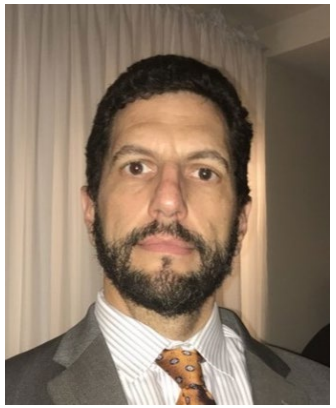
OMOP Common Data Model and Standardized Vocabularies

March 12, 2021

After the Tutorials, you will know…

1. History of OMOP, OHDSI
2. How the Standardized Vocabulary works
3. How to find codes and Concepts
4. How to navigate the concept hierarchy
5. The OMOP Common Data Model (CDM)
6. How to use the OMOP CDM

Instructors

Mui van Zandt	Qi Yang	David Sonnet	Tom Galia	Meghan Pettine	Anthony Reckard
					

Agenda

1. OHDSI Overview, Concept, Concept Mapping – 30 minutes
    - Breakout groups
        - Exercises – 45 minutes - Review – 30 minutes

2. Concept Relationship, Hierarchy – 30 minutes
    - Breakout groups
        - Exercises – 45 minutes - Review – 30 minutes

3. CDM – 30 minutes
    - Breakout groups
        - Exercises – 45 minutes - Review – 30 minutes

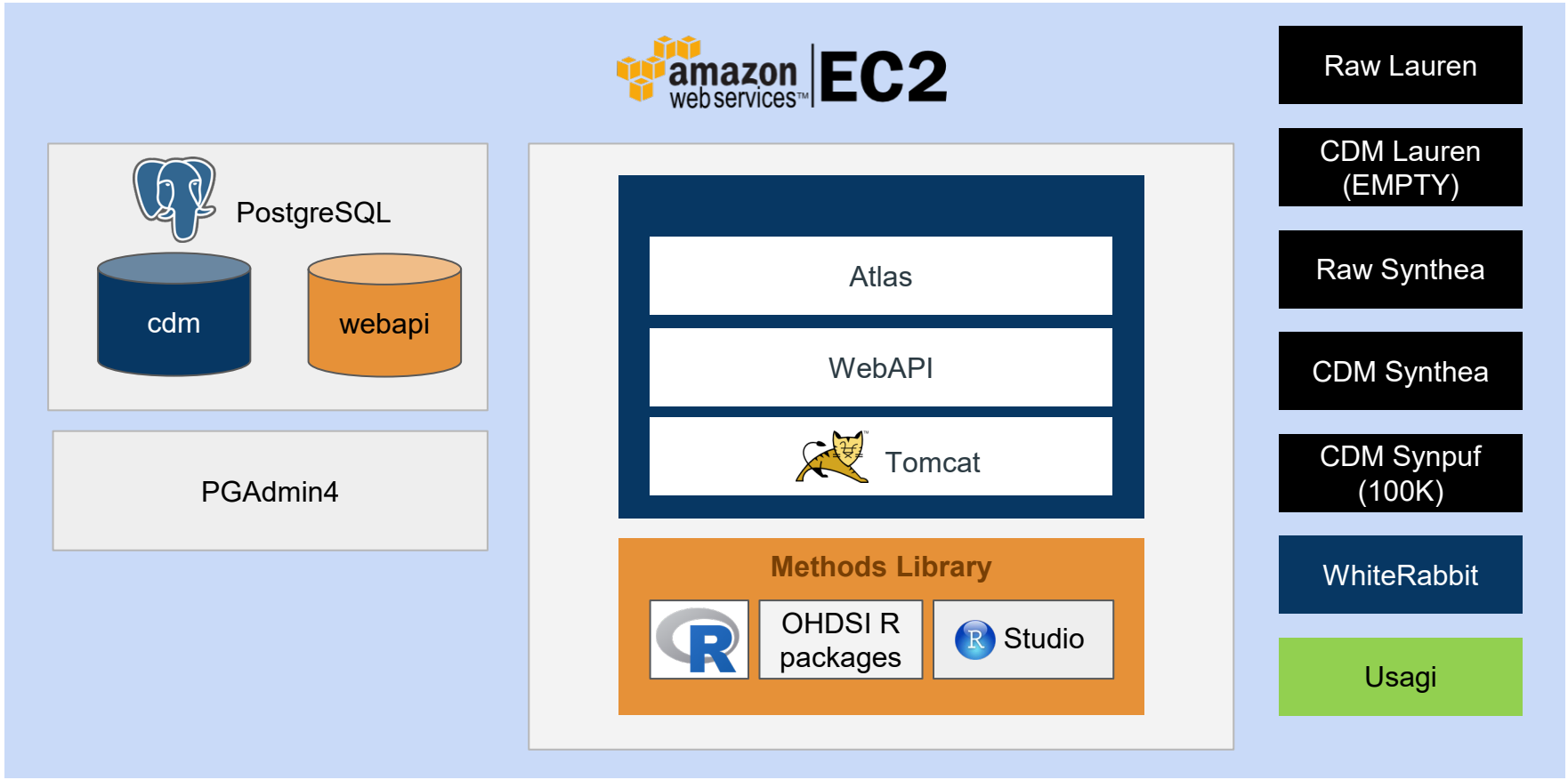
4. ETL Q&A Session – 1 hour

5. Wrap Up – 20 minutes

* 30-minute break in between each sessions

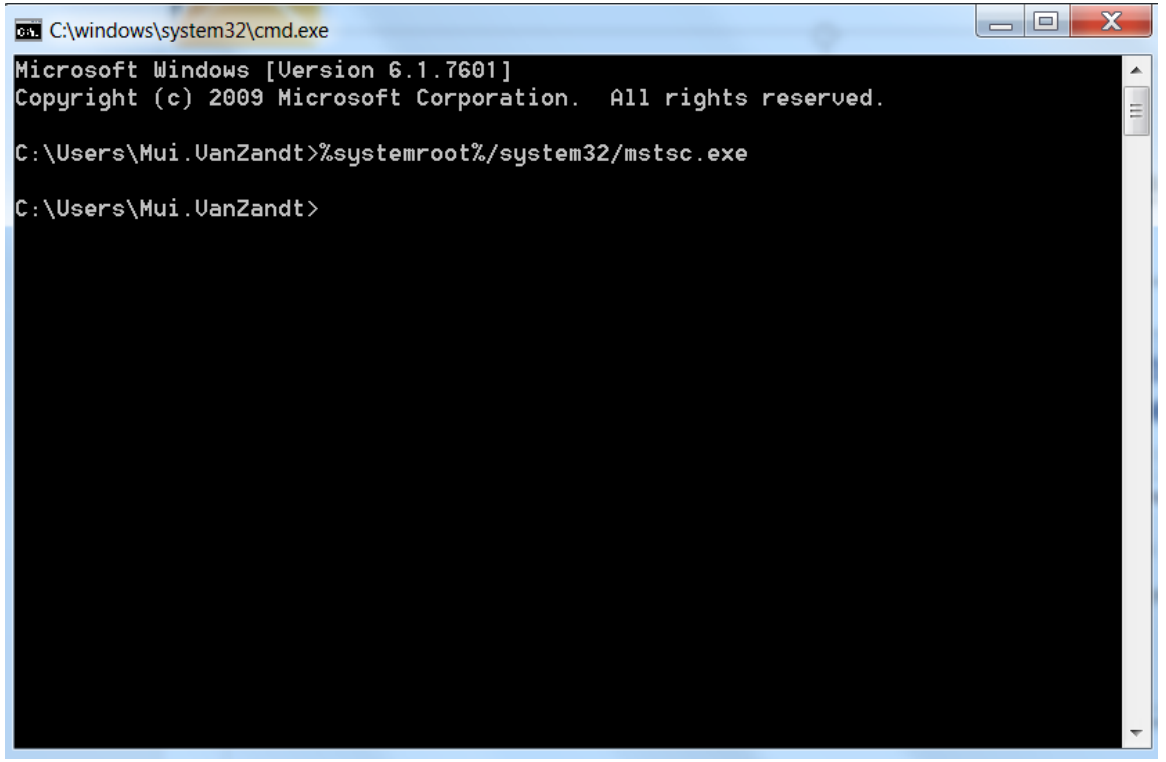
OHDSI-in-a-Box

OHDS-in-a-Box


amazon web services EC2
PostgreSQL
cdm
webapi
PGAdmin4
Atlas
WebAPI
Tomcat
Methods Library
OHDSI R packages
Studio
Raw Lauren
CDM Lauren (EMPTY)
Raw Synthea
CDM Synthea
CDM Synpuf (100K)
WhiteRabbit
Usagi

How to Sign into the Remote Desktop

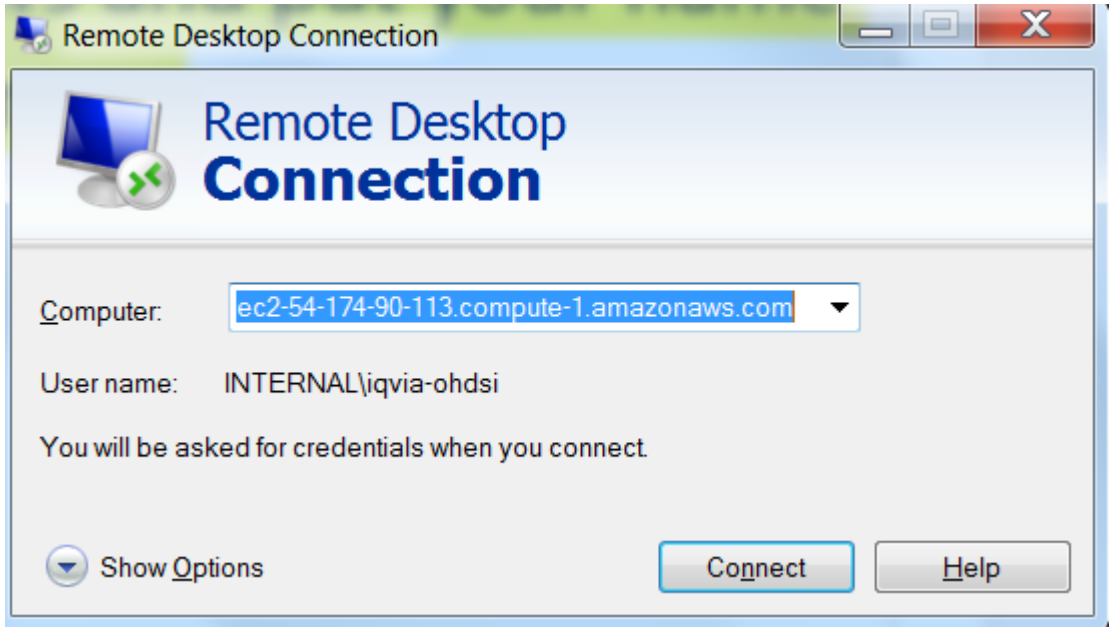
- From your command prompt, type **%systemroot%/system32/mstsc.exe** to launch Remote Desktop



How to Sign for Windows

- Use the shortcut on the desktop named “Remote Desktop”

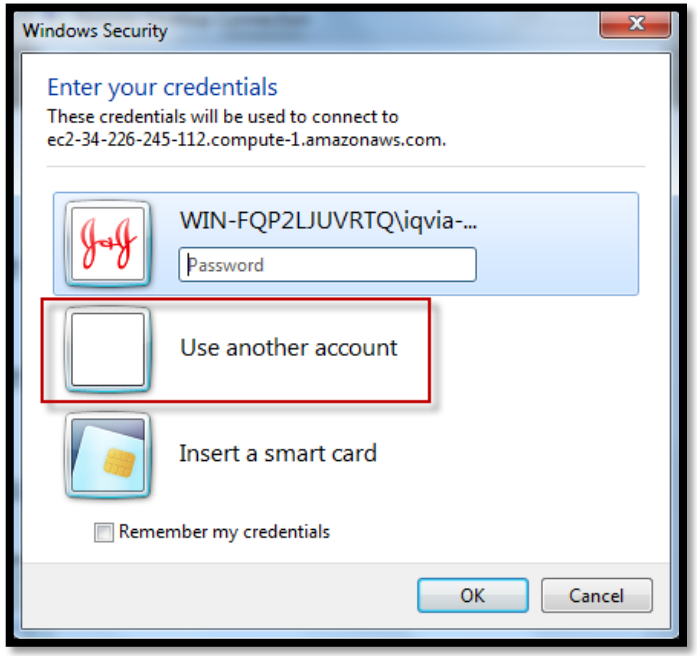
  https://docs.google.com/spreadsheets/d/1x7-6J0NBkm1cRHna0N9tmLWvirpynLSwiZsBZY7FiIQ/edit?usp=sharing
- Pick one of the rows and put your name on the second column
- Take Column A from spreadsheet and copy into the “Computer” field



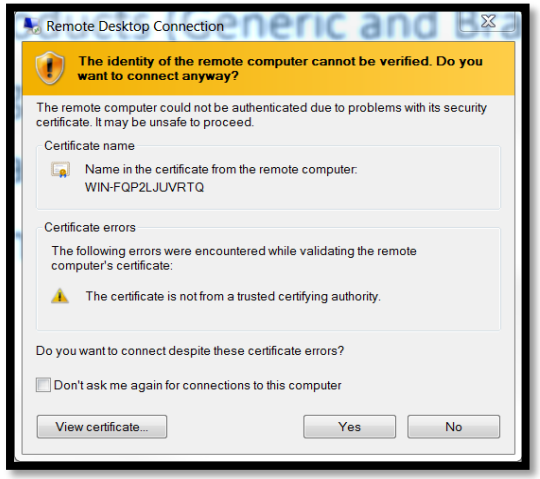
How to Sign for Windows

- Pick ‘Use Another Account’
- Copy username from Column C
- Copy password from Column D

A	B	C	D
RDP URL	**Name**	**Username**	**Password**
ec2-34-226-245-112.compute-1.amazonaws.com	Erica Voss	iqvia-ohdsi	I!QViAOH@DSI18
ec2-52-87-207-197.compute-1.amazonaws.com	Mui Van Zandt	iqvia-ohdsi	I!QViAOH@DSI18

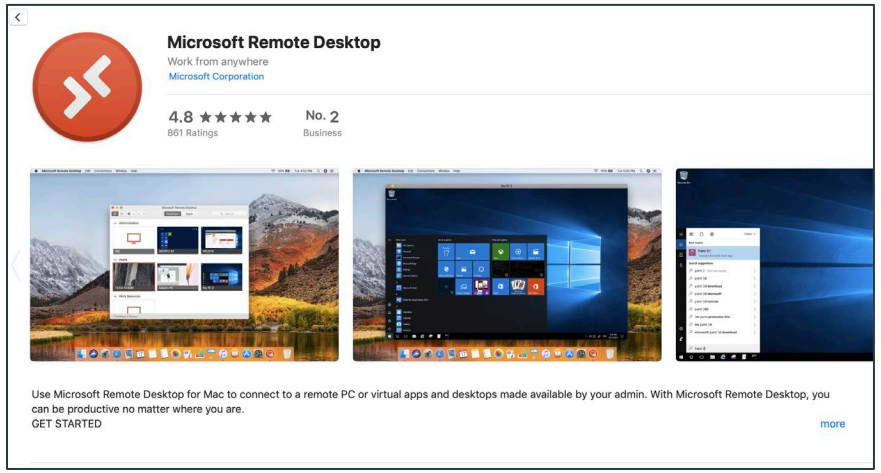


- If you get this page, select “Yes”

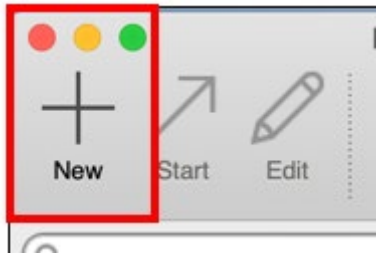


How to sign in for Apple

- Install Microsoft Remote Desktop (available in Apple store)

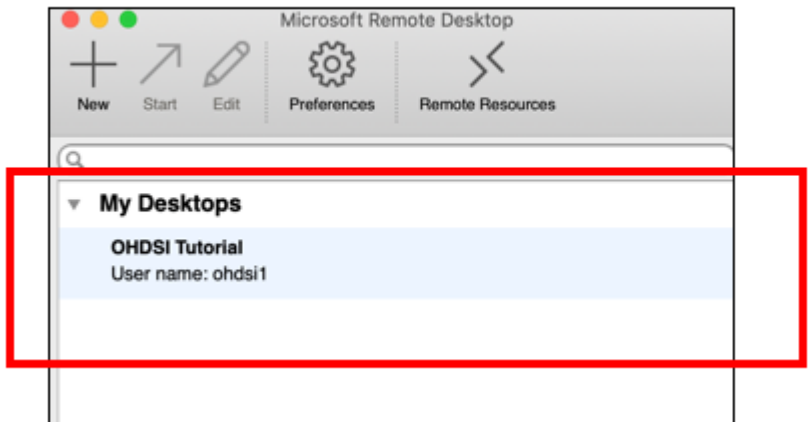


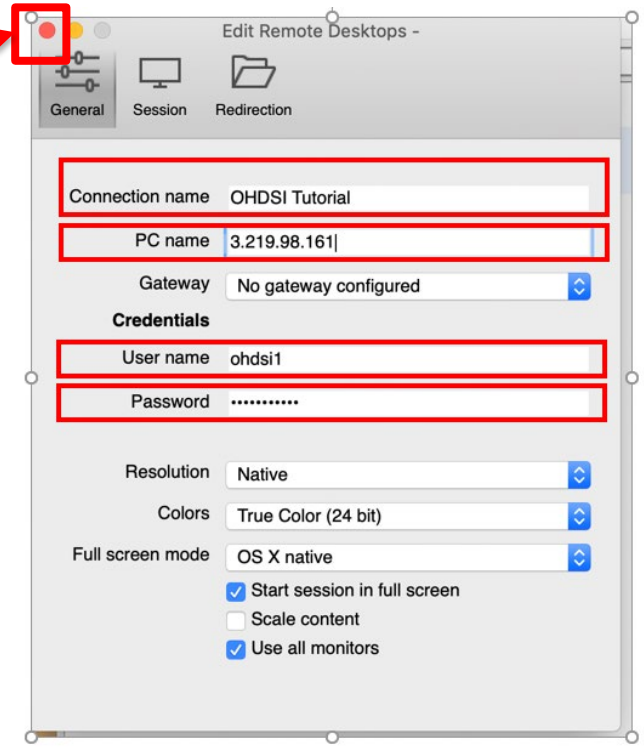
- Launch MS Remote Desktop
- Click the “New” button



How to sign in for Apple

- Enter information from spreadsheet
- Close it when you are done. There is no "Save" button
- Launch the desktop by double clicking





