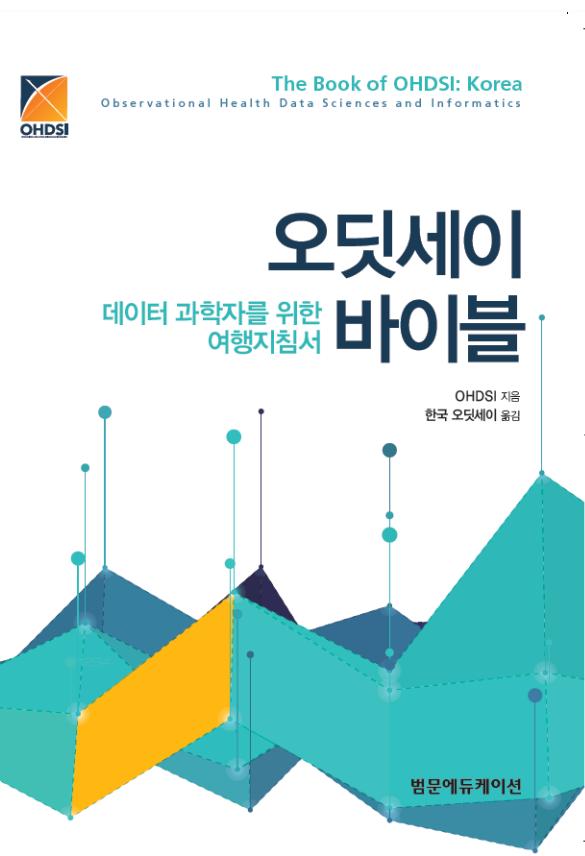




# Basic Tutorial: OMOP Common Data Model and Standardized Vocabularies

아주대학교 의료정보학과 이동윤



# The Book of OHDSI

Part I “The OHDSI community”  
Part II “Uniform Data Representation”

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



# What is OMOP/OHDSI?

Foundational



OMOP?  
OHDSI?  
Common Data Model?  
OMOP CDM?

# FDAAA calls for establishing Risk Identification and Analysis System

## SEC. 905. ACTIVE POSTMARKET RISK IDENTIFICATION AND ANALYSIS.

(a) IN GENERAL.—Subsection (k) of section 505 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 355) is amended by adding at the end the following:

“(3) ACTIVE POSTMARKET RISK IDENTIFICATION.—

“(A) DEFINITION.—In this paragraph, the term ‘data’ refers to information with respect to a drug approved under this section or under section 351 of the Public Health Service Act, including claims data, patient survey data, standardized analytic files that allow for the pooling and analysis of data from disparate data environments, and any other data deemed appropriate by the Secretary.

“(B) DEVELOPMENT OF POSTMARKET RISK IDENTIFICATION AND ANALYSIS METHODS.—The Secretary shall, not later than 2 years after the date of the enactment of the Food and Drug Administration Amendments Act of 2007, in collaboration with public, academic, and private entities—

“(i) develop methods to obtain access to disparate data sources including the data sources specified in subparagraph (C);

“(ii) develop validated methods for the establishment of a postmarket risk identification and analysis system to link and analyze safety data from multiple sources, with the goals of including, in aggregate—

“(I) at least 25,000,000 patients by July 1, 2010; and

“(II) at least 100,000,000 patients by July 1, 2012; and

“(iii) convene a committee of experts, including individuals who are recognized in the field of protecting data privacy and security, to make recommendations to the Secretary on the development of tools and methods for the ethical and scientific uses for, and communication of, postmarketing data specified under subparagraph (C), including recommendations on the development of effective research methods for the study of drug safety questions.

“(C) ESTABLISHMENT OF THE POSTMARKET RISK IDENTIFICATION AND ANALYSIS SYSTEM.—



## The Sentinel Initiative

National Strategy for Monitoring Medical Product Safety

May 2008



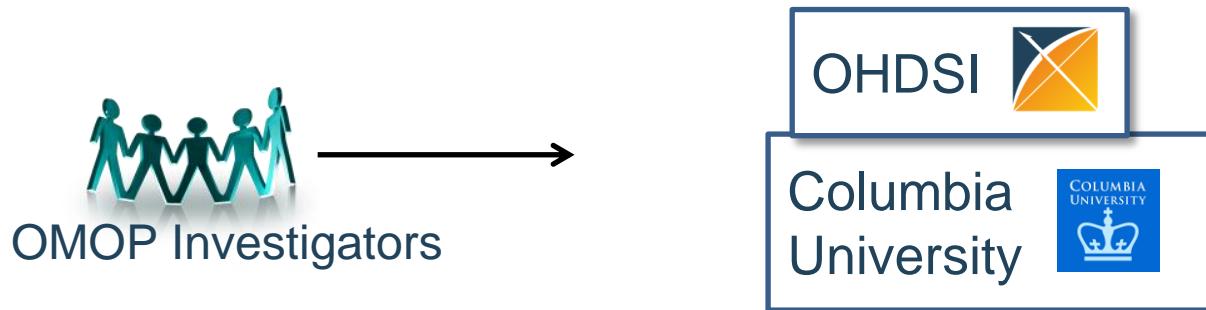
## Risk Identification and Analysis System:

a systematic and reproducible process to efficiently generate evidence to support the characterization of the potential effects of medical products from across a network of disparate observational healthcare data sources

# Main findings in OMOP experiment

- Heterogeneity in estimates due to choice of database
- Heterogeneity in estimates due to analysis choices
- Except little heterogeneity due to outcome definitions
- Good performance ( $AUC > 0.7$ ) in distinguishing positive from negative controls for optimal methods when stratifying by outcome and restricting to powered test cases
- Self controlled methods perform best for all outcomes

# Fate of OMOP - OHDSI



- The Observational Health Data Sciences and Informatics (OHDSI) program is a **multi-stakeholder, interdisciplinary collaborative** to create **open-source** solutions that bring out the value of observational health data through large-scale analytics
- OHDSI has established **an international network of researchers and observational health databases** with a central coordinating center housed at Columbia University
  - Public, Open
  - Not Pharma-funded
  - International

<http://ohdsi.org>

# OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

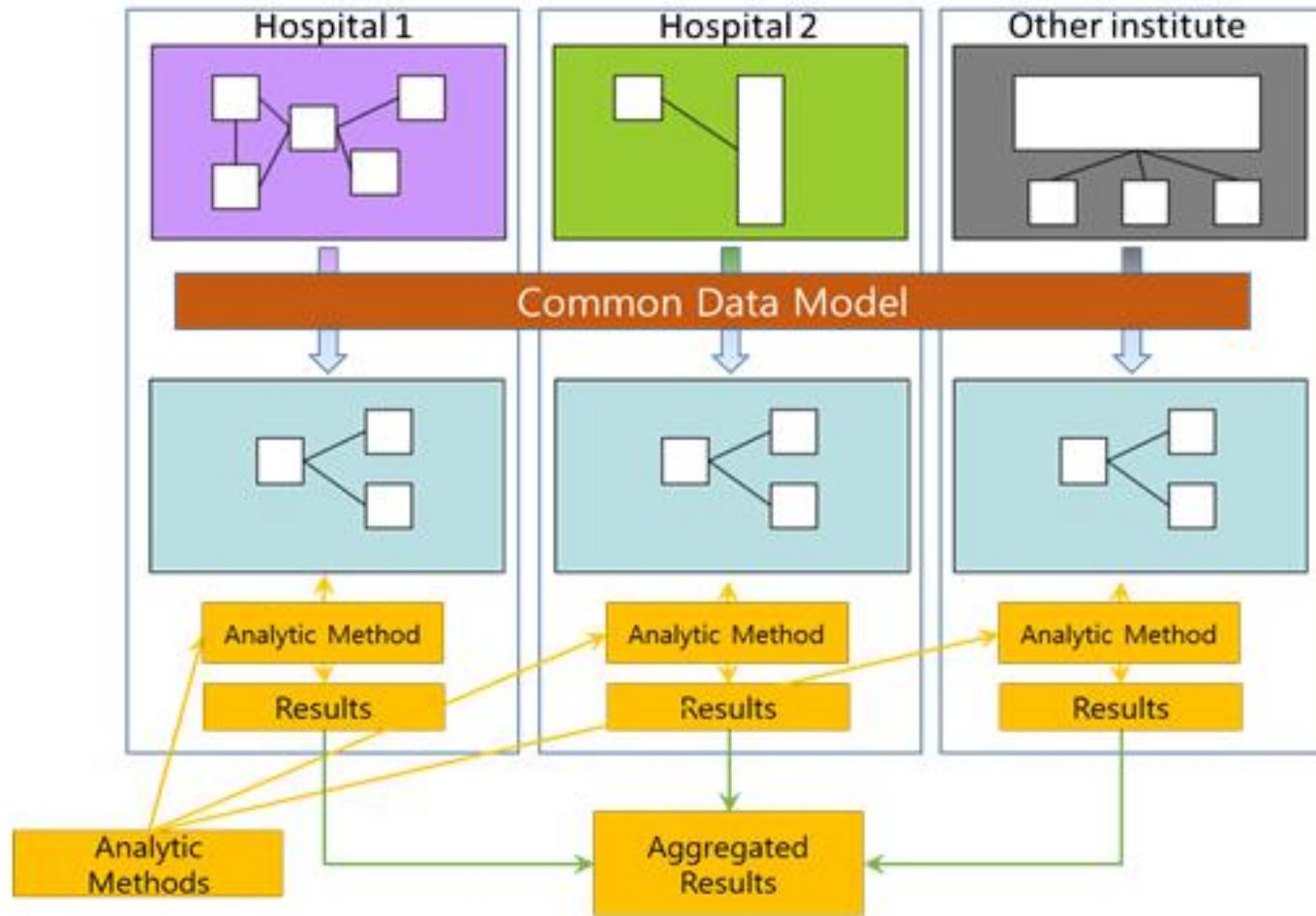


## Areas of Focus

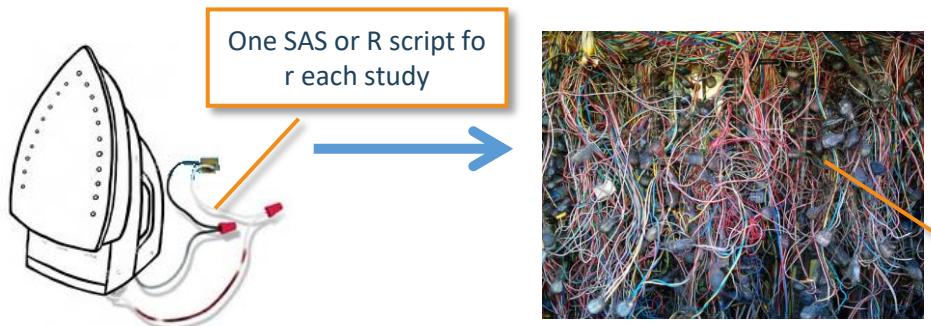
- Data Standardization
- Medical product safety surveillance
- Comparative effectiveness research
- Personalized risk prediction
- Data characterization
- Quality improvement

