



OMOP CDM and Vocabulary



Helpful Bookmarks

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

OMOP Common Data Model

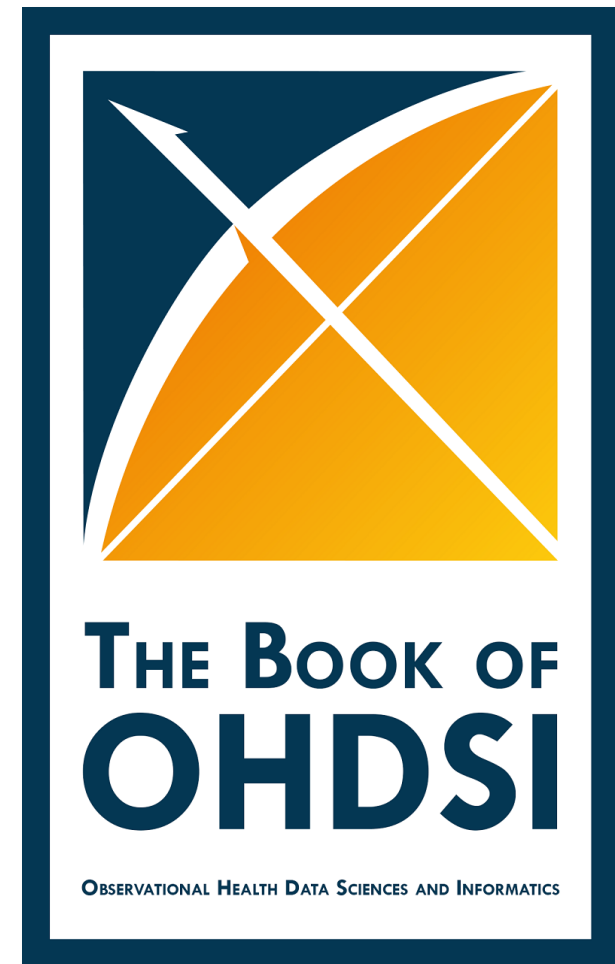
Background Conventions CDM Versions CDM Proposals How to Support

OMOP Common Data Model

The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) is an open community data standard, designed to standardize the structure and content of observational data and to enable efficient analyses that can produce reliable evidence. A central component of the OMOP CDM is the OHDSI standardized vocabularies. The OHDSI vocabularies allow organization and standardization of medical terms to be used across the various clinical domains of the OMOP common data model and enable standardized analytics that leverage the knowledge base when constructing exposure and outcome phenotypes and other features within characterization, population-level effect estimation, and patient-level prediction studies.

This website is meant to serve as a resource describing the specification of the available versions of the Common Data Model. This includes the structure of the model itself and the agreed upon conventions for each table and field as decided by the OHDSI Community. The vocabulary tables are part of the model and, as such, are detailed here. To download the vocabulary itself, please visit <https://athena.ohdsi.org>. For more information about the OHDSI suite of tools designed to implement best practices in characterization, population-level effect estimation and patient-level prediction, please visit <https://ohdsi.github.io/Hades/>.

<https://ohdsi.github.io/TheBookOfOhdsi/>



<https://athena.ohdsi.org/>

ATHENA SEARCH DOWNLOAD

Search

aspirin Search

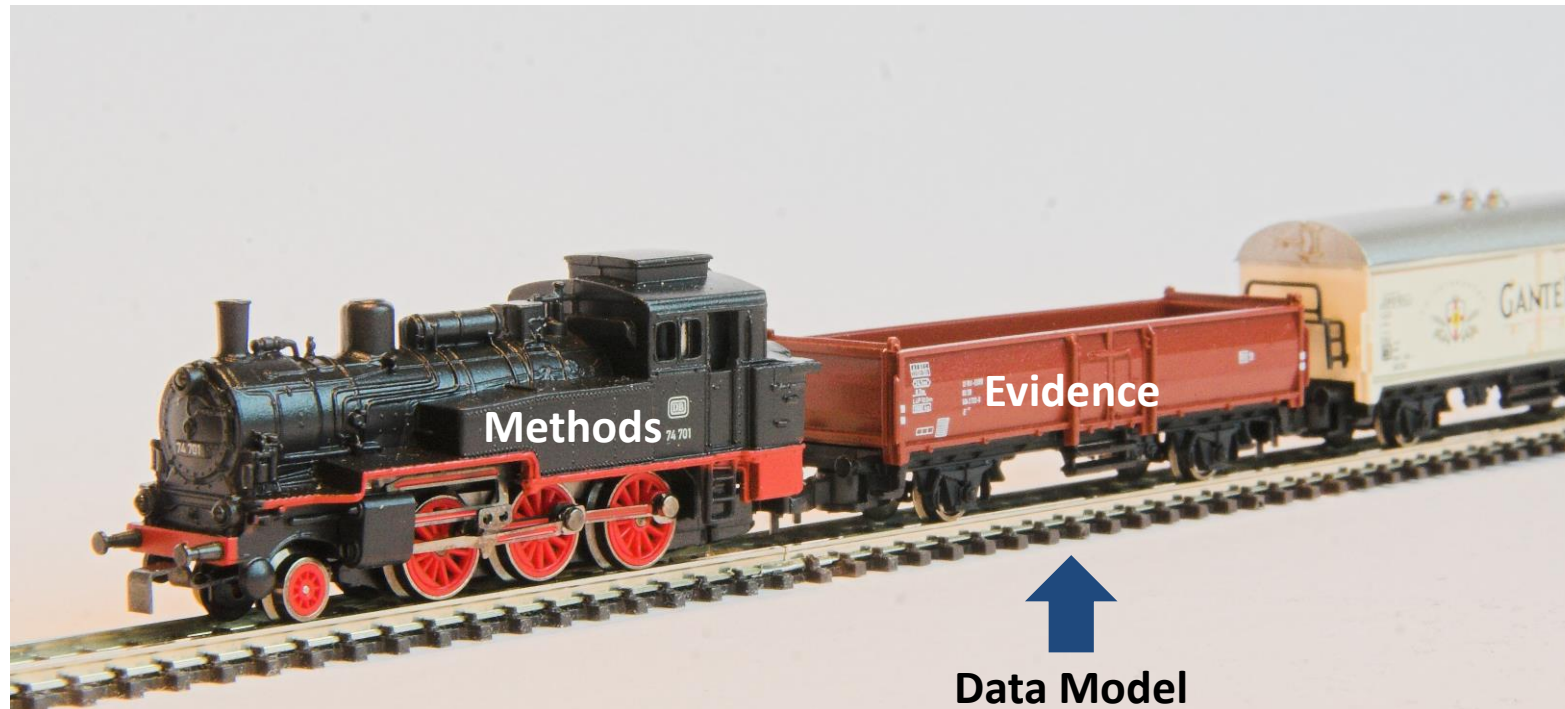
1. Usage of quotation marks forces an exact-match search
2. In case of a typo, or if there is a similar spelling of the word, the most similar result will be presented