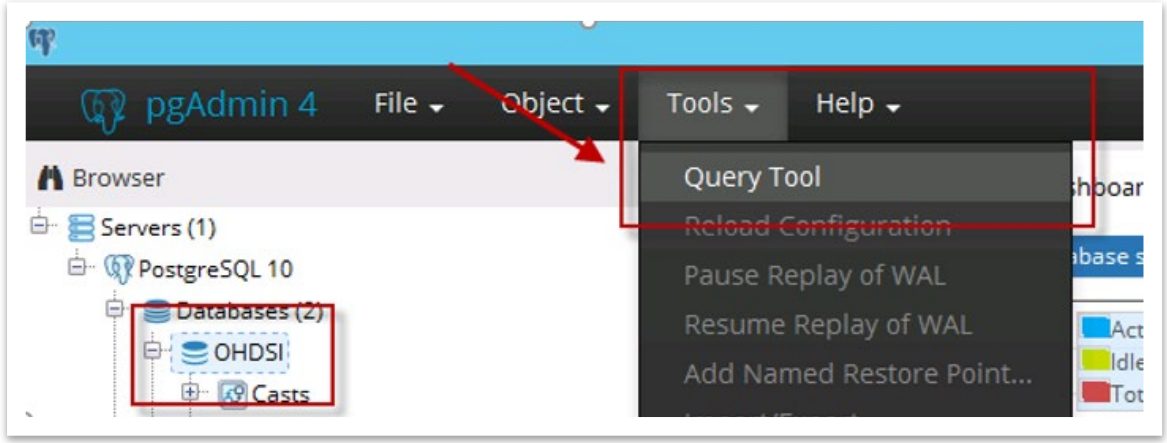
Name as you prefer

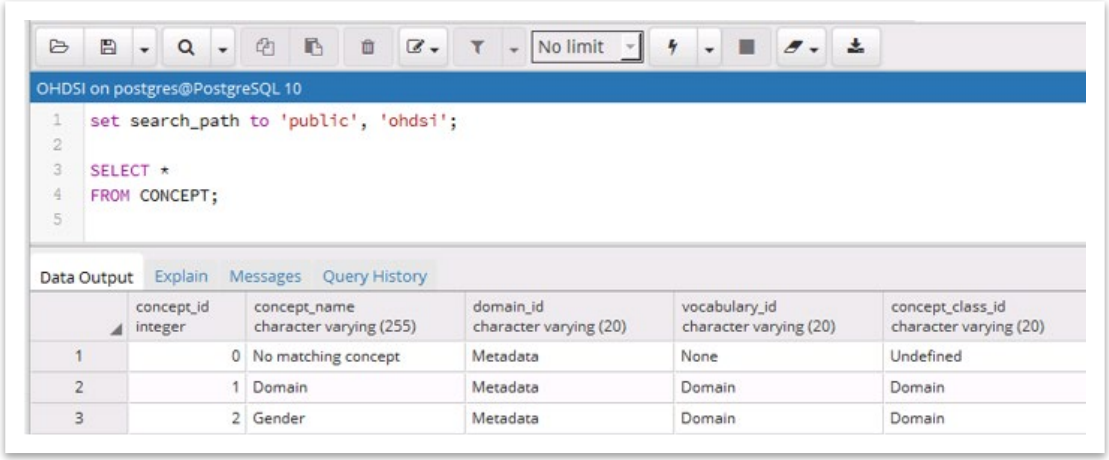
IP address

User name and PW

CDM Database:pgAdmin III New Server

- Click on PGAdmin
- Password: ohdsi
- Select the Query Tool
- Type set “search_path to 'ohdsi’;”






OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Foundation

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Opportunity for our team to embrace 'Open Science' and be part of something different

Mission: Improving health by empowering a community to collaboratively generate evidence that promotes better health decisions and better care

Vision: Creating a world in which observational research produces a comprehensive understanding of health and disease

Values: Innovation, Reproducibility, Community, Collaboration, Openness, Beneficence

OHDSI Collaborators:

- 2,770 users
- 25 workgroups
- 18,700 posts on 3,250 topics

OHDSI Network:

- >150+ databases
- 21 countries
- 2.1B patient records, 369M ex-US

http://ohdsi.org

OMOP to OHDSI

OMOP Investigators

OHDSI

Columbia University

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 centre housed at Columbia University

Public, open

Not pharma funded

International

History of OMOP/OHDSI

2009

OMOP Experiment #1

- FDAAA calls for establishing Risk Identification and Analysis System for drug surveillance
- OMOP Experiment creates a framework for evaluating 14 methods of epidemiological designs
- 10 data sources, claims and EHRs, 200M+ lives

2011

OMOP Experiment #2/European OMOP

- Focused on a subset of data (4 claims, 1 EHR) and 7 methods
- Replicating experiment findings on European databases

2013

End of OMOP Experiment

- Last of the OMOP Meetings present findings of empirical experiments
- Artifacts include the OMOP CDM, vocabularies and validated methods for analyzing real world data

Launch of OHDSI

- OHDSI's first face-to-face meeting at Columbia University

2014

2015

First OHDSI Symposium/Network Study Published

- Community begins open source work under OHDSI brand
- First global network study characterizing treatment pathways

First Hackathon at Columbia University

Formation of China Chapter

- To use data science and informatics methods to promote health and medical data research in China

Formation of Korea Chapter

2016

2017

Formation of European Chapter

- Led by the coordinating center at the Erasmus University Medical Center in Rotterdam

FDA Adoption (FDA BEST Launch)

EHDEN Initiation (Europe)

- Started under the Innovative Medicines Initiative (IMI) that will drive the adoption of the OMOP-CDM in Europe

FederNET Initiation (Korea)

First European Symposium

2018

2019

EMA Adoption

Formation of Australia, Japan and Singapore Chapters

NMPA Adoption

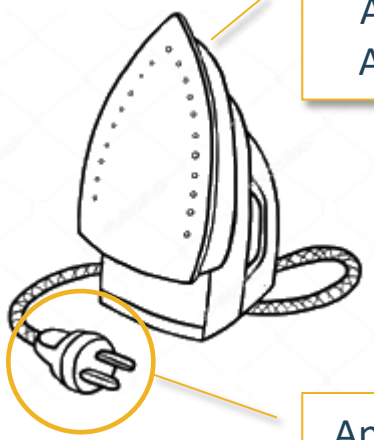
2020

Global Acceptance

- OHDSI grows to >152 databases, 18 countries, 2.1B patient records, 369M ex-US
- Regional chapters in US, Europe, China, South Korea + Asia-Pacific, Latin America
- Offering regional symposia

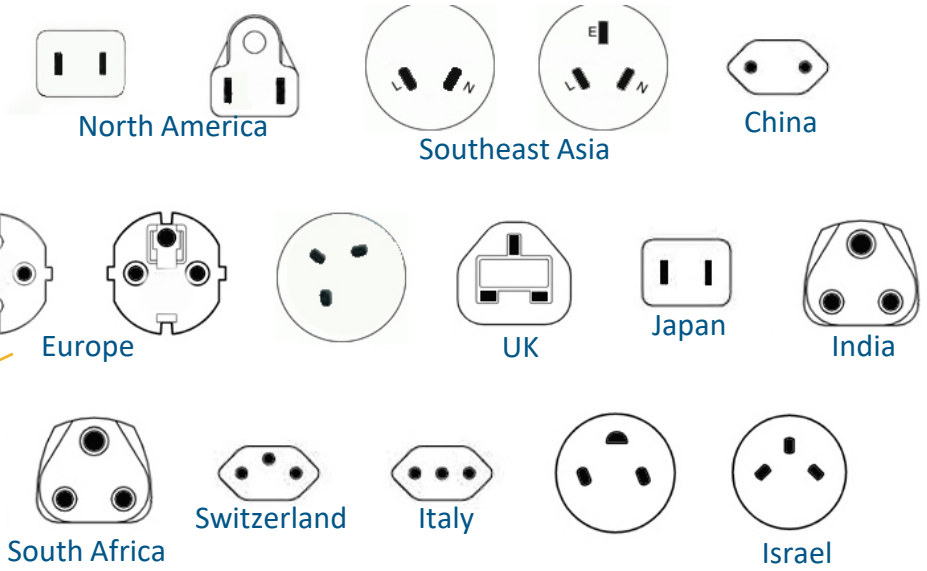
Current Approach: “One Study – One Script“

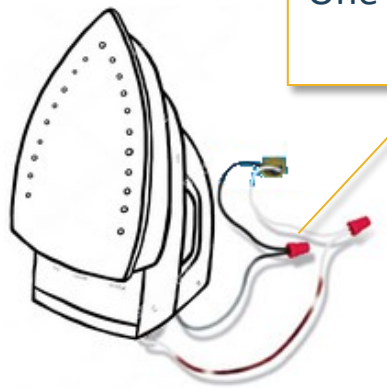
"What's the adherence to my drug in the data assets I own?"



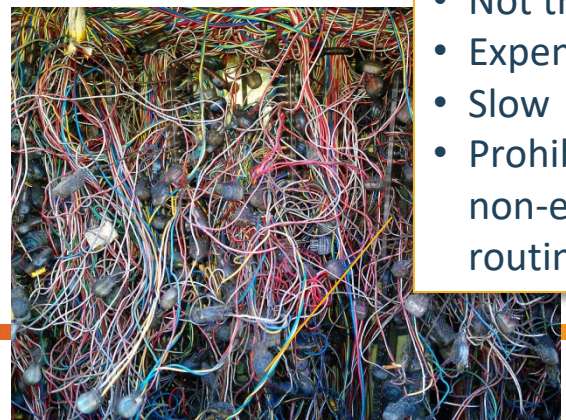
Analytical method: Adherence to Drug

Application to data



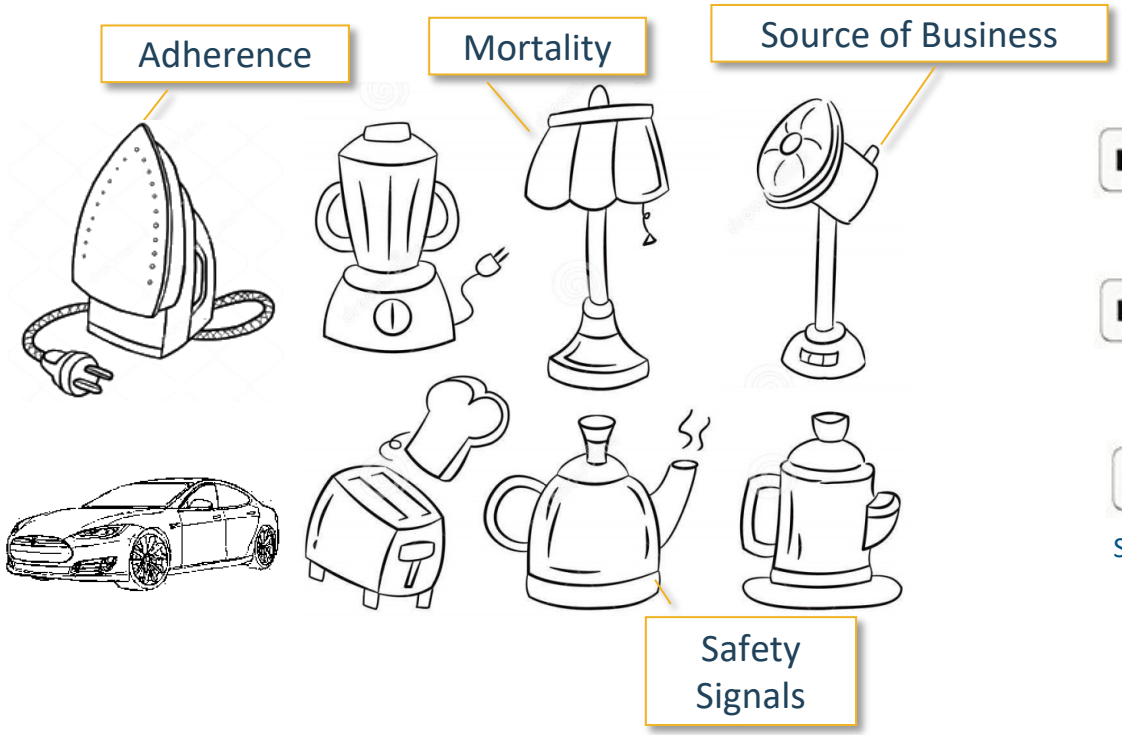


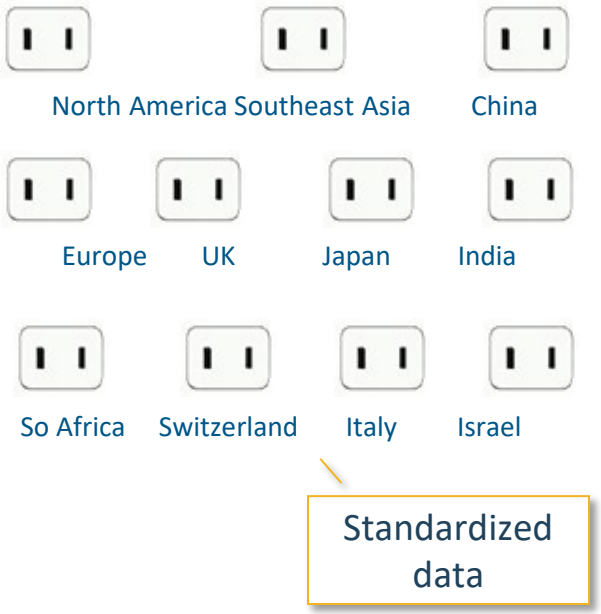
One SAS or R script for each study



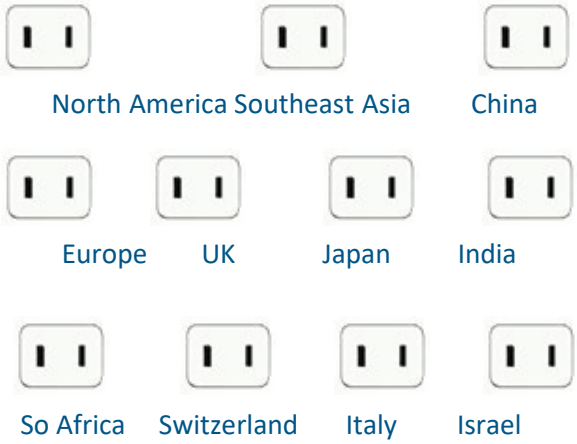
- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert routine use

Solution: Data Standardization Enables Systematic Research



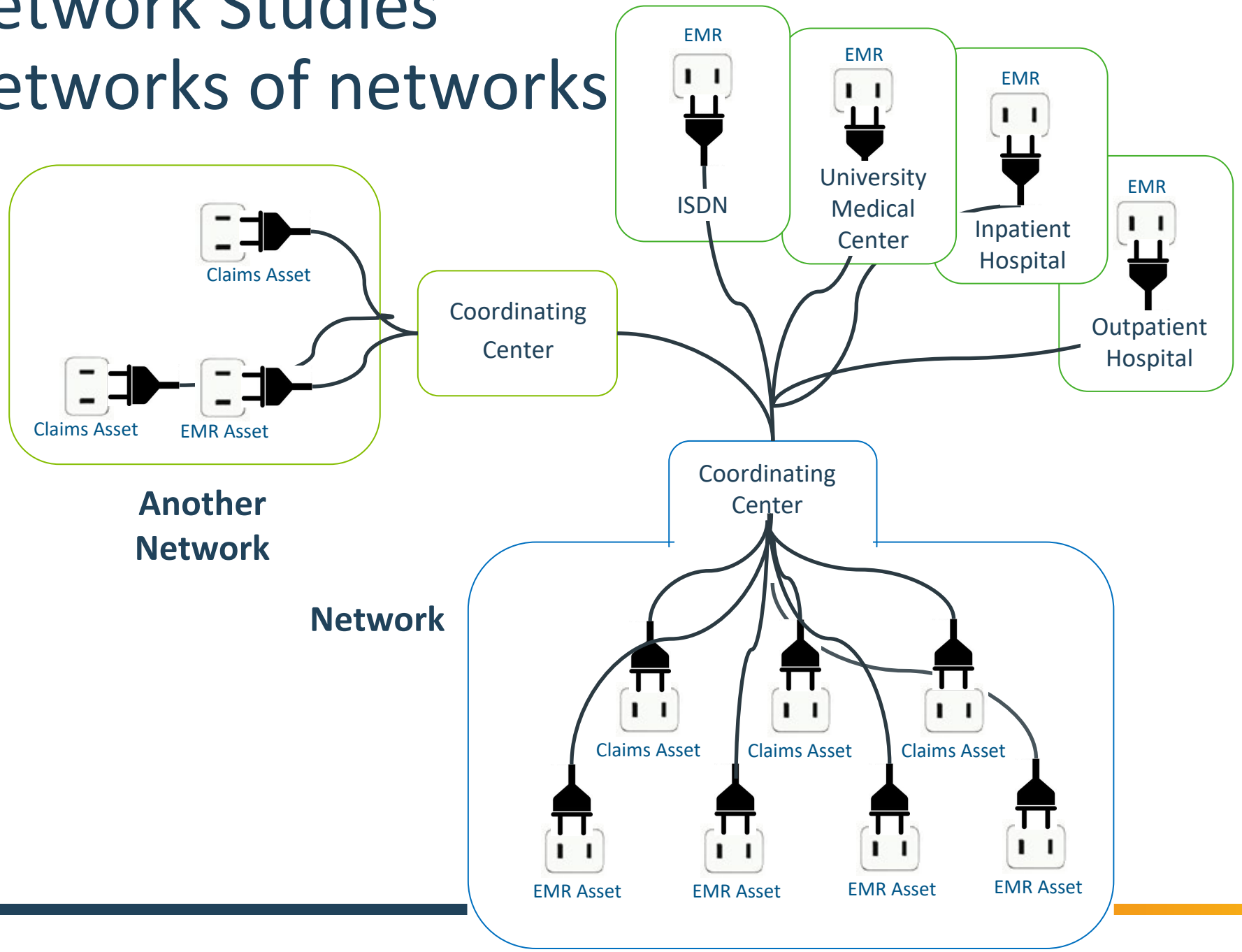


Analytics Can Be Remote



Analytics Can Be Behind Firewall

Network Studies
Networks of networks



Source Codes, Concepts and Mapping to Standard Concepts

The Source for Source Codes

1. May come from international terminology or code system
   - **SNOMED**
2. May come from a country specific terminology or code system
   - **Read, BDPM, ICD10CN, CVX**
3. May be free text strings
   - **Centimeter, Intravenous, Cigarette Smoker**
4. May come from an EHR specific code system
   - **Epic procedure codes: ‘L111’**

Different Categories of Concepts

Non-standard Concepts

Function

Unique representation of a source code

Standard Concepts

Function

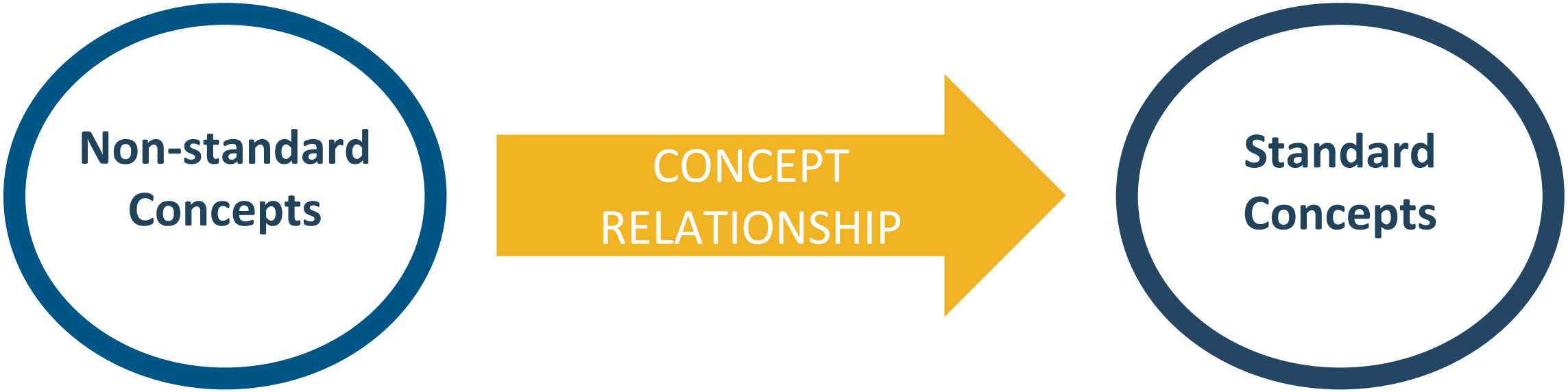
Used for standardized analytics and by OHDSI tools