21+ Countries, 114+ Databases (2019)

# Common Data Model (CDM)

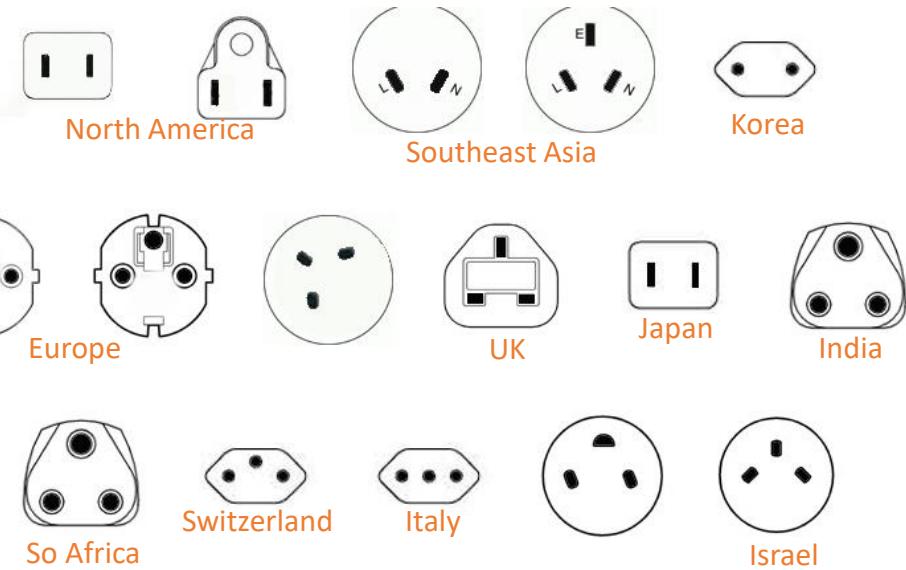
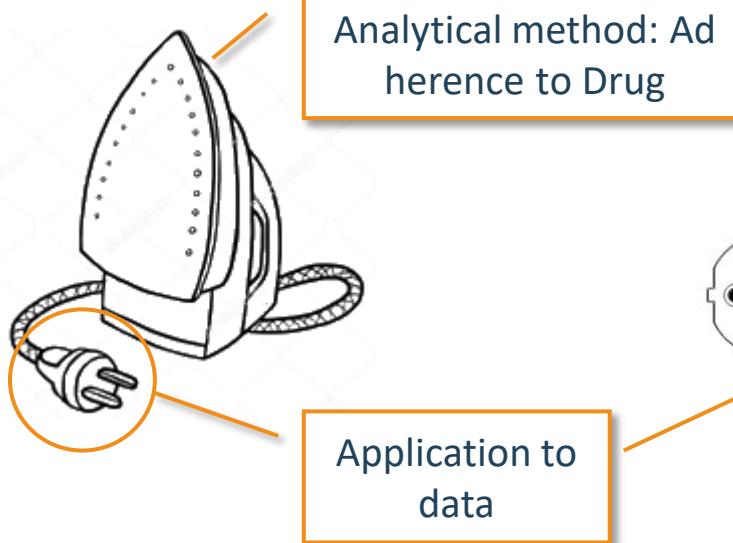