Explore domains

Drugs 5,613,135	Conditions 675,961	Procedures 738,383
Devices 518,229	Observations 973,354	Measurements 561,032

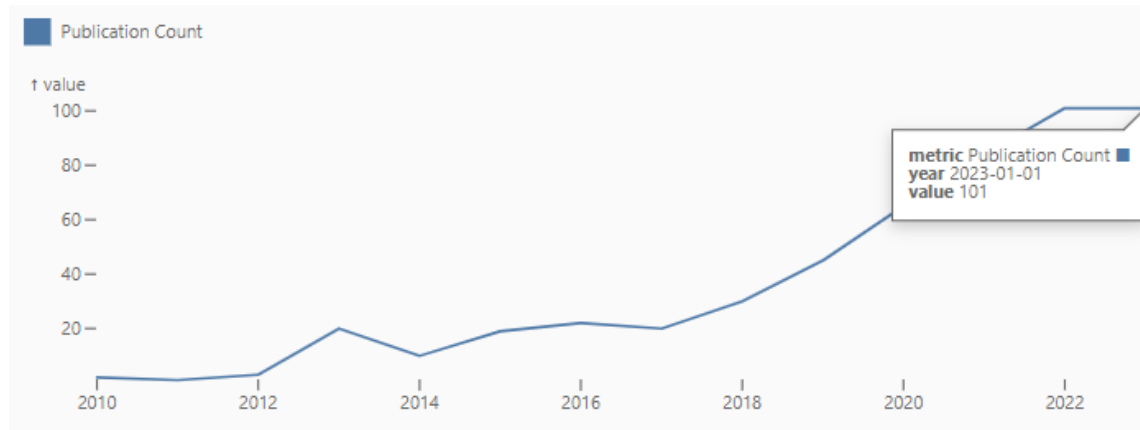


Why a Common Data Model





Why a Common Data Model



	Journal	Creation Date ↓	Authors
Predictive Models for Assessing Patients' Response to Treatment in Metastatic Prostate Cancer: A Systematic Review. 🔗	European urology open science	2024/04/10 04:15	Lawlor, Ailbhe Lin, Carol Gomez Rivas, Juan Ibanez, Laura Abad Lopez, Pablo Willemse, Peter-Paul Imran Omar, Muhammad Remmers, Sebastiaan Cornford, Philip Rajwa, Pawel Nicoletti, Rossella Gandaglia, Giorgio Yuen-Chun Teoh, Jeremy Moreno Sierra, Jesus Golozar, Asieh Bjartell, Anders Evans-Axelsson, Susan N'Dow, James Zong, Jihong Ribal, Maria J Roobol, Monique J Van Hemelrijck, Mieke Beyer, Katharina
Converge or Collide? Making Sense of a Plethora of Open Data Standards in Health Care. 🔗	Journal of medical Internet research	2024/04/09 16:53	Tsafnat, Guy Dunscombe, Rachel Gabriel, Davera Grieve, Grahame Reich, Christian
Research Protocol for an Observational Health Data Analysis on the Adverse Events of Systemic Treatment in Patients with Metastatic Hormone-sensitive Prostate Cancer: Big Data Analytics Using the PIONEER Platform. 🔗	European urology open science	2024/04/04 04:11	Rajwa, Pawel Borkowetz, Angelika Abbott, Thomas Alberti, Andrea Bjartell, Anders Brash, James T Campi, Riccardo Chitelli, Andrew Conover, Mitchell Constantinovici, Niculae Davies, Eleanor De Meulder, Bertrand Eid, Sherrine Gacci, Mauro Golozar, Asieh Hafeez, Haroon Haque, Samiul Hijazy, Ayman Hulsen, Tim Josefsson, Andreas Khalid, Sara Kolde, Raivo Kotik, Daniel Kurki, Samu Lambrecht, Mark Leung, Chi-Ho Moreno, Julia Nicoletti, Rossella Nieboer, Daan Oja, Marek Palanisamy, Soundarya Prinsen, Peter Reich, Christian Raffaele Resta, Giulio Ribal, Maria J Gomez Rivas, Juan Smith, Emma Snijder, Robert Steinbeisser, Carl Vandenberghe, Frederik Cornford, Philip Evans-Axelsson, Susan N'Dow, James Willemse, Peter-Paul M
Use of Recommended Neurodiagnostic Evaluation Among Patients With Drug-Resistant Epilepsy. 🔗	JAMA neurology	2024/04/01 16:08	Spotnitz, Matthew Ekanayake, Cameron D Ostropelets, Anna McKhann, Guy M Choi, Hyunmi Ottman, Ruth Neugut, Alfred I Hripscak, George Natarajan, Karthik Youngerman, Brett E
Increase transparency and reproducibility of real-world evidence in rare diseases through disease-specific Federated Data Networks. 🔗	Pharmacoepidemiology and drug safety	2024/04/01 02:03	van Baalen, Valerie Didden, Eva-Maria Rosenberg, Daniel Bardenheuer, Kristina van Speybroeck, Michel Brand, Monika
Correlation of Socioeconomic and Environmental Factors With Incidence of Crohn Disease in Children and Adolescents: Systematic Review and Meta-Regression. 🔗	JMIR public health and surveillance	2024/03/25 11:53	Weidner, Jens Glauche, Ingmar Manuwald, Ulf Kern, Ivana Reinecke, Ines Bathelt, Franziska Amin, Makan Dong, Fan Rothe, Ulrike Kugler, Joachim
Patterns of Comorbidities and Prescribing and Dispensing of Non-steroidal Anti-inflammatory Drugs (NSAIDs) Among Patients with Osteoarthritis in the USA: Real-World Study. 🔗	Drugs & aging	2024/03/23 12:21	Ide, Joshua Shoaibi, Azza Wagner, Kerstin Weinstein, Rachel Boyle, Kathleen E Myers, Andrew



OMOP CDM

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.



