codes are mapped to the concept ID 443767?  
  **n:1 mapping**

• What is the standard concept ID for the ICD10 code X67.0?  
  **1:n mapping**
Thank you!

Mui Van Zandt  mui.vanzandt@iqvia.com
Gyeol Song  gyeol.song@iqvia.com