# Common Data Model

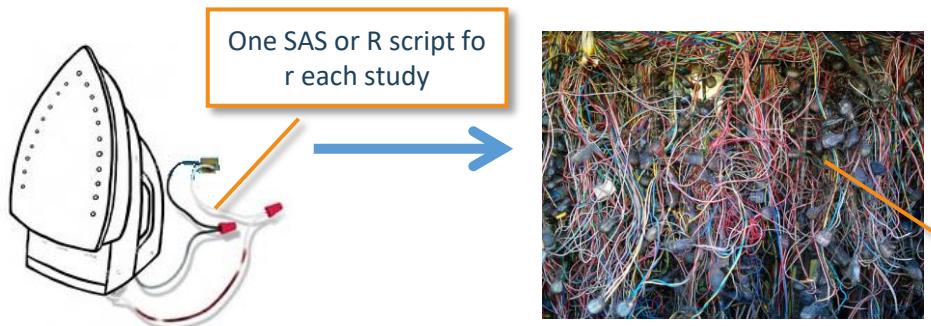


- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert routine use

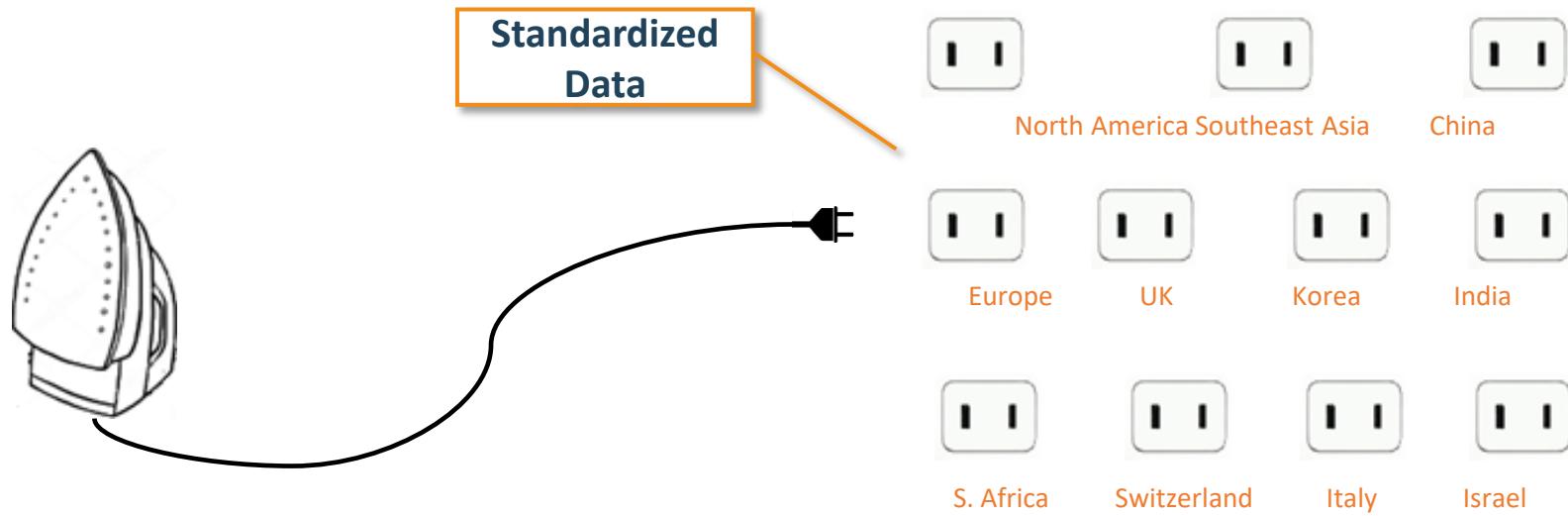
**“One Study – One Script”**



# Common Data Model



- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert routine use



# Common Data Model



North America    Southeast Asia

China



Europe

UK

Korea

India

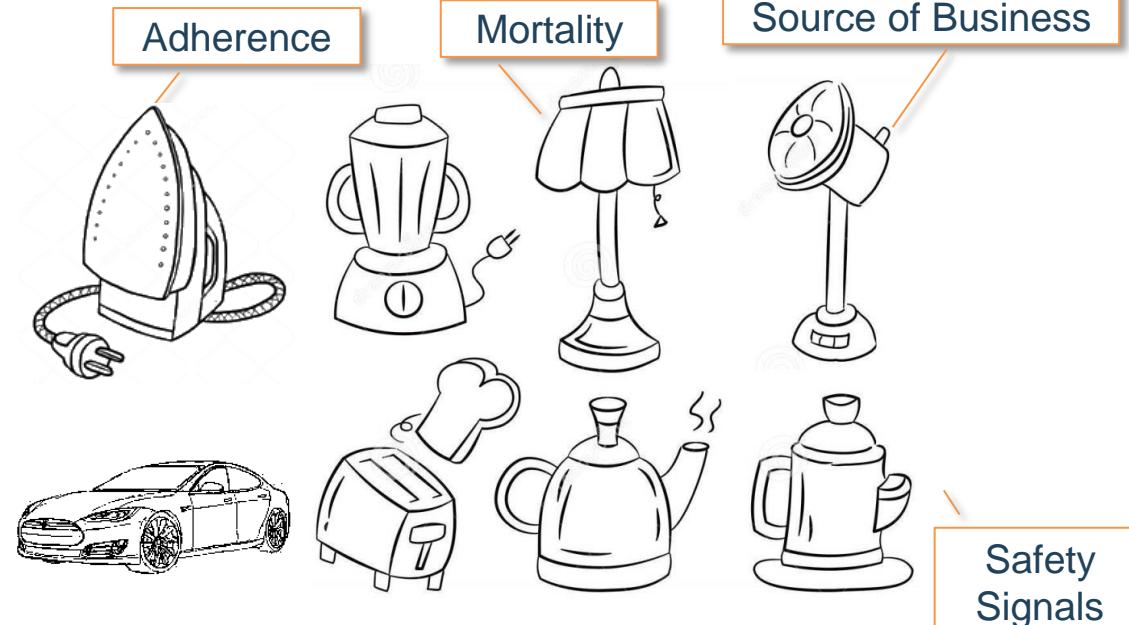


So Africa

Switzerland

Italy

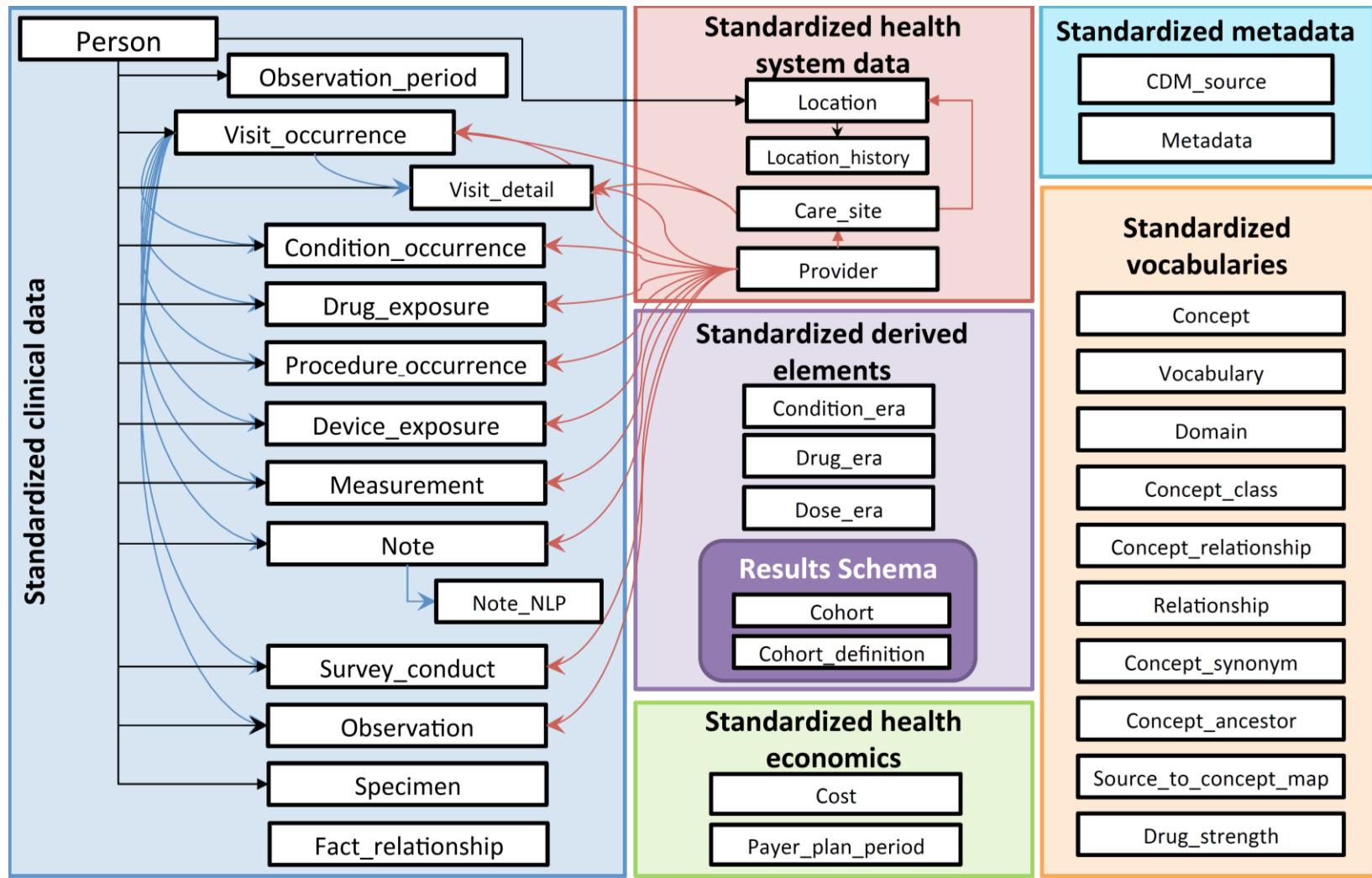
Israel



Standardized Data

OHDSI Tools

# OMOP CDM V6



# Data to evidence

## Different types of observational data:

### Populations

- Pediatric vs. elderly
- Socioeconomic disparities

### Care setting

- Inpatient vs. outpatient
- Primary vs. secondary care

### Data capture process

- Administrative claims
- Electronic health records
- Clinical registries

### Health system

- Insured vs. uninsured
- Country policies



## Types of evidence desired:

### Clinical characterization

- Clinical trial feasibility
- Treatment utilization
- Disease natural history
- Quality improvement

### Population-level effect estimation

- Safety surveillance
- Comparative effectiveness

### Patient-level prediction

- Precision medicine
- Disease interception

# Open Science and FHIR guide

Open Standards  
Open Source  
Open Data  
Open Discourse



Findability  
Accessibility  
Interoperability  
Reusability





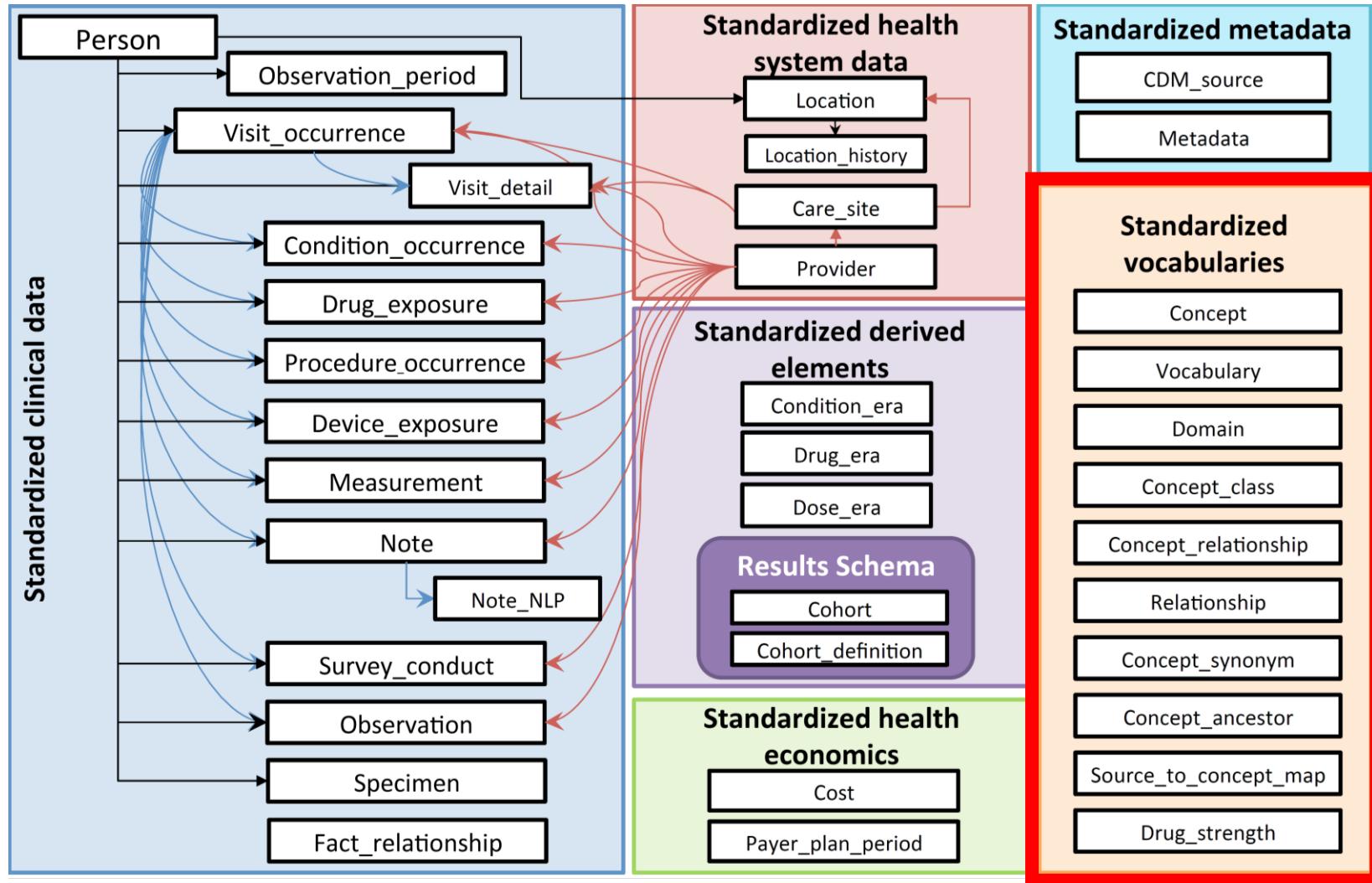
# OMOP Vocabulary

Concept  
Basic relationship,  
Ancestor & Descendant



Concept?  
Concept\_id?  
Concept\_code?  
Concept\_relationship  
Concept\_ancestor

# OMOP Common data model V6



# OMOP Standardized Vocabulary

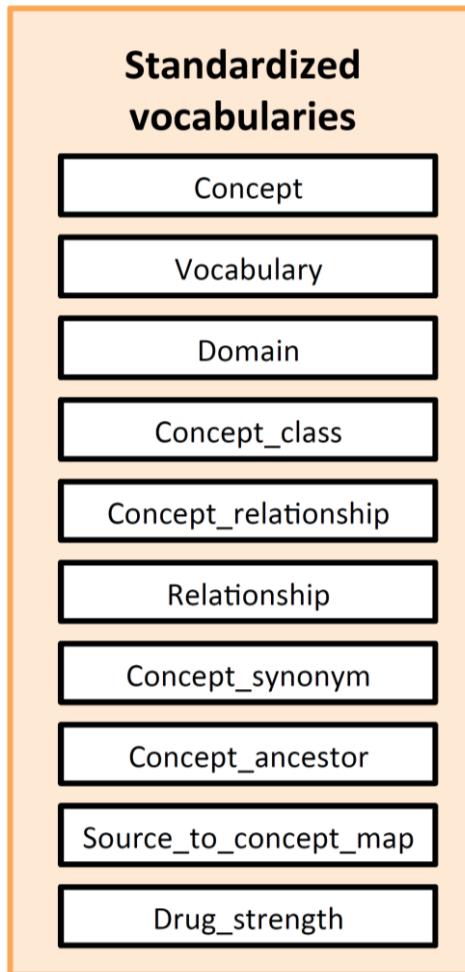
**It is**

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

**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

# OMOP Standardized Vocabulary



## Principles

- **Standardization**  
: Multiple vocabulary → Common format
- **Unique Standard Concepts**  
: Other equivalent concepts are mapped to the Standard concepts
- **Domains**  
: Each concept is assigned a Domain
- **Comprehensive coverage**  
: Patient's healthcare experience and administrative artifacts
- **Hierarchy**  
: Hierarchical structure
- **Relationships**  
: Concept - concept, Standard - Non-standard
- **Life cycle**  
: Keeping up to date

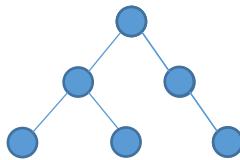
# What is concept?