OMOP CDM

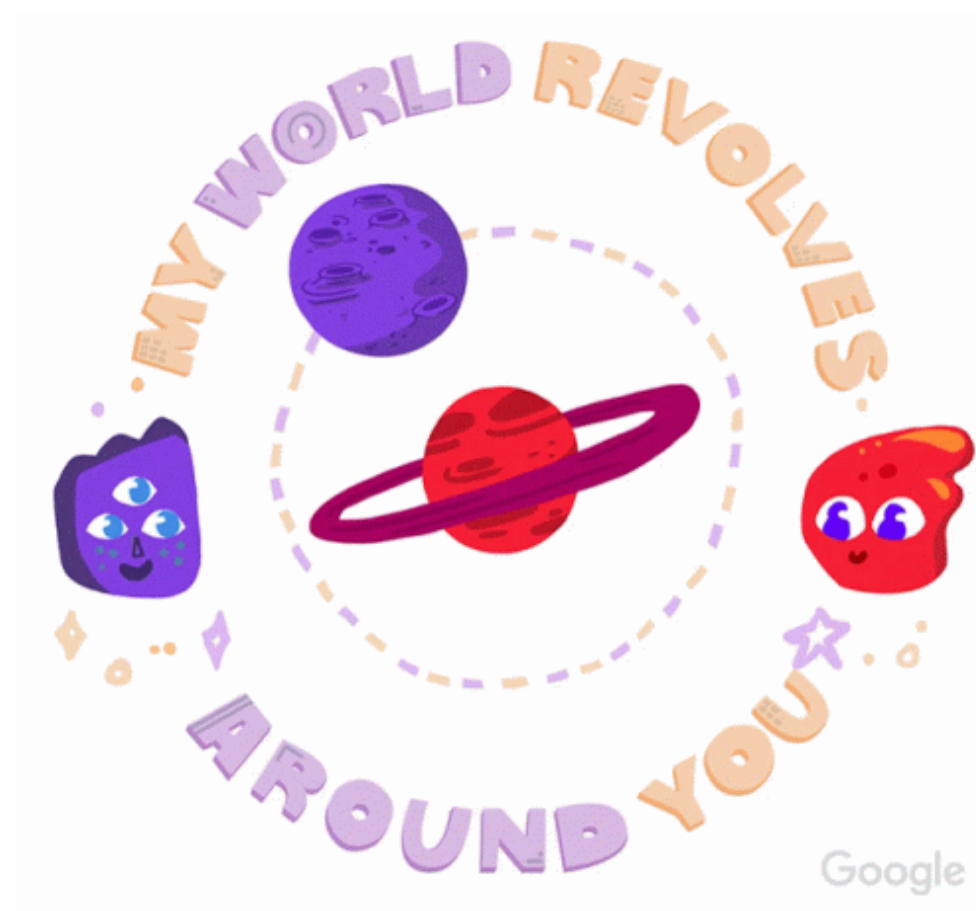
The OMOP CDM is a system of tables, vocabularies, and conventions that allow observational health data to be standardized.

