



THE DRUG DISCOVERY
ENGINE FOR **ALS**

OHDSI Training Day 1

CDM & Vocabularies

Danielle Boyce, MPH, DPA
Principal Investigator, RWE



Thank you!

- Parsa Mirhaji
- Rimma Belenkaya
- Chenyre Okpara
- Jessica Sommer
- Ferris Hussein
- Pavel Goriacko
- OHDSI Symposium Trainers

for sharing lecture materials
which were adapted for this
training

...and all of you!

HOUSEKEEPING

- We are recording so you can go back and watch later
- Please open up:
 - Athena: <https://athena.ohdsi.org/>
 - Databricks, if applicable
 - Code snippets here: https://boycelab.github.io/OHDSITrainTheTrainer-6WK/exercises/code_snippets/day-01-snippets/ -
 - CDM specs: <https://ohdsi.github.io/CommonDataModel/cdm54.html> -
 - We will drop these links in the chat
- Please put questions in the chat.
 - Check out Book of OHDSI and OHDSI Forums!
 - If we cannot get to your question today, we will follow up before next week.
- If you are having technical problems with Databricks, please follow up with Jessica and Rimma after the training.

Schedule

Time	Session Title
9:30 am – 9:45 am	Welcome and Introduction
9:45 am – 10:00 am	Value Proposition
10:00 am – 10:45 am	Overview of OMOP CDM
10:45 am – 11:30 am	Overview of Vocabulary and Athena
11:30 am – 11:45 am	Break
11:45 am – 12:45 pm	Hands-On Exercises
12:45 pm – 1:00 pm	Q & A / Homework for Next Week

Welcome and Introduction

Rimma

After today's training, you will know

1. What is OMOP and OHDSI?
2. How does the standardized vocabulary work?
3. How do I find codes and concepts?
4. How do I navigate the vocabulary hierarchy?
5. What is the OMOP CDM?
6. How to use the OMOP CDM
7. How to navigate the CDM and vocabularies using SQL and Athena.

OMOP: Observational Medical Outcomes Partnership

- Originated as a public–private research initiative (FDA, academia, industry)
- Created the **OMOP Common Data Model (CDM)**

OHDSI: Observational Health Data Sciences and Informatics

- A **global, open-science community** built on OMOP's foundation
- Uses the OMOP CDM to enable **collaborative analytics** across institutions
- Expands the mission to produce **real-world evidence at scale**
- Coordinating center: **Columbia University**

OMOP vs OHDSI

OMOP = the Common Data Model (CDM)

OHDSI = the community and tools that use OMOP CDM

The Value Proposition

The Way We Were: The Request - 2010

“I'm particularly interested in the comparisons between those patients with the [redacted] genotype and those without and comparison with the overall cohort. Demographics, functional class, six-minute walk distance, hemodynamics, pulmonary function, serology (antibodies), and survival.”

-Dr. X

E-mail 2 of...200??

- 1=Asian
- 2=Black
- 3=White
- 4=Black
- 5=White
- 6=Decline to Answer/Unknown
- 7=Other
- 8=A (I have no idea what race this is for)
- 9=M (Again, I have no idea what race this is for)
- 10=Unknown
- 11=O>All Other Races



Text View

Graphical View

JSON

SQL

Copy to clipboard



Copy

Next

no
Latino

[PhenPheb] Pulmonary arterial hypertension from Schikowski JAHA 2022

From <https://www.ahajournals.org/doi/10.1161/JAHA.122.026620> : "Individuals were identified as having PAH by ICD-9 code 416.0 and ICD-10 code I27.0 between January 1, 2015, and September 30, 2020. To reduce the risk of misclassification, we used a 2-component algorithm requiring all individuals to have both diagnostic coding for PAH and use of disease-specific treatment with pulmonary vasodilators (ERAs, PDE5, prostanoids, or sGCS)."

Cohort Entry Events

People may enter the cohort when observing any of the following:

1. condition occurrences of 'Pulmonary arterial hypertension diagnosis from Schikowski JAHA 2022'.
2. drug exposures of 'PAH treatments - ERAs, PDE5, prostanoids, or sGCS'.

Limit cohort entry events to the earliest event per person.

Inclusion Criteria

1. has ≥ 1 Rx and ≥ 1 Dx

Entry events with all of the following criteria:

1. having at least 1 condition occurrence of 'Pulmonary arterial hypertension diagnosis from Schikowski JAHA 2022', starting between 0 days before and all days after cohort entry start date.
2. having at least 1 drug exposure of 'PAH treatments - ERAs, PDE5, prostanoids, or sGCS', starting between 0 days before and all days after cohort entry start date.

Cohort Exit

The person exits the cohort at the end of continuous observation.

Cohort Eras

Remaining events will be combined into cohort eras if they are within 0 days of each other.

Informatics Grand Rounds



JOHNS HOPKINS
M E D I C I N E

The Road to Reproducible Research on Medical Records: Welcome to the JH OHDSI Community

Presented by: Paul Nagy, PhD, FSIIM

Play [K]

Activity Director: Dr. Harold Lehmann, MD, PhD
September 9th, 2021

Dr. Paul Nagy & the OHDSI Community Informatics Grand Rounds 9/9/21



Johns Hopkins Medicine
479K subscribers

Subscribe

18

18

Share

Download

...

Biomedical Informatics & Data Science Grand Rounds



**Lived Experience in Observational Research:
Data Scientists as Patient Advocates**

Presented by: Danielle Boyce, MPH, DPA

Activity Director: Paul Nagy, PhD
April 13th, 2023

**Dr. Danielle Boyce & the Lived Experience in Observational Research Informatics Grand
Rounds 4/13/23**



Johns Hopkins Medicine
479K subscribers

Subscribe



Share

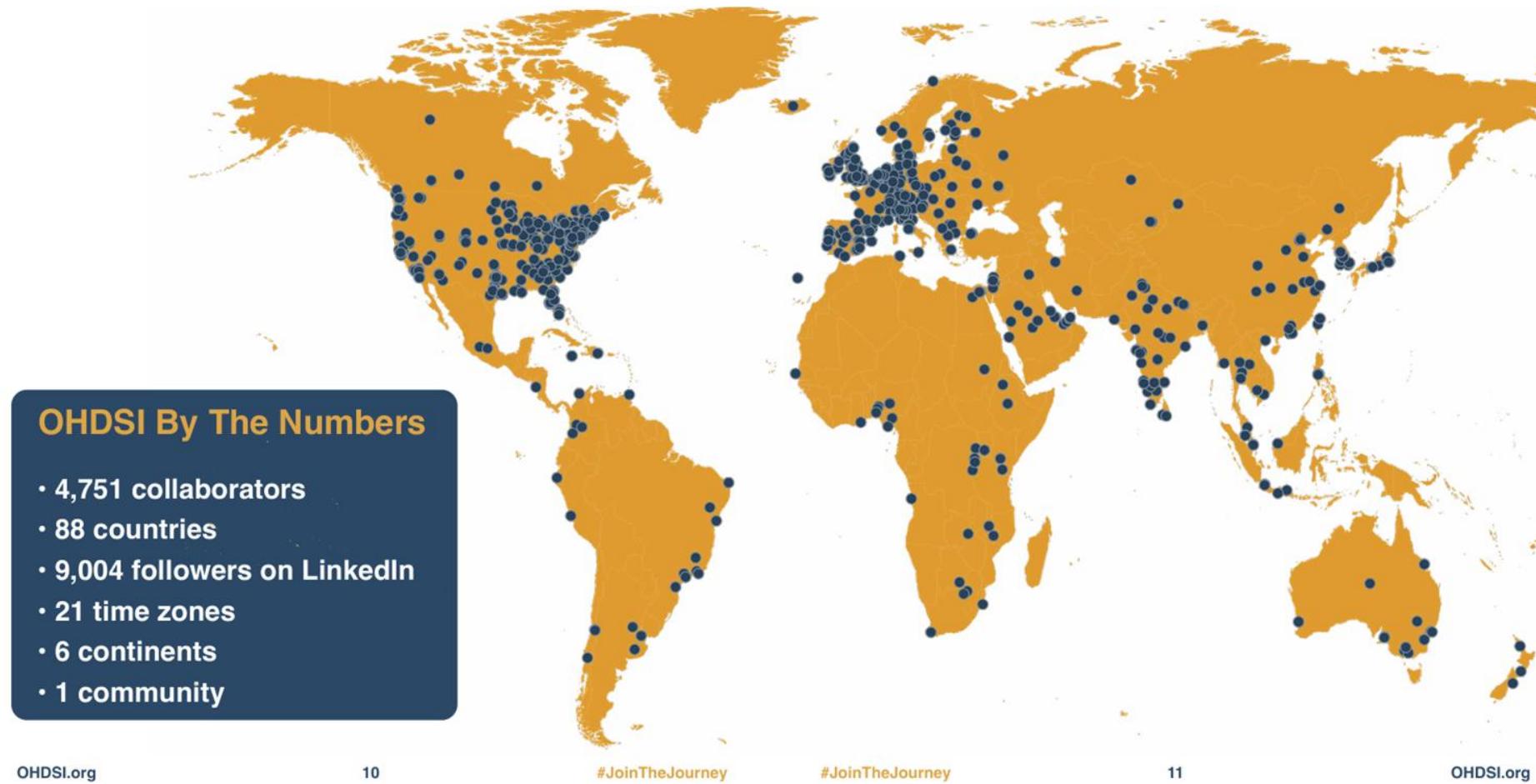
Download

...

No One Does OHDSI Alone

- It takes a village
 - Clinicians
 - Researchers
 - Engineers
 - Analysts
 - Data scientists
 - Statisticians
- I don't know everything about OMOP, and that's OK.
- I will teach you not only about OMOP but "How to OHDSI."

OHDSI: A Global Community



We are in Good Company!



Our Programs

Research Funding ▾

News & Media ▲

New Program Planning ▼

Evaluation & Assessment ▼

About Us ▾

Bridge to Artificial Intelligence (Bridge2AI)

Watch the Video ►

[Home](#) » Bridge To Artificial Intelligence (Bridge2AI)

Bridge to Artificial Intelligence (Bridge2AI)

For the Public

Health Relevance

Program Snapshot

The NIH Common Fund's **Bridge to Artificial Intelligence (Bridge2AI)** program will propel biomedical research forward by setting the stage for widespread adoption of artificial intelligence (AI) that tackles complex biomedical challenges beyond human intuition. A key step in this process is generating new "flagship" data sets and best practices for machine learning (ML) analysis. The biomedical research community

Program Updates

The Bridge2AI Voice A project shares the first round of voice data

My Big Break: CURE ID

- J Clin Transl Sci. 2024 Apr 3;8(Suppl 1):20. doi: [10.1017/cts.2024.77](https://doi.org/10.1017/cts.2024.77) ↗

76 Lessons learned during implementation of OMOP common data model across multiple health systems

William Garneau¹, Benjamin Martin², Kelly Gebo³, Paul Nagy³, Johns Hopkins¹, Danielle Boyce², Michael Cook², Matthew Robinson³

- Author information
- Article notes
- Copyright and License information

PMCID: PMC11023828

STARDUSTT Framework

Secure Data — De-identification, privacy protection, and regulatory compliance.

Technology Awareness — Familiarity with Docker, SQL, and OHDSI tools.

Active Clinician Representation — Inclusion of clinicians in decision-making and workflow mapping.

Relational Databases — Understanding schema, joins, and OMOP table relationships.

Data Quality — Processes for completeness, accuracy, and plausibility.

Utilization of Git — Version control, collaboration, and issue tracking.

Standardized Vocabularies — e.g. SNOMED, LOINC, RxNorm, and use of Athena for mapping.

Training & Documentation — SOPs, MOPs, and curated learning paths.

Translation & Communication — Plain-language messaging, feedback loops, and artifacts for different personas.



ALS TDI Real World Evidence Resources

Home New to RWE? Start Here ALS TDI OMOP Data Set STARDUSTT Approach OMOP/OHDSI Resources Select Publications Authors

ALS TDI OMOP Data Set



Release Notes: Version 0.1.0

The first release of the [ALS TDI ARC Study](#), mapped to the [Observational Medical Outcomes Partnership Common Data Model \(OMOP CDM\)](#), restructures a subset of the ARC Natural History Study into the OMOP CDM structure and maps a subset to standardized vocabularies.

This is part of a larger harmonization effort with [Answer ALS](#) and the [Critical Path Institute](#).

Note about EHR data: We are actively working to include electronic health record (EHR) data in future releases. Tools in use for EHR integration:

- [CDAtransformer](#) — Parse C-CDA and FHIR files into structured tables.
- [RWDExchange](#) — Evaluate exchangeability of real-world data (EHR, registries) for external comparator trials.



requesting data, follow the link below.

[Learn More](#)

2,563

Participants

1,491

Whole Genome

1,039

Epigenomics

1,217

Transcriptomics

1,209

Proteomics

ENGINE FOR ALS

Hi, I'm Lou! Ask me anything...