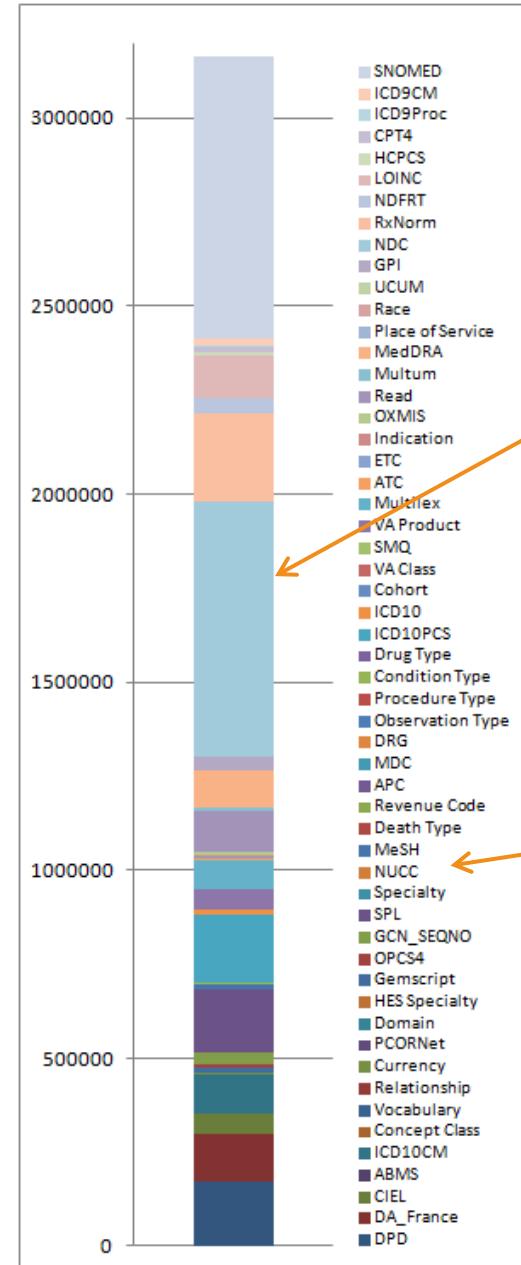
All content: concepts in  
**concept**



Direct relationships  
between concepts in  
**concept\_relationship**



Multi-step hierarchical  
relationships pre-processed  
into **concept\_ancestor**



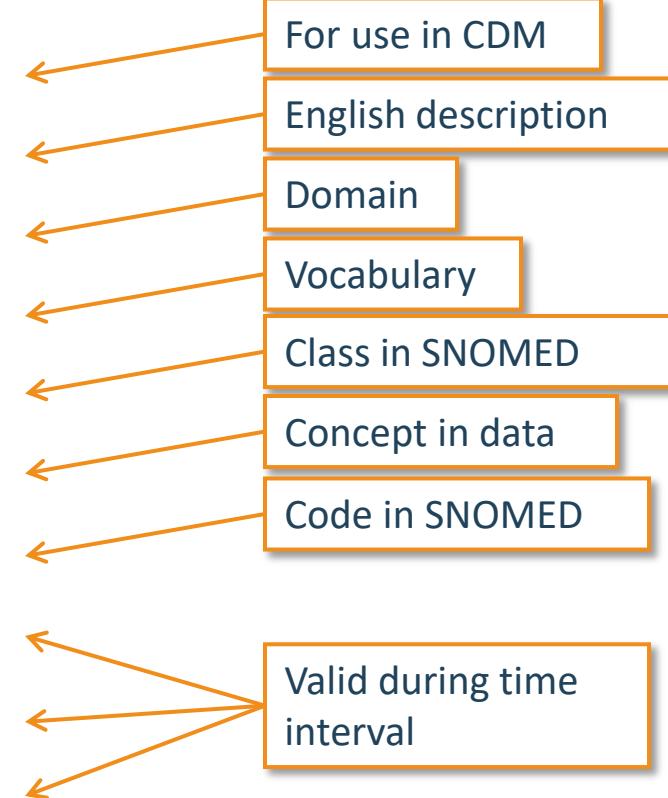
All vocabularies stacked  
up in one table

Vocabulary ID

[https://github.com/  
OHDSI/Vocabulary-  
v5.0/releases](https://github.com/OHDSI/Vocabulary-v5.0/releases)

# What is 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-1970
VALID_END_DATE	31-Dec-2099
INVALID_REASON	



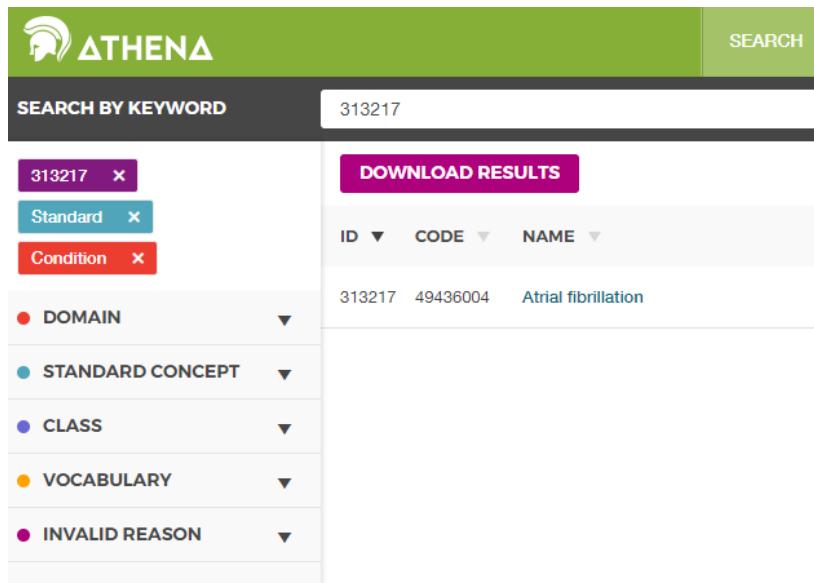
Domain	for Standard Concepts	for source concepts	for classification concepts
Condition	SNOMED, ICD03	SNOMED Veterinary	MedDRA
Procedure	SNOMED, CPT4, HCPCS, ICD10PCS, ICD9Proc, OPCS4	SNOMED Veterinary, HemOnc, NAACCR	None at this point
Measurement	SNOMED, LOINC	SNOMED Veterinary, NAACCR, CPT4, HCPCS, OPCS4, PPI	None at this point
Drug	RxNorm, RxNorm Extension, CVX	HCPCS, CPT4, HemOnc, NAACCR	ATC
Device	SNOMED	Others, currently not normalized	None at this point
Observation	SNOMED	Others	None at this point
Visit	CMS Place of Service, ABMT, NUCC	SNOMED, HCPCS, CPT4, UB04	None at this point

# What is concept?

- Concepts are referred to by concept\_id
- All details are in the CONCEPT table:

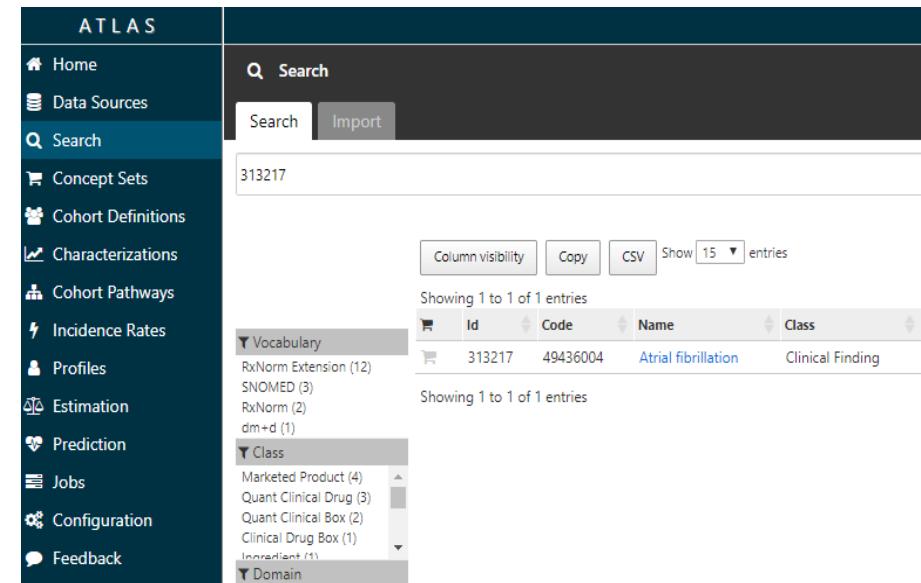
```
SELECT * FROM concept
WHERE concept_id = 313217
```

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	1970-01-01	2099-12-31	NULL



The ATHENA search interface displays the following results for concept\_id 313217:

- SEARCH BY KEYWORD: 313217
- FILTERS:
  - 313217 (selected)
  - Standard
  - Condition
  - DOMAIN
  - STANDARD CONCEPT
  - CLASS
  - VOCABULARY
  - INVALID REASON
- DOWNLOAD RESULTS: ID, CODE, NAME
- RESULTS: 313217, 49436004, Atrial fibrillation



The ATLAS search interface displays the following results for concept\_id 313217:

- ATLAS sidebar:
  - Home
  - Data Sources
  - Search (selected)
  - Concept Sets
  - Cohort Definitions
  - Characterizations
  - Cohort Pathways
  - Incidence Rates
  - Profiles
  - Estimation
  - Prediction
  - Jobs
  - Configuration
  - Feedback
- SEARCH:
  - Search: 313217
  - Column visibility, Copy, CSV, Show 15 entries
  - Showing 1 to 1 of 1 entries
  - RESULTS: Id, Code, Name, Class
  - RESULTS: 313217, 49436004, Atrial fibrillation, Clinical Finding
  - EXPANDED Vocabularies:
    - RxNorm Extension (12)
    - SNOMED (3)
    - RxNorm (2)
    - dm-d (1)
  - EXPANDED Class:
    - Marketed Product (4)
    - Quant Clinical Drug (3)
    - Quant Clinical Box (2)
    - Clinical Drug Box (1)
    - Ingredient (1)
  - EXPANDED Domain

# Concept ID vs Concept Code

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

Unique ID

SNOMED-CT  
399,109

LOINC  
72,625

RxNorm  
218,011

MedDRA  
94,176

Multilex  
77,177

VA Product  
52,297

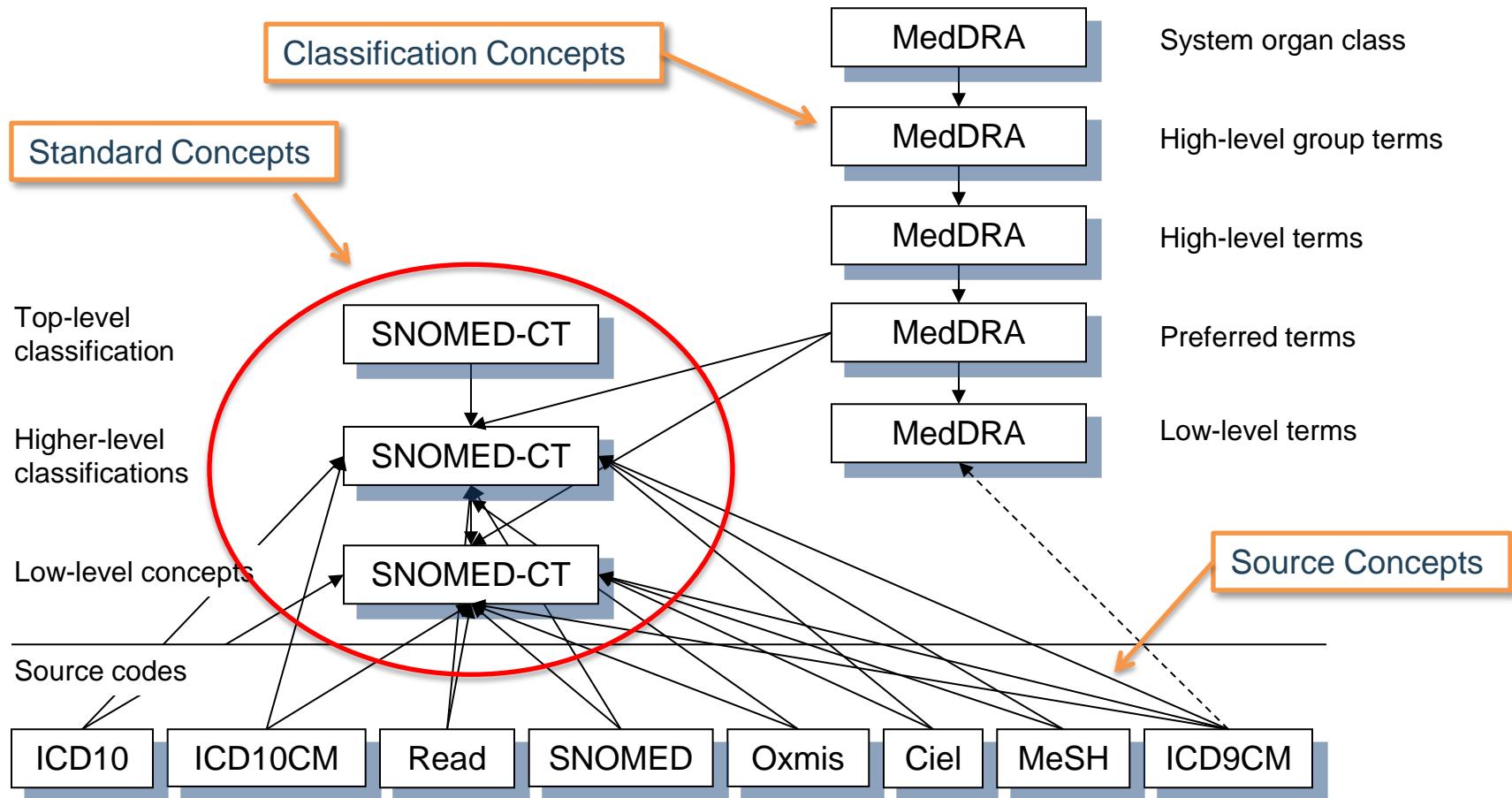
concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1970-01-01	2099-12-31	NULL

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

Concept_ID	Concept_Code	Concept_Name	Concept Class	Vocabulary_ID
1036059	1001	Antipyrine	Ingredient	RxNorm
43228317	1001	Aceprometazine maleate	Ingredient	BDPM
45912144	1001	Serum	Specimen	CIEL
45417187	1001	Brompheniramine Maleate, 10 mg/mL injectable solution	Multum	Multum
38003544	1001	Residential Treatment – Psychiatric	Revenue Code	Revenue Code

Same code

# Condition Concepts



# Finding the Right Concept #1

1. ..if I know the **concept\_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	

2. ..if I know the **concept\_code**

SNOMED 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	

# Finding the Right Concept #2

## 3. ..if I know the name

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

CONCEPT_ID	CONCEPT_NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT	CONCEPT_CODE
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31
35204953	Atrial fibrillation	Condition	MedDRA	PT	C	10003658
45500085	Atrial fibrillation	Condition	Read	Read		G573000
45883018	Atrial fibrillation	Meas Value	LOINC	Answer	S	LA17084-7

# Finding the Right Concept #3

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

ICD-9 is not a Standard Concept

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

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

1	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	

# Mapping = Translating

## Step 1. Lookup the Source Concept

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

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

## Step 2. Translate to Standard

```
SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957 AND relationship_id = 'Maps to';
```

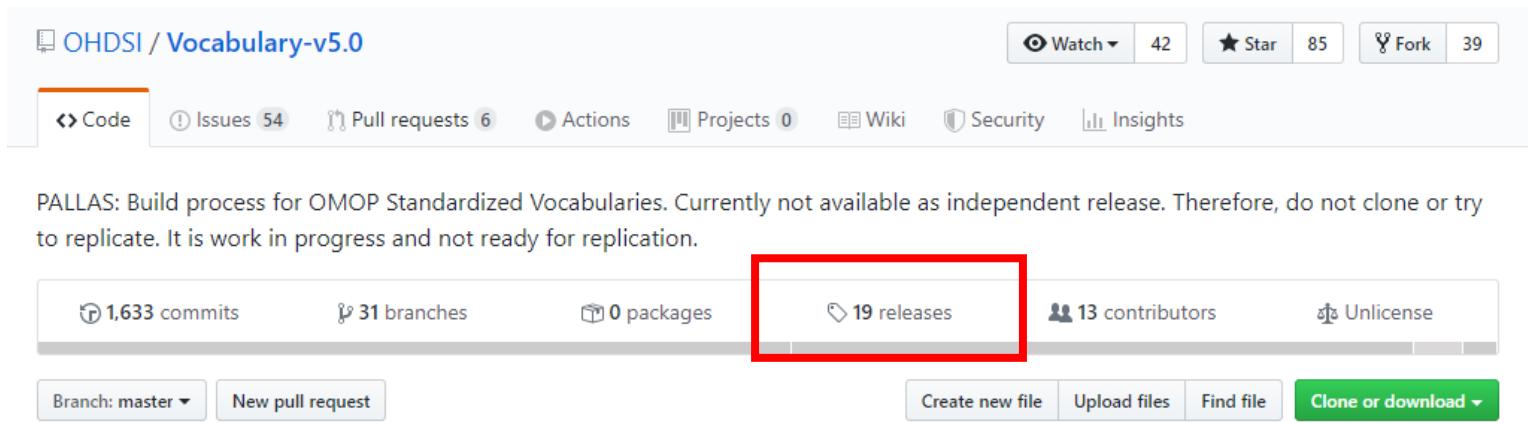
CONCEPT_ID_1	CONCEPT_ID_2	RELATIONSHIP_ID	VALID_START_DATE	VALID_END_DATE	INVALID_REASON
44821957	313217	Maps to	01-Jan-1970	31-Dec-2099	



## Step 3. Check out the translated Concept

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

- <https://github.com/OHDSI/vocabulary-v5.0>



PALLAS: Build process for OMOP Standardized Vocabularies. Currently not available as independent release. Therefore, do not clone or try to replicate. It is work in progress and not ready for replication.

1,633 commits 31 branches 0 packages 19 releases 13 contributors Unlicense

Branch: master ▾ New pull request Create new file Upload files Find file Clone or download ▾

## Vocabulary

This repo contains the build process for the Standard Vocabularies, which is part of ATHENA. We will consolidate this with the download page at some point in the future.

You do not have to clone and run this to create a vocabulary. Instead, download all the necessary files from <http://athena.ohdsi.org>

If you do want to recreate the vocabulary please find the step-by-step description in each vocabulary's README.