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



Tables

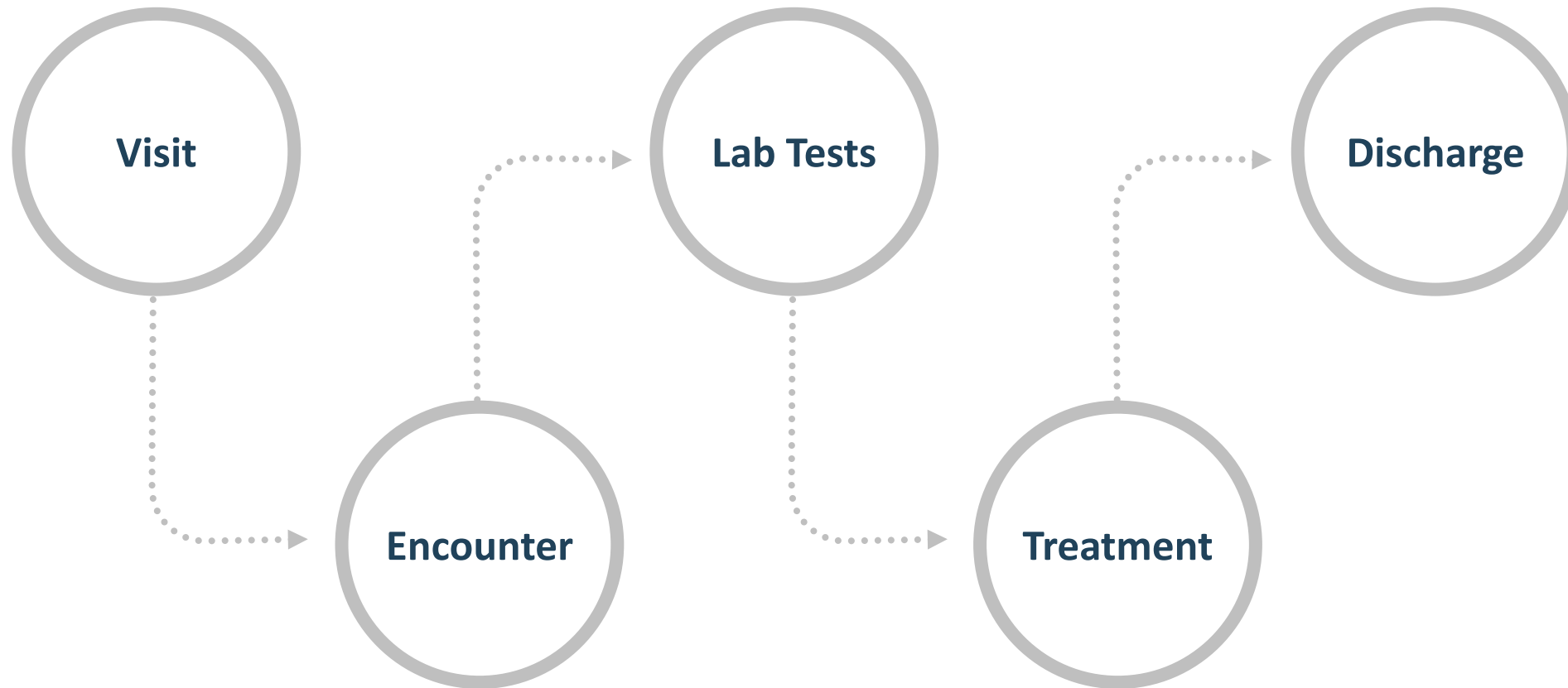
The OMOP CDM is a person-centric model





Tables

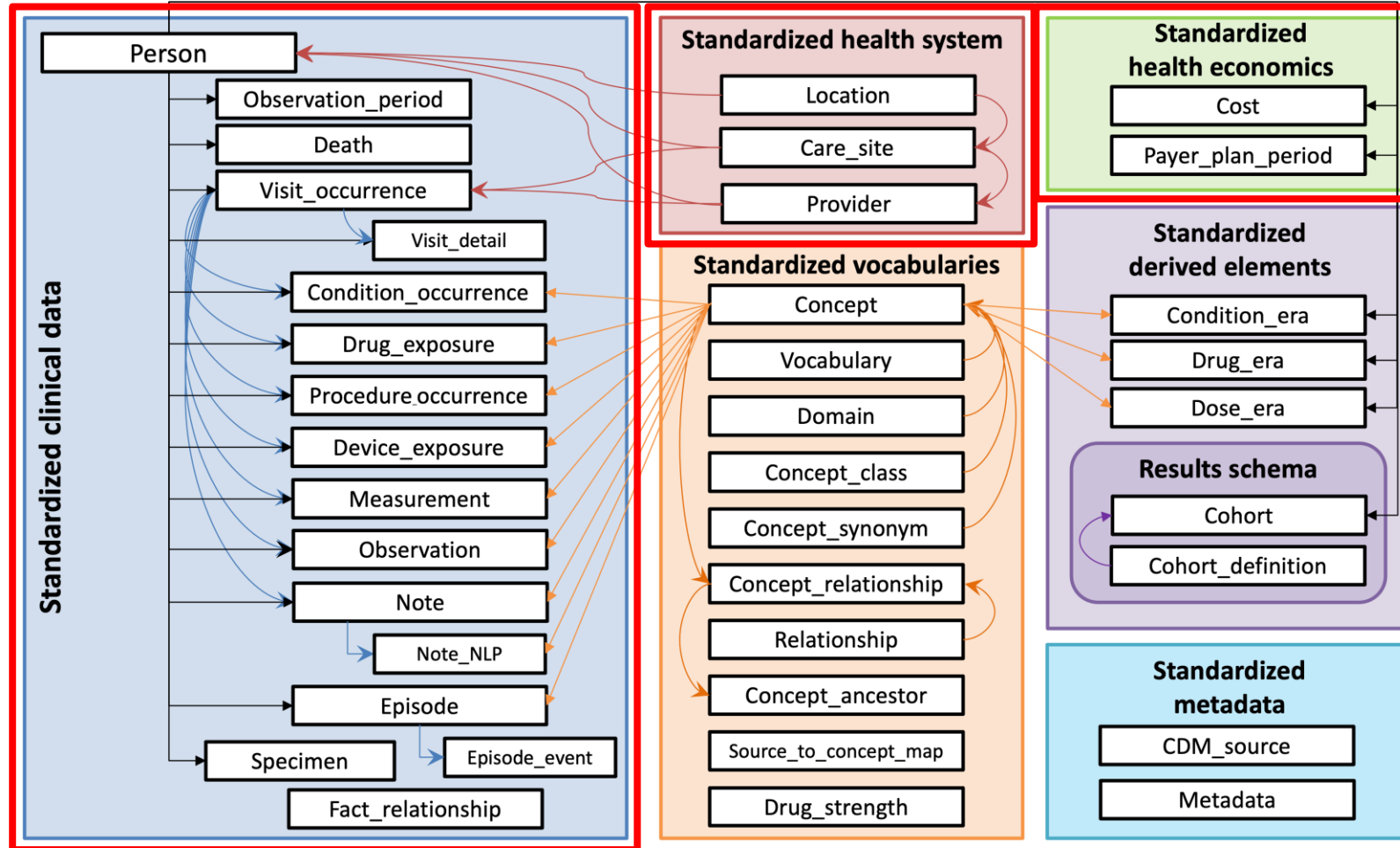
A typical patient journey within a healthcare system