**THERAPY DEVELOPMENT
INSTITUTE**

NEWS & STORIES

Pat Dolan: Mapping the future of ALS Care, Cures, and Community





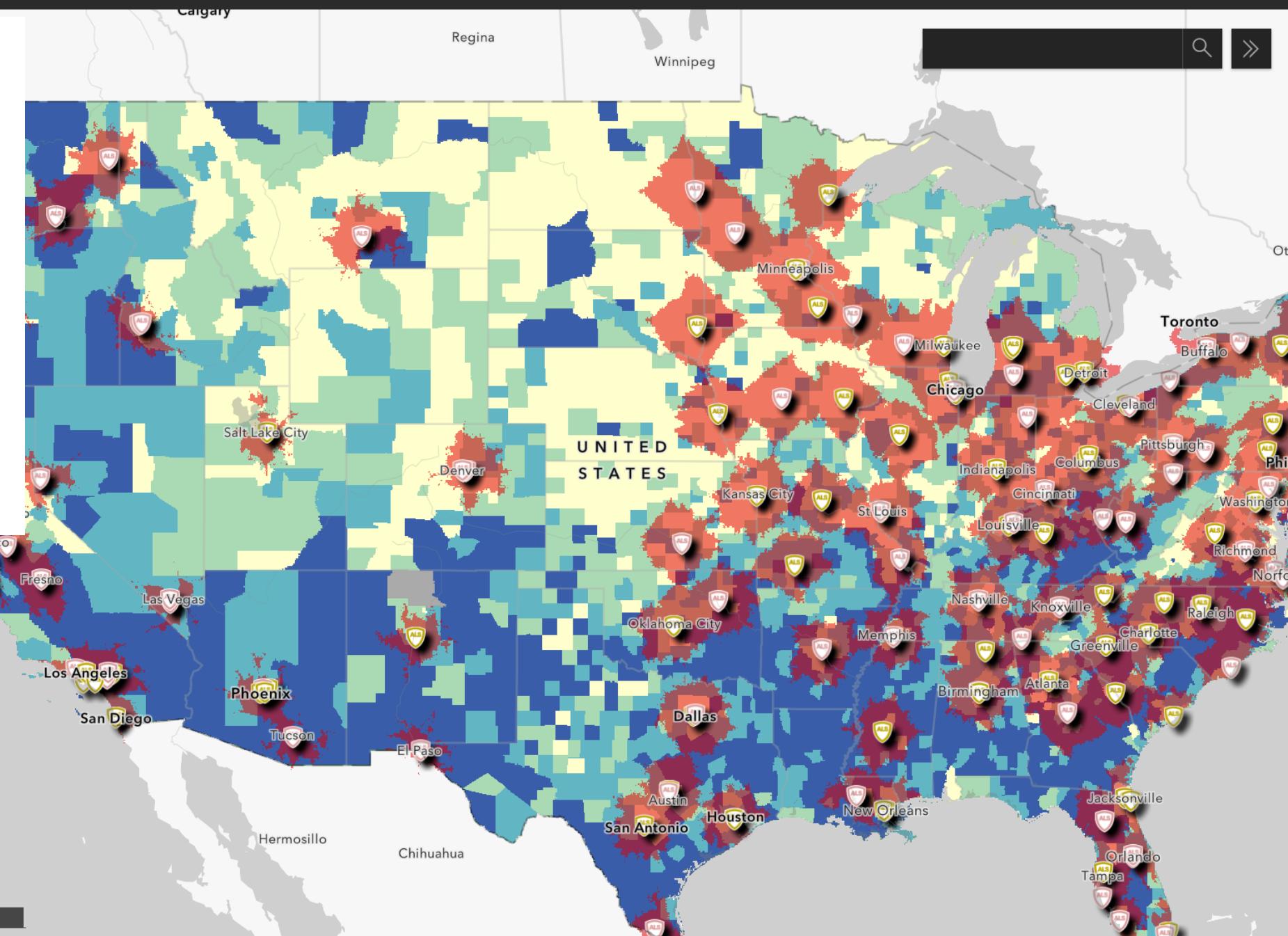
Individuals over 40 and military
Veterans are 2x-10x risk of
ALS. [NIH & Military Medicine](#)

Source

Social Vulnerability Index (SVI)
uses U.S. Census 2018 [data](#)
ALS clinics data from [ALSA](#)
(updated 08/2022)

ALS Geospatial Hub

Discover more [maps](#) for the ALS
community.





WHAT WE DO PROGRAMS IMPACT EUROPE EVENTS MEDIA

TOOLS & PLATFORMS

GIVE



PRESS RELEASES

C-Path Receives Largest Data Transfer to Date from UK's National Neonatal Research Database

ILAE Big Data Commission

ILAE International League Against Epilepsy

About ILAE Guidelines Congresses Regions & Countries Journals

Common Data Model Task Force



Colin Josephson
Chair



Danielle Boyce
Member



Spiros Denaxas
Member



James Mitchell
Member



Samuel Wiebe
Member



Brett Youngerman
Member

UNIVERSITY OF CALGARY

CENTRE FOR HEALTH INFORMATICS



OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



JOHNS HOPKINS
SCHOOL OF MEDICINE

Biomedical Informatics & Data Science (BIDS)

OHDSI = Team Science

- Interdisciplinary by design
- No one knows it all
- Collaboration drives progress
- We learn from each other
- Open community, open science
- Every skill and perspective matters
- Mentoring emphasized

[Home](#)

OHDSI/OMOP Train-the-Trainer

Learn how to navigate the OHDSI ecosystem—from data model to advanced analytics.

This multi-day training series provides participants with a comprehensive, hands-on introduction to the OHDSI ecosystem and its suite of analytic tools. Through guided sessions, attendees will gain familiarity with the OMOP Common Data Model (CDM), standardized vocabularies, and data quality principles before progressing to cohort development, data extraction, and advanced analytics in ATLAS and HADES. The program is designed to support participants at all levels—from those new to OHDSI tools to analysts and statisticians looking to deepen their technical expertise. Optional advanced sessions are available for those interested in treatment pathway analysis and applying OHDSI's R-based HADES framework for research and publication.



How to use this site

This site is designed as a **living companion** to the OHDSI/OMOP *Train-the-Trainer* program.

Each week builds on the last — start with the earliest modules and progress through the sequence.

How to get the most out of it: - **Follow the modules in order** — each one includes objectives, slides, and hands-on exercises.

- **Explore the “Exercises” section** for SQL labs, Atlas walkthroughs, and vocabulary practice.
- **Check “Resources”** for curated references, external tools, and reading materials.

The e

[Contact Us](#)[Submit RFI/RFP](#)[Trust Center](#)[Ethics Hotline \(report\)](#)

LIFE SCIENCES

HEALTHCARE

RESOURCES ▾

ABOUT

Cosmos

TRIAL DESIGN

Power
With
Hub

Using Slicer Dicer to Identify Participants

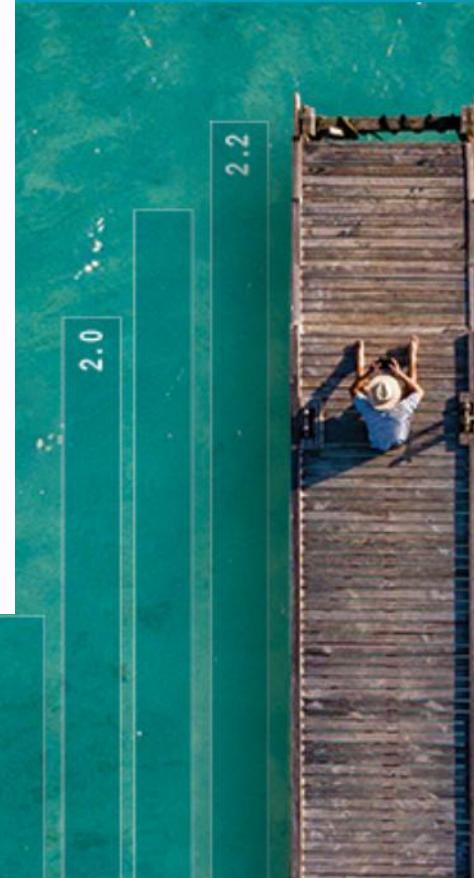
In addition to DataCore's three services, Slicer Dicer is a self-service cohort exploration tool embedded in Epic that is available to researchers. This tool allows researchers to identify patients based on criteria entered into the user interface without needing to write code. The user interface displays counts of patients that meet specified criteria and allows for these aggregate counts to be displayed in several ways.

If IRB approval is obtained, researchers can also use Slicer Dicer to identify a cohort of interest and request the list of patients with contact information to be used for trial recruitment.

We envision a future where data and intelligence accelerate innovation across the global healthcare ecosystem.

[Explore TriNetX LIVE™](#)

WE GENERATION





X 1 selected

Summarize this file



Name	Owner	Date modified	File type
874_Nephrectomy_adults_inpt_stay_no_ED.json	me	Oct 22	25
372_Otitis_media.sql	me	Oct 22	14
244_Dizziness_or_giddiness_including_motion_sickness_.json	me	Oct 22	13
803_Fascial_dehiscence_and_evisceration.sql	me	Oct 22	13
84_SARS-CoV-2_test_1pos_or_COVID-19_diagnosis_with_.json	me	Oct 22	19
395_Dysmenorrhea.json	me	Oct 22	21
1064_New_users_of_Cephalosporin_systemetic_nested_in_.json	me	Oct 22	33
1223_Birdshot_chorioretinitis.json	me	Oct 22	10
782_Chronic_Thromboembolic_Pulmonary_Hypertension....json	me	Oct 22	37

The Common Data Model (CDM)

OMOP CDM Principles

- OMOP model is an information model
 - Vocabulary(Conceptual) and Data Model are blended
 - Domain--oriented concepts
- Patient centric
- Accommodates data from various sources
- Preserves data provenance
- Extendable
- Evolving

Current Support for CDM v5.4

person

observation_period

visit_occurrence

visit_detail

condition_occurrence

drug_exposure

procedure_occurrence

device_exposure

measurement

observation

death

note

note_nlp

the provider who implanted
the device.

put here based on what
is available this may or
may not be different
than the provider
associated with the
overall
VISIT_OCCURRENCE
record.

visit_occurrence_id	The Visit during which the device was prescribed or given.	To populate this field device exposures must be explicitly initiated in the visit.	integer	No	No	Yes
visit_detail_id	The Visit Detail during which the device was prescribed or given.	To populate this field device exposures must be explicitly initiated in the visit detail record.	integer	No	No	Yes
device_source_value	This field houses the verbatim value from the source data representing the device exposure that occurred. For example, this could be an NDC or Gemscript code.	This code is mapped to a Standard Device Concept in the Standardized Vocabularies and the original code is stored here for reference.	varchar(50)	No	No	No

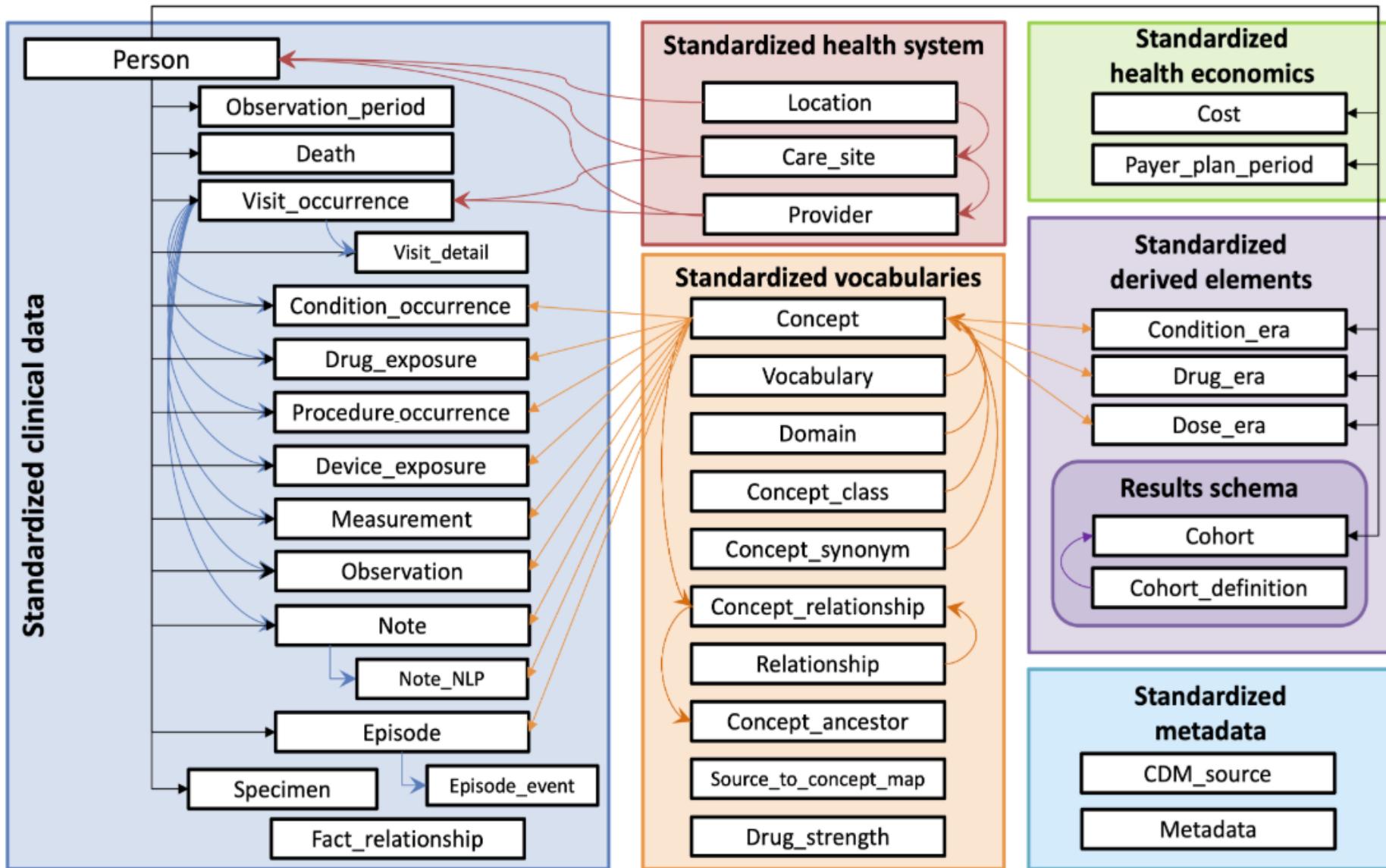
OMOP CDM Standard Domain Features

Feature	Description & Purpose	Field Name Convention	Example
Patient centric	Every domain table has patient identifier . Patient data can be retrieved independently from other domains.	person_id	person_id 123
Unique domain identifiers	Ever domain table has a unique primary key to identify domain entities.	<entity>_id	condition_occurrence_id 470985
Standard concept from a respective vocabulary domain	Integration with the Vocabulary. Foreign key into the Standard Vocabulary for Standard Concept .	<entity>_concept_id	condition_concept_id 313217 (SNOMED “Atrial Fibrillation”)
Source value	Provenance. Verbatim information from the source data, not to be used by any standard analytics.	<entity>_source_value	condition_source_value 427.31 (ICD9CM “Atrial Fibrillation”)
Source concept from a respective vocabulary domain	Provenance. Foreign key into Standard Vocabulary for Source Concept .	<entity>_source_concept_id	condition_source_concept_id 44821957 (ICD9CM “Atrial Fibrillation”)
Source type	Provenance. Foreign key into Vocabulary for the origin of the data .	<entity>_type_concept_id	condition_type_concept_id 38000199 (“Inpatient header – primary”)

Standard Variable Name Conventions

OMOP Common Data Model	Background	Conventions	CDM Versions	CDM Additions	How to	Help	
CONDITION_START_DATE and the CONDITION_END_DATE.							
Current Support for CDM v5.4							
person							
observation_period							
visit_occurrence							
visit_detail							
condition_occurrence		condition_end_datetime					
drug_exposure		condition_type_concept_id	This field can be used to determine the provenance of the Condition record, as in whether the condition was from an EHR system, insurance claim, registry, or other sources.	Choose the CONDITION_TYPE_CONCEPT_ID that best represents the provenance of the record. Accepted Concepts . A more detailed explanation of each Type Concept can be found on the vocabulary wiki .	integer	Yes	No
procedure_occurrence		condition_status_concept_id	This concept represents the point during the visit	Choose the Concept in the Condition Status domain that best	integer	No	No
device_exposure							
<entity>_id	Unique identifiers for entities (row numbers, or IDs imported from source)	person_id 1234567 visit_occurrence_id 7654321 could be a person identifier or an autogenerated number by the CDM builder					

CDM Version 5.4 Key Domains



PERSON

Person	
💡	person_id
	gender_concept_id
	year_of_birth
	month_of_birth
	day_of_birth
	birth_datetime
	race_concept_id
	ethnicity_concept_id
	location_id
	provider_id
	care_site_id
	person_source_value
	gender_source_value
	gender_source_concept_id
	race_source_value
	race_source_concept_id
	ethnicity_source_value
	ethnicity_source_concept_id