Classification Concepts

Function

Used to perform hierarchical queries

Mapping Non-Standard Source concept_IDs to Standard concept_IDs



Mapping Source Codes to Standard Concept_IDs

Source Codes Mapping – Scenario 1

Scenario

- Source code is available in an OHDSI supported Vocabulary

Solution

- Use using the following condition to perform the mapping:

  Where <source code> = CONCEPT.concept_code and <source vocabulary> = CONCEPT.vocabulary_id

Source code	Source vocabulary	Code description	CONCEPT.concept_id
61462000	SNOMED	Malaria	438067
A663D00	Read	Zika Fever	45489770
A92.3	ICD10CN	West Nile Virus Infection	1404276

ATHENA

SEARCH | DOWNLOAD | LOGIN

SEARCH BY KEYWORD

61462000

61462000 ×

SNOMED ×

DOMAIN

CONCEPT

CLASS

VOCAB

filter

- [ ] Nebraska Lexicon (42)
- [ ] Read (1)
- [ ] RxNorm Extension (7)
- [x] SNOMED (42)

VALIDITY

CLEAR FILTERS

DOWNLOAD RESULTS

Show by 15 items

Total 42 items

1 2 3

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
438067	61462000	Malaria	Clinical Finding	Standard	Valid	Condition	SNOMED
4040837	16562000	Apolonia	Organism	Standard	Valid	Observation	SNOMED
4016020	114262000	Citrobacter braakii	Organism	Standard	Valid	Observation	SNOMED
4013208	11262000	Lonchocarpus floribundus	Organism	Standard	Valid	Observation	SNOMED
4016302	11462003	Ostomy appliance adhesive	Substance	Standard	Valid	Device	SNOMED
4058278	161468000	H/O: schizophrenia	Context-dependent	Standard	Valid	Observation	SNOMED
4041011	16642000	Glutamate-ethylamine ligase	Substance	Standard	Valid	Observation	SNOMED
4059194	161482000	H/O: trigeminal neuralgia	Context-dependent	Standard	Valid	Observation	SNOMED
4017258	114620002	Salmonella II 3,10:g,t:-	Organism	Standard	Valid	Observation	SNOMED
4015110	114623000	Salmonella Tibati	Organism	Standard	Valid	Observation	SNOMED

Source Codes Mapping – Scenario 1

Scenario

- Source code is available in an OHDSI supported Vocabulary

Solution

- Use using the following condition to perform the mapping:

  Where <source code> = CONCEPT.concept_code and <source vocabulary> = CONCEPT.vocabulary_id

Source code	Source vocabulary	Code description	CONCEPT.concept_id
61462000	SNOMED	Malaria	438067
A663D00	Read	Zika Fever	45489770
A92.3	ICD10CN	West Nile Virus Infection	1404276

ATHENA

SEARCH | DOWNLOAD | LOGIN

SEARCH BY KEYWORD

A663D00

A663D00 ×

Read ×

DOMAIN

CONCEPT

CLASS

VOCAB

- Read (32)
- SNOMED (1)
- ABMS (0)
- AMT (0)
- APC (0)
- ATC (0)

VALIDITY

CLEAR FILTERS

DOWNLOAD RESULTS

Show by 15 items

Total 32 items

1 2 3

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
45489770	A663D00	Zika fever	Read	Non-standard	Valid	Condition	Read
45499671	A663000	Bunyamwera fever	Read	Non-standard	Valid	Condition	Read
45449567	A663100	Bwamba fever	Read	Non-standard	Valid	Condition	Read
45442925	A663200	Chikungunya fever	Read	Non-standard	Valid	Condition	Read
45426396	A663300	Guama fever	Read	Non-standard	Valid	Condition	Read
45483143	A663400	Mayaro fever	Read	Non-standard	Valid	Condition	Read
45456273	A663500	Mucambo fever	Read	Non-standard	Valid	Condition	Read
45486481	A663600	O'nyong-nyong fever	Read	Non-standard	Valid	Condition	Read
45506275	A663700	Oropouche fever	Read	Non-standard	Valid	Condition	Read
45506276	A663800	Pixuna fever	Read	Non-standard	Valid	Condition	Read

Source Codes Mapping – Scenario 1

Scenario

- Source code is available in an OHDSI supported Vocabulary

Solution

- Use using the following condition to perform the mapping:

  Where <source code> = CONCEPT.concept_code and <source vocabulary> = CONCEPT.vocabulary_id

Source code	Source vocabulary	Code description	CONCEPT.concept_id
61462000	SNOMED	Malaria	438067
A663D00	Read	Zika Fever	45489770
A92.3	ICD10CN	West Nile Virus Infection	1404276

ATHENA

SEARCH | DOWNLOAD | LOGIN

SEARCH BY KEYWORD

A92.3

A92.3 ×

DOMAIN

CONCEPT

CLASS

VOCAB

filter

- CIM10 (20)
- HCPCS (2)
- ICD10 (21)
- ICD10CM (24)
- ICD10CN (19)

VALIDITY

CLEAR FILTERS

DOWNLOAD RESULTS

Show by 15 items

Total 167 items

1 2 3 4 5 ... 12

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
37085667	A92.3	West Nile virus infection	ICD10 code	Non-standard	Valid	Condition	ICD10GM
42484133	A92.3	West Nile virus infection	KCD7 code	Non-standard	Valid	Condition	KCD7
45595599	A92.3	West Nile virus infection	ICD10 code	Non-standard	Valid	Condition	ICD10
1567345	A92.3	West Nile virus infection	4-char nonbill code	Non-standard	Valid	Condition	ICD10CM
1404276	A92.3	West Nile virus infection	ICD10 code	Non-standard	Valid	Condition	ICD10CN
37611463	A92.3	West Nile virus infection	ICD10 code	Non-standard	Valid	Condition	CIM10
44508037	A62.3	Microsurgical graft to peripheral nerve NEC	Procedure	Standard	Valid	Procedure	OPCS4
44514524	T92.3	Excision of lymphoedematous tissue of leg and buried flaps HFQ	Procedure	Standard	Valid	Procedure	OPCS4
2615319	A9283	Foot pressure off loading/supportive device, any type, each	HCPCS	Standard	Valid	Device	HCPCS
40664451	A9273	Cold or hot fluid bottle, ice cap or collar, heat and/or cold wrap, any type	HCPCS	Standard	Valid	Device	HCPCS

Source Codes Mapping – Scenario 2

Scenario

- Source code is a text string

Solution

- Use using the following condition to perform the mapping:

  Where <source string> = CONCEPT.concept_name and <source domain> = CONCEPT.domain_id

Source string	Source domain	Source table/field	CONCEPT.concept_id
Centimeter	Unit	Unit for height measurement	8582
Intravenous	Route	Route for drug administration	4171047
Female	Gender	Demographics	8532

ATHENA

SEARCH | DOWNLOAD | Melanie Philofsky

SEARCH BY KEYWORD

centimeter

centimeter ×

DOMAIN

CONCEPT

CLASS

VOCAB

VALIDITY

CLEAR FILTERS

DOWNLOAD RESULTS

Show by 15 items

Total 704 items

1 2 3 4 5 ... 47

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
8582	cm	centimeter	Unit	Standard	Valid	Unit	UCUM
3195806	258672001	centimeter	Qualifier Value	Non-standard	Valid	Unit	Nebraska Lexicon
44777590	cm[H2O]	centimeter watercolumn	Unit	Standard	Valid	Unit	UCUM
44777662	cm3	cubic centimeter	Unit	Standard	Valid	Unit	UCUM
9483	cm2	square centimeter	Unit	Standard	Valid	Unit	UCUM
32738	cm/s	centimeter per second	Unit	Standard	Valid	Unit	UCUM
3191324	259054002	Joules/square centimeter	Qualifier Value	Non-standard	Valid	Unit	Nebraska Lexicon
3195908	259019009	centimeter of water	Qualifier Value	Non-standard	Valid	Unit	Nebraska Lexicon
937594	Q4176	Neopatch, per square centimeter	HCPCS	Standard	Valid	Device	HCPCS
937595	Q4182	Transcyte, per square centimeter	HCPCS	Standard	Valid	Device	HCPCS

Source Codes Mapping – Scenario 2

Scenario

- Source code is a text string

Solution

- Use using the following condition to perform the mapping:

  Where <source string> = CONCEPT.concept_name and <source domain> = CONCEPT.domain_id

Source string	Source domain	Source table/field	CONCEPT.concept_id
Centimeter	Unit	Unit for height measurement	8582
Intravenous	Route	Route for drug administration	4171047
Female	Gender	Demographics	8532

ATHENA

SEARCH | DOWNLOAD | Melanie Philofsky

SEARCH BY KEYWORD

intravenous

intravenous ×

DOMAIN

CONCEPT

CLASS

VOCAB

VALIDITY

CLEAR FILTERS

DOWNLOAD RESULTS

Show by 15 items

Total 36,527 items

1 2 3 4 5 ... 2436

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
45880861	LA9437-0	Intravenous	Answer	Standard	Valid	Meas Value	LOINC
4112421	255560000	Intravenous	Qualifier Value	Standard	Valid	Observation	SNOMED
35807199	44957	Intravenous	Route	Non-standard	Valid	Drug	HemOnc
4171047	47625008	Intravenous	Qualifier Value	Standard	Valid	Route	SNOMED
3297201	255560000	Intravenous	Qualifier Value	Non-standard	Valid	Observation	Nebraska Lexicon
4302613	78432000	Intravenous anesthesia	Procedure	Standard	Valid	Procedure	SNOMED
4303422	386340006	Intravenous therapy	Procedure	Standard	Valid	Observation	SNOMED
4138256	32265006	Intravenous pyelogram	Procedure	Standard	Valid	Procedure	SNOMED
4146535	265760000	Intravenous chemotherapy	Procedure	Standard	Valid	Procedure	SNOMED

Source Codes Mapping – Scenario 2

Scenario

- Source code is a text string

Solution

- Use using the following condition to perform the mapping:

  Where <source string> = CONCEPT.concept_name and <source domain> = CONCEPT.domain_id

Source string	Source domain	Source table/field	CONCEPT.concept_id
Centimeter	Unit	Unit for height measurement	8582
Intravenous	Route	Route for drug administration	4171047
Female	Gender	Demographics	8532

ATHENA

SEARCH | DOWNLOAD | Melanie Philofsky

SEARCH BY KEYWORD

female

female ×

- DOMAIN
- CONCEPT
- CLASS
- VOCAB
- VALIDITY

CLEAR FILTERS

DOWNLOAD RESULTS

Show by 15 items

Total 22,761 items

1 2 3 4 5 ... 1518 >

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
8532	F	FEMALE	Gender	Standard	Valid	Gender	Gender
45878463	LA3-6	Female	Answer	Standard	Valid	Meas Value	LOINC
442986	248152002	Female	Clinical Finding	Standard	Valid	Condition	SNOMED
35919260	220@2	Female	NAACCR Value	Standard	Valid	Meas Value	NAACCR
35933457	peritoneum_ female_gen @2879@002	Female	NAACCR Value	Standard	Valid	Meas Value	NAACCR
35935550	peritoneum @2879@002	Female	NAACCR Value	Standard	Valid	Meas Value	NAACCR
44814665	Sex-F	Female	Gender	Non-standard	Valid	Observation	PCORNet
45421841	1K1..00	Female	Read	Non-standard	Valid	Condition	Read
			Clinical	Non-			Nebraska

Source Codes Mapping – Scenario 3

Scenario

- Source data does not map to standard vocabulary

Solution

- Ask OHDSI to incorporate the Vocabulary
- Create custom mapping
  - Use concept_id > 2 billion
    - Create a concept record for the source code
    - Create Concept_Relationship records to link the source concept_id to the standard concept_id
    - See Melanie's poster for a full description