Tables

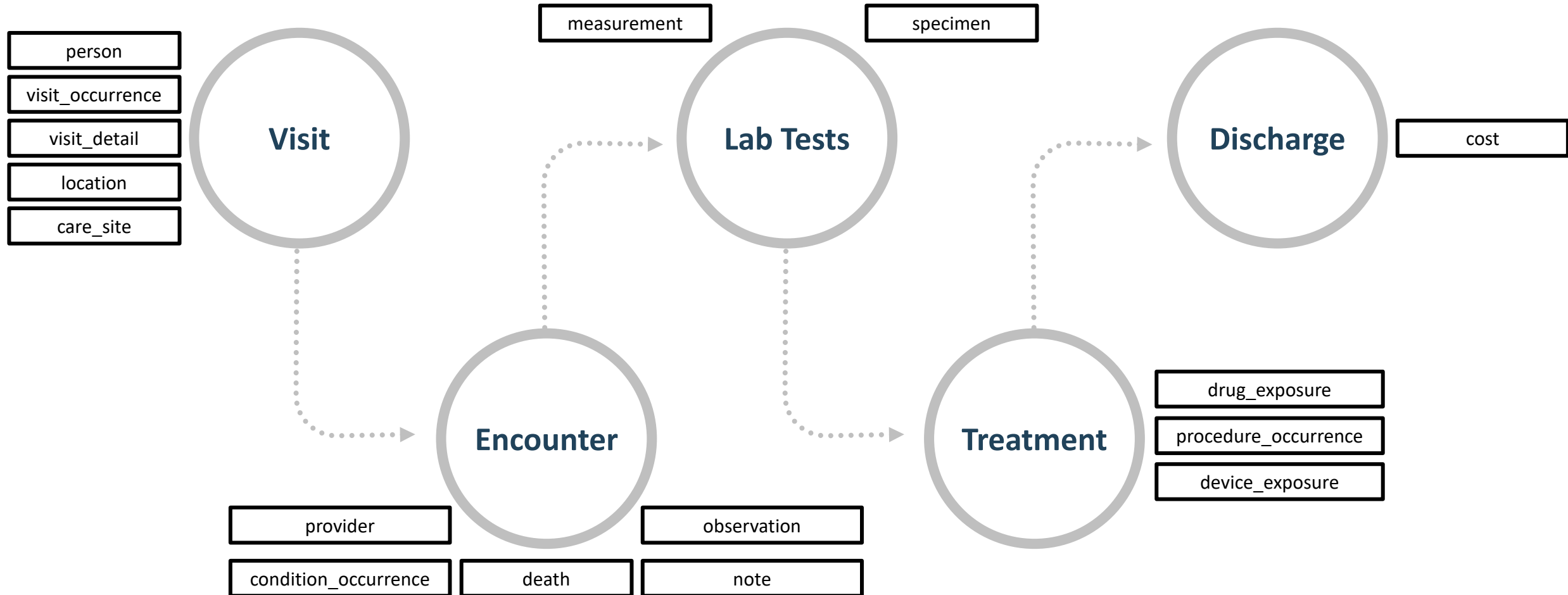
A typical patient journey within a healthcare system **into data**





Tables

A typical patient journey within a healthcare system **into data**





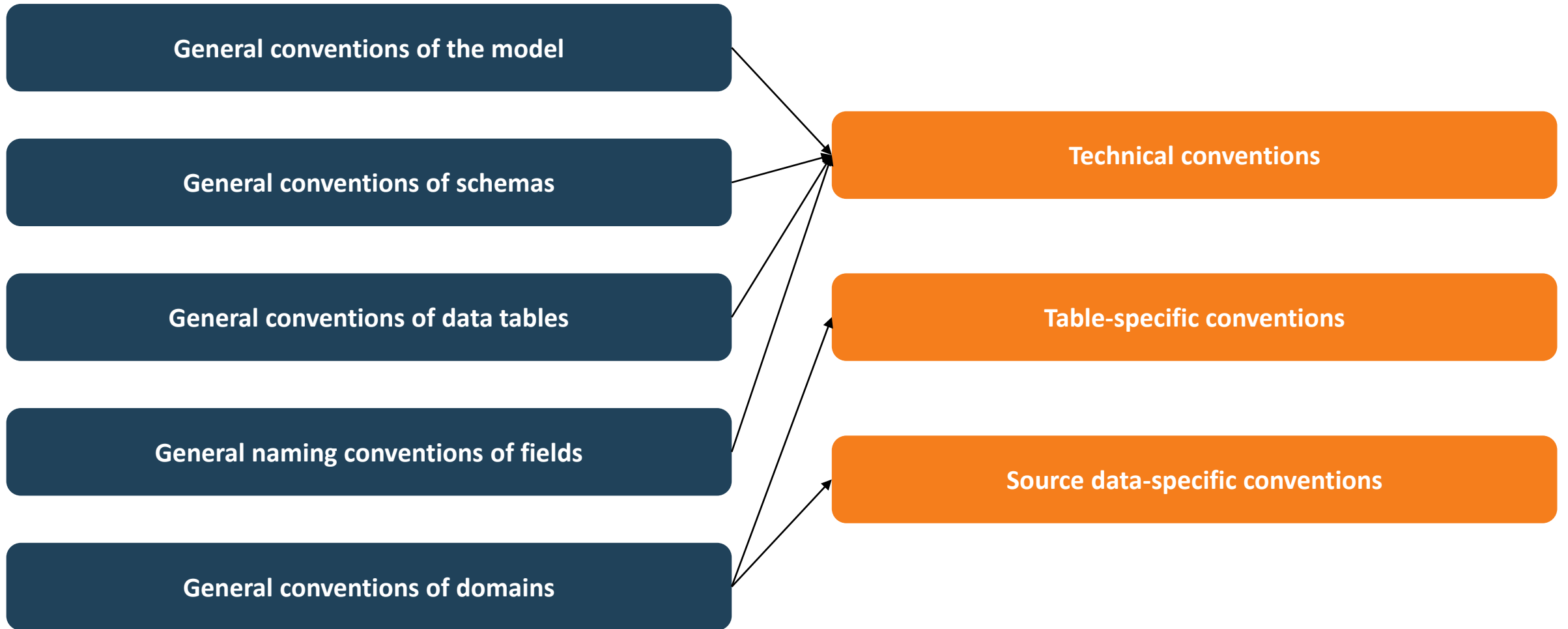
OMOP CDM

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





Technical Conventions

Fields

Variable names across all tables follow one convention:

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



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.

CDM Field	User Guide	ETL Conventions	Datatype	Required	Primary Key	Foreign Key	FK Table	FK Domain
person_id	It is assumed that every person with a different unique identifier is in fact a different person and should be treated independently.	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 autogenerated number.	integer	Yes	Yes	No		