Additional info form drug vocabularies: [http://www.ohdsi.org/web/wiki/doku.php?id=documentation:international\\_drugs](http://www.ohdsi.org/web/wiki/doku.php?id=documentation:international_drugs)

Christian and the Odysseus Vocabulary Team. [reich@ohdsi.org](mailto:reich@ohdsi.org)

OHDSI / Vocabulary-v5.0

Code Issues 54 Pull requests 6 Actions Projects 0 Wiki Security Insights

Releases Tags

Latest release v20191114\_15737... e11a965

## Release notes v20191114

OHDSIVocabularyReleaseRobot released this 11 days ago · 3 commits to master since this release

### Domain changes

vocabulary_id	old_domain_id	new_domain_id	count
ICD10	Condition	Observation	1
ICD10	Observation	Condition	1
ICD10CM	Condition	Observation	22
ICD10CM	Observation	Condition	2339
ICD10CM	Procedure	Observation	306
ICD9CM	Condition	Observation	3
ICD9CM	Measurement	Observation	1
ICD9CM	Procedure	Observation	10
KCD7	Condition	Observation	1
SNOMED	Condition	Observation	4
SNOMED	Condition	Race	150
SNOMED	Drug	Device	987
SNOMED	Drug	Observation	309
SNOMED	Observation	Condition	2
SNOMED	Observation	Device	2123
SNOMED	Observation	Drug	139
SNOMED	Observation	Procedure	2
SNOMED	Observation	Provider	1
SNOMED	Procedure	Measurement	14

# Concept Relationships

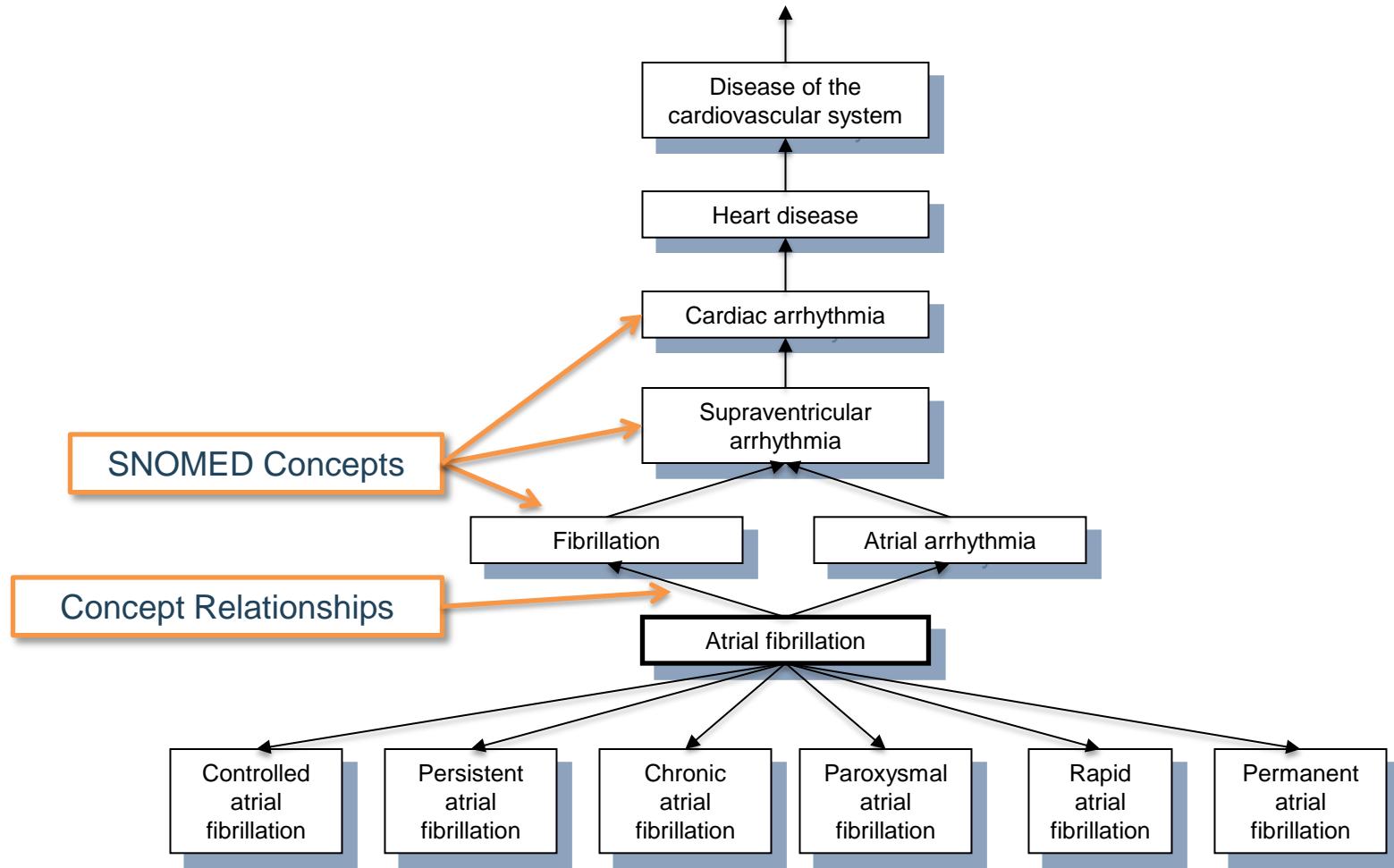
```
SELECT *
FROM concept_relationship
WHERE concept_id_1 = 313217
```

Related Concepts

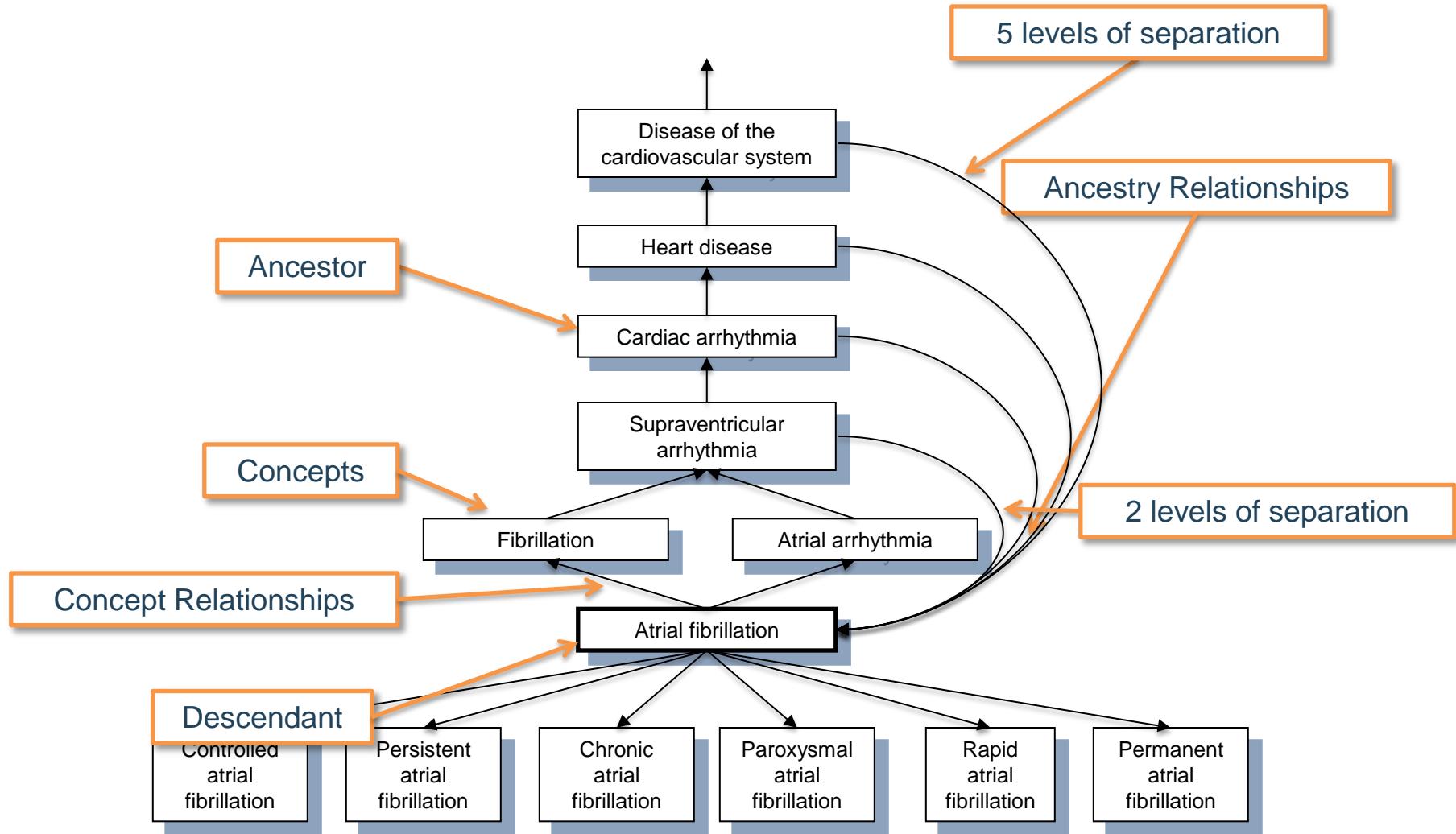
Relationship ID

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/CI
313217	40248987	SNOMED - ind/CI
313217	21001551	SNOMED - ind/CI
313217	21001540	SNOMED - ind/CI
313217	45576876	Mapped from
313217	44807374	Asso finding of
313217	21013834	SNOMED - ind/CI
313217	21001572	SNOMED - ind/CI
313217	21001606	SNOMED - ind/CI
313217	21003176	SNOMED - ind/CI
313217	4226399	Is a
313217	500001801	SNOMED - HOI
313217	500002401	SNOMED - HOI
313217	411962	Subsumes
313217	40631039	Subsumes
313217	4108832	Subsumes
313217	21013671	SNOMED - ind/CI
313217	21013390	SNOMED - ind/CI