- Need to create one unique record per person
- No history of location/demographics: need to select latest available
- Year of birth required...day/month optional
- Foreign key to the LOCATION, PROVIDER, and CARE_SITE table that contains one record

OBSERVATION_PERIOD

Observation_Period	
!	observation_period_id
	person_id
	observation_period_start_date
	observation_period_end_date
	period_type_concept_id

- Spans of time where data source has capture of data
- One person may have multiple periods if there is interruption in data capture
- Required to run analytical methods
- Challenge: determine observation periods based on the source data

VISIT OCCURRENCE

OMOP Common Data Model



Background ▾

Conventions ▾

CDM Versions ▾

+ CDM Additions ▾

? How to ▾

Support ▾



Current Support for CDM v5.4

person
observation_period
visit_occurrence
visit_detail
condition_occurrence
drug_exposure
procedure_occurrence
device_exposure
measurement
observation
death
note
note_nlp
specimen
fact_relationship
location
care_site
provider
payer_plan_period
cost
drug_era
dose_era
condition_era
episode

absence of inpatient procedures/drugs. - Non-hospital institution Visits: Particularly for claims data, if end dates are not provided assume the visit is for the duration of month that it occurs. For Inpatient Visits ongoing at the date of ETL, put date of processing the data into visit_end_datetime and visit_type_concept_id with 32220 "Still patient" to identify the visit as incomplete. - All other Visits: visit_end_datetime = visit_start_datetime. If this is a one-day visit the end date should match the start date.

visit_end_datetime	If a Person is still an inpatient in the hospital at the time of the data extract and does not have a visit_end_datetime, then set the visit_end_datetime to the datetime of the data pull.	If no time is given for the end date of a visit, set it to midnight (00:00:000).	datetime	No	No	No	
visit_type_concept_id	Use this field to understand the provenance of the visit record, or where the record comes from.	Populate this field based on the provenance of the visit record, as in whether it came from an EHR record or billing claim. Accepted Concepts . A more detailed explanation of each Type Concept can be found on the vocabulary wiki .	integer	Yes	No	Yes	CONCEPT
provider_id	There will only be one provider per visit record and the ETL document should clearly state how they were chosen (attending, admitting, etc.). If there are multiple providers associated with a visit in the	If there are multiple providers associated with a visit, you will need to choose which one to put here. The additional providers can be stored in the VISIT_DETAIL table.	integer	No	No	Yes	PROVIDER

CONDITION_OCCURRENCE

Condition_Occurrence	
!	condition_occurrence_id
	person_id
	condition_concept_id
	condition_start_date
	condition_start_datetime
	condition_end_date
	condition_end_datetime
	condition_type_concept_id
	stop_reason
	provider_id
	visit_occurrence_id
	condition_source_value
	condition_source_concept_id
	condition_status_source_value
	condition_status_concept_id

- **Records suggesting the presence of a disease or medical condition stated as a diagnosis, a sign or a symptom**
- **Examples:**
 - Billing diagnosis
 - Problem list

DRUG_EXPOSURE

Drug_Exposure	
!	drug_exposure_id
	person_id
	drug_concept_id
	drug_exposure_start_date
	drug_exposure_start_datetime
	drug_exposure_end_date
	drug_exposure_end_datetime
	verbatim_end_date
	drug_type_concept_id
	stop_reason
	refills
	quantity
	days_supply
	sig
	route_concept_id
	lot_number
	provider_id
	visit_occurrence_id
	drug_source_value
	drug_source_concept_id
	route_source_value
	dose_unit_source_value

- **Records about the utilization of a drug when ingested or otherwise introduced into the body**
- **Data sources:**
 - Pharmacy dispensing
 - Prescriptions written
 - Medication history
- **If drug is represented as a procedure, the OMOP Vocabulary realigns as drug**

PROCEDURE_OCCURRENCE

Procedure_Occurrence	
procedure_occurrence_id	key
person_id	
procedure_concept_id	
procedure_date	
procedure_datetime	
procedure_type_concept_id	
modifier_concept_id	
quantity	
provider_id	
visit_occurrence_id	
procedure_source_value	
procedure_source_concept_id	
modifier_source_value	

- Contains records of activities or processes ordered by, or carried out by, a healthcare provider on the patient to have a diagnostic or therapeutic purpose
- Vocabularies include CPT-4, HCPCS, ICD-9 Procedures, ICD-10 Procedures, LOINC, SNOMED
- Procedures have the least standardized vocabularies that causes some redundancy

MEASUREMENT

Measurement	
measurement_id	PK
person_id	
measurement_concept_id	
measurement_date	
measurement_datetime	
measurement_type_concept_id	
operator_concept_id	
value_as_number	
value_as_concept_id	
unit_concept_id	
range_low	
range_high	
provider_id	
visit_occurrence_id	
measurement_source_value	
measurement_source_concept_id	
unit_source_value	
value_source_value	

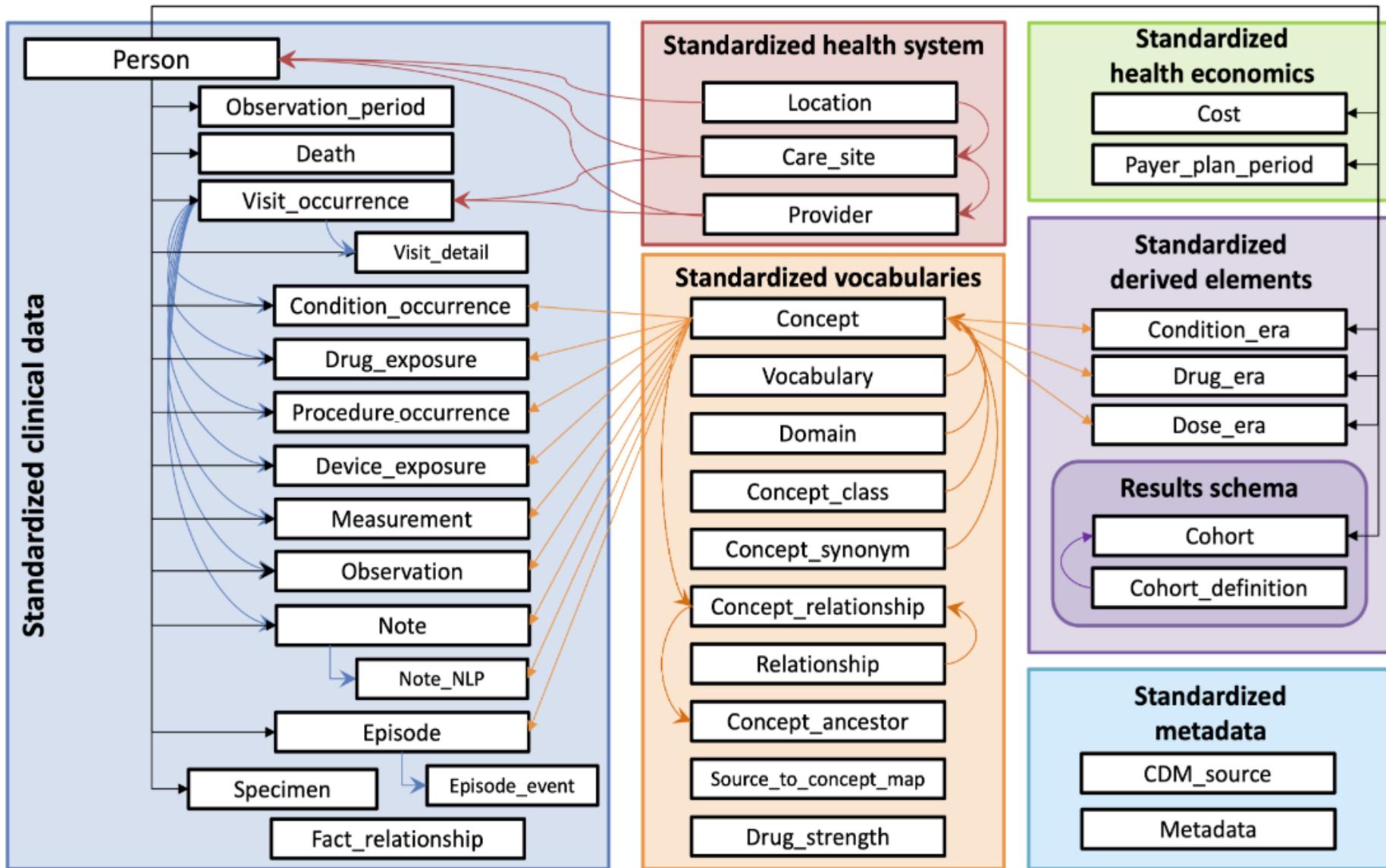
- Contains records of Measurement, i.e. structured values (numerical or categorical) obtained through systematic and standardized examination or testing of a Person or Person's sample
- Data sources: structured, quantitative measures, such as laboratory tests
- Measures have associated units

OBSERVATION

Observation	
observation_id	PK
person_id	
observation_concept_id	
observation_date	
observation_datetime	
observation_type_concept_id	
value_as_number	
value_as_string	
value_as_concept_id	
qualifier_concept_id	
unit_concept_id	
provider_id	
visit_occurrence_id	
observation_source_value	
observation_source_concept_id	
unit_source_value	
qualifier_source_value	

- Captures clinical facts about a Person obtained in the context of examination, questioning or a procedure
- Any data that cannot be represented by any other domains, such as social and lifestyle facts, medical history, family history, etc. are recorded here
- Instrument for CDM extension, playpen

CDM Version 5.4 Key Domains

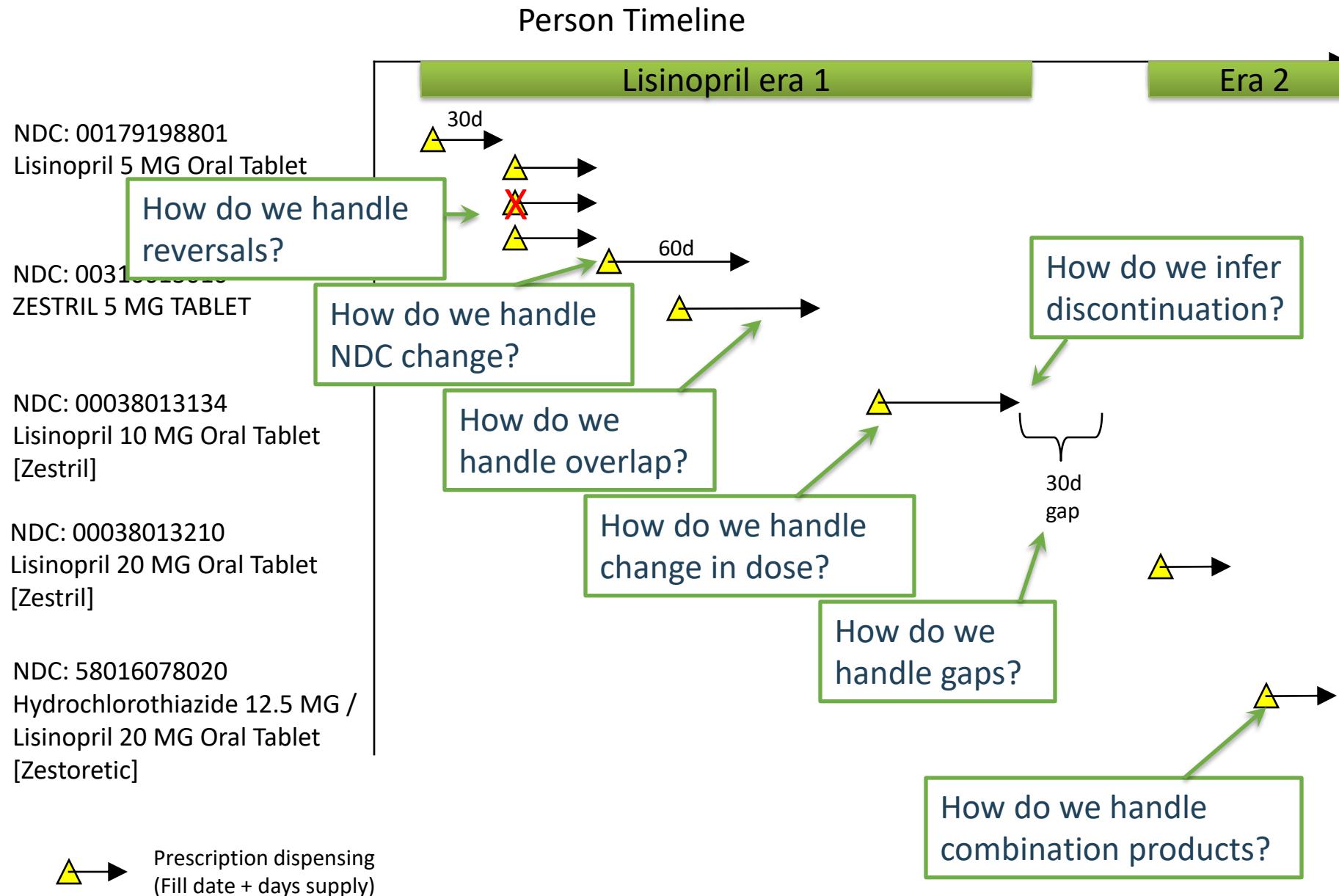


DRUG_ERA

Drug_Era
drug_era_id
person_id
drug_concept_id
drug_era_start_date
drug_era_end_date
drug_exposure_count
gap_days

- **Standardized inference of length of exposure to product for all active ingredients**
- **Derived from records in DRUG_EXPOSURE under certain rules to produce continuous Drug Eras**

Illustrating inferences needed within longitudinal pharmacy claims data for one patient



https://cran.r-project.org/web/packages/DrugUtilisation/DrugUtilisation.pdf

ORIGINAL ARTICLE | Open Access |

Calculating daily dose in the Observational Medical Outcomes Partnership Common Data Model

Theresa Burkard, Kim López-Güell, Artem Gorbachev, Lucía Bellas, Annika M. Jödicke, Edward Burn, Maria de Ridder, Mees Mosseveld, Jasmine Gratton, Sarah Seager, Dina Vojinovic, Miguel Angel Mayer, Juan Manuel Ramírez-Anguita, Angela Leis Machín, Marek Oja, Raivo Kolde, Klaus Bonadt, Daniel Prieto-Alhambra , Christian Reich, Martí Català ... See fewer authors ^

First published: 21 May 2024 | <https://doi.org/10.1002/pds.5809> | Citations: 3

Christian Reich and Martí Català contributed equally.
Preliminary results of this work have been presented at the Observational Health Data Sciences and Informatics (OHDSI) Symposium held in East Brunswick, USA, on October 20–22, 2023.