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**



OMOP CDM

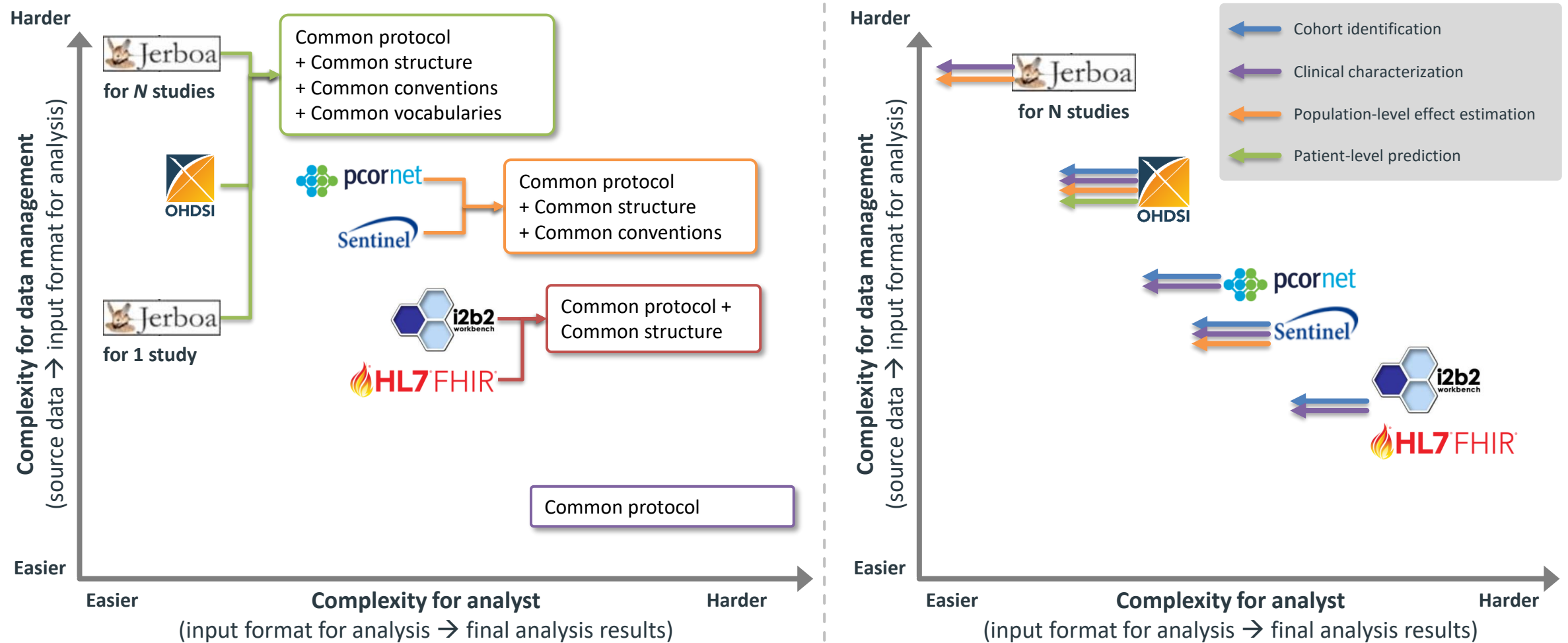
The OMOP CDM is a system of tables, vocabularies, and conventions that allow observational health data to be standardized.