Breakout Session 1
Exercises 45 minutes – Review 30 minutes

Other Relationships: Hierarchical and Part-of Relationships

Exploring Relationships

```
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

Ancestor concepts

Relationships are bi-directional: 'Maps to' and 'Mapped from'; 'Is a' and 'Subsumes' etc.

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	4194288		Observation	SNOMED	Context-dependent	S	312442005	1/1/1970 0:00	12/31/2099 0:00	NULL
Asso finding of	4203375		Observation	SNOMED	Context-dependent	S	433276002	1/31/2009 0:00	12/31/2099 0:00	NULL
Asso finding of	42689685		Observation	SNOMED	Context-dependent	S	1.06706E+15	4/1/2017 0:00	12/31/2099 0:00	NULL
Asso finding of	44807374	Atrial fibrillation excluded	Observation	SNOMED	Context-dependent	S	8.16401E+14	4/1/2014 0:00	12/31/2099 0:00	NULL
Concept poss_eq from	40323929	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	155364009	1/1/1970 0:00	3/11/2016 0:00	U
Concept poss_eq from	40345197	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	266306001	1/1/1970 0:00	3/11/2016 0:00	U
Due to of	4139517	Transient cerebral ischemia due to atrial fibrillation	Condition	SNOMED	Clinical Finding	S	426814001	1/1/1970 0:00	12/31/2099 0:00	NULL
Focus of	42709991	Insertion of pacemaker for control of atrial fibrillation	Procedure	SNOMED	Procedure	S	449863006	1/31/2012 0:00	12/31/2099 0:00	NULL
Has finding site	4242112	Atrial structure	Spec Site							NULL
Is a	4226399	Fibrillation	Cond							NULL
Is a	4068155	Atrial arrhythmia	Cond							NULL
Mapped from	40323929	Fibrillation - atrial	Cond							U
Mapped from	2617597	Patient with heart failure and atrial fibrillation documented to be on warfarin therapy	Obse							D
Mapped from	45576876	Unspecified atrial fibrillation	Cond							NULL
Mapped from	45500085	Atrial fibrillation	Cond							NULL
Mapped from	45611600	Atrial Fibrillation	Cond							NULL
Mapped from	40345197	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	266306001	1/1/1970 0:00	3/11/2016 0:00	U
Mapped from	45951191	Atrial Fibrillation	Condition	CIEL	Diagnosis	NULL	148203	11/3/2007 0:00	12/31/2099 0:00	NULL
Mapped from	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code	NULL	427.31	1/1/1970 0:00	12/31/2099 0:00	NULL
Maps to	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - HOI	500002401	OMOP Atrial Fibrillation 1	Condition	Cohort	Cohort	C	500002401	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - HOI	500001801	OMOP Qt Prolongation/Torsade De Pointes 1	Condition	Cohort	Cohort	C	500001801	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21005673	Prevention of Thromboembolism in Chronic Atrial Fibrillation	Drug	Indication	Indication	C	5673	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21003176	Tachyarrhythmia	Drug	Indication	Indication	C	3176	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21001542	Supraventricular Tachycardia	Drug	Indication	Indication	C	1542	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21001594	Disease of Cardiovascular System	Drug	Indication	Indication	C	1594	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - MedDRA eq	35204953	Atrial fibrillation	Condition	MedDRA	PT	C	10003658	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4117112	Controlled atrial fibrillation	Condition	SNOMED	Clinical Finding	S	300996004	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4119601	Lone atrial fibrillation	Condition	SNOMED	Clinical Finding	S	233910005	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4232697	Persistent atrial fibrillation	Condition	SNOMED	Clinical Finding	S	440059007	1/31/2009 0:00	12/31/2099 0:00	NULL
Subsumes	4141360	Chronic atrial fibrillation	Condition	SNOMED	Clinical Finding	S	426749004	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	44782442	Atrial fibrillation with rapid ventricular response	Condition	SNOMED	Clinical Finding	S	1.20041E+14	1/31/2014 0:00	12/31/2099 0:00	NULL
Subsumes	4199501	Rapid atrial fibrillation	Condition	SNOMED	Clinical Finding	S	314208002	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4119602	Non-rheumatic atrial fibrillation	Condition	SNOMED	Clinical Finding	S	233911009	1/1/1970 0:00	12/31/2099 0:00	NULL

Descendant concepts

Type of Relationships

Entity Relationships

- ‘Maps to’
- ‘Maps to Value’

All source vocabularies, provided by source or built by Vocab Team
Standard are mapped to themselves

Part-of Relationships

- ‘Has direct procedure site’
- ‘Has method’
- ‘Occurs after’

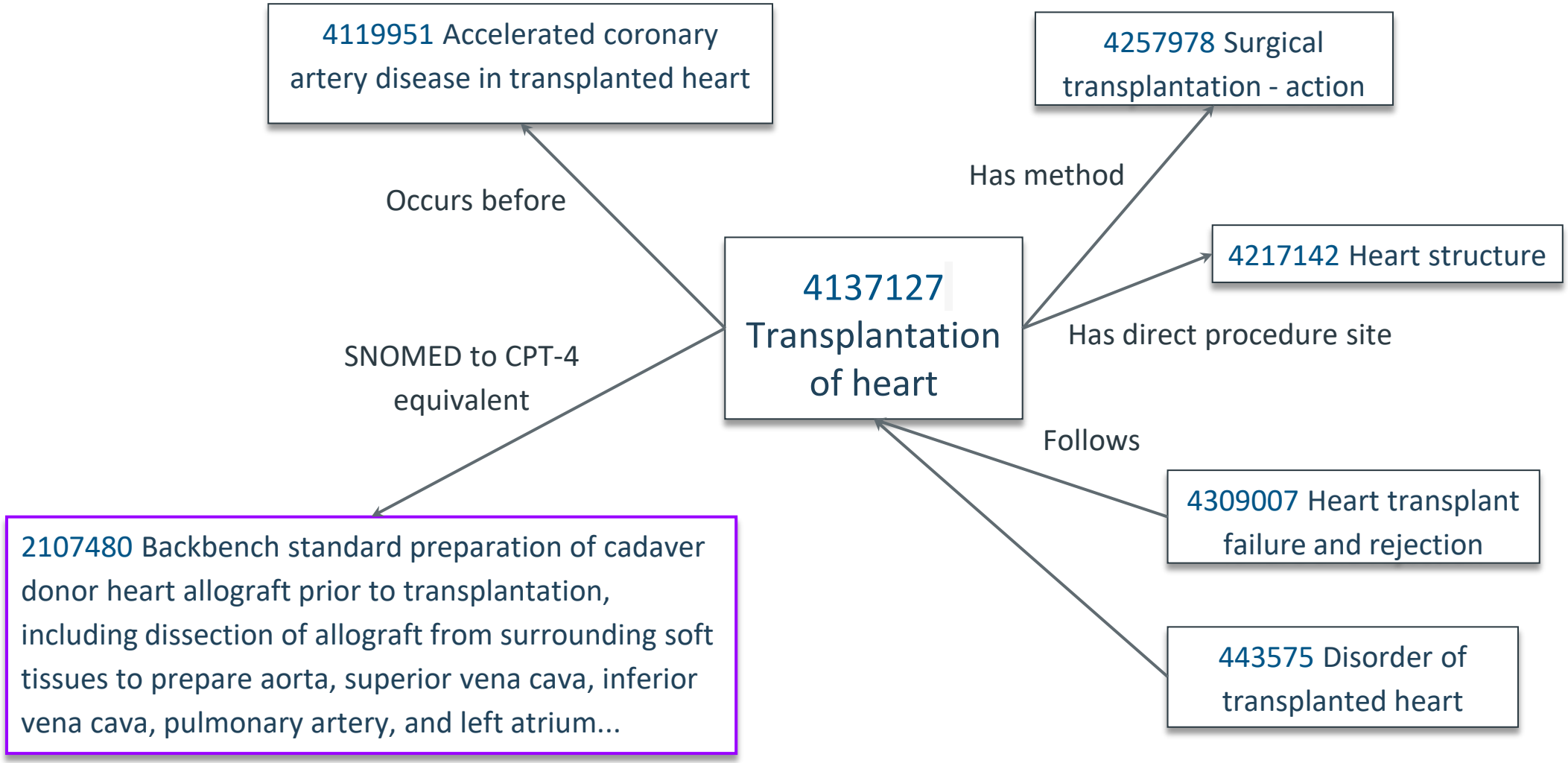
Controlled well-structured ontologies: RxNorm, SNOMED, LOINC, ICD10PCS

Hierarchical Relationships

- ‘Is a’
- ‘CVX-RxNorm’
- ‘Panel Contains’

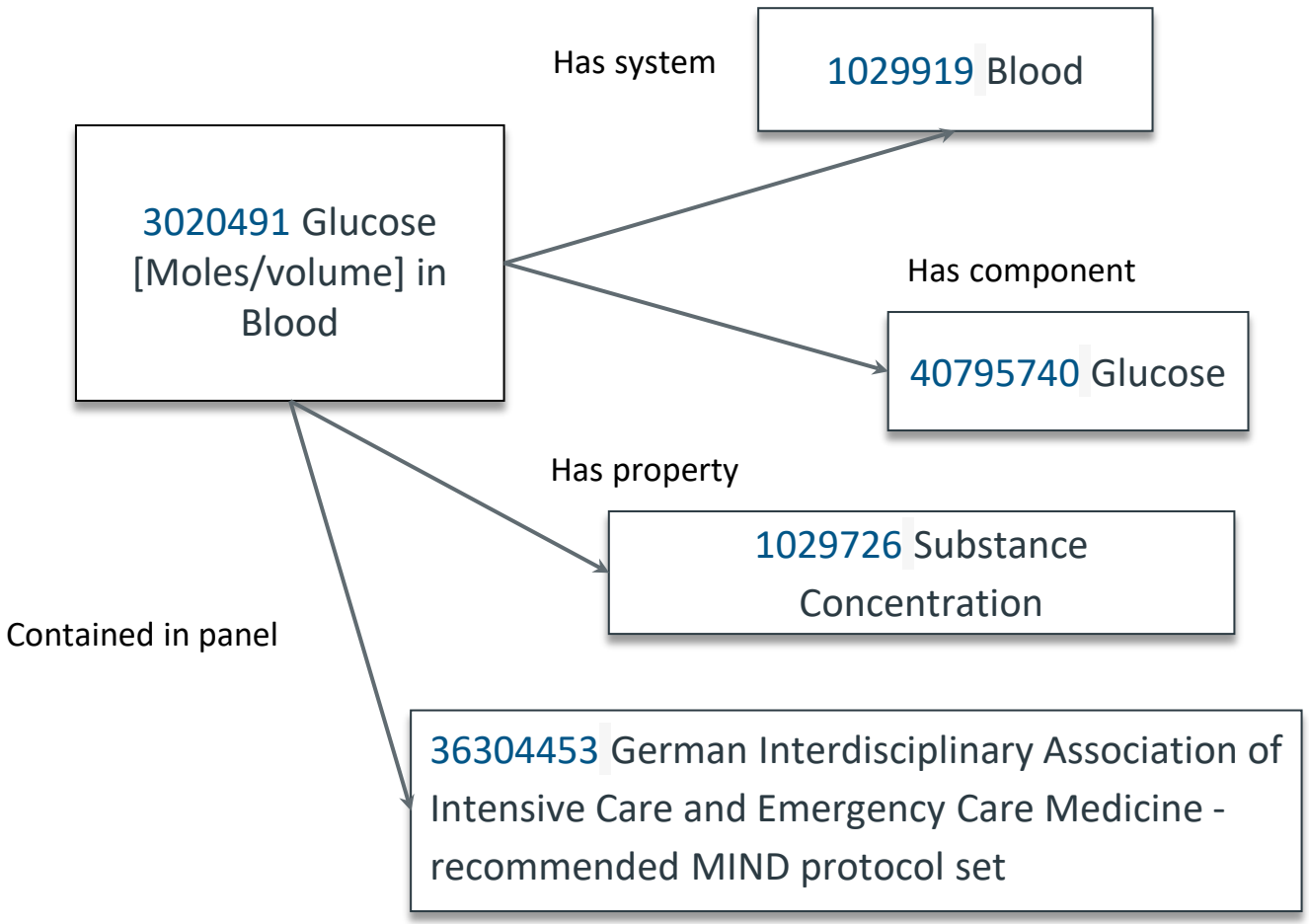
Standard and classificational terminologies (ATC, ETC, CVX, NDFRT)

SNOMED - Transplantation of Heart

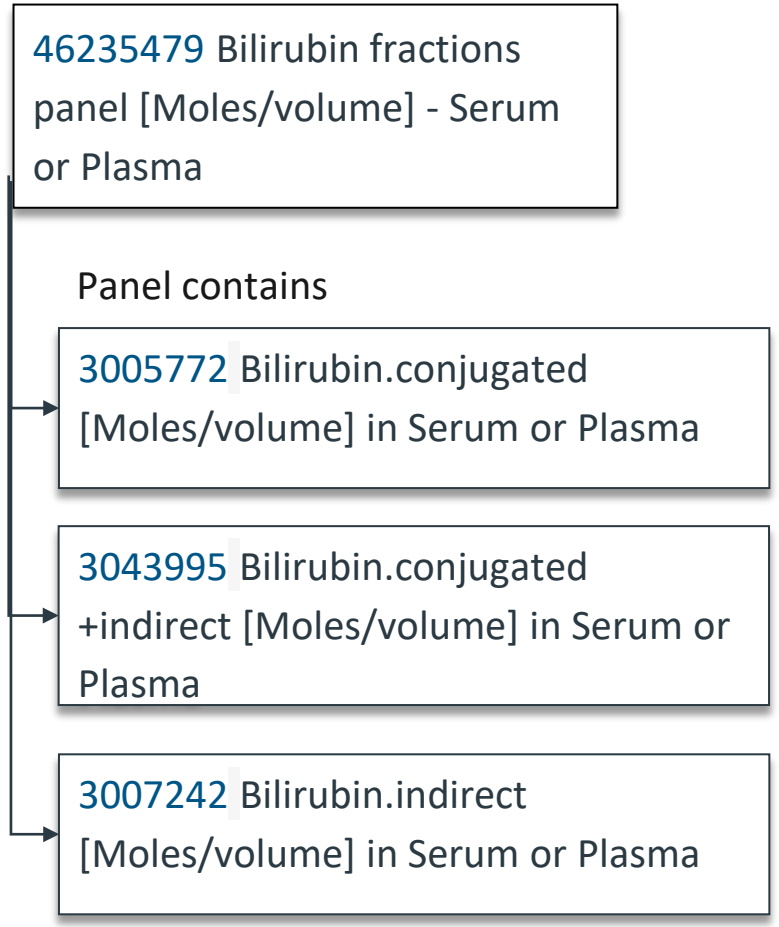


LOINC

Individual Tests

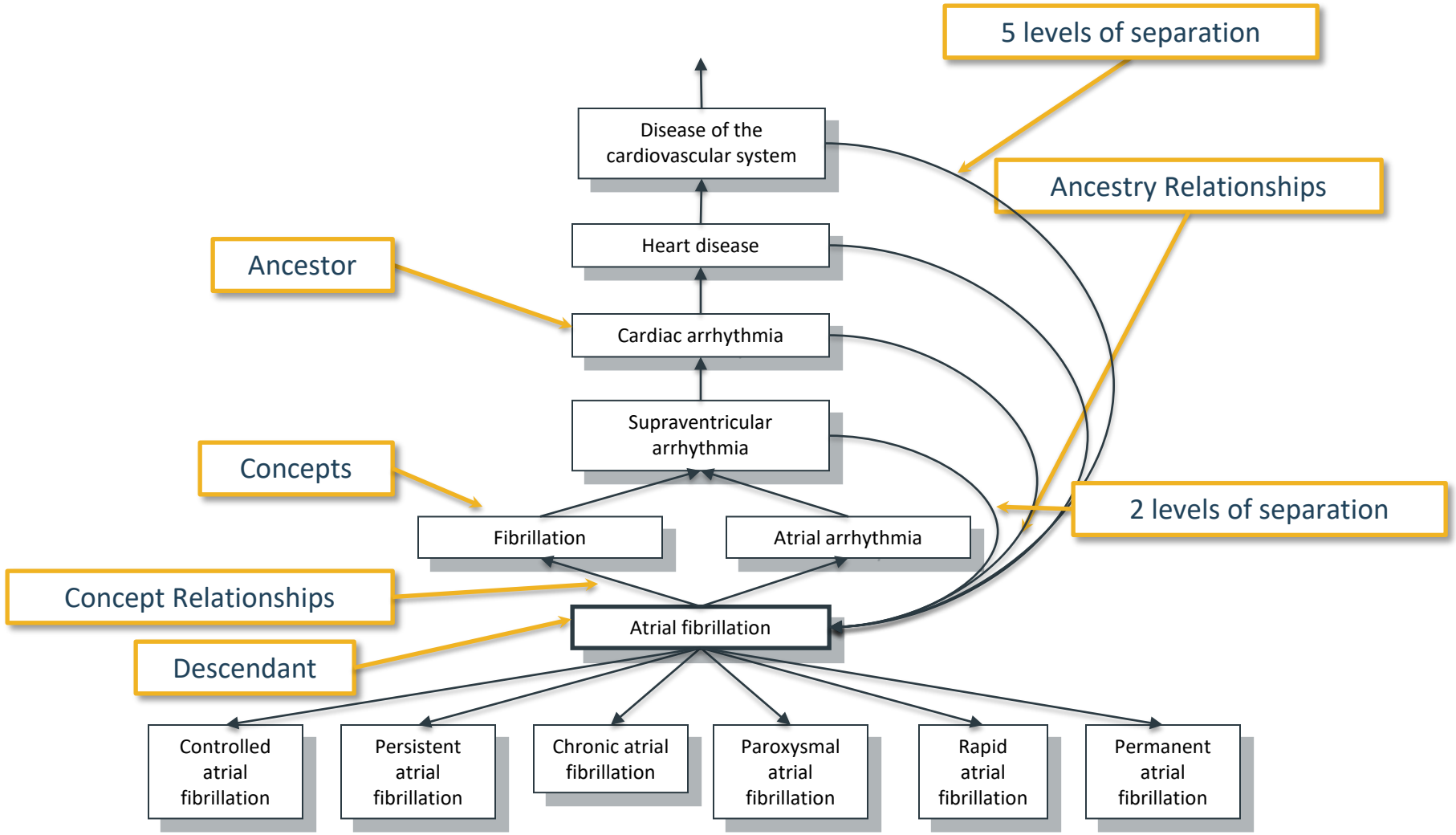


Panel



Hierarchy: Ancestors and Descendants

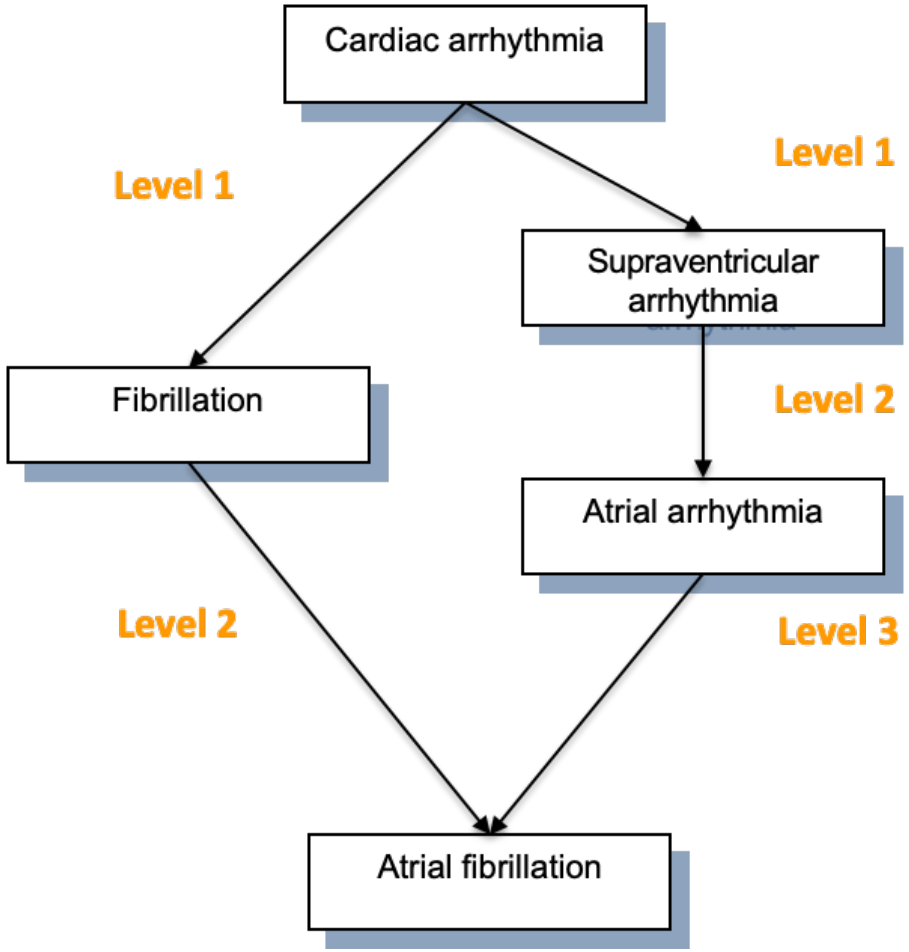
Ancestry Relationships: Higher-Level Relationships



Concept Ancestor: Structure

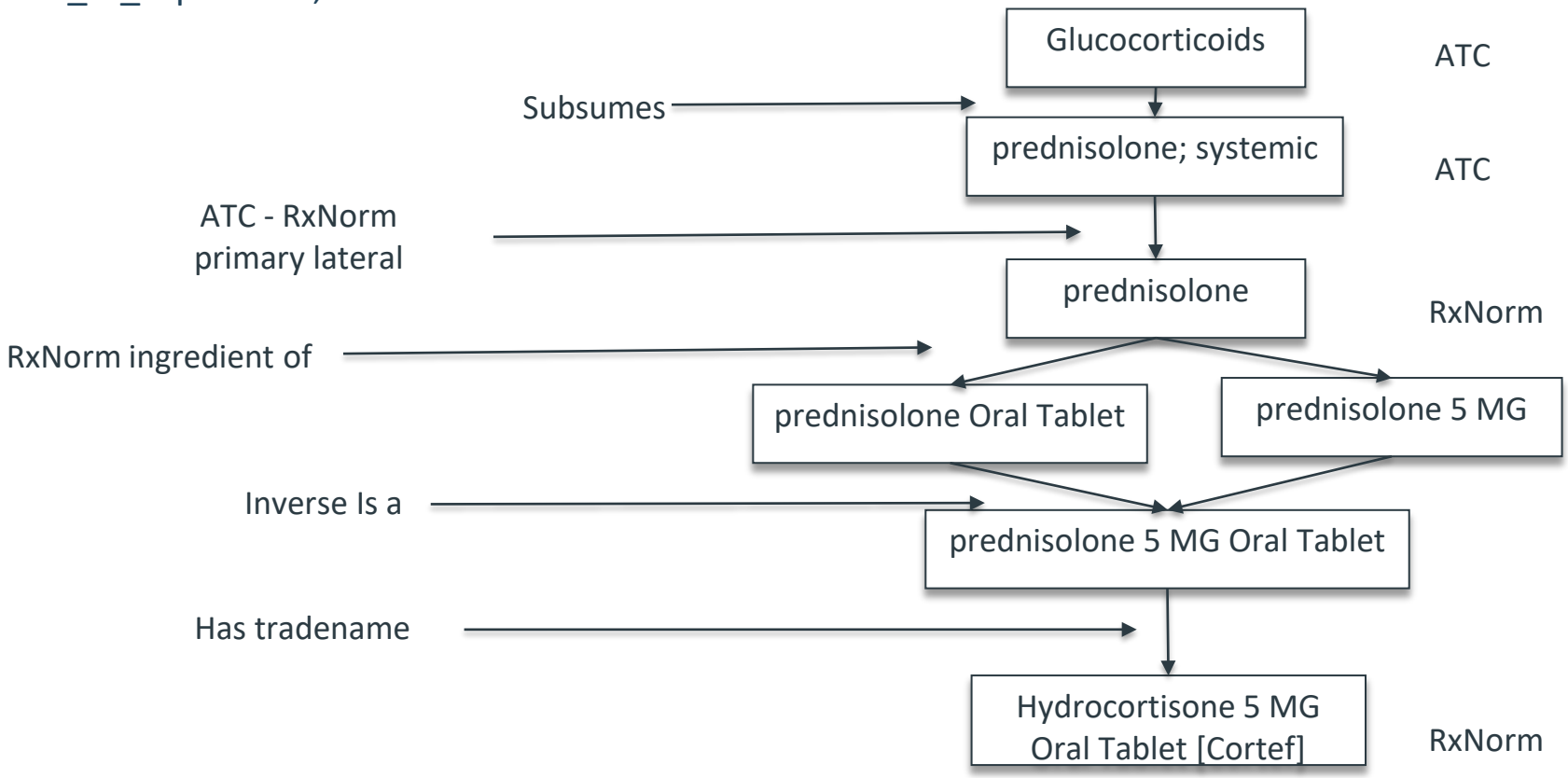
Self

ancestor	descendant	min_levels	max_levels	concept_name
44784217	44784217	0	0	**Cardiac arrhythmia**
44784217	4086313	1	1	Withdrawal arrhythmia
44784217	4185572	1	1	Ventricular arrhythmia
44784217	4248028	1	1	Supraventricular arrhythmia
44784217	44784234	1	1	Cardiac arrhythmia associated with genetic disorder
44784217	4226399	1	1	Fibrillation
44784217	4164083	1	1	Ectopic rhythm
44784217	45757098	1	1	Cardiac arrhythmia in mother complicating childbirth
44784217	4088986	1	1	Atrial escape complex
44784217	4068155	2	2	Atrial arrhythmia
44784217	4306984	2	2	Cardiac arrest due to trauma
44784217	4301015	2	2	Cardiac arrest due to pacemaker failure
44784217	4128968	2	2	Circulatory arrest
44784217	4256374	2	2	Cardiorespiratory arrest
44784217	4254116	2	5	Tachycardia-bradycardia
44784217	313217	2	3	**Atrial fibrillation**

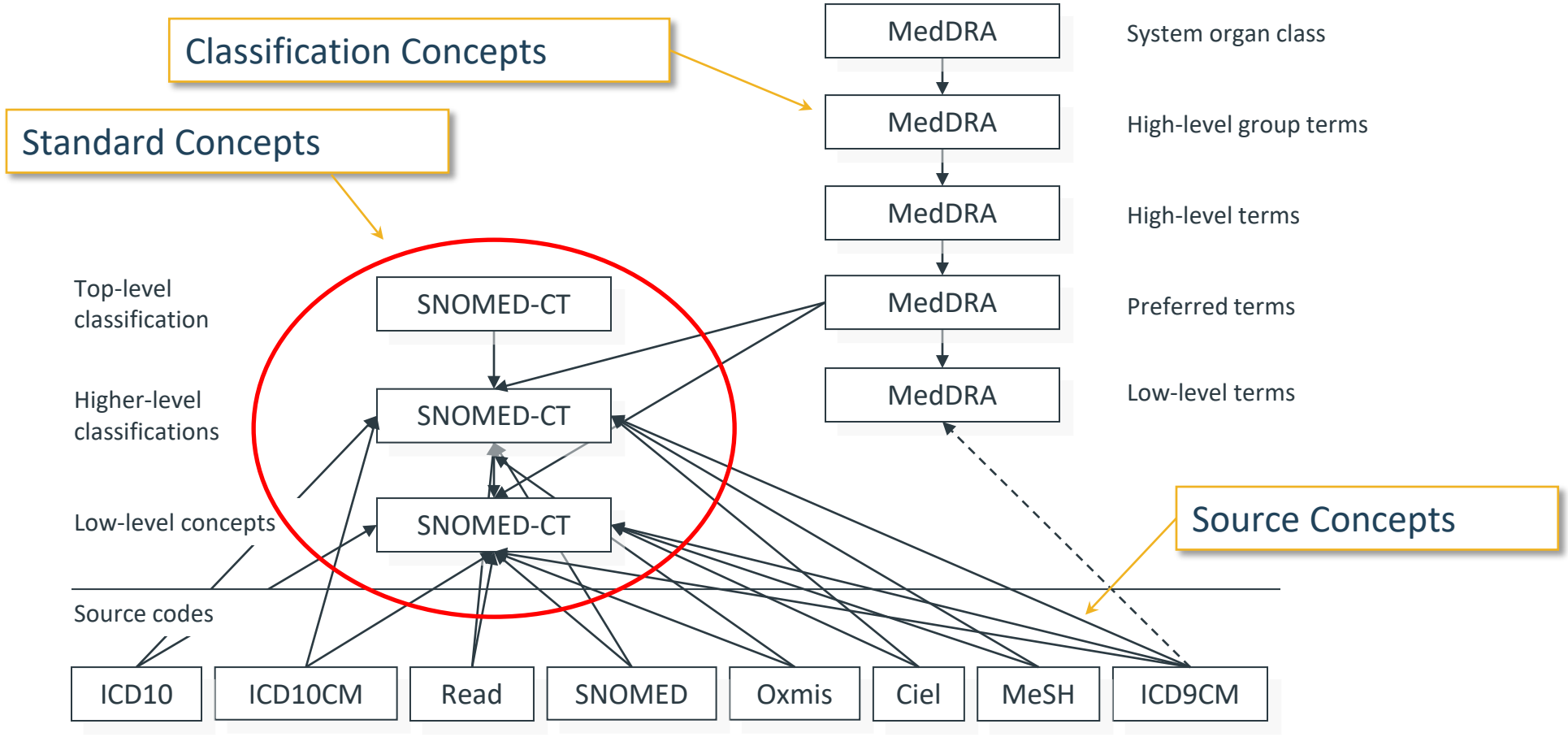


Concept Ancestor vs. Concept Relationship