SECTIONS PDF TOOLS SHARE

Abstract

Purpose

We aimed to develop a standardized method to calculate daily dose (i.e., the amount of drug a patient was exposed to per day) of any drug on a global scale using only drug information of typical observational data in the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) and a single reference table from NeedsCompilation no

Author Martí Català [aut, cre] (ORCID: <https://orcid.org/0000-0003-3308-9905>),
Mike Du Jethl (ORCID: <https://orcid.org/0000-0002-9517-8834>)

CDM Tables Not Covered in Detail

- VISIT_DETAIL
- SPECIMEN
- DEATH
- DEVICE_EXPOSURE
- NOTE
- NOTE_NLP
- FACT_RELATIONSHIP
- LOCATION
- CARE_SITE
- PROVIDER
- PAYER_PLAN_PERIOD
- COST
- COHORT
- COHORT_ATTRIBUTES
- CONDITION_ERA
- DOSE_ERA
- CDM_SOURCE

FACT_RELATIONSHIP Table - Domain Mappings

 M

MPhilofsky Melanie Philofsky

Aug 2022

Welcome to OHDSI, [@JohnOsborne](#) !

JohnOsborne:

using the FACT_RELATIONSHIP table to represent clinical reasoning.



This would be an “off label” use of the Fact Relationship table. And not a very good off label use. The Fact Relationship table is really good to connect [mom-baby relationships](#), hierarchical care site relationships, systolic & diastolic blood pressure measurements, or other clinical or health system records in the CDM. It shouldn’t be used to connect concept_ids. We have the Vocabulary tables, specifically Concept Ancestor and Concept Relationship to represent relationships between concepts. Concept Ancestor “rolls up” concepts.

Please explain your use case a little more and we (the community) will help you find a better way to represent or query your data. Just off the top of my head, creating concept sets and then phenotypes for Type 2 Diabetes might serve your use case. And for drugs, I always advocate the use of the Concept Ancestor table when looking for specific ingredients. Then narrow down the results based on the drug’s attributes (dose, form, etc.). Do you have Atlas installed?

Vocabularies

Vocabulary Housekeeping

- Open up ohdsi.athena.org and Databricks if applicable
- Feel free to follow along by writing the code, or just watch and learn
- I will pause from time to time
- Question/comment in chat

← ⌂ ⌂ https://boycelab.github.io/OHDSITrainTheTrainer-6WK/exercises/code_snippets/day-01-snippets/ ⌂



OMOP Train the Trainer with Dr. Danielle Boyce

Search

[Home](#) [Syllabus](#) [Personas & Learning Paths](#) [Modules](#) [Exercises](#) [Resources](#) [Community & Support](#) [Ackno](#)

Exercises

[Day 1 · Athena & CDM](#)

[Day 1 · Code Snippets](#)

[Day 2 · Vocab & DQD](#)

Day 1 — OMOP / OHDSI Code Snippets

These SQL examples accompany **Day 1 – OMOP Common Data Model**. Learners can use them within Databricks, DBeaver, or any SQL client connected to an OMOP-formatted database.

Atrial Fibrillation — Concept Exploration

Finding the Right Concept#1

```
SELECT * FROM concept WHERE concept_id = 313217;
```

Finding the Right Concept#2

```
SELECT * FROM concept WHERE concept_name = 'Atrial fibrillation';
```

OMOP Common Vocabulary Model

What it is

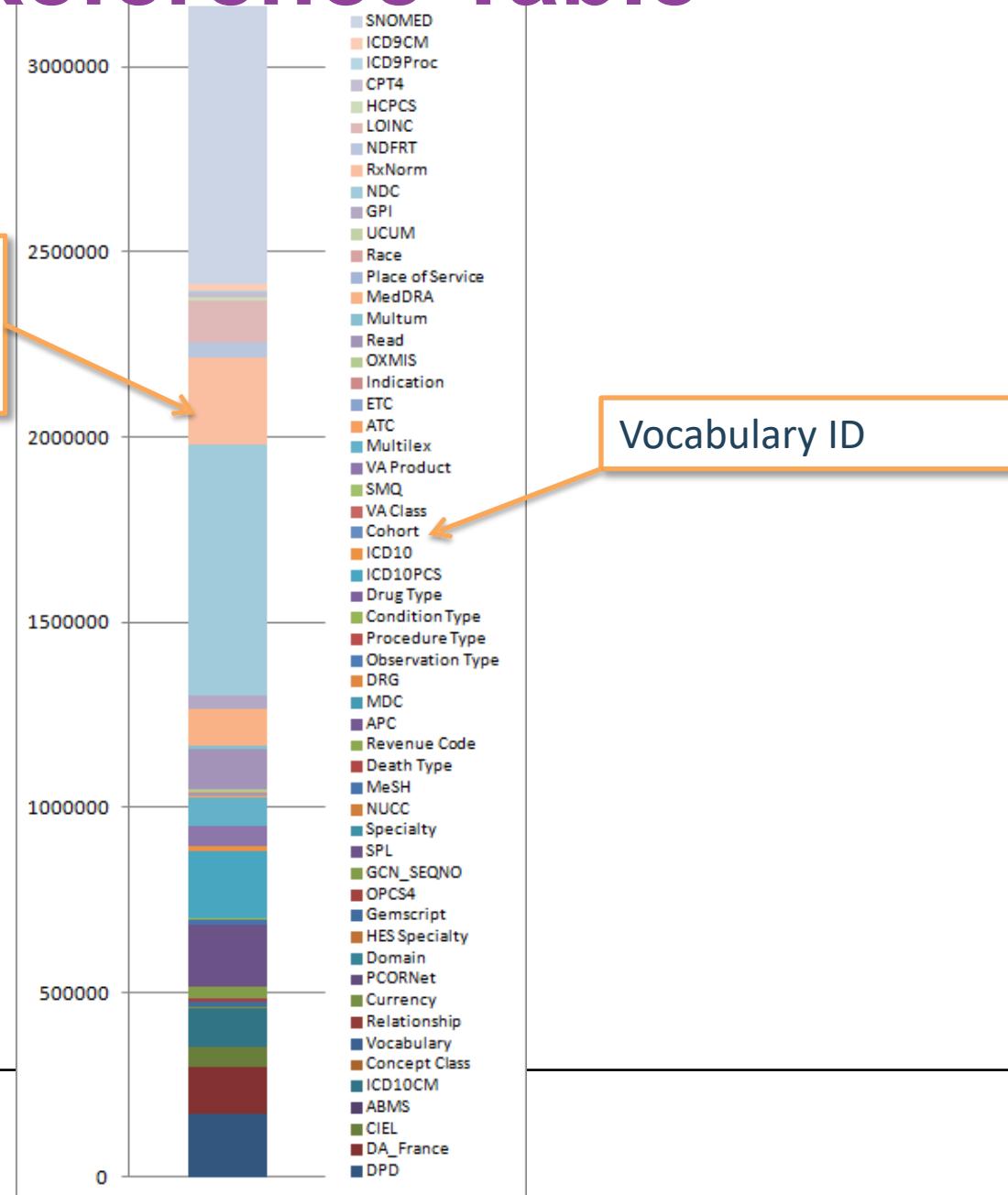
- **Standardized structure** to house existing vocabularies used in the public domain
- **Compiled standards** from disparate public and private sources and some OMOP--grown concepts

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

Single Concept Reference Table

All vocabularies
stacked up in
one table



https://athena.ohdsi.org/vocabulary/list

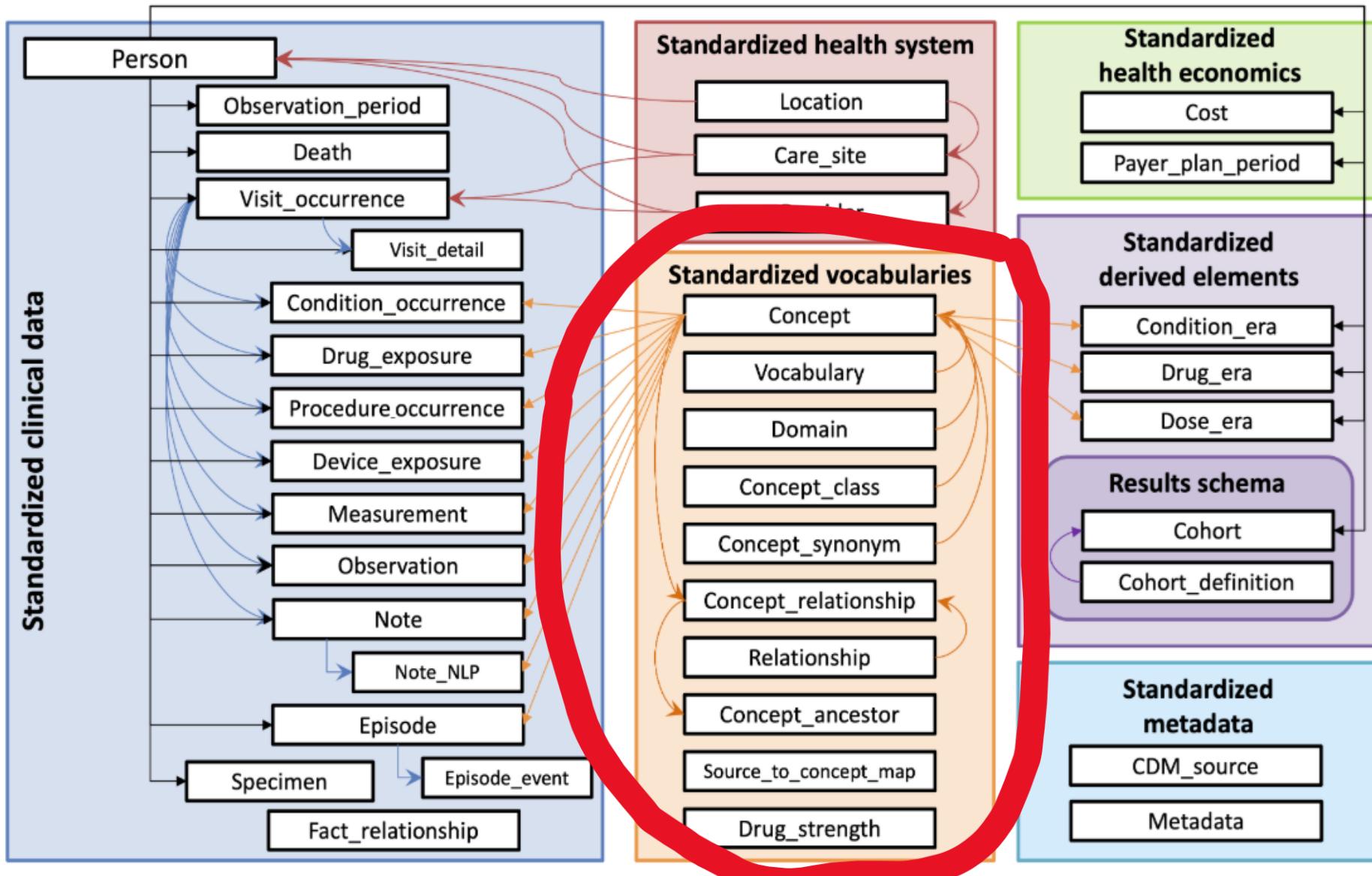
ATHENA

SEARCH DOWNLOAD Danielle Boyce DOWNLOAD HISTORY DOWNLOAD VOCABULARIES

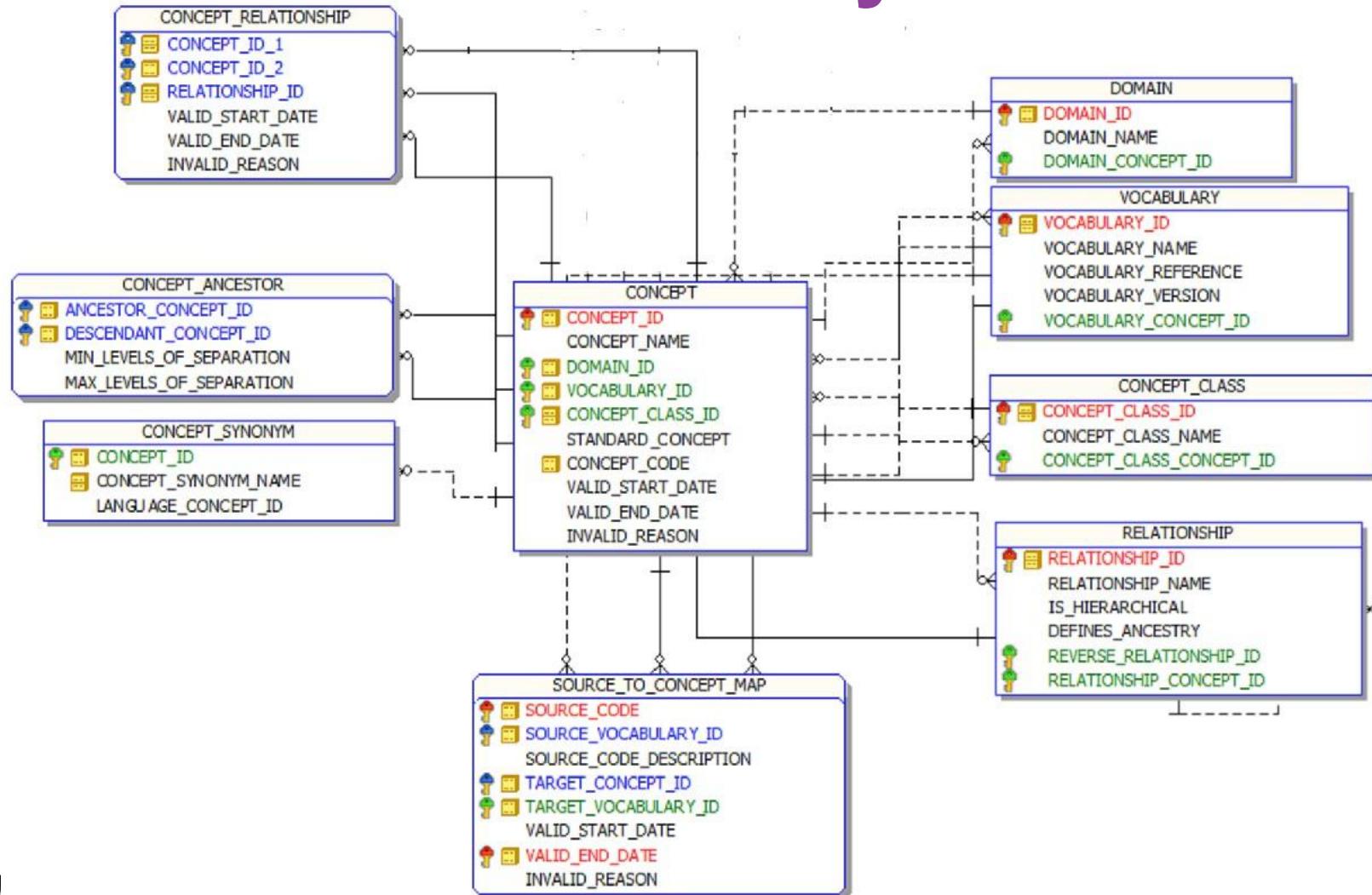
Show all ▾

ID (CDM V4.5)	CODE (CDM V5)	NAME	REQUIRED	LATEST UPDATE
<input type="checkbox"/>	1	SNOMED	Systematic Nomenclature of Medicine - Clinical Terms (IHTSDO)	28-Feb-2025
<input checked="" type="checkbox"/>	2	ICD9CM	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 1 and 2 (NCHS)	30-Sep-2014
<input checked="" type="checkbox"/>	3	ICD9Proc	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 3 (NCHS)	30-Sep-2014
<input checked="" type="checkbox"/>	4	CPT4	Current Procedural Terminology version 4 (AMA)	EULA required 04-May-2025
<input checked="" type="checkbox"/>	5	HCPCS	Healthcare Common Procedure Coding System (CMS)	30-Jun-2025
<input checked="" type="checkbox"/>	6	LOINC	Logical Observation Identifiers Names and Codes (Regenstrief Institute)	25-Feb-2025
<input type="checkbox"/>	7	NDFRT	National Drug File - Reference Terminology (VA)	05-Aug-2018
<input checked="" type="checkbox"/>	8	RxNorm	RxNorm (NLM)	01-Jun-2025
<input checked="" type="checkbox"/>	9	NDC	National Drug Code (FDA and manufacturers)	16-Aug-2025