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



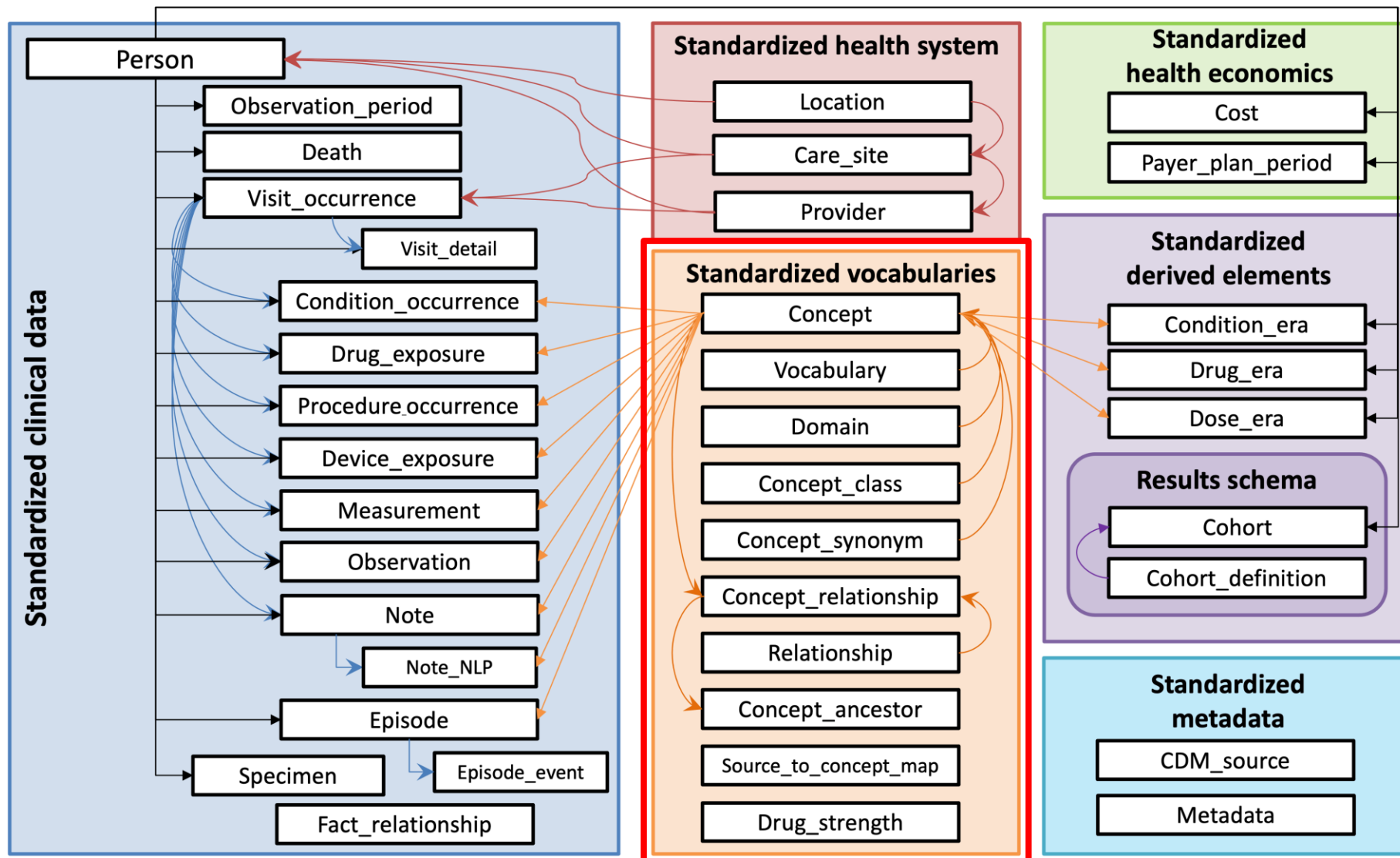
Comparison of common data models

Balancing trade-offs in data management vs. analysis complexity



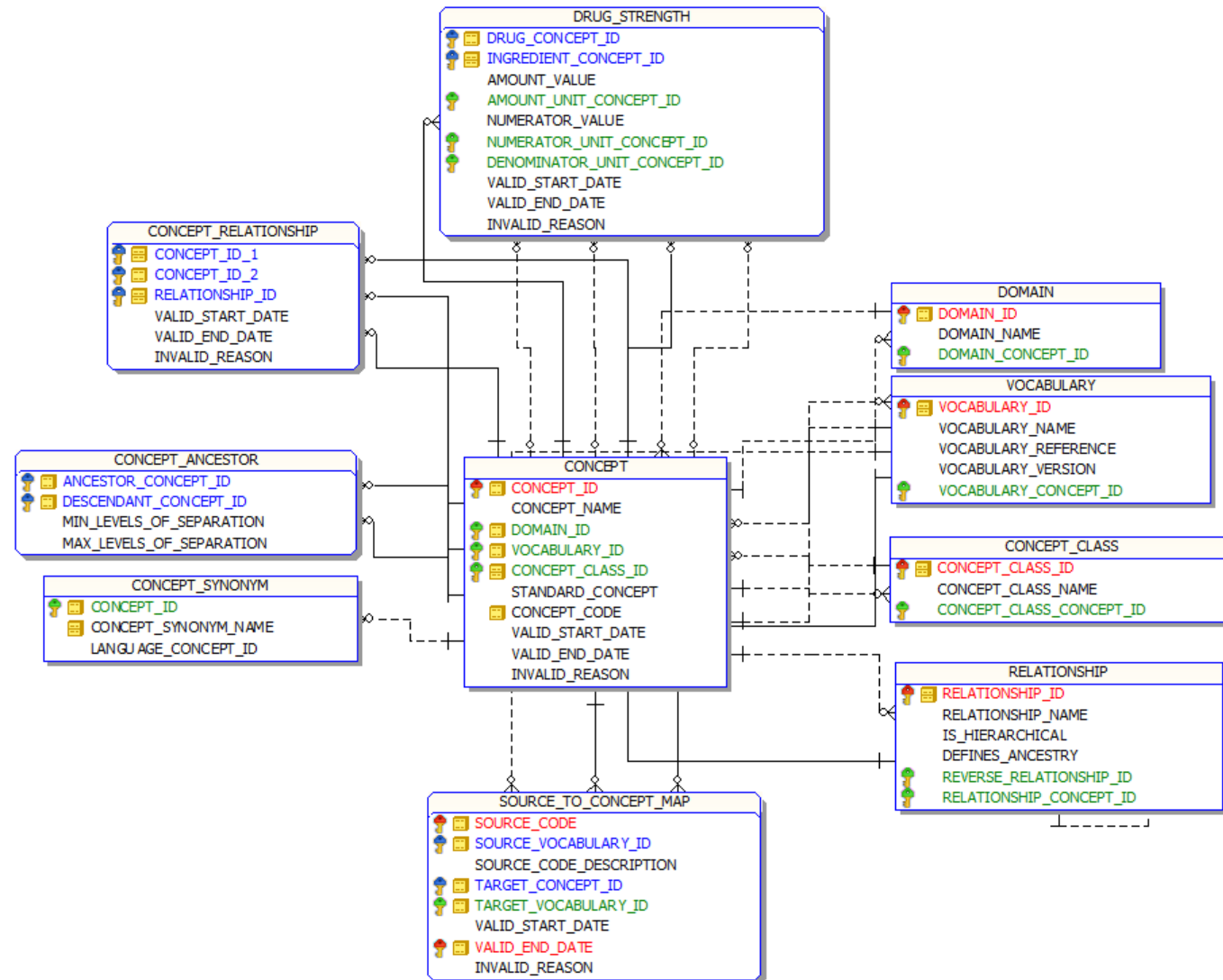


Vocabularies



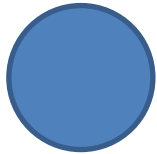


OMOP Standardized Vocabularies





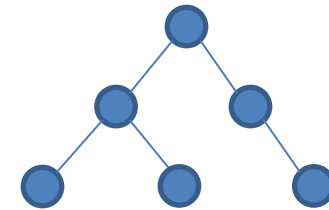
OMOP Standardized Vocabularies



All content: concepts in
concept



Direct relationships between
concepts in
concept_relationship

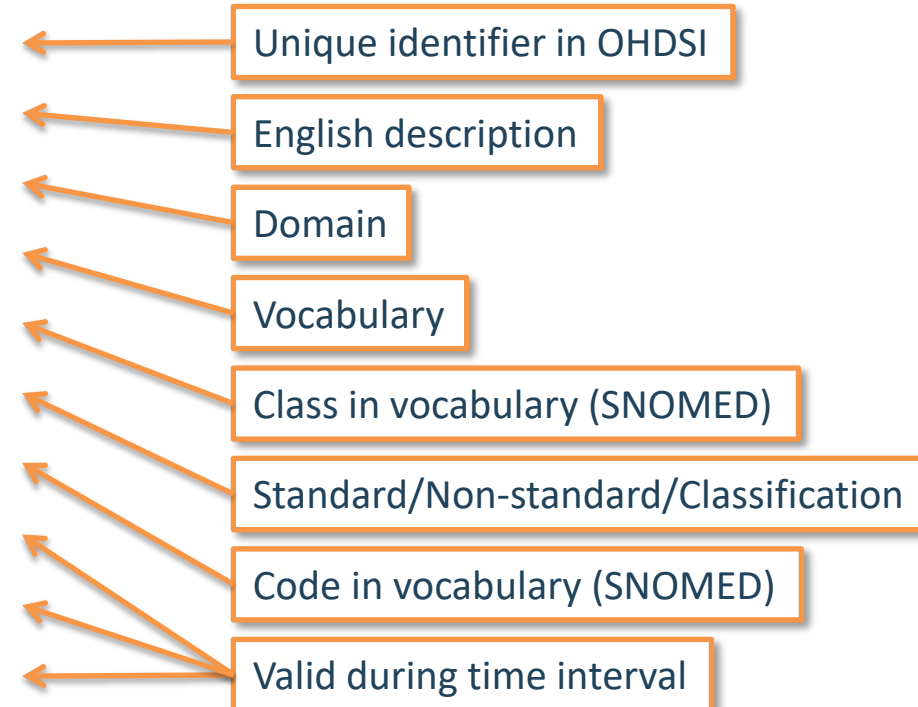


Multi-step hierarchical
relationships pre-processed
into
concept_ancestor



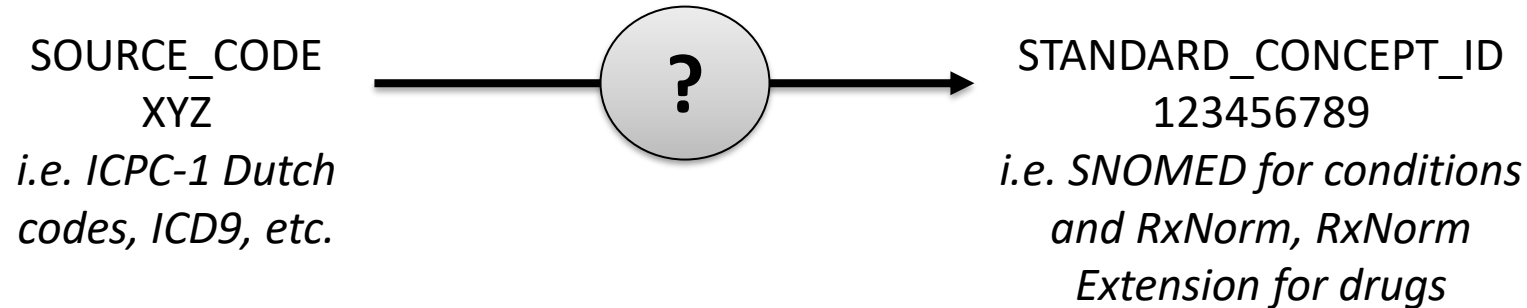
Concept

CONCEPT_ID	313217
CONCEPT_NAME	Atrial fibrillation
DOMAIN_ID	Condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	49436004
VALID_START_DATE	01-Jan-2002
VALID_END_DATE	31-Dec-2099
INVALID_REASON	





Mapping to OMOP Standardized Vocabularies

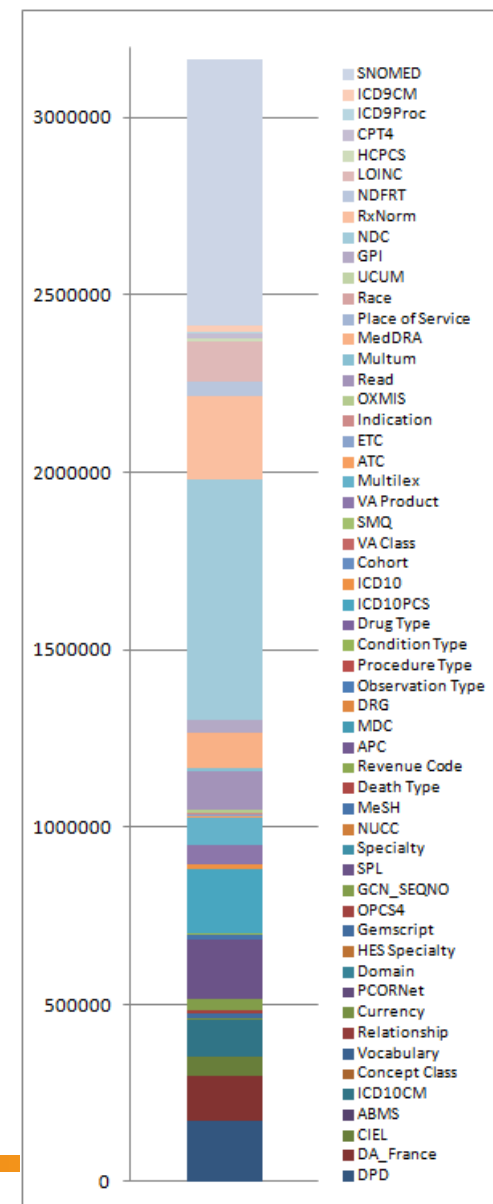


- What is standardized:
 - TABLE_CONCEPT_ID: standard concept the source code maps to, **used for analysis**
 - TABLE_SOURCE_CONCEPT_ID: concept representation of the source code, **helps maintain tie to raw data**
 - TABLE_SOURCE_VALUE: original source code as given in the source table, **helps to review 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 → SNOMED or NDC → 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/>