# Concept Hierarchy - Disease



# 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

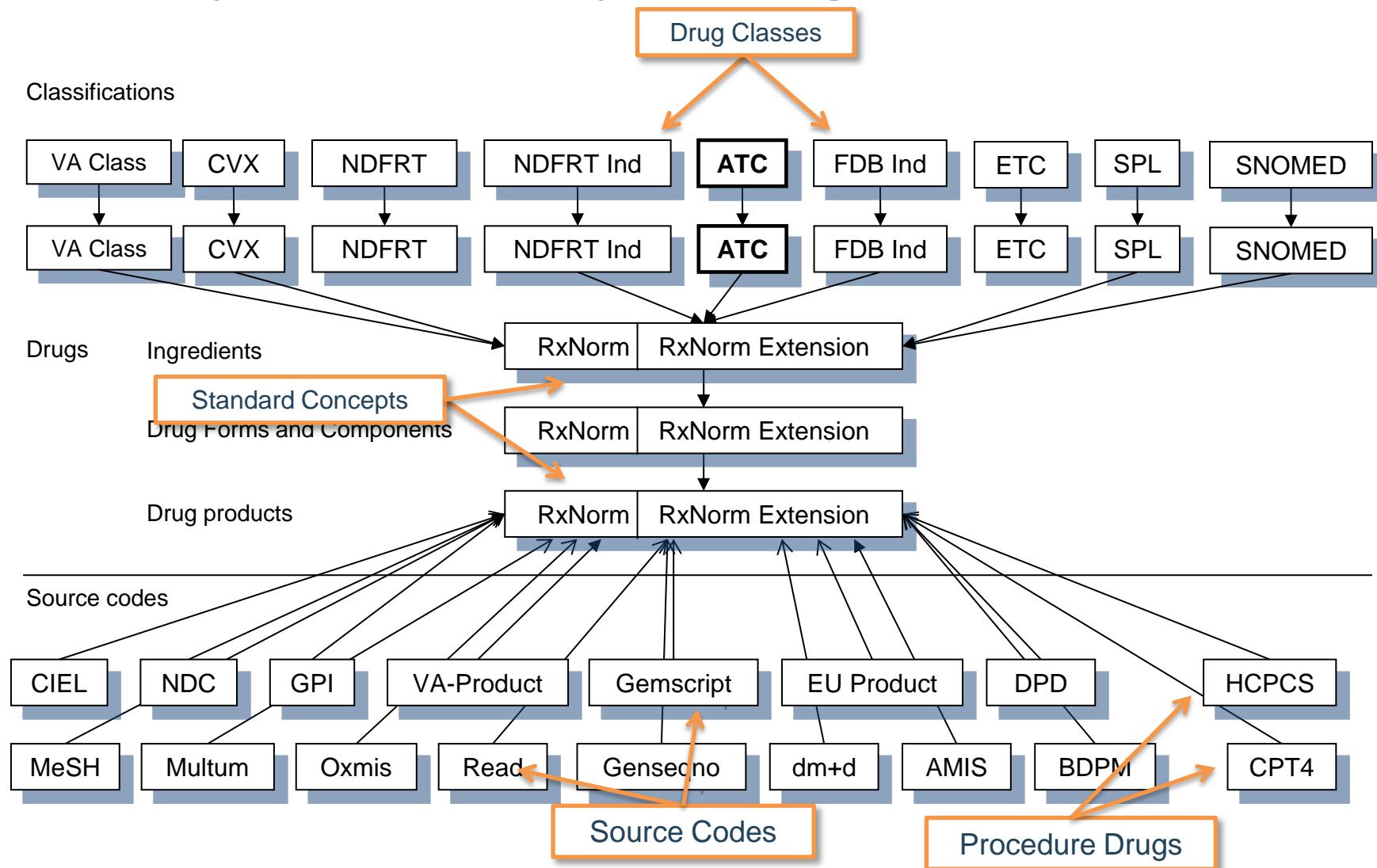
# Exploring Descendants of a Concept

```
SELECT max_levels_of_separation, concept.*  
FROM concept_ancestor  
JOIN concept ON descendant_concept_id = concept_id  
WHERE ancestor_concept_id = 44784217 /* cardiac arrhythmia */  
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

# Concept Hierarchy - Drug





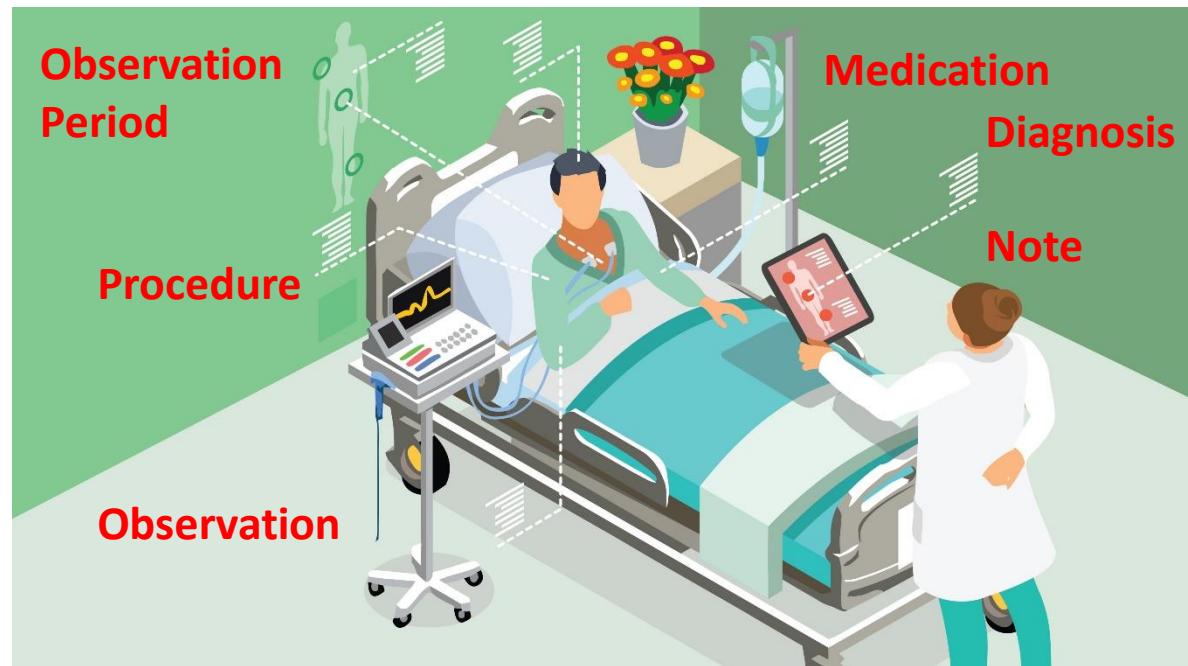
# Common Data Model Structure

OMOP CDM

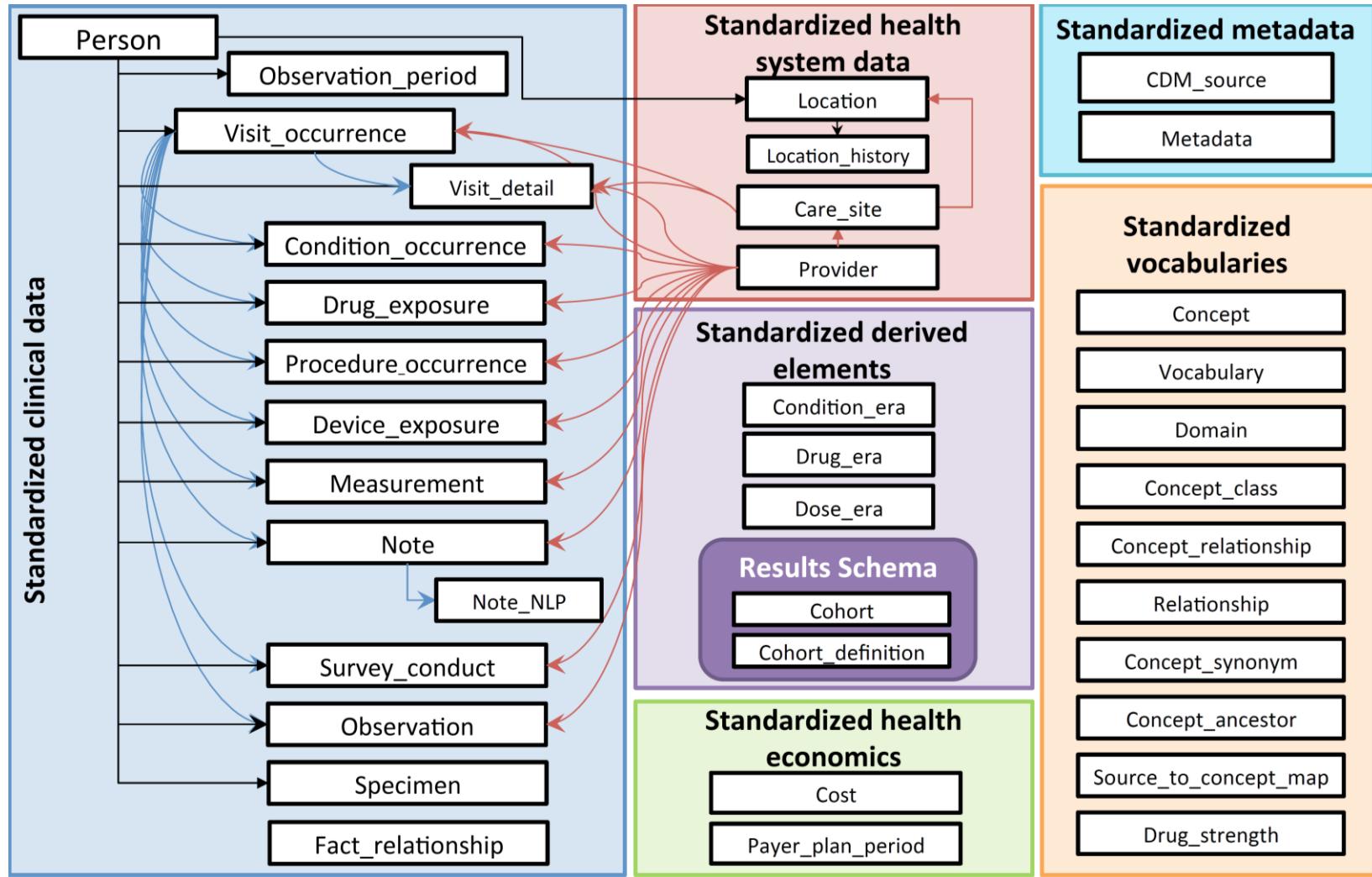


OHDSI

## Where is the patient's information ?



# OMOP Common data model V6



# OMOP CDM Principles

- Suitability for purpose
- Data protection
- Design of domains
- Rationale for domains
- Standardized  
Vocabularies
- Reuse of existing  
vocabularies
- Maintaining source  
codes
- Technology neutrality
- Scalability
- Backwards compatibility

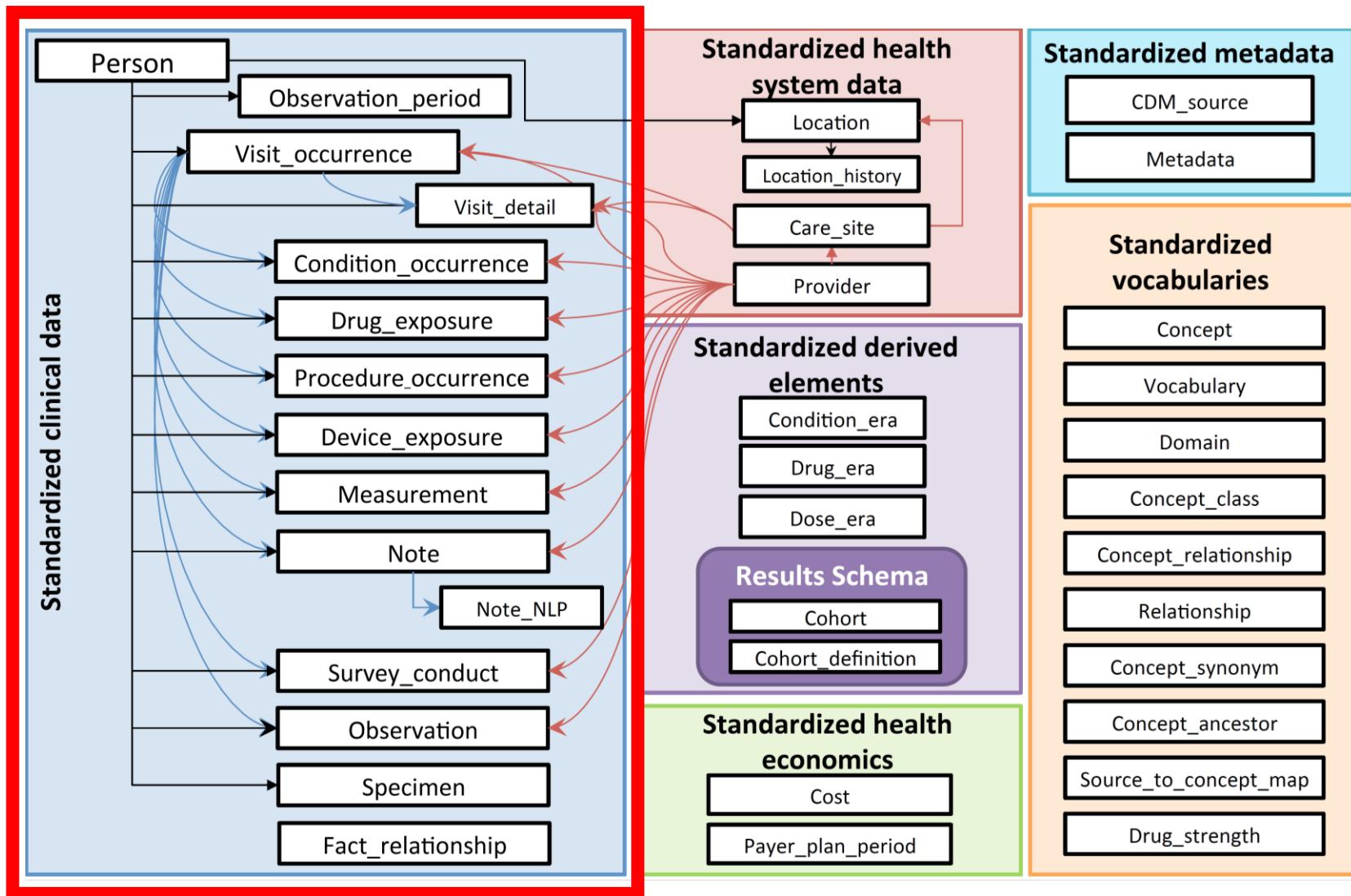
# Standard concepts belonging to each domain

Concept Count	Domain ID	Concept Count	Domain ID	Concept Count	Domain ID
1731378	Drug	1046	Unit	50	Race
477597	Device	944	Metadata	13	Plan Stop Reason
257000	Procedure	538	Revenue Code	11	Plan
163807	Condition	336	Type Concept	6	Episode
145898	Observation	194	Relationship	6	Sponsor
89645	Measurement	183	Route	5	Meas Value Operator
33759	Spec Anatomic Site	180	Currency	3	Spec Disease Status
17302	Meas Value	158	Payer	2	Gender
1799	Specimen	123	Visit	2	Ethnicity
1215	Provider Specialty	51	Cost	1	Observation Type

# OMOP CDM Standard Domain Features

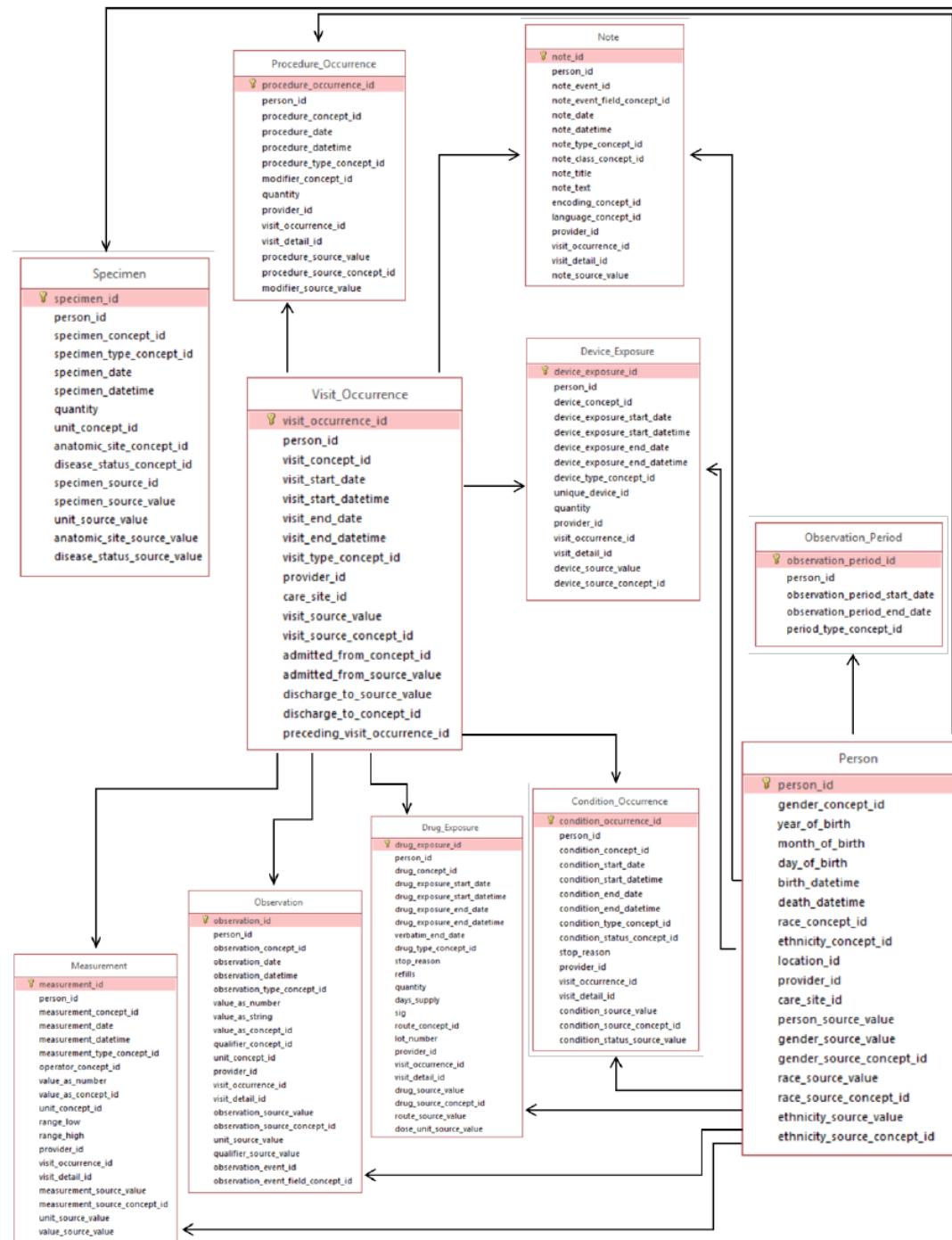
Field Name Convention	Example	Feature	Description & Purpose
person_id	person_id 123	Patient centric	Every domain table has <b>patient identifier</b> . Patient data can be retrieved independently from other domains.
<entity>_id	condition_occurrence_id 470985	Unique domain identifiers	Ever domain table has a unique primary key to identify domain entities.
<entity>_concept_id	condition_concept_id 313217 (SNOMED “Atrial Fibrillation”)	Standard concept from a respective vocabulary domain	Integration with the Vocabulary. Foreign key into the Standard Vocabulary for <b>Standard Concept</b> .
