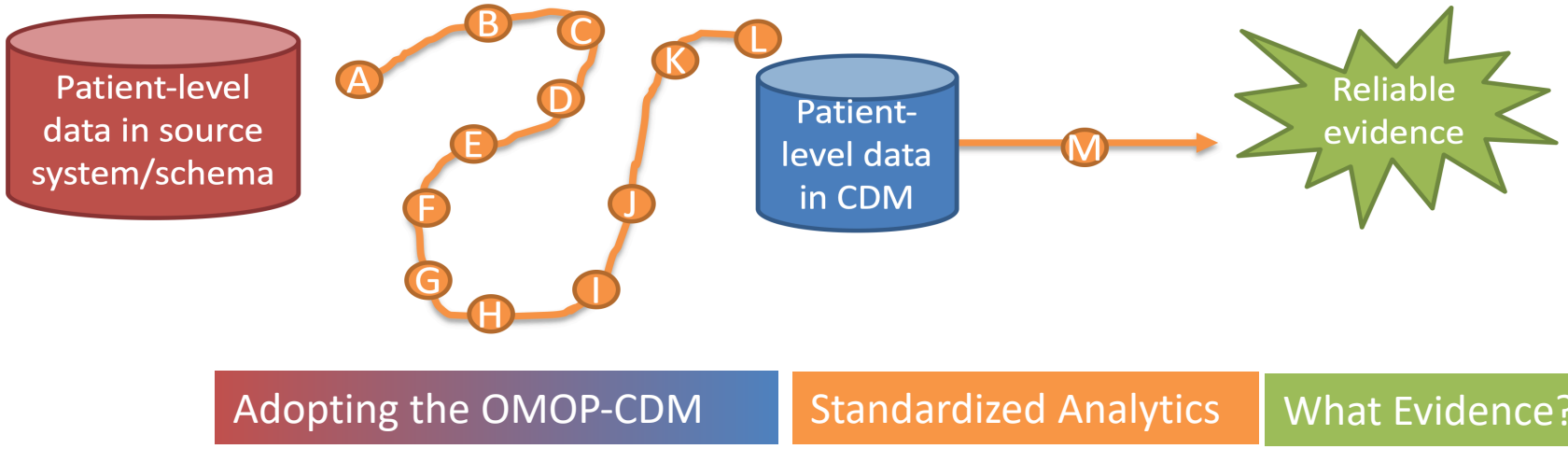


OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Adopting the OMOP CDM to enable standardized analytics