CDM Version 5.4 Key Domains



OMOP Vocabulary CDM



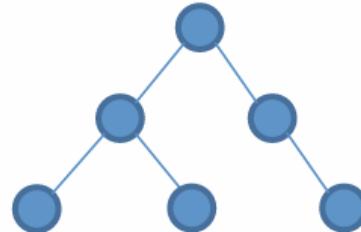
Structure of OMOP Vocabulary



All content: concepts in
concept



Direct relationships between
concepts in
concept_relationship

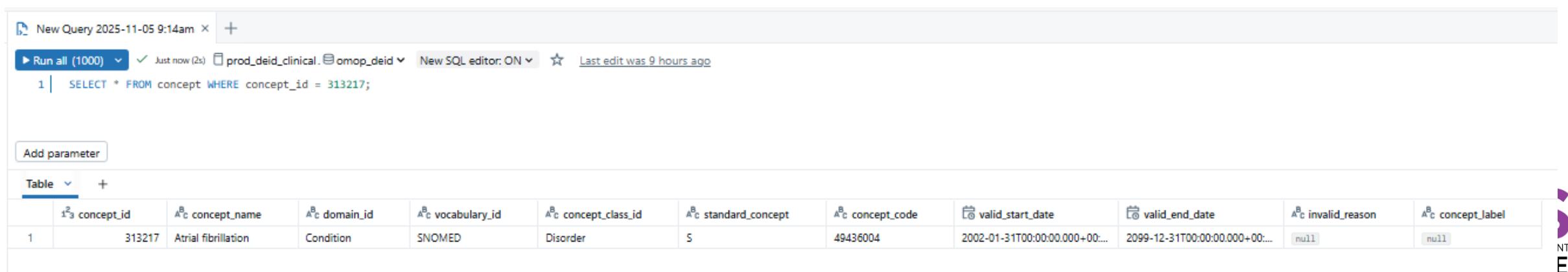


Multi-step hierarchical
relationships pre-processed
into
concept_ancestor

All Content in CDM is Coded as Concepts

- Concepts are referred to by concept_id
- All details are in the **CONCEPT** table:

```
SELECT *
FROM concept
WHERE concept_id = 313217
```



The screenshot shows a database query editor interface with the following details:

- Query:** New Query 2025-11-05 9:14am
- Status:** Run all (1000) Just now (2s)
- Database:** prod_deid_clinical. (omop_deid)
- Editor:** New SQL editor: ON
- Last edit:** Last edit was 9 hours ago
- SQL Statement:** `SELECT * FROM concept WHERE concept_id = 313217;`
- Add parameter:** Add parameter
- Table:** (selected)
- Results:** A table with 1 row and 12 columns. The columns are: concept_id, concept_name, domain_id, vocabulary_id, concept_class_id, standard_concept, concept_code, valid_start_date, valid_end_date, invalid_reason, and concept_label.
- Row Data:** concept_id: 313217, concept_name: Atrial fibrillation, domain_id: Condition, vocabulary_id: SNOMED, concept_class_id: Disorder, standard_concept: S, concept_code: 49436004, valid_start_date: 2002-01-31T00:00:00.000+00:00, valid_end_date: 2099-12-31T00:00:00.000+00:00, invalid_reason: null, concept_label: null.

Doing the same thing in Athena.ohsi.org

The screenshot shows the Athena search interface at the URL <https://athena.ohdsi.org/search-terms/terms?standardConcept=Standard&domain=Condition&page=1&pageSize=15&query=313217>. The search term "313217" is highlighted in red circles in three locations: the search bar, the filter dropdown, and the result table.

SEARCH BY KEYWORD

- 313217 X
- Standard X
- Condition X

DOMAIN

filter

Condition (1)
 Device (1)
 Drug (28)
 Geography (3)
 Measurement (1)
 Procedure (1)

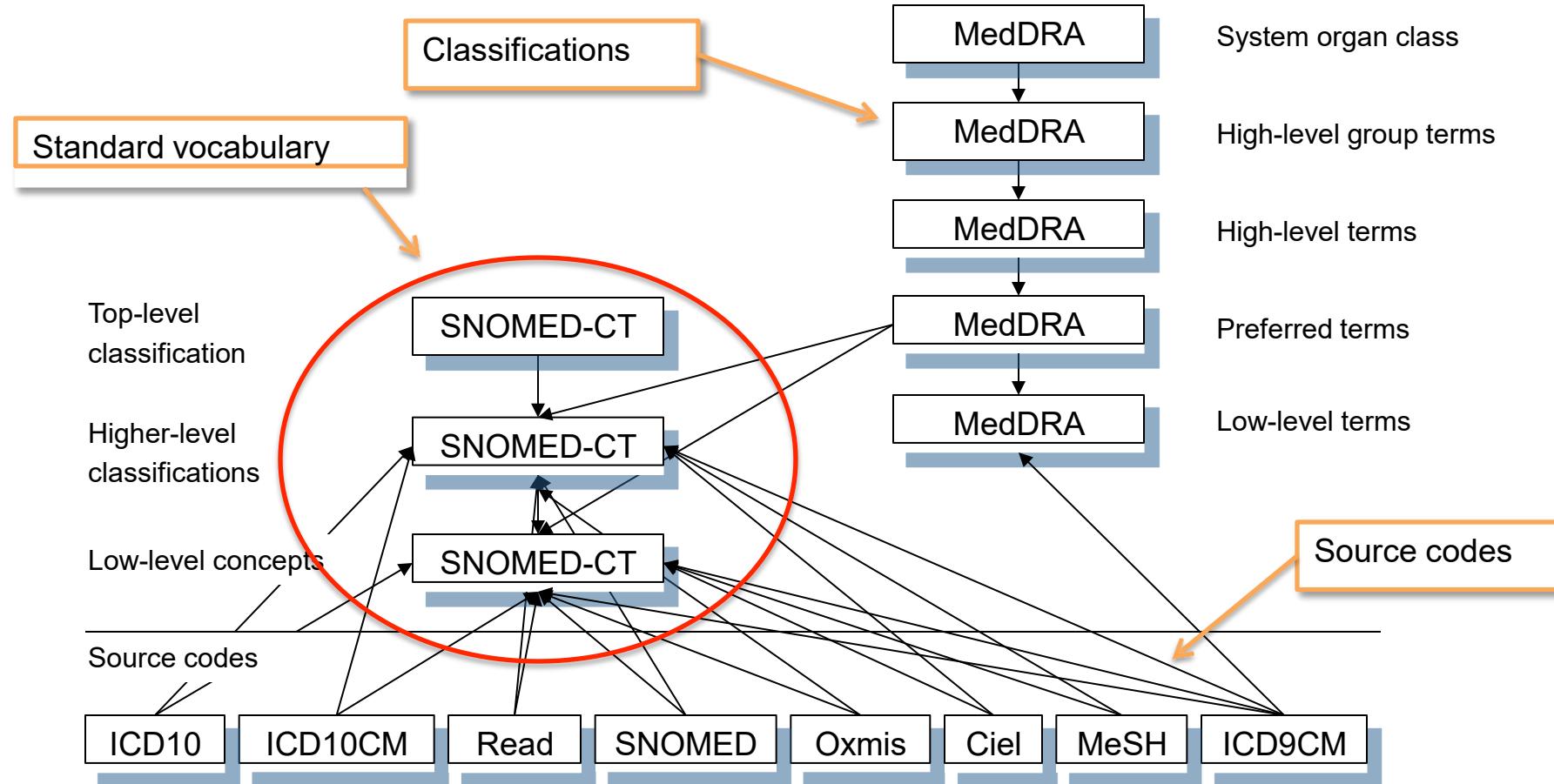
DOWNLOAD RESULTS

ID	CODE	NAME	CLASS	CONCEPT
313217	49436004	Atrial fibrillation	Disorder	Standard

Dozens of schemes, formats, rules

LOINC_248_MULTI--AXIAL_HIERARCHY.CSV													
PATH_TO_ROOT			SEQUENC	IMMEDIATE_PARENT	CODE	CODE_TEXT							
			1		LP31755-9	Microbiology							
LP317	loinc.csv				1 LP31755-9	Microorganism							
LP317					1 LP14559-6	Bacteria							
LP317					1 LP98185-9	Bacteria							
LOINC_NL_COMPONENT			PROPERTY	TIME_ASPECT	SYSTEM	SCALE_TYP	METHOD_TYP	CLASS	SOURCE				
10454-7	Xylose^2H post 25 g xylose PO		MCnc	Pt	Ser/Plas	Qn		CHAL	SH				
10455-4	Xylose^30M post 25 g xylose PO		MCnc	Pt	Ser/Plas	Qn		CHAL	SH				
10456-2	Xylose^post 6H CFst		MCnc	Pt	Ser/Plas	Qn		CHAL	SH				
10457-0	Actin Ag		ACnc	Pt	Tiss	Ord	Immune stain	PATH	SH;DL-M				
10458-1						Drd	Immune stain	PATH	DL-M				
10459-1						Drd	Immune stain	PATH	DL-M				
10460-1	CMS32_DESC_LONG_SHORT_DX.xlsx					SHORT DESCRIPTION							
DIAGNOSIS_CODE		LONG DESCRIPTION			SHORT DESCRIPTION								
10461-1	0010	Cholera due to vibrio cholerae			Cholera d/t vib cholerae								
10462-1	0011	Cholera due to vibrio cholerae el tor			Cholera d/t vib el tor								
10462-1	0019	Cholera, unspecified			Cholera NOS								
10463-1	0020	Typhoid fever			Typhoid fever								
10464-1	0021	Paratyphoid fever A			Paratyphoid fever a								
10465-1	0022	Paratyphoid fever B			Paratyphoid fever b								
	0023	Paratyphoid fever C			Paratyphoid fever c								
	0029	Paratyphoid fever, unspecified			Paratyphoid fever NOS								
	0030	Salmonella gastroenteritis			Salmonella enteritis								
	0031	Salmonella septicemia			Salmonella septicemia								
	00320	Localized salmonella infection, unspecified			Local salmonella inf NOS								
	00321	Salmonella meningitis			Salmonella meningitis								
	00322	Salmonella pneumonia			Salmonella pneumonia								
	00323	Salmonella arthritis			Salmonella arthritis								
	00324	Salmonella osteomyelitis			Salmonella osteomyelitis								
	00329	Other localized salmonella infections			Local salmonella inf NEC								

Condition Concepts



Finding the Right Concept: #1

1. ..if I know the ID

```
SELECT * FROM concept WHERE concept_id = 313217;
```

CONCEPT_ID	CONCEPT_NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT	CONCEPT_CODE	VALID_START_DATE	VALID_END_DATE	INVALID_REASON
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	01-Jan--1970	31-Dec--2099	

SNOMED code

2. ..if I know the code

```
SELECT * FROM concept WHERE concept_code = '49436004';
```

CONCEPT_ID	CONCEPT_NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT	CONCEPT_CODE	VALID_START_DATE	VALID_END_DATE	INVALID_REASON
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	01-Jan--1970	31-Dec--2099	



Concept code 49436004 in SNOMED Browser

The SNOMED CT Browser - SNOMED Clinical Terms version: 20130131 [R] (January 2013 Release) - Mozilla Firefox

File Edit View History Bookmarks Tools Help

The SNOMED CT Browser - SNOMED Clinical ... +

www.medicalclassifications.com/SNOMEDbrowser/ Google

The SNOMED CT Browser ©

Search term/concept: atrial fibrillation Go

Refine your search in the list below:

Atrial fibrillation (disorder)

- Atrial fibrillation
- AF - Atrial fibrillation

Atrial fibrillation monitoring (regime/therapy)

- Atrial fibrillation monitoring

Electrocardiogram: atrial fibrillation (finding)

- ECG: atrial fibrillation
- Electrocardiogram: atrial fibrillation

Atrial fibrillation (disorder)

Concept codes & terms:

Atrial fibrillation (disorder)

conceptid	49436004
snomedid	D3-31520
ctv3id	G5730

Preferred term descriptionid

Atrial fibrillation 82343012

Synonym(s) descriptionid(s)

AF - Atrial fibrillation 1230726010

Concept definition / attribute relations:

Is a

- Atrial arrhythmia (disorder)
- Fibrillation (disorder)

group 0 Finding site Atrial structure (body structure)

Finding site Cardiac conducting system structure (body structure)

5 inverse relation(s):

Associated finding Family history of atrial fibrillation (situation)

History of - atrial fibrillation (situation)

Due to Transient cerebral ischemia due to atrial fibrillation (disorder)

Has focus Insertion of pacemaker for control of atrial fibrillation (procedure)

Maze procedure for atrial fibrillation (procedure)

Subtypes:

9 DIRECT SUBTYPES and 0 ADDITIONAL SUBTYPES:

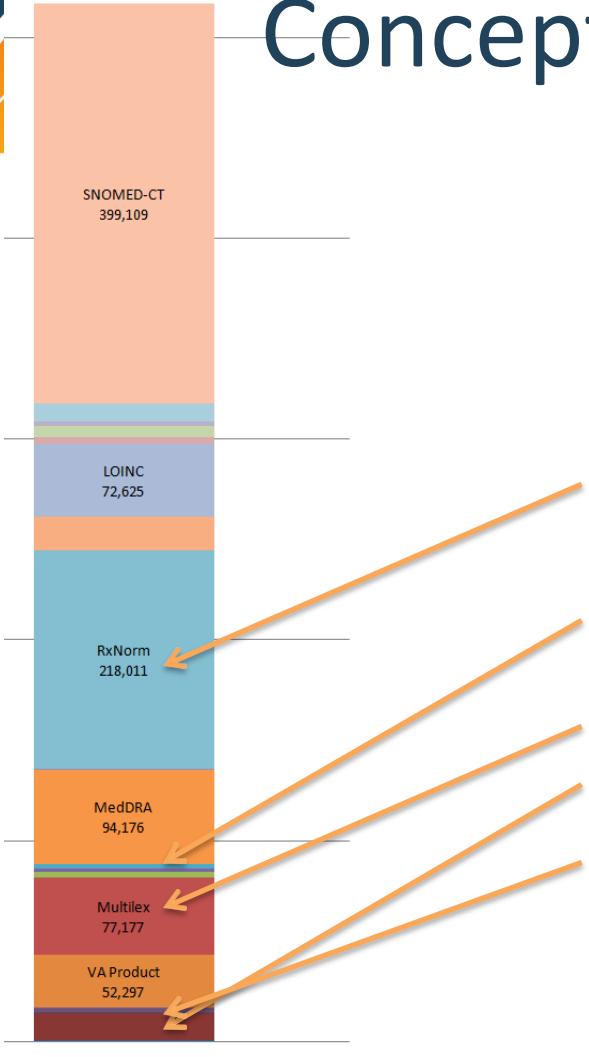
- Atrial fibrillation and flutter (disorder)
- Chronic atrial fibrillation (disorder)
- Controlled atrial fibrillation (disorder)
- Lone atrial fibrillation (disorder)
- Non-rheumatic atrial fibrillation (disorder)
- Paroxysmal atrial fibrillation (disorder)
- Permanent atrial fibrillation (disorder)
- Persistent atrial fibrillation (disorder)
- Rapid atrial fibrillation (disorder)

SNOMED code



Concept ID versus Concept Code

```
SELECT *
FROM concept
WHERE concept_code = '1001';
```



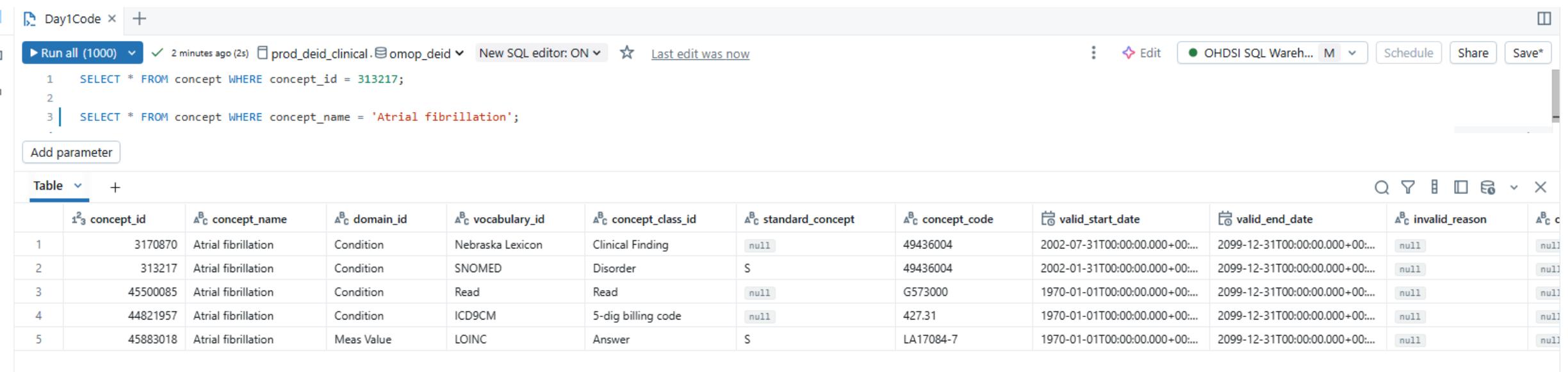
Concept_Name	Concept Class	Vocabulary_ID	Concept_Code
Antipyrine	Ingredient	RxNorm	1001
Aceprometazine maleate	Ingredient	BDPM	1001
Serum	Specimen	CIEL	1001
methixene hydrochloride	Ingredient	Multilex	1001
Brompheniramine Maleate, 10 mg/mL injectable solution	Multum	Multum	1001
ABBOTT COLD SORE BALM 4%/0.06% W/	Drug Product	LPD_Australia	1001
Residential Treatment - Psychiatric	Revenue Code	Revenue Code	1001

Same code

Finding the Right Concept: #2

3. ..if I know the name

```
SELECT * FROM concept WHERE concept_name = 'Atrial fibrillation';
```



The screenshot shows a SQL editor interface with the following details:

- Title Bar:** Day1Code
- Toolbar:** Run all (1000), 2 minutes ago (2s), prod_deid_clinical, omop_deid, New SQL editor: ON, Last edit was now, Edit, OHDSI SQL Ware..., Schedule, Share, Save*
- Code Area:**

```
1  SELECT * FROM concept WHERE concept_id = 313217;
2
3  SELECT * FROM concept WHERE concept_name = 'Atrial fibrillation';
4
```
- Parameter Input:** Add parameter
- Table View:** Shows the results of the third query in a grid format. The columns are: concept_id, concept_name, domain_id, vocabulary_id, concept_class_id, standard_concept, concept_code, valid_start_date, valid_end_date, invalid_reason, and code. The data rows are:

	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason	code
1	3170870	Atrial fibrillation	Condition	Nebraska Lexicon	Clinical Finding	null	49436004	2002-07-31T00:00:00.000+00:00	2099-12-31T00:00:00.000+00:00	null	null
2	313217	Atrial fibrillation	Condition	SNOMED	Disorder	S	49436004	2002-01-31T00:00:00.000+00:00	2099-12-31T00:00:00.000+00:00	null	null
3	45500085	Atrial fibrillation	Condition	Read	Read	null	G573000	1970-01-01T00:00:00.000+00:00	2099-12-31T00:00:00.000+00:00	null	null
4	44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code	null	427.31	1970-01-01T00:00:00.000+00:00	2099-12-31T00:00:00.000+00:00	null	null
5	45883018	Atrial fibrillation	Meas Value	LOINC	Answer	S	LA17084-7	1970-01-01T00:00:00.000+00:00	2099-12-31T00:00:00.000+00:00	null	null

Doing the same thing in Athena.ohsi.org

The screenshot shows the Athena search interface at the URL <https://athena.ohdsi.org/search-terms/terms?query=Atrial+fibrillation>. A red circle highlights the search term 'Atrial fibrillation' in the search bar. The results table displays five entries:

ID	CODE	NAME	CLASS	CONCEPT
313217	49436004	Atrial fibrillation	Disorder	Standard
45611600	D001281	Atrial Fibrillation	Main Heading	Non-standard
44807374	816401000000105	Atrial fibrillation excluded	Context-dependent	Standard
42689664	1066831000000104	Atrial fibrillation detected	Clinical Finding	Standard

Finding the Right Concept #3

1. if don't know any of this, but I know the code in another vocabulary

```
SELECT * FROM concept WHERE concept_code = '427.31';
```

ICD-9 is not a Standard Concept

CONCEPT_ID	CONCEPT_NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT	CONCEPT_CODE
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31

```
SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957;
```

Mapping to different vocabularies

Kind of relationship

	CONCEPT_ID_2	RELATIONSHIP_ID	VALID_START_DATE	VALID_END_DATE	INVALID_REASON
44821957	21001551	ICD9CM - FDB Ind	01-Oct-13	31-Dec-2099	
44821957	35204953	ICD9CM - MedDRA	01-Jan-70	31-Dec-2099	
44821957	44824248	Is a	01-Oct-14	31-Dec-2099	
44821957	44834731	Is a	01-Oct-14	31-Dec-2099	
44821957	313217	Maps to	01-Jan-70	31-Dec-2099	

[← Atrial fibrillation](#)

DETAILS

Domain ID	Condition
Concept Class ID	5-dig billing code
Vocabulary ID	ICD9CM
Concept ID	44821957
Concept code	427.31
Validity	Valid
Concept	Non-standard
Valid start	01-Jan-1970
Valid end	31-Dec-2099

TERM CONNECTIONS (3)

RELATIONSHIP	RELATES TO	CONCEPT ID	VOCABULARY
Is a	Atrial fibrillation and flutter	44824248	ICD9CM
	Cardiac dysrhythmias	44834731	ICD9CM
Non-standard to Standard map (OMOP)	Atrial fibrillation	313217	SNOMED

Pěrshéndetjee

Здравей

Zpravev

Buná

Bok

Bonuá

Buna'styte

Здравствуйт

Hei

Ha

Salua

Bonjour

Hello

Olá

O lá

Helló

Suavé

Dia dhuit

Labas

Halló

Ciao

Hello

[Atrial fibrillation](#)

DETAILS

Domain ID Condition

Concept Class ID 5-dig billing code

Vocabulary ID ICD9CM



Concept ID 44821957

Concept code 427.31



Validity Valid

Concept Non-standard

Valid start 01-Jan-1970

Valid end 31-Dec-2099

TERM CONNECTIONS (3)

RELATIONSHIP	RELATES TO	CONCEPT ID	VOCABULARY
Is a	Atrial fibrillation and flutter	44824248	ICD9CM
Non-standard to Standard map (OMOP)	Cardiac dysrhythmias	44834731	ICD9CM
	Atrial fibrillation	313217	SNOMED



Paroxysmal atrial fibrillation

DETAILS	
Domain ID	Condition
Concept Class ID	Disorder
Vocabulary ID	SNOMED
Concept ID	4154290
Concept code	282825002
Validity	Valid
Concept	Standard
LANGUAGE	SYNONYM CONCEPT
English	AF - Paroxysmal atrial fibrillation
English	Intermittent atrial fibrillation
English	PAF - Paroxysmal atrial fibrillation

TERM CONNECTIONS (29)		HIERARCHY	RELATED CONCEPTS	
RELATIONSHIP	RELATES TO		CONCEPT ID	VOCABULARY
Active same_as inactive (SNOMED)	Paroxysmal atrial fibrillation		40392263	SNOMED
Has Module	SNOMED CT core		40642539	SNOMED
Has clinical course (SNOMED)	Sudden onset AND/OR short duration		4312510	SNOMED
Has finding site (SNOMED)	Atrial structure		4242112	SNOMED
	Cardiac conducting system structure		4093357	SNOMED
Has interpretation (SNOMED)	Increased		4146067	SNOMED
Has interprets (SNOMED)	Heart rate		4239408	SNOMED
Has status	Primitive		40642538	SNOMED
Is a	Acute heart disease		4132088	SNOMED
	Atrial fibrillation	313217		SNOMED

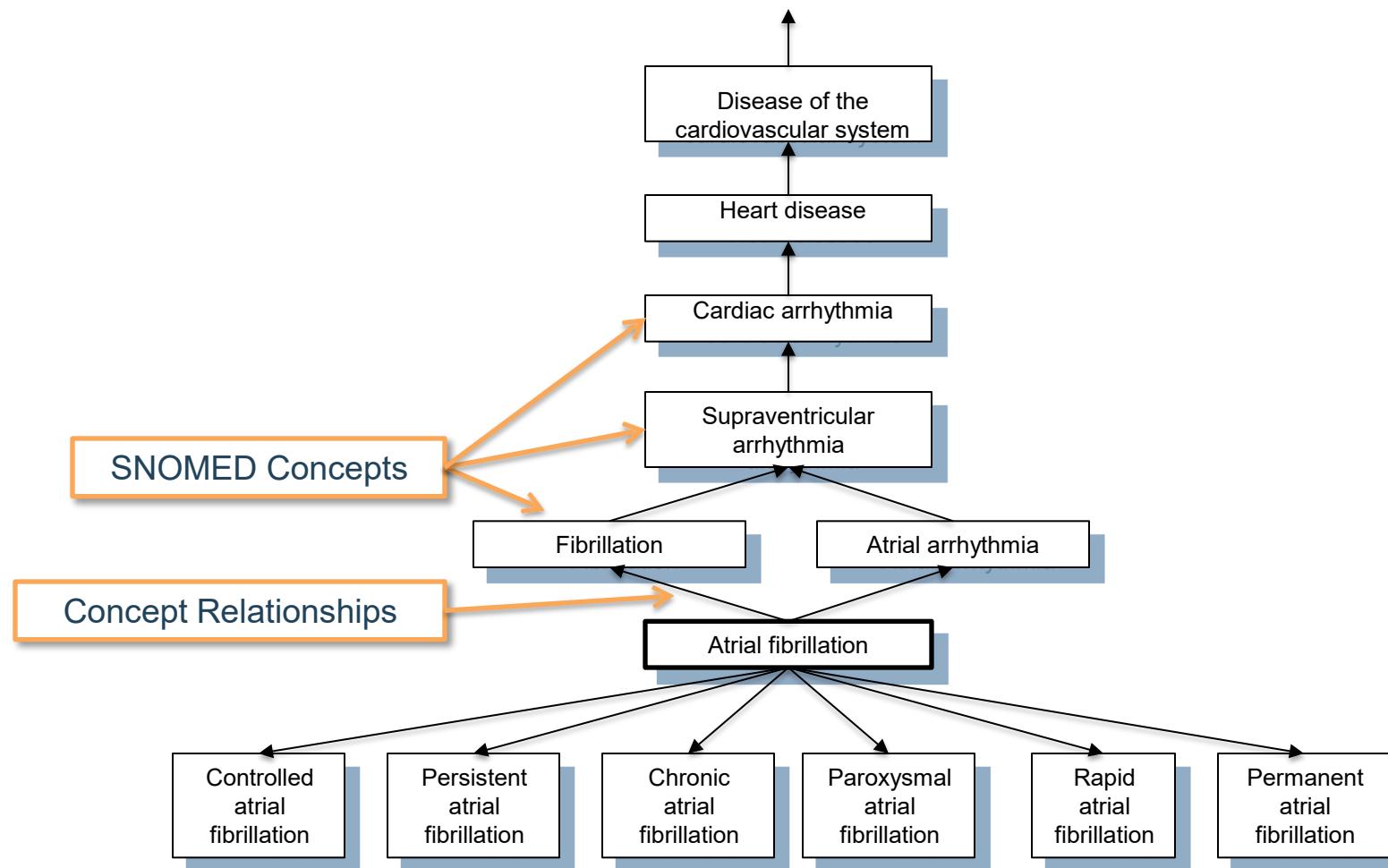
Codes Used in the World



- Conditions
 - READ, OXMIS, ICD--9--CM, ICD--10--CM, ICPC, MedDRA, free text in different languages

- Drugs
 - RxNorm, MulHlex, dm+d, BDPM, AMIS, AMT, ATC, NPI, NDC, free text in different languages

Disease Hierarchy



Exploring Relationships

```
SELECT *
FROM concept_relationship
WHERE concept_id_1 = 313217
```