<entity>_source_value	condition_source_value 427.3 1 (ICD9CM “Atrial Fibrillation”)	Source value	Provenance. Verbatim information from the source data, <b>not to be used</b> by any standard analytics.
<entity>_source_concept_id	condition_source_concept_id 44821957 (ICD9CM “Atrial Fibrillation”)	Source concept from a respective vocabulary domain	Provenance. Foreign key into Standard Vocabulary for <b>Source Concept</b> .
<entity>_type_concept_id	condition_type_concept_id 38000199 ("Inpatient header – primary")	Source type	Provenance. Foreign key into Vocabulary for the <b>origin of the data</b> .

# OMOP Common data model V6



# OMOP CDM

## Standardized Clinical Data



# 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

Visit_Occurrence	
visit_occurrence_id	PK
person_id	
visit_concept_id	
visit_start_date	
visit_start_datetime	
visit_end_date	
visit_end_datetime	
visit_type_concept_id	
provider_id	
care_site_id	
visit_source_value	
visit_source_concept_id	
admitting_source_concept_id	
admitting_source_value	
discharge_to_concept_id	
discharge_to_source_value	
preceding_visit_occurrence_id	

- Visits are ‘Encounters’
- Contains spans of time where a person receives medical services
- Visit Types
  - Inpatient (9201)
  - Outpatient (9202)
  - Emergency room (9203)
  - Emergency room/Inpatient (262)
  - Long-term care (42898160)

# 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

- 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	PK
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

# 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



Join the OHDSI forum

OHDSI Tip: Follow topics to receive emails when new posts are added

Introduce yourself!

Let the community know you're here by introducing yourself in the forum or at a community meeting

Join an OHDSI meeting

Sit in on our weekly community meetings



Join the OHDSI research network

By leading a study across the network

OR

By converting data to the OMOP Common Data Model

Join a working group

Or start your own work group!



Join the Journey

Help improve medical decision making today!

Provide feedback

Identify and evaluate ways to use real-world evidence to inform decision making



# Thank you