A screenshot of the ATHENA web application interface. The top navigation bar is green and contains the ATHENA logo, a search bar, and buttons for SEARCH, DOWNLOAD, LOGIN, and a help icon. The main content area is light gray and features a search section with a search bar containing the text 'aspirin' and a 'Search' button. Below the search bar are two numbered instructions: 1. Usage of quotation marks forces an exact-match search; 2. In case of a typo, or if there is a similar spelling of the word, the most similar result will be presented. The 'Explore domains' section is below, showing six domain cards: Drugs (5,391,909), Conditions (698,141), Procedures (737,007), Devices (493,782), Observations (585,559), and Measurements (368,765). Each card includes an icon representing the domain and a document icon with the count.

ATHENA SEARCH DOWNLOAD LOGIN ?

Search

aspirin **Search**

1. Usage of quotation marks forces an exact-match search
2. In case of a typo, or if there is a similar spelling of the word, the most similar result will be presented

Explore domains

Drugs 5,391,909	Conditions 698,141	Procedures 737,007
Devices 493,782	Observations 585,559	Measurements 368,765



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?
- 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?

1:1 mapping

Source domain = OMOP domain

1:1 mapping

Source domain \neq OMOP domain

n:1 mapping

1:n mapping



Thank you!

Mui Van Zandt

mui.vanzandt@iqvia.com

Gyeol Song

gyeol.song@iqvia.com