```
SELECT ca.*, concept.concept_name FROM concept_ancestor ca
JOIN concept ON  descendant_concept_id = concept_id
WHERE ancestor_concept_id = 21602728 /* Glucocorticoids */
ORDER BY min_levels_of_separation;
```



Condition Concepts



Exploring the Hierarchy

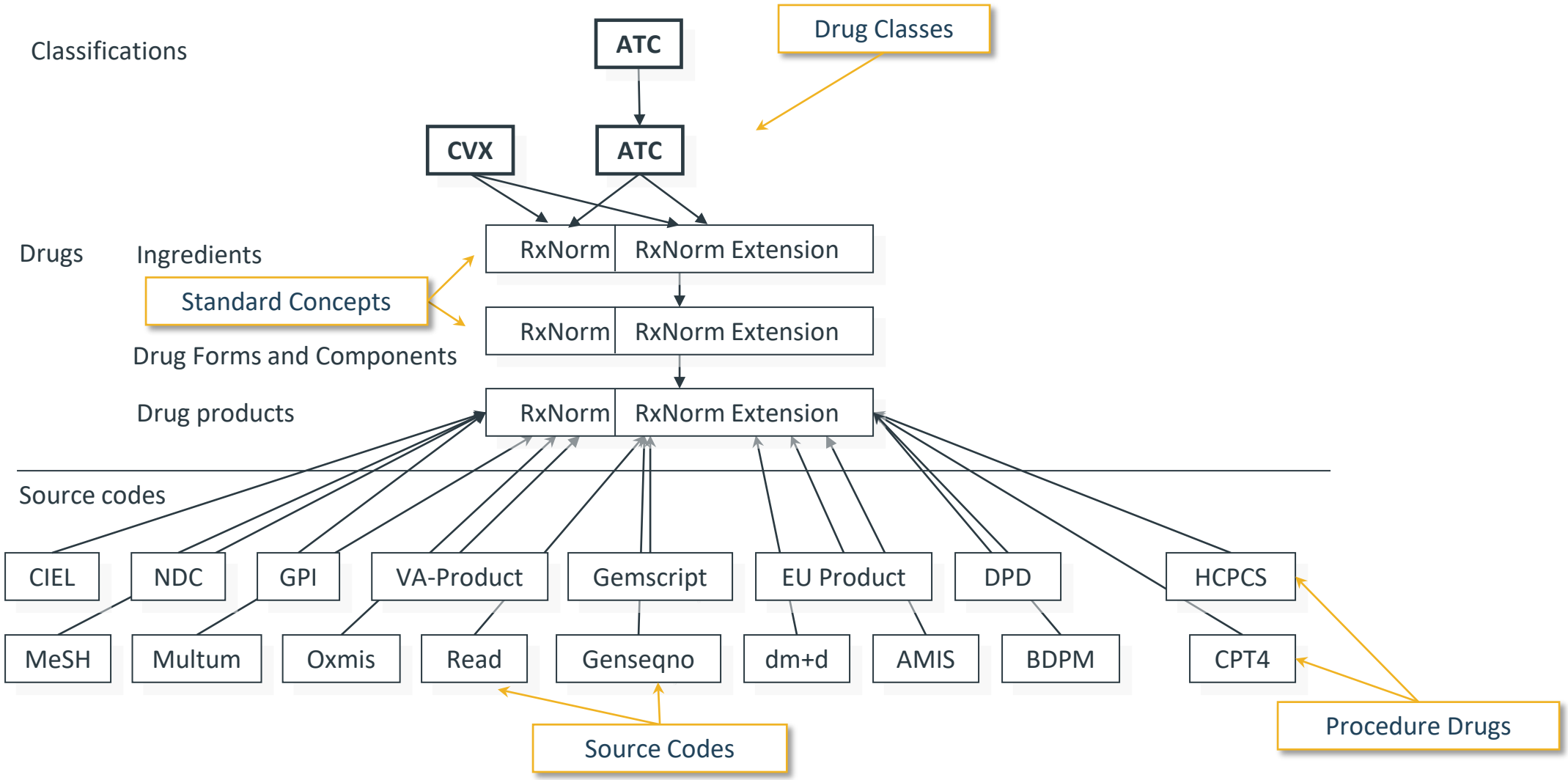
```
SELECT max_levels_of_separation, concept.*
FROM concept_ancestor
JOIN concept ON descendant_concept_id = concept_id
WHERE ancestor_concept_id = 44784217
/* cardiac arrhythmia */
ORDER BY max_levels_of_separation
```

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
1	4248028	Supraventricular arrhythmia	Condition	SNOMED	Clinical Finding	S
1	4262389	Tic-tac rhythm	Condition	SNOMED	Clinical Finding	S

```
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	MedDRA	PT	C
1	4226399	Fibrillation	Condition	SNOMED	Clinical Finding	S
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

Drug Hierarchy



Finding Drugs with an Ingredient

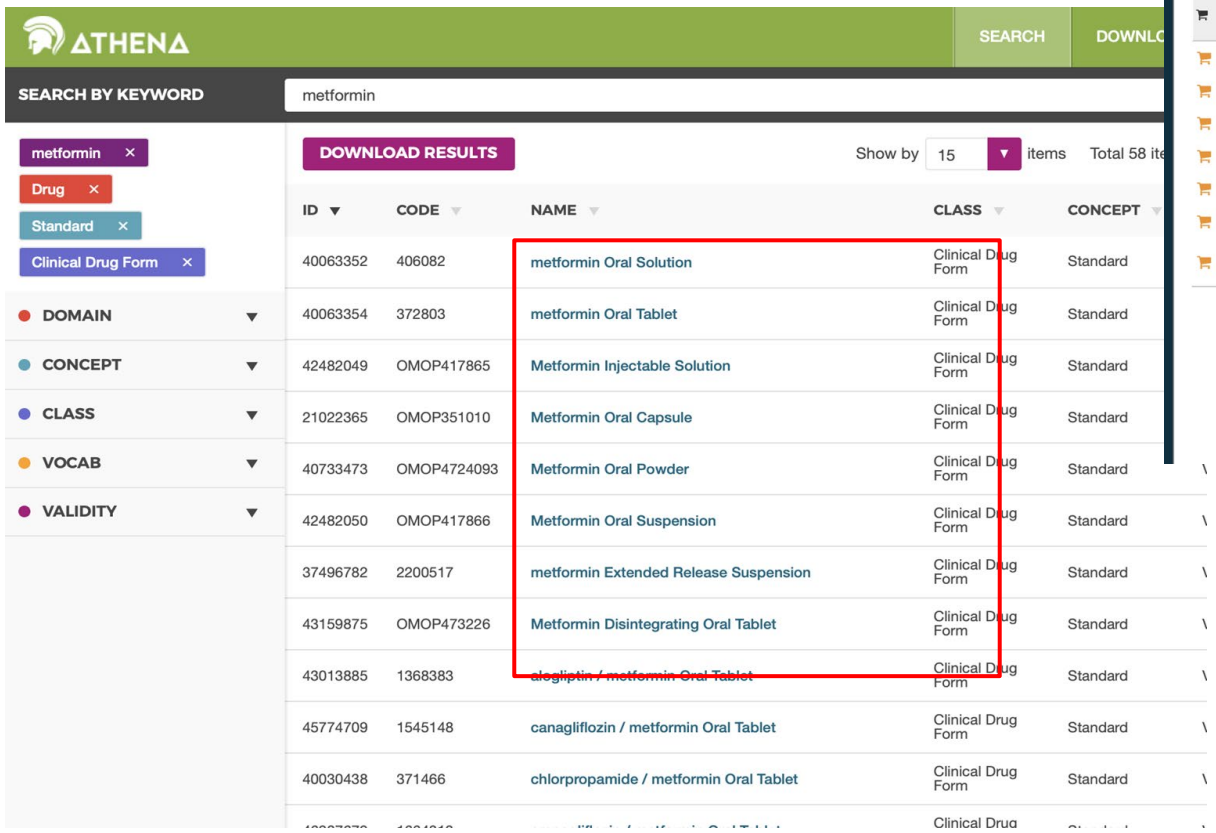
```
SELECT max_levels_of_separation, concept.*
FROM concept_ancestor
JOIN concept ON descendant_concept_id = concept_id
WHERE ancestor_concept_id = 1310149 /* Warfarin or 1322184 Clopidogrel*/
ORDER BY max_levels_of_separation
```

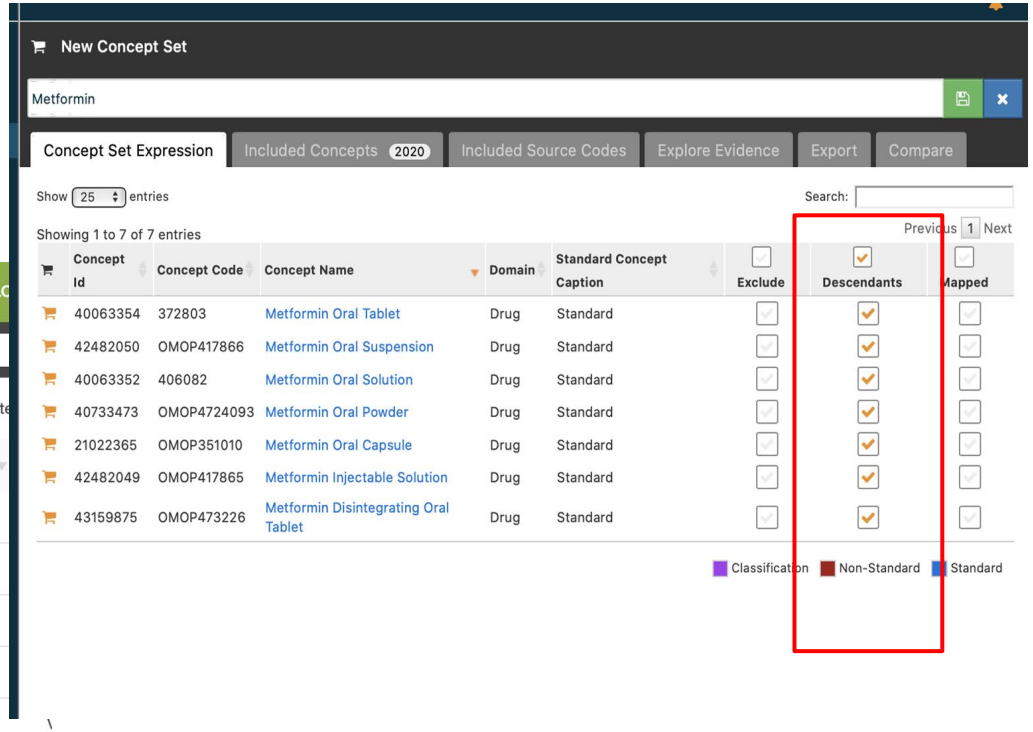
concept_id	concept_name	vocabulary_id	concept_class_id
1310149	Warfarin	RxNorm	Ingredient
36221229	Jantoven Pill	RxNorm	Branded Dose Group
40163559	Warfarin Sodium 6 MG	RxNorm	Clinical Drug Comp
40163544	Warfarin Sodium 3 MG [Jantoven]	RxNorm	Branded Drug Comp
21134746	Warfarin 0.2 MG/ML	RxNorm Extension	Clinical Drug Comp
21105414	Warfarin 5 MG/ML	RxNorm Extension	Clinical Drug Comp
36221228	Jantoven Oral Product	RxNorm	Branded Dose Group
40163565	Warfarin Sodium 7.5 MG	RxNorm	Clinical Drug Comp
21115236	Warfarin 0.3 MG/ML	RxNorm Extension	Clinical Drug Comp
40163509	Warfarin Sodium 1 MG	RxNorm	Clinical Drug Comp
21156284	1 ML Warfarin 0.02 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug
21095537	Warfarin 0.3 MG/ML Oral Solution	RxNorm Extension	Clinical Drug
21105427	Warfarin 0.4 MG/ML Oral Solution	RxNorm Extension	Clinical Drug
21046557	Warfarin 1 MG/ML Oral Solution	RxNorm Extension	Clinical Drug
40093133	Warfarin Oral Tablet [Coumadin]	RxNorm	Branded Drug Form
40093134	Warfarin Oral Tablet [Jantoven]	RxNorm	Branded Drug Form
21077698	1 ML Warfarin 1 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug
40163534	Warfarin Sodium 2.5 MG Oral Tablet	RxNorm	Clinical Drug
40163530	Warfarin Sodium 2 MG/ML Injectable Solution	RxNorm	Clinical Drug
21066136	Warfarin 5 MG Oral Tablet [Marevan]	RxNorm Extension	Branded Drug
40163542	Warfarin Sodium 3 MG Oral Tablet [Jantoven]	RxNorm	Branded Drug
21116822	1 ML Warfarin 0.6 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
21175784	1 ML Warfarin 0.1 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug
21175783	1 ML Warfarin 0.832 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug

concept_id	concept_name	vocabulary_id	concept_class_id
1322184	clopidogrel	RxNorm	Ingredient
21043471	clopidogrel Oral Suspension	RxNorm Extension	Clinical Drug Form
36229332	Plavix Pill	RxNorm	Branded Dose Group
21043470	clopidogrel Oral Solution	RxNorm Extension	Clinical Drug Form
21023802	clopidogrel Injectable Solution	RxNorm Extension	Clinical Drug Form
21023806	clopidogrel 5 MG	RxNorm Extension	Clinical Drug Comp
1322187	clopidogrel 75 MG	RxNorm	Clinical Drug Comp
21141600	clopidogrel 1 MG/ML	RxNorm Extension	Clinical Drug Comp
36222254	clopidogrel Oral Product	RxNorm	Clinical Dose Group
21092477	clopidogrel 5 MG/ML	RxNorm Extension	Clinical Drug Comp
21177192	100 ML clopidogrel 1 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
21047899	1 ML clopidogrel 5 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
21121870	clopidogrel 5 MG/ML Oral Suspension	RxNorm Extension	Clinical Drug
21063106	clopidogrel 75 MG Oral Tablet [Grepid]	RxNorm Extension	Branded Drug
1322190	clopidogrel 300 MG Oral Tablet [Plavix]	RxNorm	Branded Drug
21121869	clopidogrel 75 MG Injectable Solution	RxNorm Extension	Clinical Drug
21053280	clopidogrel 6 MG Injectable Solution	RxNorm Extension	Clinical Drug
21023810	clopidogrel 4 MG Injectable Solution	RxNorm Extension	Clinical Drug
21106783	1 ML clopidogrel 1 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug
19075601	clopidogrel 75 MG Oral Tablet	RxNorm	Clinical Drug
21102364	clopidogrel 1 MG/ML Oral Suspension	RxNorm Extension	Clinical Drug
40095879	clopidogrel Oral Tablet [Plavix]	RxNorm	Branded Drug Form
40095878	clopidogrel Oral Tablet	RxNorm	Clinical Drug Form
21088717	100 ML clopidogrel 15 MG/ML Oral Suspension	RxNorm Extension	Quant Clinical Drug

Finding Mono-Therapy Drugs

- Count ingredients in DRUG_STRENGTH
- Count ingredients in CONCEPT_ANCESTOR
- Select descendants of Clinical Dose Form





Find Members of Drug Classes

```
SELECT max_levels_of_separation, concept.*
FROM concept_ancestor
JOIN concept ON descendant_concept_id = concept_id
WHERE ancestor_concept_id = 21600961 /* ATC Antithromboic Agent */
  AND concept_class_id = 'Ingredient'
ORDER BY max_levels_of_separation
```

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id
46275677	cangrelor	Drug	RxNorm	Ingredient
45892847	edoxaban	Drug	RxNorm	Ingredient
1322184	clopidogrel	Drug	RxNorm	Ingredient
44818499	vorapaxar	Drug	RxNorm	Ingredient
43013024	apixaban	Drug	RxNorm	Ingredient
42898933	defibrotide	Drug	RxNorm	Ingredient
42801108	Protein C	Drug	RxNorm	Ingredient
40241331	rivaroxaban	Drug	RxNorm	Ingredient
1310149	Warfarin	Drug	RxNorm	Ingredient
40241186	Ticagrelor	Drug	RxNorm	Ingredient
40228152	dabigatran etexilate	Drug	RxNorm	Ingredient
40163718	prasugrel	Drug	RxNorm	Ingredient
35604848	selexipag	Drug	RxNorm	Ingredient
19136187	Streptokinase	Drug	RxNorm	Ingredient
19129274	reviparin	Drug	RxNorm	Ingredient