Christian Reich, Erica Voss, Mui van Zandt

Adopting the OMOP Common Data Model (CDM) to Enable Standardized Analytics

Christian Reich

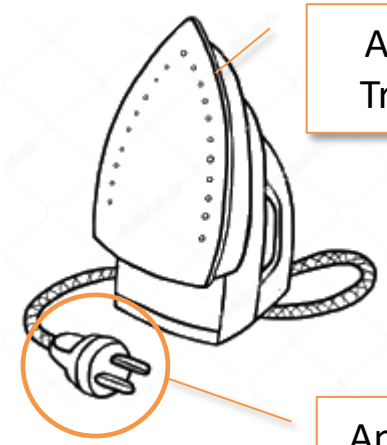
Erica A. Voss

Mui van Zandt

OMOP Common Data Model & OMOP Vocabularies

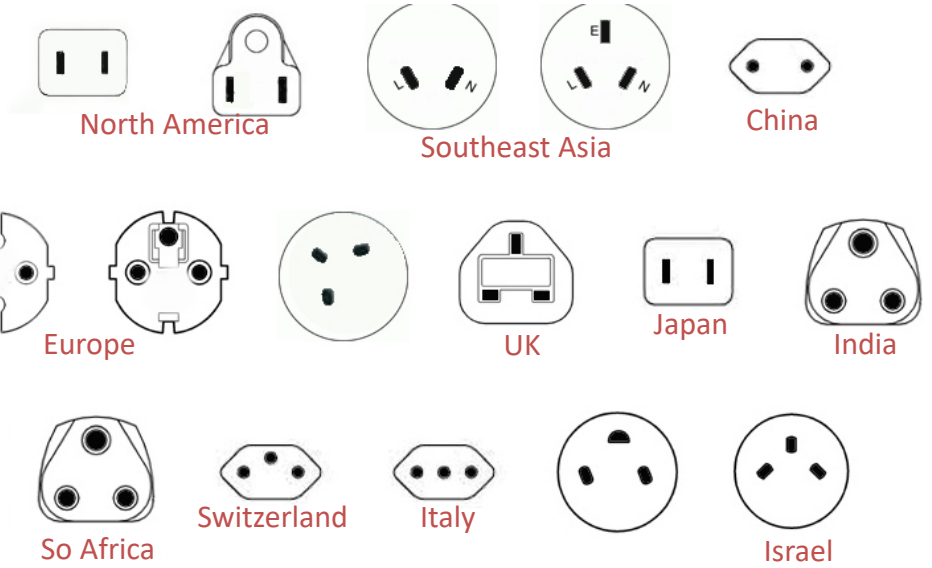
Traditional RWE Approach

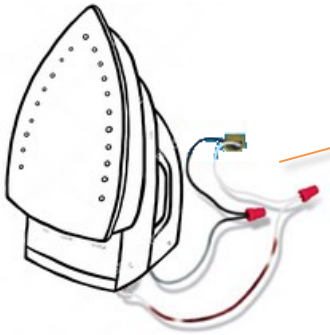
"What treatment pathway is used in treatment of blood pressure in different healthcare settings?"