Related Concepts

Relationship ID

CONCEPT_ID_1	CONCEPT_ID_2	RELATIONSHIP_ID
313217	4232697	Subsumes
313217	4181800	Focus of
313217	35204953	SNOMED - MedDRA eq
313217	4203375	Asso finding of
313217	4141360	Subsumes
313217	4119601	Subsumes
313217	4117112	Subsumes
313217	4232691	Subsumes
313217	4139517	Due to of
313217	4194288	Asso finding of
313217	44782442	Subsumes
313217	44783731	Focus of
313217	21003018	SNOMED - ind/Cl
313217	40248987	SNOMED - ind/Cl
313217	21001551	SNOMED - ind/Cl
313217	21001540	SNOMED - ind/Cl
313217	45576876	Mapped from
313217	44807374	Asso finding of
313217	21013834	SNOMED - ind/Cl
313217	21001572	SNOMED - ind/Cl
313217	21001606	SNOMED - ind/Cl
313217	21003176	SNOMED - ind/Cl
313217	4226399	Is a
313217	500001801	SNOMED - HOI
313217	500002401	SNOMED - HOI
313217	4119602	Subsumes
313217	40631039	Subsumes
313217	4108832	Subsumes
313217	21013671	SNOMED - ind/Cl
313217	21013390	SNOMED - ind/Cl
313217	313217	Maps to
313217	44821957	Mapped from
313217	2617597	Mapped from
313217	45500085	Mapped from
313217	313217	Mapped from
313217	45951191	Mapped from
313217	21013856	SNOMED - ind/Cl
313217	21001575	SNOMED - ind/Cl
313217	21001594	SNOMED - ind/Cl

Athena

English	AF - Atrial fibrillation	Family history of atrial fibrillation
English	Atrial fibrillation (disorder)	H/O: atrial fibrillation
Spanish	fibrilación auricular	Due to of (SNOMED) Hypercoagulable state due to atrial fibrillation
Spanish	fibrilación auricular (trastorno)	Transient cerebral ischemia due to atrial fibrillation
Valid start	31-Jan-2002	Focus of (SNOMED) Cardiac ablation for atrial fibrillation
Valid end	31-Dec-2099	Insertion of pacemaker for contralateral atrial fibrillation Maze procedure for atrial fibrillation
		Provision of written information about atrial fibrillation
		Has Module SNOMED CT core
		Has finding site (SNOMED) Atrial structure
		Cardiac conducting system structure
		Has interpretation (SNOMED) Increased
		Has interprets (SNOMED) Heart rate
		Has status Defined
	Is a	Atrial arrhythmia
		Fibrillation

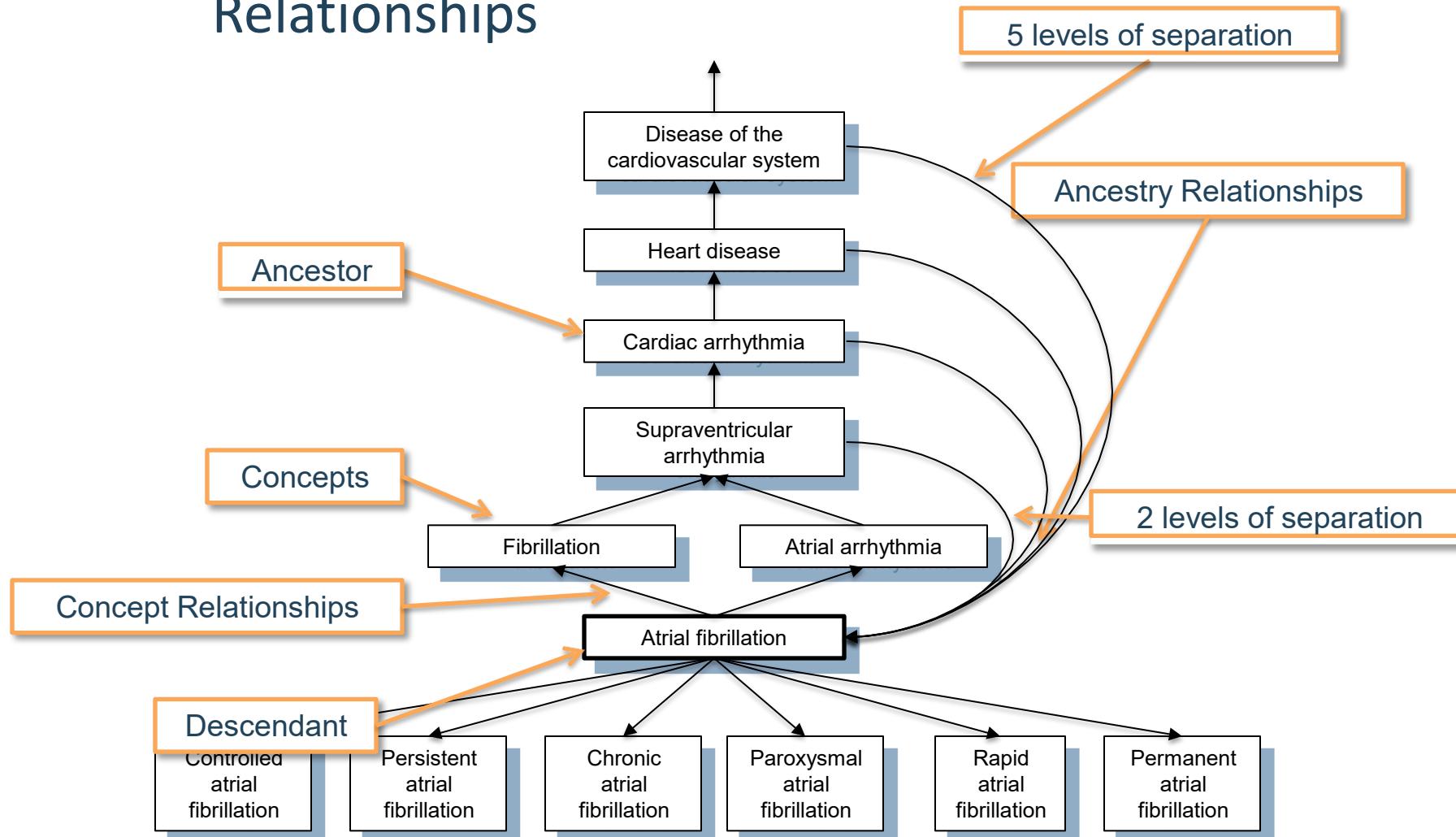
Exploring Relationships #2

```
SELECT cr.relationship_id, c.*  
FROM concept_relationship cr  
JOIN concept c ON cr.concept_id_2 = c.concept_id  
WHERE cr.concept_id_1 = 313217;
```

Find out related concept

relationship_id	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
Asso finding of	4203375	Family his	Observation	SNOMED	Context-dependent	S	433276002	31-Jan-2009	31-Dec-2099	
Asso finding of	44807374	Atrial fibr	Observation	SNOMED	Context-dependent	S	816401000000	1-Apr-2014	31-Dec-2099	
Asso finding of	4194288	H/O: atrial fibrillation	Observation	SNOMED	Context-dependent	S	312442005	1-Jan-1970	31-Dec-2099	
Due to of	4139517	Transient cerebral ischemia due to atrial fibrillation	Condition	SNOMED	Clinical Finding	S	426814001	1-Jan-1970	31-Dec-2099	
Focus of	42709991	Insertion of pacemaker for control of atrial fibrillation	Procedure	SNOMED	Procedure	S	449863006	31-Jan-2012	31-Dec-2099	
Focus of	4181800	Maze procedure for atrial fibrillation	Procedure	SNOMED	Procedure	S	429211003	31-Jan-2008	31-Dec-2099	
Focus of	44783731	Provision of written information about atrial fibrillation	Procedure	SNOMED	Procedure	S	699833006	31-Jan-2014	31-Dec-2099	
Has finding site	4242112	Atrial structure	Spec Anatomic Site	SNOMED	Body Structure	S	59652004	1-Jan-1970	31-Dec-2099	
Has finding site	4093357	Cardiac conducting system structure	Spec Anatomic Site	SNOMED	Body Structure	S	24964005	1-Jan-1970	31-Dec-2099	
Is a	4226399	Fibrillation	Condition	SNOMED	Clinical Finding	S	40593004	1-Jan-1970	31-Dec-2099	
Is a	4068155	Atrial arrhythmia	Condition	SNOMED	Clinical Finding	S	17366009	1-Jan-1970	31-Dec-2099	
Mapped from	45576876	Unspecified atrial fibrillation	Condition	ICD10CM	5-char billing code		I48.91	30-Dec-2006	31-Dec-2099	
Mapped from	45611600	Atrial Fibrillation	Condition	MeSH	Main Heading		D001281	1-Jan-1970	31-Dec-2099	
Mapped from	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1-Jan-1970	31-Dec-2099	
Mapped from	45951191	Atrial Fibrillation	Condition	CIEL	Diagnosis		148203	3-Nov-2007	31-Dec-2099	
Mapped from	44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31	1-Jan-1970	31-Dec-2099	
Mapped from	2617597	Patient with heart failure and atrial fibrillation documented to be on warfarin therapy	Observation	HCPCS	HCPCS		G8183	1-Jan-1970	11-Nov-2014	D
Mapped from	45500085	Atrial fibrillation	Condition	Read	Read		G573000	1-Jan-1970	31-Dec-2099	
Maps to	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1-Jan-1970	31-Dec-2099	
SNOMED - HOI	500002401	OMOP Atrial Fibrillation 1	Condition	Cohort	Cohort	C	500002401	1-Jan-1970	31-Dec-2099	
SNOMED - HOI	500001801	OMOP Qt Prolongation/Torsade De Pointes 1	Condition	Cohort	Cohort	C	500001801	1-Jan-1970	31-Dec-2099	
SNOMED - ind/CI	4344544	Atrial Fibrillation	Drug	NDFRT	Ind / CI	C	N0000000507	1-Jan-1970	31-Dec-2099	
SNOMED - MedDRA	35204953	Atrial fibrillation	Condition	MedDRA	PT	C	10003658	1-Jan-1970	31-Dec-2099	
Subsumes	4232697	Persistent atrial fibrillation	Condition	SNOMED	Clinical Finding	S	440059007	31-Jan-2009	31-Dec-2099	
Subsumes	4119601	Lone atrial fibrillation	Condition	SNOMED	Clinical Finding	S	233910005	1-Jan-1970	31-Dec-2099	
Subsumes	4117112	Controlled atrial fibrillation	Condition	SNOMED	Clinical Finding	S	300996004	1-Jan-1970	31-Dec-2099	
Subsumes	4232691	Permanent atrial fibrillation	Condition	SNOMED	Clinical Finding	S	440028005	31-Jan-2009	31-Dec-2099	
Subsumes	4154290	Paroxysmal atrial fibrillation	Condition	SNOMED	Clinical Finding	S	282825002	1-Jan-1970	31-Dec-2099	
Subsumes	4119602	Non-rheumatic atrial fibrillation	Condition	SNOMED	Clinical Finding	S	233911009	1-Jan-1970	31-Dec-2099	
Subsumes	4199501	Rapid atrial fibrillation	Condition	SNOMED	Clinical Finding	S	314208002	1-Jan-1970	31-Dec-2099	
Subsumes	40631039	Atrial flutter fibrillation	Condition	SNOMED	Clinical Finding		81216002	1-Jan-1970	11-Mar-2016	U
Subsumes	4119603	Chronic atrial fibrillation	Condition	SNOMED	Clinical Finding	S	195080001	1-Jan-1970	31-Dec-2099	
Subsumes	4119604	Transient atrial fibrillation	Condition	SNOMED	Clinical Finding	S	1.20041E+14	31-Jan-2014	31-Dec-2099	
Subsumes	4119605	Supraventricular tachycardia	Condition	SNOMED	Clinical Finding	S	426749004	1-Jan-1970	31-Dec-2099	

Ancestry Relationships: Higher--Level Relationships



Exploring Ancestors of a Concept

```
SELECT max_levels_of_separation, concept.*  
FROM concept_ancestor  
JOIN concept ON ancestor_concept_id = concept_id  
WHERE descendant_concept_id = 313217 /* Atrial fibrillation */  
ORDER BY max_levels_of_separation
```

max_levels_of_separation	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept
0	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S
0	35204953	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S
1	4226399	Fibrillation	Condition	MedDRA	PT	C
1	4068155	Atrial arrhythmia	Condition	SNOMED	Clinical Finding	S
1	35204969	Cardiac fibrillation	Condition	MedDRA	PT	C
2	4248028	Supraventricular arrhythmia	Condition	SNOMED	Clinical Finding	S
2	35204952	Arrhythmia supraventricular	Condition	MedDRA	PT	C
2	35202454	Rate and rhythm disorders NEC	Condition	MedDRA	HLT	C
3	44784217	Cardiac arrhythmia	Condition	SNOMED	Clinical Finding	S
3	35202455	Supraventricular arrhythmias	Condition	MedDRA	HLT	C
4	321588	Heart disease	Condition	SNOMED	Clinical Finding	S
4	35204989	Cardiac disorder	Condition	MedDRA	PT	C
4	35202050	Cardiac arrhythmias	Condition	MedDRA	HLGT	C
5	4103183	Cardiac finding	Condition	SNOMED	Clinical Finding	S
5	440142	Disorder of mediastinum	Condition	SNOMED	Clinical Finding	S
5	134057	Disorder of cardiovascular system	Condition	SNOMED	Clinical Finding	S
5	35204998	Cardiovascular disorder	Condition	MedDRA	PT	C
5	37219970	Mediastinal disorder	Condition	MedDRA	PT	C
5	37622411	Phlebosclerosis	Condition	MedDRA	PT	C
5	35202457	Cardiac disorders NEC	Condition	MedDRA	HLT	C
6	4115390	Mediastinal finding	Condition	SNOMED	Clinical Finding	S
6	4023995	Cardiovascular finding	Condition	SNOMED	Clinical Finding	S

Hold the descendant

Exploring Descendants of a Concept

```
SELECT max_levels_of_separation, c.*  
FROM concept_ancestor ca, concept c  
WHERE ca.ancestor_concept_id = 44784217 /* cardiac arrhythmia */  
AND ca.descendant_concept_id = c.concept_id  
ORDER BY max_levels_of_separation  
;
```

Hold the ancestor

MAX_LEVELS_OF_SEPARATION	CONCEPT_ID	CONCEPT_NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT
0	44784217	Cardiac arrhythmia	Condition	SNOMED	Clinical Finding	S
1	313224	Anomalous atrioventricular excitation	Condition	SNOMED	Clinical Finding	S
1	315643	Tachyarrhythmia	Condition	SNOMED	Clinical Finding	S
1	316429	Premature beats	Condition	SNOMED	Clinical Finding	S
1	316999	Conduction disorder of the heart	Condition	SNOMED	Clinical Finding	S
1	321042	Cardiac arrest	Condition	SNOMED	Clinical Finding	S
1	4030583	Pacemaker twiddler's syndrome	Condition	SNOMED	Clinical Finding	S
1	4057008	Accelerated atrioventricular conduction	Condition	SNOMED	Clinical Finding	S
1	4086313	Withdrawal arrhythmia	Condition	SNOMED	Clinical Finding	S
1	4088507	Ventricular escape complex	Condition	SNOMED	Clinical Finding	S
1	4088986	Atrial escape complex	Condition	SNOMED	Clinical Finding	S
1	4091901	Aberrant premature complexes	Condition	SNOMED	Clinical Finding	S
1	4092011	Aberrantly conducted complex	Condition	SNOMED	Clinical Finding	S
1	4124704	Postoperative sinoatrial disease	Condition	SNOMED	Clinical Finding	S
1	4143042	Ectopic beats	Condition	SNOMED	Clinical Finding	S
1	4164083	Ectopic rhythm	Condition	SNOMED	Clinical Finding	S
1	4172863	Fetal dysrhythmia	Condition	SNOMED	Clinical Finding	S
1	4173170	Neonatal dysrhythmia	Condition	SNOMED	Clinical Finding	S
1	4175473	Atrioventricular dissociation	Condition	SNOMED	Clinical Finding	S
1	4185572	Ventricular arrhythmia	Condition	SNOMED	Clinical Finding	S
1	4217221	Nodal rhythm disorder	Condition	SNOMED	Clinical Finding	S
1	4226399	Fibrillation	Condition	SNOMED	Clinical Finding	S
1	4228448	Bradyarrhythmia	Condition	SNOMED	Clinical Finding	S