Find Members of Drug Classes

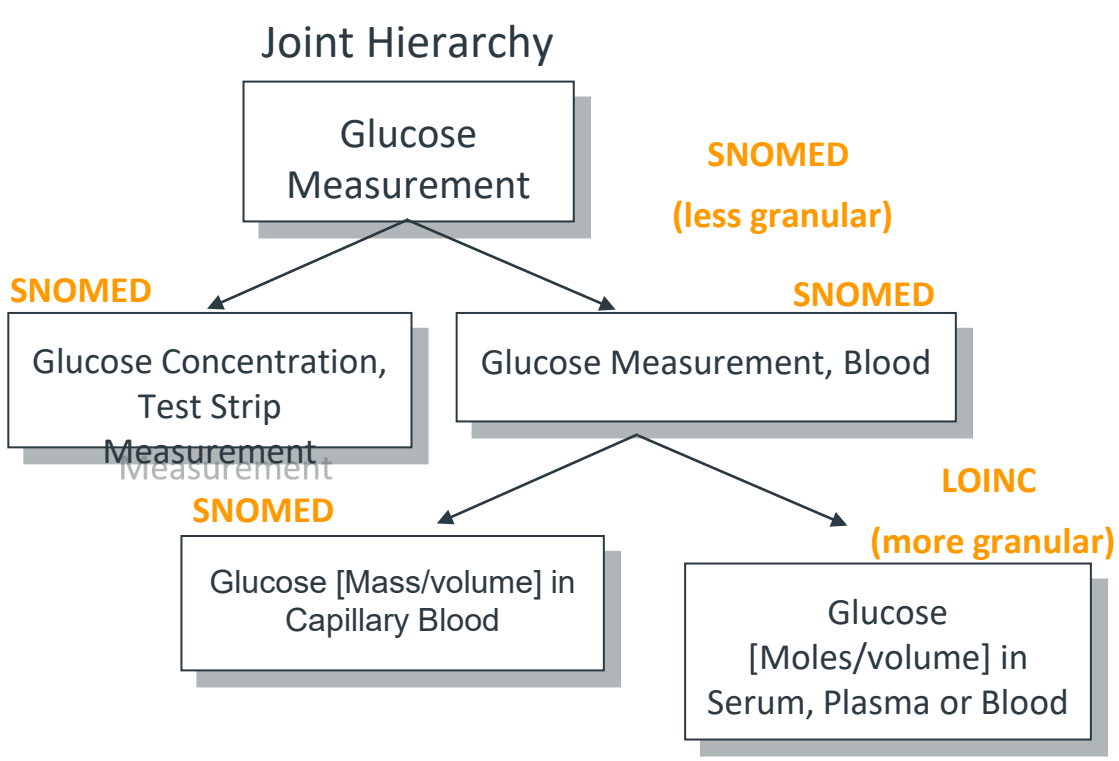
```
SELECT concept.*
FROM concept_ancestor
JOIN concept
ON descendant_concept_id = concept_id
WHERE ancestor_concept_id = 21600033
/* ATC Corticosteroids for local oral treatment */
```

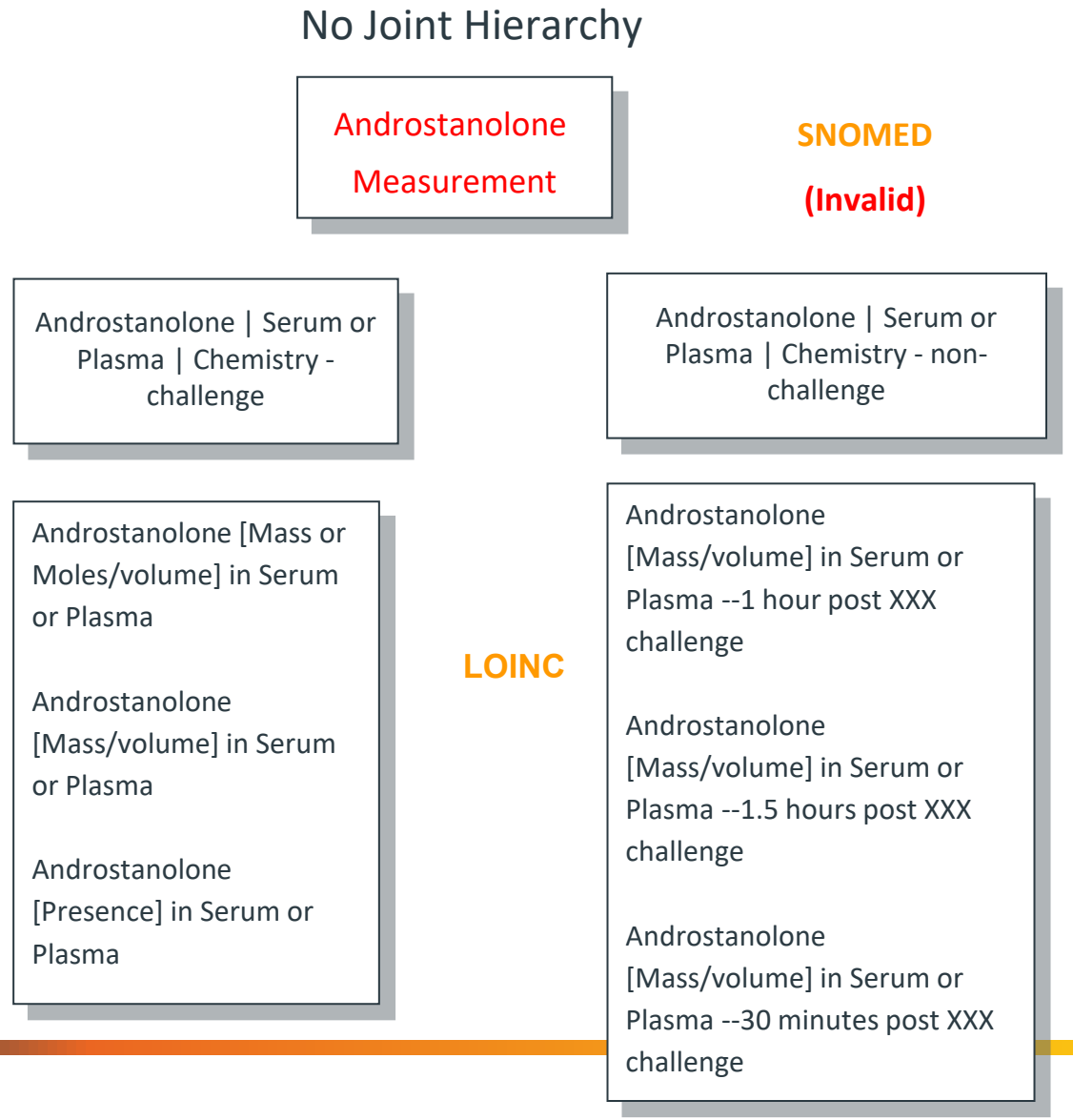
concept_id	concept_name	vocabulary_id	concept_class_id
976018	hydrocortisone 0.005 MG/MG Toothpaste	RxNorm	Clinical Drug
1518491	dexamethasone 0.2 MG/ML Oral Solution	RxNorm	Clinical Drug
1518851	dexamethasone 0.1 MG/ML Oral Solution [Decadron]	RxNorm	Branded Drug
1518872	dexamethasone 0.1 MG/ML Oral Solution [Hexadrol]	RxNorm	Branded Drug
40049693	hydrocortisone Oral Lozenge	RxNorm	Clinical Drug Form
40049747	hydrocortisone Toothpaste	RxNorm	Clinical Drug Form
40085513	triamcinolone Oral Paste	RxNorm	Clinical Drug Form
40085514	triamcinolone Oral Paste [Adcortyl]	RxNorm	Branded Drug Form
43739364	100 ML prednisolone 2 MG/ML / Salicylic Acid 4 MG/ML Oral Solution	RxNorm Extension	Quant Clinical Drug

```
SELECT concept.*
FROM concept_ancestor
JOIN concept
ON descendant_concept_id = concept_id
WHERE ancestor_concept_id = 21602722
/* ATC CORTICOSTEROIDS FOR SYSTEMIC USE */
```

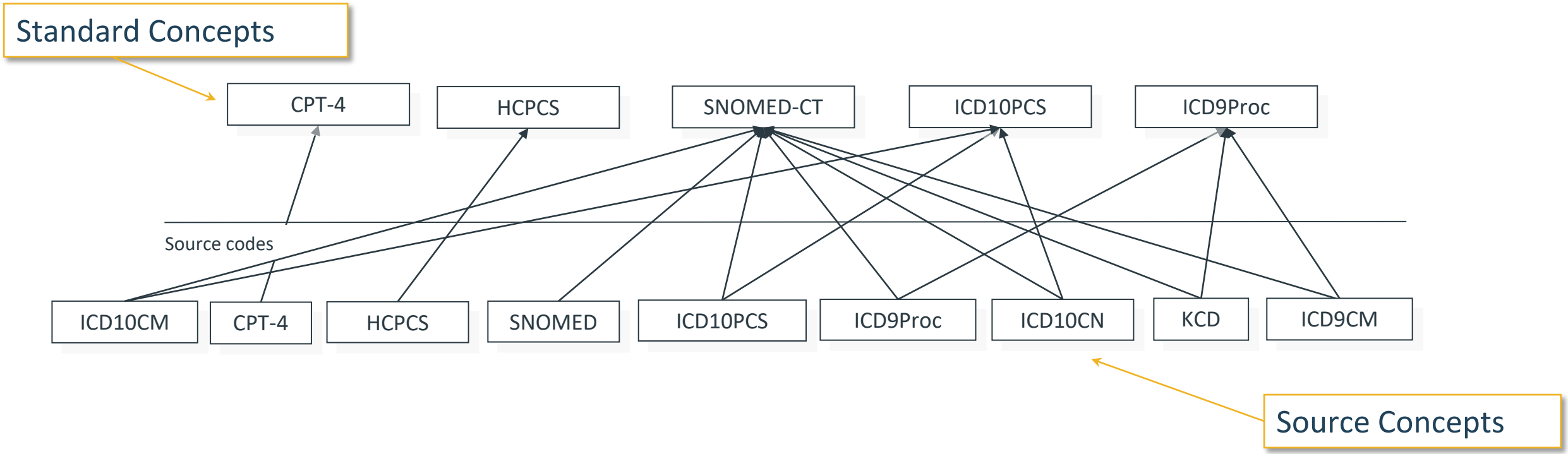
concept_id	concept_name	vocabulary_id	concept_class_id
590101	4 ML Hydrocortisone 125 MG/ML Injection	RxNorm Extension	Quant Clinical Drug
792424	triamcinolone Injection [Zilretta]	RxNorm	Branded Drug Form
1506479	methylprednisolone 16 MG Oral Tablet	RxNorm	Clinical Drug
1507707	cortisone 5 MG Oral Tablet	RxNorm	Clinical Drug
1559869	triamcinolone acetonide 1.5 MG Oral Tablet [Vetalog]	RxNorm	Branded Drug
1592182	{27 (dexamethasone 1.5 MG Oral Tablet) } Pack	RxNorm	Clinical Pack
19034806	methylprednisolone 125 MG Injection [A-MethaPred]	RxNorm	Branded Drug
19034807	methylprednisolone 62.5 MG/ML Injectable Solution [Solu-Medrol]	RxNorm	Branded Drug

Measurement Hierarchy





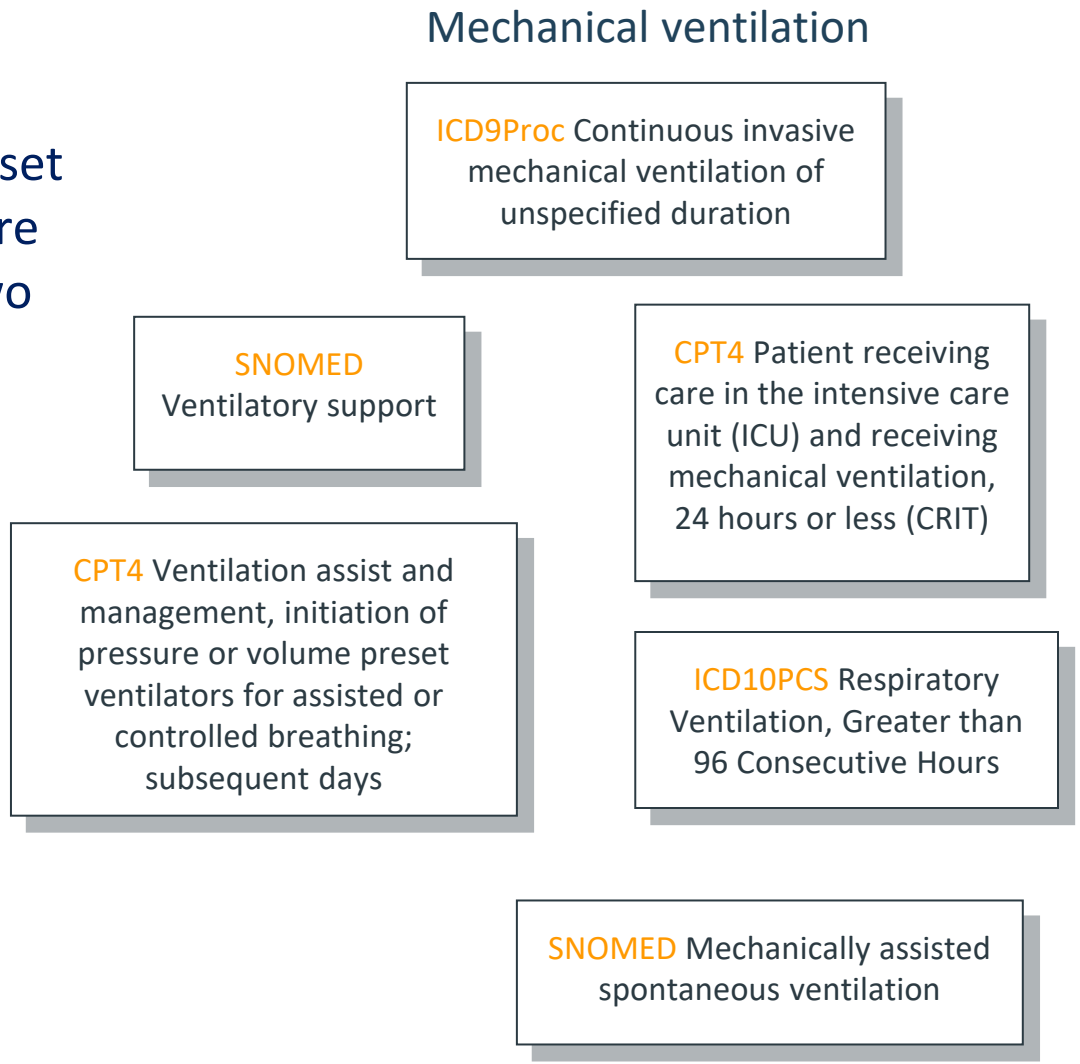
Procedure Hierarchy



- SNOMED and ICD-9(10) Proc codes can be within each other's hierarchy
- HCPCS and CPT-4 are not embedded in the SNOMED - ICD-9(10) Proc hierarchy

Procedure Hierarchy

When creating a comprehensive concept set for procedures, make sure that you have at least two vocabularies in your set: SNOMED and CPT-4 or HCPCS!



Breakout Session 2
Exercises 45 minutes – Review 30 minutes

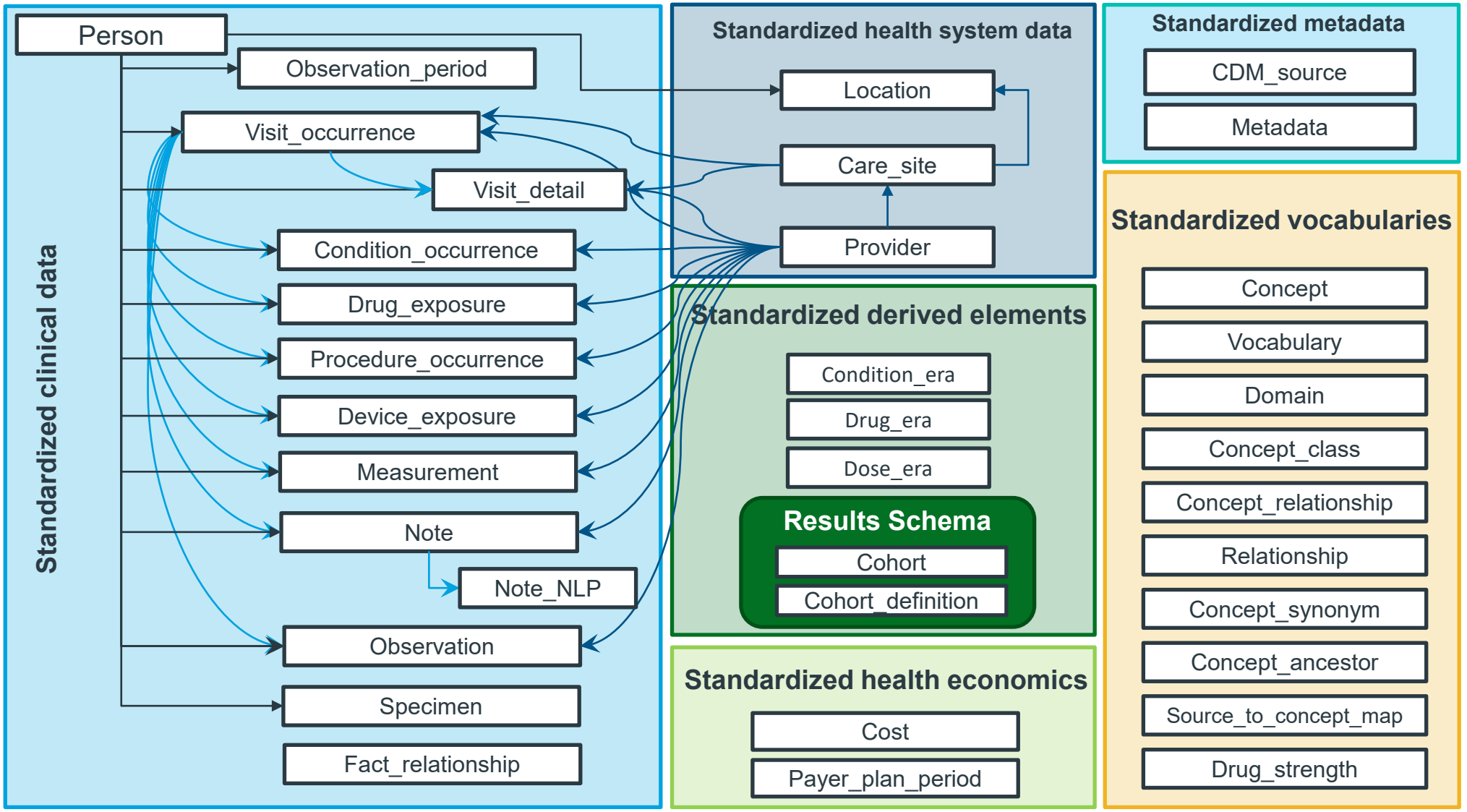

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

OMOP CDM

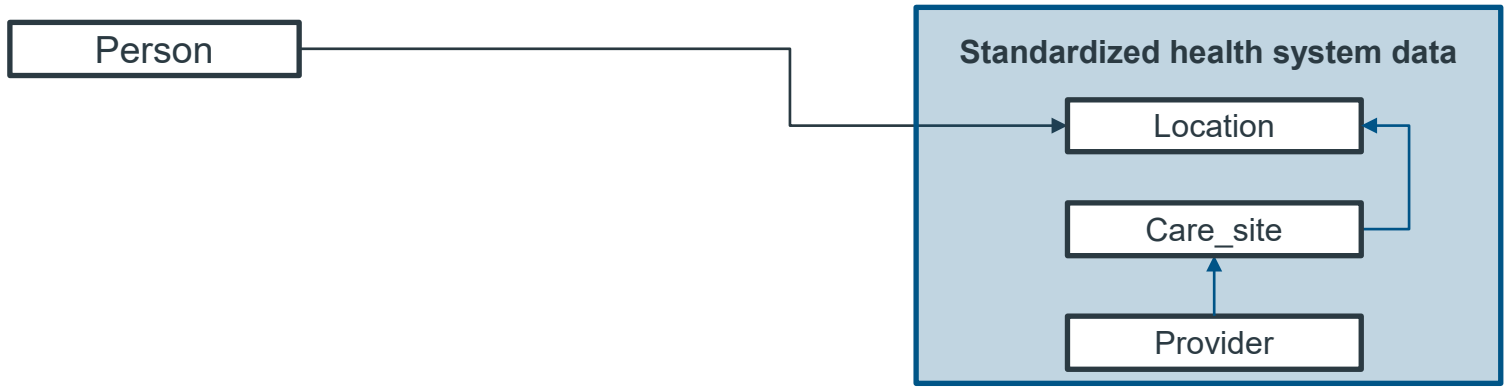
OMOP CDM v5.3.1

- Map
  - Who – Person and Providers
  - Where – Locations and Care Site
  - What – Exposures, Measurements, Specimens
  - When - Occurrences
  - How – Vocabulary
  - Why – Research, Improve Health

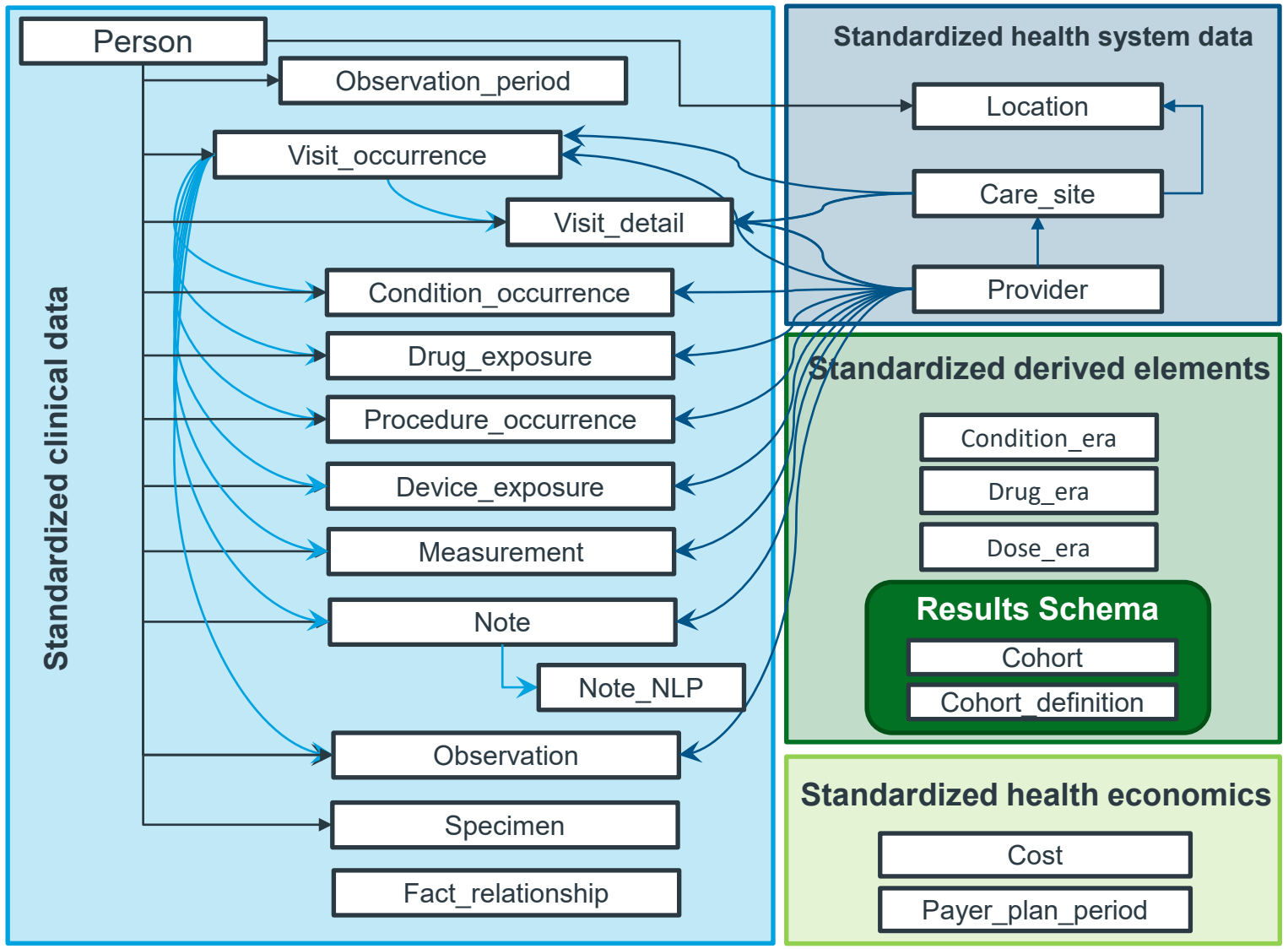
OMOP CDM v5.3.1



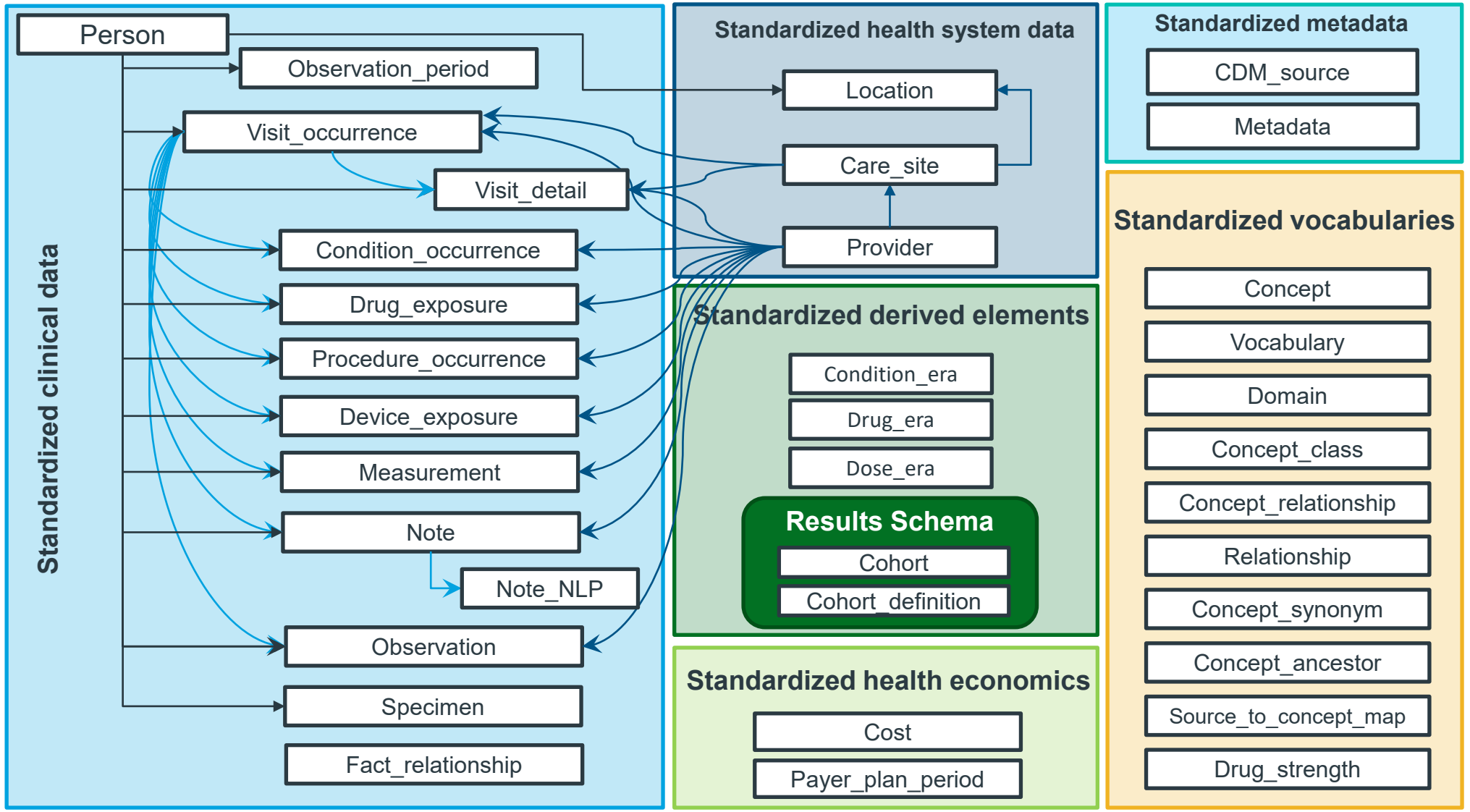
OMOP CDM v5.3.1 Who and Where



OMOP CDM v5.3.1 – What and When



OMOP CDM v5.3.1 - How



OMOP CDM Principles

- Patient centric
- Vocabulary and Data Model are standardized
- Domain-oriented concepts
- Accommodates data from various sources
- Preserves data provenance & source data
- Extendable & Evolving
- Database Platform Independent

CDM Version Control

- Working group meets once a month to discuss proposed changes to the CDM
- All CDM documentation, versions, and proposals located on GitHub
  - https://ohdsi.github.io/CommonDataModel/index.html
  - Proposals tracked and discussed as GitHub issues
  - https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html
- Meeting information can be found on the working group wiki page
- Please contact Clair Blacketer (mblacke@its.jnj.com) for more information

Common Myths

Myth Busters

Common Doubts

1. **Myth #1**
   “Loss of Data”
2. **Myth #2**
   “Loss of Accuracy in Conversion”
3. **Myth #3**
   “Loss of Accuracy in Vocab Mapping”
4. **Myth #4**
   “It Takes Too Much Time”
5. **Myth #5**
   “You Don’t Have My Use Case in OMOP”
6. **Myth #6**
   “I Have to Learn New Medical Terminology”

OHDSI Standards

Myth #1

“Loss of Data”

Myth

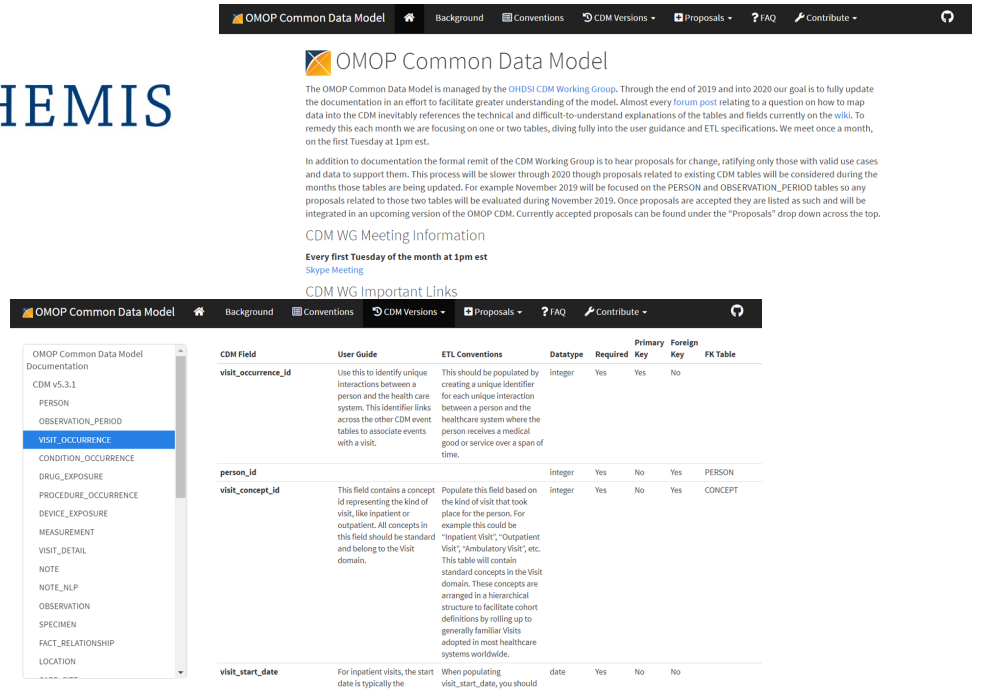
Some believe that converting to a CDM will result in “losing” data because it does not map to the standard. There is skepticism about why the total number of patients in the source data does not match the CDM data.

Truth

- Source data is preserved within the CDM, even if no standard mapping exist
- OHDSI has standard conventions that enforce rules data must follow, including addressing:
  - Patients without transactions
  - Cleaning dirty data
  - Patient IDs reused
  - And more

Resources

THEMIS



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

Data Quality

Myth #1

“Loss of Data”

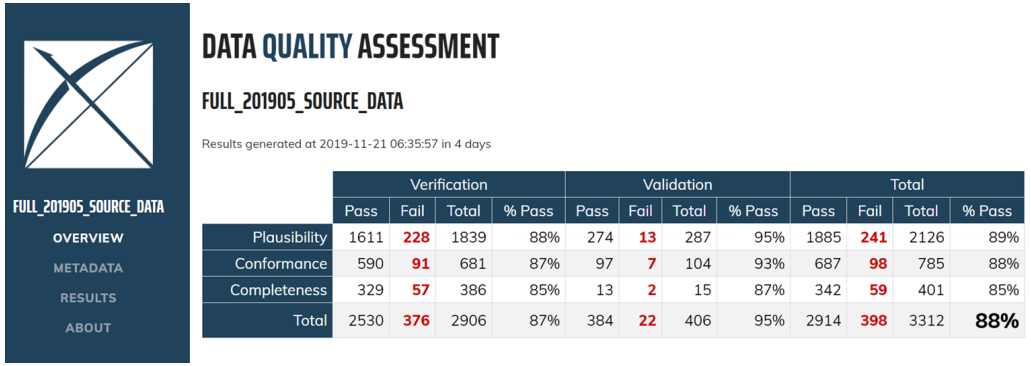
Myth

Some believe that converting to a CDM will result in “losing” data because it does not map to the standard. There is skepticism about why the total number of patients in the source data does not match the CDM data.

Truth

- After converting data from source to OMOP, have standard quality control including:
  - OHDSI Data Quality Dashboard
  - Achilles checks

Resources



FULL_201905_SOURCE_DATA

OVERVIEW

METADATA

RESULTS

ABOUT

DATA QUALITY ASSESSMENT

FULL_201905_SOURCE_DATA

Results generated at 2019-11-21 06:35:57 in 4 days

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	1611	228	1839	88%	274	13	287	95%	1885	241	2126	89%
Conformance	590	91	681	87%	97	7	104	93%	687	98	785	88%
Completeness	329	57	386	85%	13	2	15	87%	342	59	401	85%
Total	2530	376	2906	87%	384	22	406	95%	2914	398	3312	**88%**

Retaining the Accuracy of Source Data

Myth #2

"Loss of Accuracy in Conversion"

Myth

Some believe that using OMOP standards can degrade the accuracy of the data. There could be issues in the conversions ability to accurately reflect a data set.

Truth

- Validation studies have found minimal differences in the source to OMOP data
- DA France / LPD France validation study found consistency between native and OMOP data sets

Examples

CONVERSION OF A FRENCH ELECTRONIC MEDICAL RECORD DATABASE INTO THE OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP COMMON DATA MODEL

Marie-Sophie Schwalm[a], Thomas Raoult[a], Diana Chu[b], Urvi Shah[c], Meghana Potdar[d], Mui Van Zandt[e], Gabriel Coffin[a], Sophie L Jouaville[a,1]
[a]IQVIA, Boulogne Billancourt, France ; [c,d]IQVIA, New Jersey, USA ; [b]QuintilesIMS, Pennsylvania, USA ; [e]IQVIA, California, USA

[1] Corresponding Author : sophie.jouaville@iqvia.com

IMS Health & Quintiles are now IQVIA™

Table 3 : Patients profile : comparison between the 3 data sources : DA FR Native / DA FR OMOP / LPD Native

	DA FR NATIVE	DA FR OMOP	LPD NATIVE
	N=12 302*	N=12 382*	N=15 623
Males	7 179 (58.4)	7 231 (58.4)	9 291 (59.5)
Age (in year))	74.6 (± 11.1)	74.4 (± 11.1)	74.6 (± 11.1)
Age ≥75 years	7 055 (57.3)	7 032 (56.8)	8 981 (57.5)
BMI	28.1 (± 5.3)	28.2 (± 5.3)	27.9 (± 5.3)
Diastolic blood pressure	76.2 (± 12)	79.3 (± 20)	76 (± 9)
Systolic blood pressure	131.9 (± 15)	132.7 (± 15)	133 (± 15)
Co-treatments			
NSAIDs	5 320 (43.3)	5 341 (43.1)	7 492 (48.0)
Anti-arrhythmic drug	6 014 (48.9)	6 018 (48.6)	7 425 (47.5)
Injectable anticoagulants	373 (3.0)	377 (3.0)	482 (3.1)
$CHADS_2$ Score[2]			
0	2 262 (18.4)	2 311 (18.7)	2 638 (16.9)
1	3 997 (32.5)	4 027 (32.5)	5 026 (32.2)
≥2	6 043 (49.1)	6 044 (48.8)	7 959 (50.9)
CHA_2DS_2-Vasc Score[3]			
0	822 (6.7)	844 (6.8)	998 (6.4)
1	1 559 (12.7)	1 591 (12.8)	1 774 (11.4)
≥2	9 921 (80.6)	9 947 (80.3)	12 851 (82.3)

**OMOP model assign an occurrence date to all events including clinical measures and lab results. As a result there is a slight difference in visit number between DA FR OMOP and DA FR NATIVE, which explains a slightly elevated number of included patients in DA FR OMOP.*

Evaluating the Accuracy of Vocabulary Mapping

Myth #3

"Loss of Accuracy in Vocab Mapping"

Myth

Some believe that using OMOP vocabulary mappings are incorrect. There could be issues in the preservation of source information as it is translated to standard concepts.

Truth

- Validation studies have found minimal differences in the source to OMOP data
- EMA Validation study of IQVIA IMRD UK found consistency between source and OMOP CDM data

Examples

Clin Pharmacol Ther. 2020 Apr;107(4):915-925. doi: 10.1002/cpt.1785. Epub 2020 Mar 2.