Let's find Upper Gastrointestinal Bleeding

1. Find some imitation concept

```
SELECT * FROM concept WHERE concept_name = 'Upper gastrointestinal bleeding';
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
42891225	Upper gastrointestinal bleeding	Condition	MedDRA	LLT	C	10071910

2. Find standard concepts

```
SELECT * FROM concept WHERE lower(concept_name) LIKE '%upper gastrointestinal%'  
AND domain_id = 'Condition' AND standard_concept = 'S';
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
4000609	Disorder of upper gastrointestinal tract	Condition	SNOMED	Clinical Finding	S	119291004
4012503	Excessive upper gastrointestinal gas	Condition	SNOMED	Clinical Finding	S	162076009
4103011	Chronic upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	25349007
4115581	Finding of upper gastrointestinal gas	Condition	SNOMED	Clinical Finding	S	300370006
4291649	Upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	37372002
4308202	Acute upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	38938002
Upper gastrointestinal hemorrhage						
4332645	associated with hypercoagulability state	Condition	SNOMED	Clinical Finding	S	430349003

Going up the hierarchy: Finding the right concept

```
SELECT max_levels_of_separation,  
c.* FROM concept_ancestor ca,  
WHERE pta.Cdescendant_concept_id = 4332645 /* Upper gastrointestinal  
hemorrhage associated...*/  
AND ca.ancestor_concept_id = c.concept_id
```

Hold the descendant

max_levels_of_separation	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
0	4332645	Upper gastrointestinal hemorrhage associated with hypercoag	Condition	SNOMED	Clinical Finding	S	430349003
1	35708054	Gastritis haemorrhagic	Condition	MedDRA	PT	C	10017866
1	4291649	Upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	37372002
1	35707871	Upper gastrointestinal haemorrhage	Condition	MedDRA	PT	C	10046274
2	35707864	Gastrointestinal haemorrhage	Condition	MedDRA	PT	C	10017955
2	4000609	Disorder of upper gastrointestinal tract	Condition	SNOMED	Clinical Finding	S	119291004
2	35707858	Intestinal haemorrhage	Condition	MedDRA	PT	C	10059175
2	35702752	Gastritis (excl infective)	Condition	MedDRA	HLT	C	10017854
2	192671	Gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	74474003
3	37604042	Gastrointestinal haemorrhages	Condition	MedDRA	HLT	C	10052742
3	37622518	Haemorrhage	Condition	MedDRA	PT	C	10055798
3	437312	Bleeding	Condition	SNOMED	Clinical Finding	S	131148009
3	4198525	Disorder of upper digestive tract	Condition	SNOMED	Clinical Finding	S	50410009
3	37622515	Extravasation blood	Condition	MedDRA	PT	C	10015867
3	4000610	Disorder of gastrointestinal tract	Condition	SNOMED	Clinical Finding	S	119292006
3	35702116	Gastrointestinal inflammatory conditions	Condition	MedDRA	HLGT	C	10017969
3	35702743	Intestinal haemorrhages	Condition	MedDRA	HLT	C	10022653
3	35702744	Non-site specific gastrointestinal haemorrhages	Condition	MedDRA	HLT	C	10017958
4	35702114	Gastrointestinal haemorrhages NEC	Condition	MedDRA	HLGT	C	10017959
4	4304916	Gastrointestinal tract finding	Condition	SNOMED	Clinical Finding	S	386618008
4	35702767	Nausea and vomiting symptoms	Condition	MedDRA	HLT	C	10028817

Going down: Checking the right content

```
SELECT max_levels_of_separation, c.* FROM  
concept_ancestor ca, concept c  
  
WHERE ca.ancestor_concept_id = 4291649 /* Upper gastrointestinal hemorrhage */  
AND ca.descendant_concept_id = c.concept_id  
  
ORDER BY max_levels_of_separation;
```

max_levels_of_separation	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code
0	4291649	Upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	37372002
1	4318535	Duodenal hemorrhage					
1	23245	Esophageal bleeding					
1	4308202	Acute upper gastrointestinal hemorrhage					
1	4271696	Peptic ulcer with hemorrhage	Condition	SNOMED	Clinical Finding	S	64121000
1	4103011	Chronic upper gastrointestinal hemorrhage	Condition	SNOMED	Clinical Finding	S	25349007
1	26727	Hematemesis	Condition	SNOMED	Clinical Finding	S	8765009
1	4332645	Upper gastrointestinal hemorrhage associated with hypercoag	Condition	SNOMED	Clinical Finding	S	430349003
1	193250	Gastric hemorrhage	Condition	SNOMED	Clinical Finding	S	61401005
2	4131525	Hemorrhagic gastropathy	Condition	SNOMED	Clinical Finding	S	413218001
2	4204041	Hematemesis - cause unknown	Condition	SNOMED	Clinical Finding	S	308904008
2	4134808	Hemorrhagic duodenopathy	Condition	SNOMED	Clinical Finding	S	413212000
2	4260059	Hemorrhagic gastroenteritis	Condition	SNOMED	Clinical Finding	S	409506009
2	4099014	Duodenal ulcer with hemorrhage	Condition	SNOMED	Clinical Finding	S	27281001
2	46270145	Gastric hemorrhage due to atrophic gastritis	Condition	SNOMED	Clinical Finding	S	1.5072E+14
2	4096032	Duodenal hematoma	Condition	SNOMED	Clinical Finding	S	262843005
2	4174044	Chronic peptic ulcer with hemorrhage	Condition	SNOMED	Clinical Finding	S	49232000
2	4095555	Esophageal hematoma	Condition	SNOMED	Clinical Finding	S	262790002
2	46269904	Hemorrhage of duodenum co-occurrent and due to diverticul	Condition	SNOMED	Clinical Finding	S	1.0866E+15
2	45768629	Gastric hemorrhage due to erosive gastritis	Condition	SNOMED	Clinical Finding	S	7.071E+12

Concept 4291649 and all its descendants comprise Upper GI Bleeding

Athena: Upper Gastrointestinal Bleeding

ATHENA

Upper gastrointestinal bleeding

DETAILS		TERM CONNECTIONS (36)		HIERARCHY	RELATED CONCEPTS	
Domain ID	Condition	RELATIONSHIP	RELATES TO	CONCEPT ID	VOCABULARY	
Concept Class ID	Disorder	Active possibly_equivalent_to inactive (SNOMED)	Blood in vomit: [symptom] or [fresh blood]	40325778	SNOMED	
Vocabulary ID	SNOMED	Active same_as inactive (SNOMED)	Upper gastrointestinal hemorrhage	40394270	SNOMED	
Concept ID	4291649	Associated finding of (SNOMED)	History of upper gastrointestinal tract hemorrhage	4207485	SNOMED	
Concept code	37372002	Has Module	SNOMED CT core	40642539	SNOMED	
Validity	Valid	Has associated morphology (SNOMED)	Hemorrhage	4179955	SNOMED	
Concept	Standard	Has finding site (SNOMED)	Upper gastrointestinal tract structure	4271189	SNOMED	
LANGUAGE	SYNONYM CONCEPT	Has status	Defined	40642537	SNOMED	
English	Upper GI - gastrointestinal haemorrhage	Is a	Disorder of upper gastrointestinal tract	4000609	SNOMED	
English	Upper GI - gastrointestinal hemorrhage					

LS
DEVELOPMENT
INSTITUTE

BACK
15 IN
MINS



Hands-on Exercises

Learning Goals

- By the end of this exercise, participants will be able to:
- Navigate **Athena** and locate clinical concepts across vocabularies.
- Distinguish between **standard** and **non-standard** concepts.
- Interpret concept **relationships** (“Maps to,” “Is a,” “Has ancestor,” etc.).
- Recognize the structure and purpose of **OMOP vocabularies** and **domains**.
- Understand how vocabulary choice impacts analytic consistency and data quality.

Warm up Exercise: Find Standard Concept ID for Conditions

Asthma

Plague

Ingrown toenail

Your favorite condition here

Search for a Clinical Condition

- Open Athena.
- Search for “**Type 2 Diabetes Mellitus.**”
- Identify:
 - The **standard concept**
 - A related **non-standard concept**
 - The **domain, vocabulary, and concept class**

Volunteers?



SEARCH BY KEYWORD

Type 2 Diabetes Mellitus



SEARCH

DOWNLOAD

LOGIN



Type 2 Diabetes Melli... X

Standard X

DOWNLOAD RESULTS

Show by 15 items Total 210,539 items

1 2 3 4 5 ... 14036 >

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
201826	44054006	Type 2 diabetes mellitus	Disorder	Standard	Valid	Condition	SNOMED
4063043	199230006	Pre-existing type 2 diabetes mellitus	Disorder	Standard	Valid	Condition	SNOMED
40485020	444110003	Well controlled type 2 diabetes mellitus	Disorder	Standard	Valid	Condition	SNOMED
4129519	237627000	Pregnancy and type 2 diabetes mellitus	Disorder	Standard	Valid	Condition	SNOMED
4130162	237599002	Insulin treated type 2 diabetes mellitus	Disorder	Standard	Valid	Condition	SNOMED
37016349	36805100011 9109	Hyperglycemia due to type 2 diabetes mellitus	Disorder	Standard	Valid	Condition	SNOMED
13793100011							

Review Concept Details and Hierarchies

- Open the **Concept Details** page for your chosen concept.
- Explore relationships, ancestors / descendants, and concept class.

Volunteers?

https://athena.ohdsi.org/search-terms/terms/201826

The screenshot shows the ATHENA search interface at the URL https://athena.ohdsi.org/search-terms/terms/201826. The top navigation bar includes links for SEARCH, DOWNLOAD, LOGIN, and HELP. The main search bar contains the query "Type 2 diabetes mellitus". Below the search bar, there are two tabs: "DETAILS" and "TERM CONNECTIONS (423)". The "DETAILS" tab displays basic concept information: Domain ID (Condition), Concept Class ID (Disorder), Vocabulary ID (SNOMED), Concept ID (201826), Concept code (44054006), Validity (Valid), and Concept (Standard). The "TERM CONNECTIONS" tab lists 423 related concepts across four columns: RELATIONSHIP, RELATED TO, CONCEPT ID, and VOCABULARY. Relationships include "Active possibly equivalent_to inactive (SNOMED)" and "Active same_as inactive (SNOMED)". Related terms include "Diabetes mellitus: [adult onset, with no mention of complication] or [maturity onset] or [non-insulin dependent]", "Type II diabetes mellitus", and "Family history of diabetes mellitus type 2". All related concepts are from the SNOMED vocabulary.

Explore Relationships

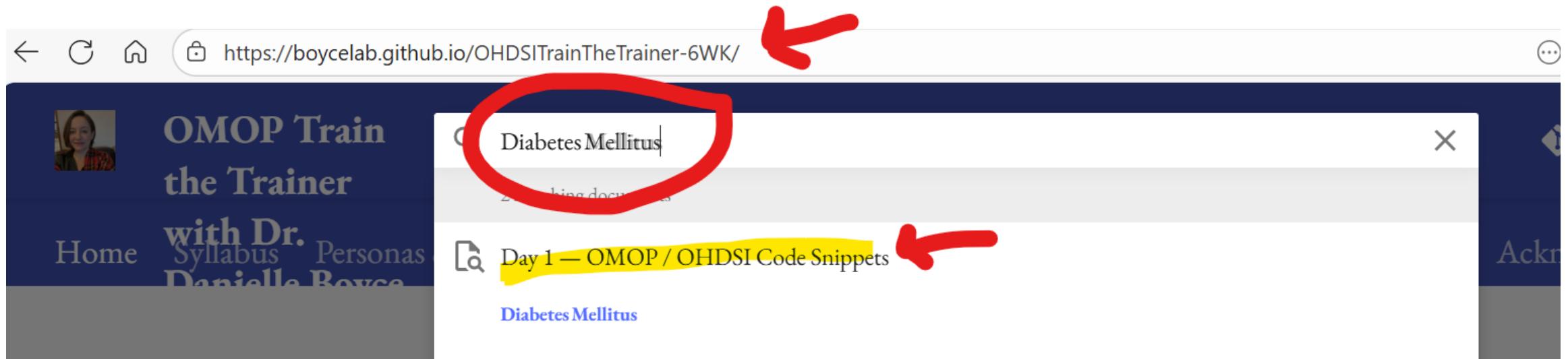
Choose a **non-standard ICD10CM** code for Type 2 Diabetes and inspect its mappings.

THE DRUG DISCOVERY ENGINE FOR ALS

ALS
THERAPY DEVELOPMENT
INSTITUTE

Volunteers?

Repeat in SQL



Conclusions

OMOP CDM

- Is used to **standardize structure and queries**
- Integrated with **Controlled Vocabulary**
- Consolidates data from **heterogeneous data sources**: EMR, claims, registries
- Patient centric
- Domain (subject area) based: concepts decide what table each piece of data lands on
- Preserves data provenance
- Database platform independent

OMOP Vocabularies

- Is used to **standardize terminology**
- **Compiles standards** from disparate public and private sources and some OMOP-grown concepts
- Has **one uniform structure** to house multiple vocabularies used in the public domain
- Is designed to **facilitate efficient queries**
- Is **regularly updated, maintained, and improved**

What Makes OMOP CDM Unique

- Supports collaborative research across data sources both within and outside of US
- Developed based on analytic use cases by community of collaborators
- Specialized: reflective of clinical domain, granular, well structured
- Integrated with Vocabulary that is uniformly structured and well curated
- Extendable: new concepts and attributes can be added
- Supported by Community of interdisciplinary developers and researchers

Homework

Readings: The Book of OHDSI Chapter 10, Defining Cohorts

Exercises

[Chapter 5 Standardized Vocabularies | The Book of OHDSI](#)

Only the first two!

Exercise 5.1 What is the Standard Concept ID for “Gastrointestinal hemorrhage”?

Exercise 5.2 Which ICD-10CM codes map to the Standard Concept for “Gastrointestinal hemorrhage”? Which ICD-9CM codes map to this Standard Concept?



THE DRUG DISCOVERY
ENGINE FOR **ALS**

Thank you. Questions?

Danielle Boyce

dboyce@als.net