Can We Rely on Results From IQVIA Medical Research Data UK Converted to the Observational Medical Outcome Partnership Common Data Model?: A Validation Study Based on Prescribing Codeine in Children

Gianmario Candore [1], Karin Hedenmalm [1], Jim Slattery [2], Alison Cave [2], Xavier Kurz [2], Peter Arlett [2]
Affiliations – collapse

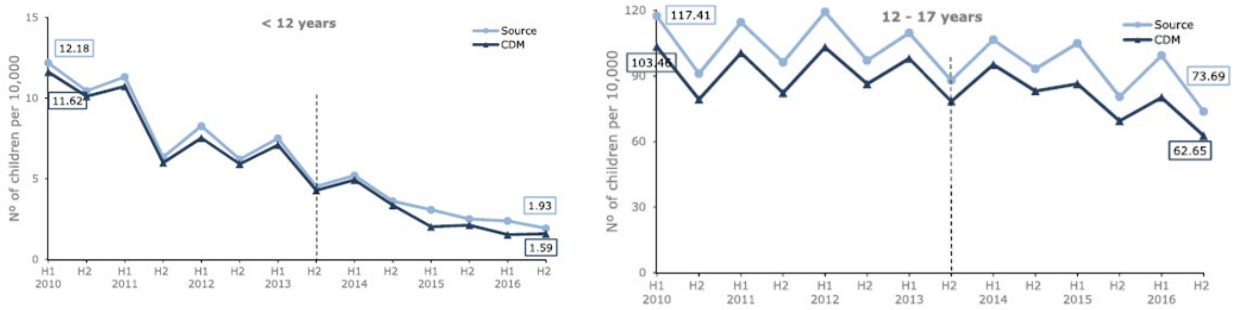


Figure 1-2: Six-monthly prevalence (per 10,000) of codeine prescribing for pain in 0–17 years

OMOP Conversion Overview

Myth #4

"It Takes Too Much Time"

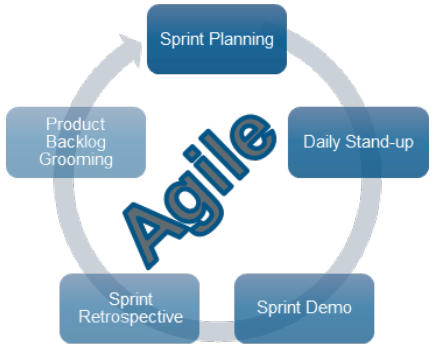
Myth

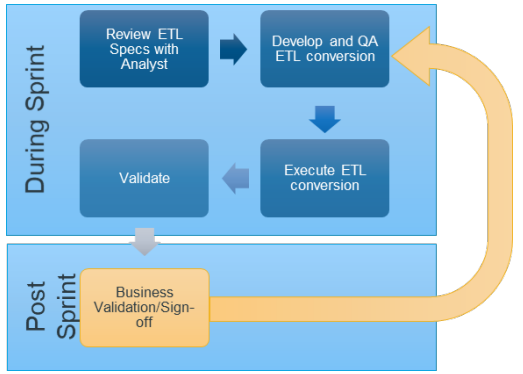
Taking data from source format to OMOP common data model is tedious and time consuming.

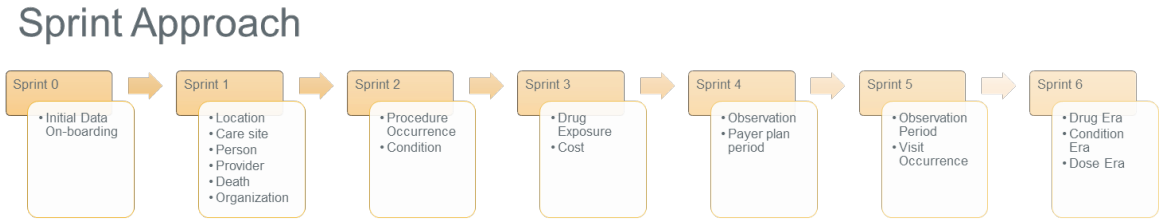
Truth

- It's true, it takes time to convert data into the OMOP CDM.
- We spend time cleaning the data and removing data that cannot contribute to analytical use cases.
- We push down common business rules into the ETL process.
- Studies execute faster due to standardization and common model.

Examples







OMOP Evolves to Meet Analytical Needs

Myth #5

"You Don't Have My Use Case in OMOP"

Myth

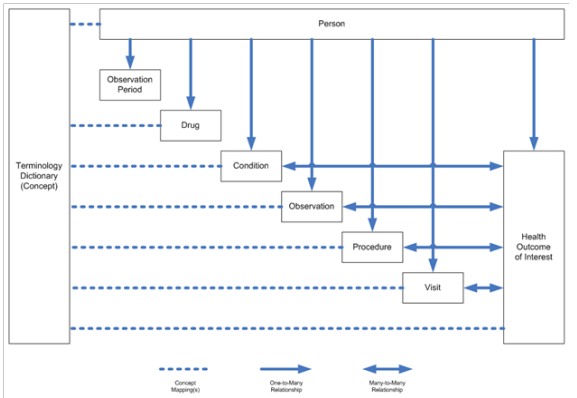
OMOP is not good enough for my analytical use case **or** covers the therapeutic area that I want to study.

Truth

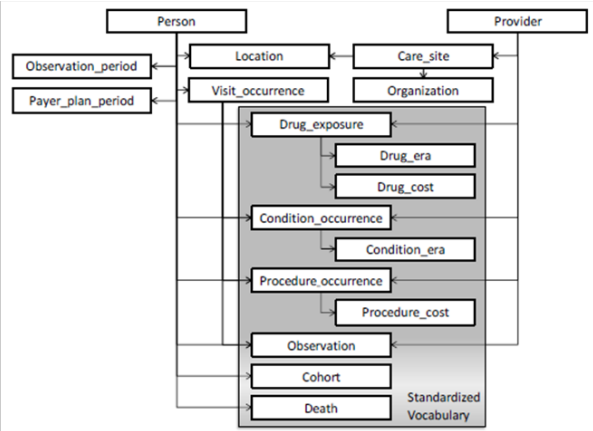
- It's true, OMOP was not built for every use case, but it has evolved over the years to support a variety of analytic needs.
- If you have a use case, bring it up on the CDM working group. If there's enough of a common need, OMOP will evolve to support additional use cases.

Examples

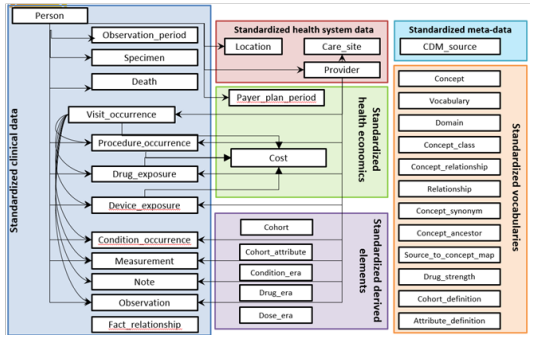
OMOP CDMv2



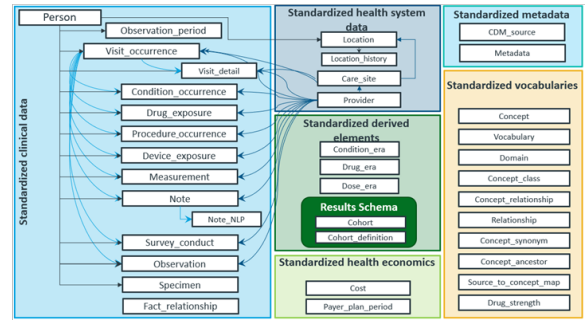
OMOP CDMv4



OMOP CDMv5



OMOP CDMv6



OMOP Oncology Extension

Myth #5

"You Don't Have My Use Case in OMOP"

Myth

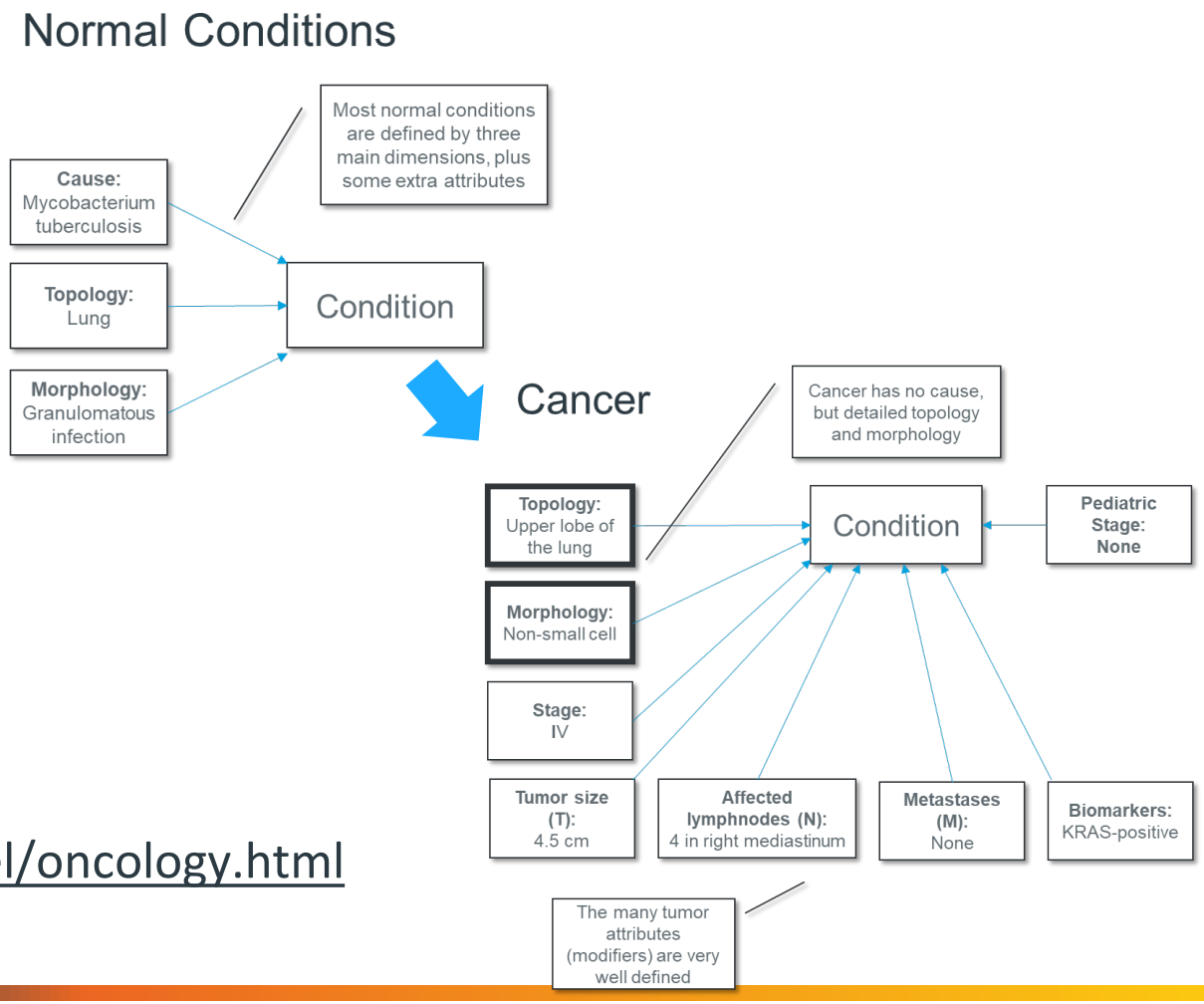
OMOP cannot support oncology data.

Truth

- OMOP Oncology experts created an Oncology working group.
- The Oncology working group designed an Oncology extension to house oncology-specific information in the OMOP CDM.

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

Examples



OMOP Vocabulary Hierarchy

Myth

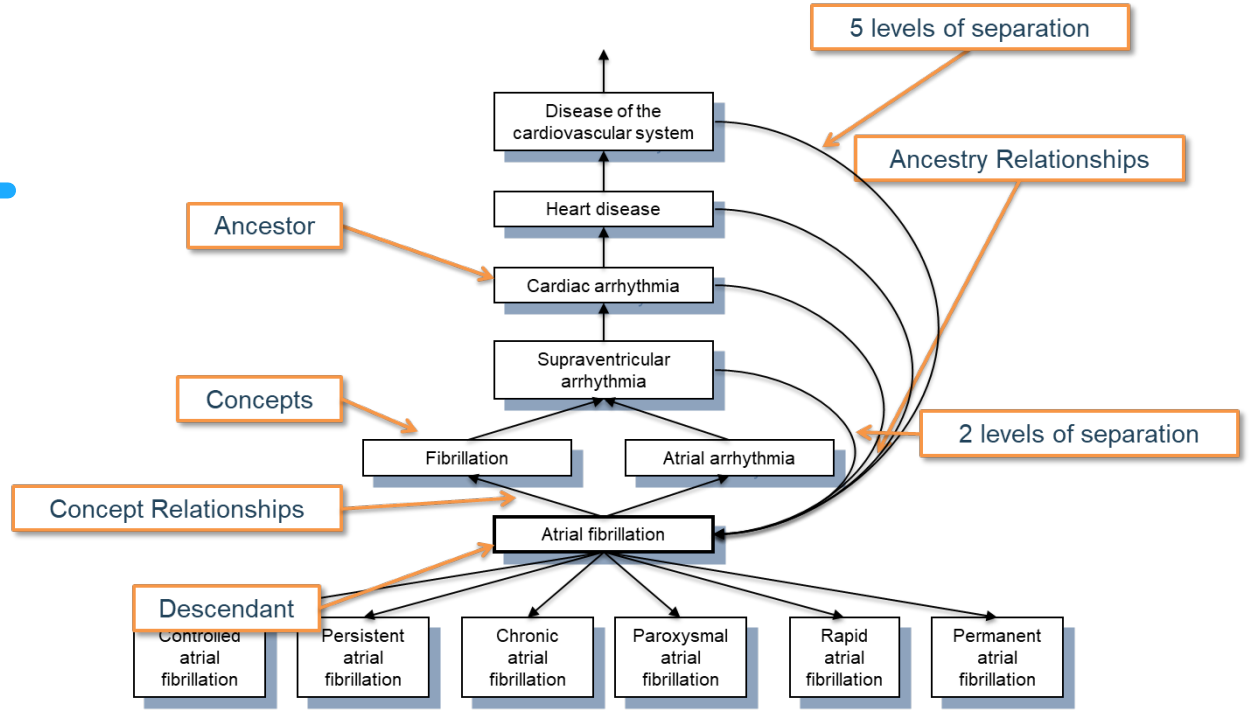
OMOP is forcing me to speak in SNOMED, RxNorm and LOINC codes.

Truth

- OMOP CDM preserves the source codes from the native data and creates a map to a standard concept that is interoperable across all data assets
- You can always start with source codes (e.g. ICD-9/ICD-10) and use the maps to relationships to find standards.
- The hierarchy structure in the standard vocabulary is easily navigated using ancestors and descendants.

Examples

Ancestry Relationships



Common Issues

Common OMOP CDM issues

Non-standard Vocabulary
Codes mapped to OMOP vocabulary aren't mapped to a 'Standard'

Multiple Input on Records
Some records will contain multiple coding systems and text. A hierarchy must be selected to avoid duplicate records

Non-Clinical Events
Due to text options in EHR Data, many options are not clinical events (e.g. 'Tuesday' or 'XXYZ'). These records will be scrubbed to ensure quality of data converted to OMOP.

Multiple records for one concept mapping
Picking one of the multiple standard vocabulary mapping to create the OMOP CDM record instead of one record per mapping

Abnormal values
Unconventional values in data asset (i.e. Negative or 0 as value_as_number)

Incorrect logic - Observation_Period
Observation_Period table populated incorrectly. Observation period does not cover the entire period of time where events are recorded for a person

Wrong type_concept_id
Use of the wrong type_concept_id or misunderstanding the definition of this field

Missing CDM tables
OMOP CDM tables missing due to misunderstanding on how to populate the table.

Incorrect logic - Visit_Occurrence
Visit_Occurrence table populated incorrectly

No Standard Vocabulary

Issue

- Text fields
- Duplicate and unclear values in source concept names
- Proprietary coding system
- No OMOP standard vocabulary mapping available even though vocabulary is in Athena

Solution

- Own Mapping Team
  Mapped translated terms to OMOP standard vocabulary
- OMOP Vocabulary Team
  - Prioritized terms for mapping
  - Verify translated terms
  - Confirm translation with medical team
  - Downloaded latest vocabularies

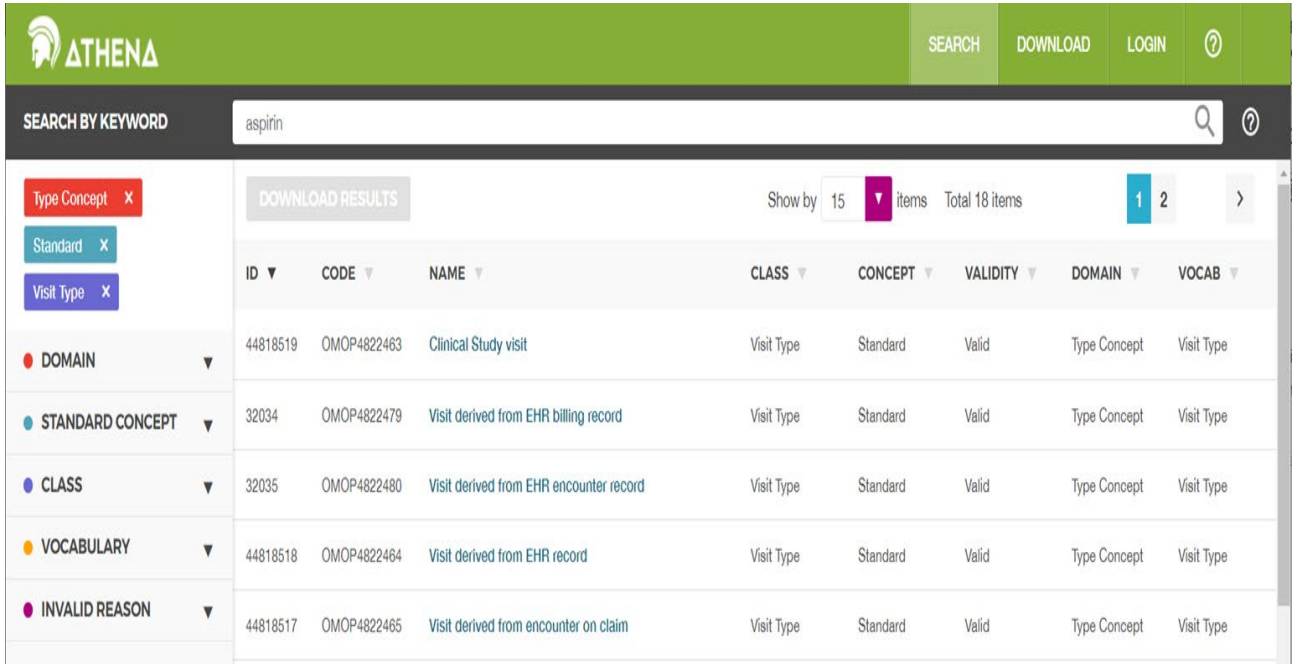
Wrong Type_Concept_IDs

Issue

- Wrong meanings were assigned to type_concept_ids

Solution

- Standardize all type_concept_ids in each table
- Find correct concept_id using ATHENA



Abnormal Values

Issue

- Negative, 0, null values or decimals in measurement and drug_exposure domain

	EVENT_TABLE	VALUE_AS_NUMBER_	COUNT(*)
1	measurement	> 0	173140
2	measurement	< 0	90
3	measurement	0	6236
4	measurement	null	6572

Solution

- Check source data for related domains and check if it's reasonable from medical perspective

Missing CDM Tables

Issue

- Incomplete OMOP CDM tables
- Potential Missing Tables:
  - Procedure_occurrence
  - Device_exposure
  - Visit_detail
  - Observation
  - Payer_plan_period
  - Dose_era
  - Location
  - Cost

Solution

- Check source data for related domains
- Provide mapping rules from source data to OMOP CDM, and populate the missing tables

Breakout Session 3
Exercises 45 minutes – Review 30 minutes

ETL Q&A Session


OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Thank You