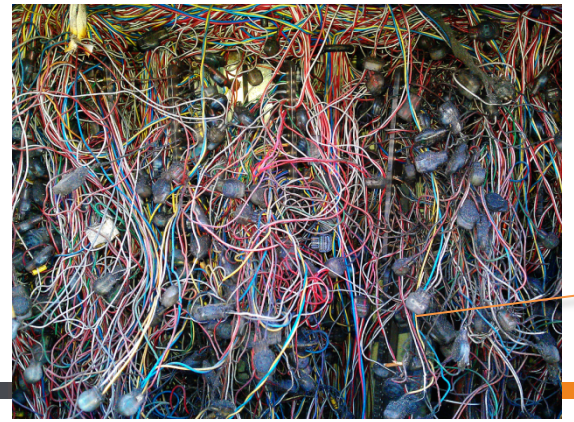
Analytical method: Treatment Pathway

Application to data



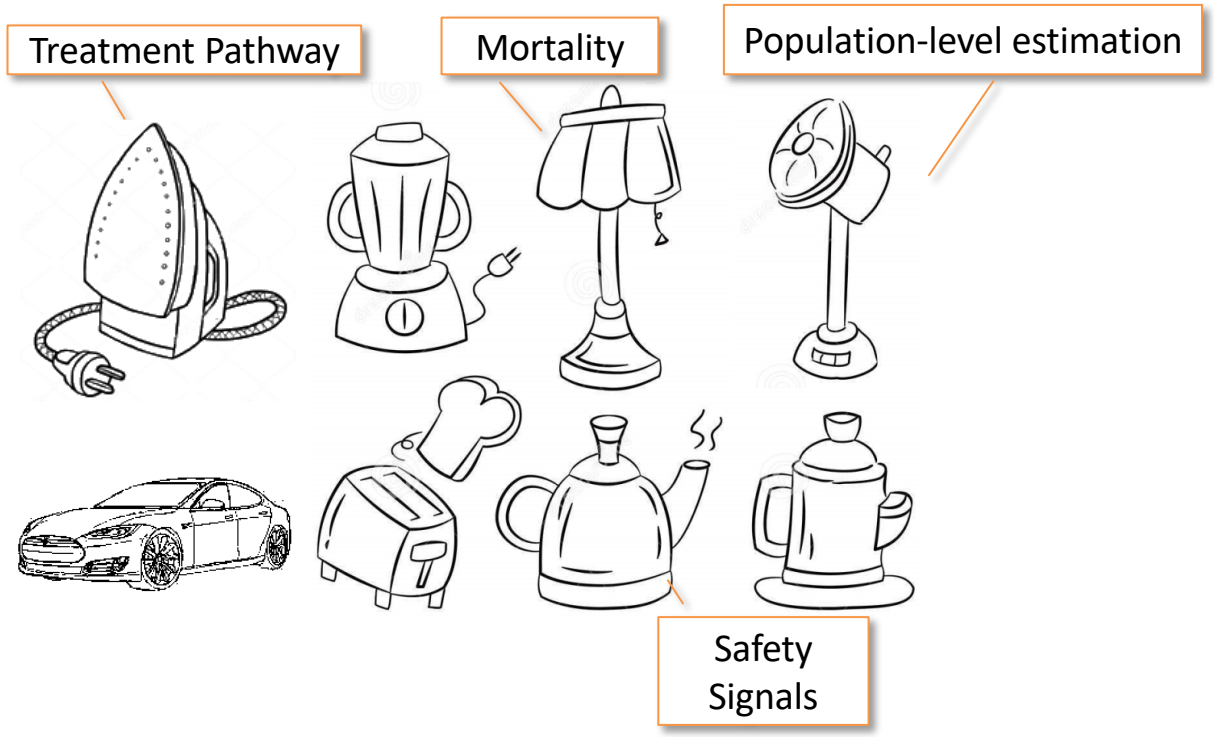


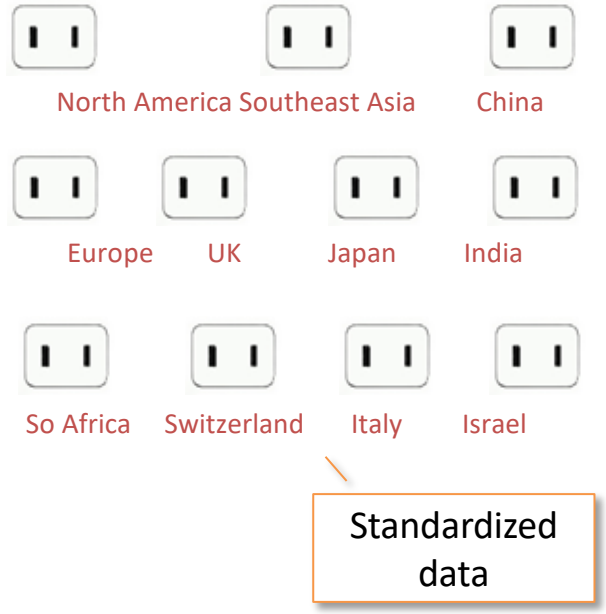
One study – one script model (often in SAS!)



- Reliant on partner capabilities
- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert users

Solution: Systematic and Standardized Research

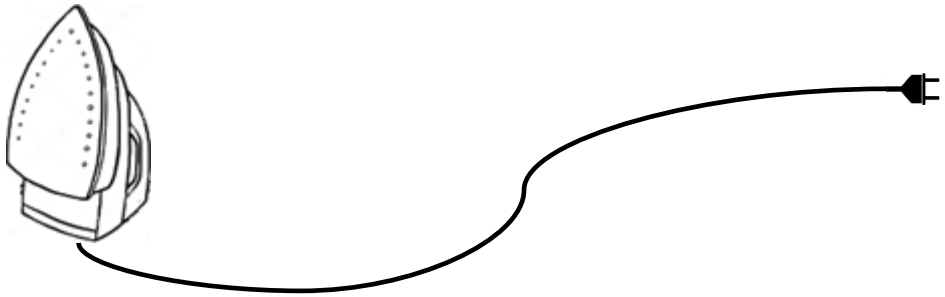


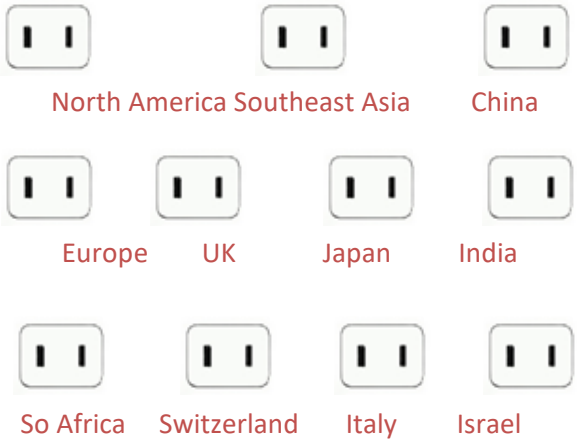


OHDSI Tools

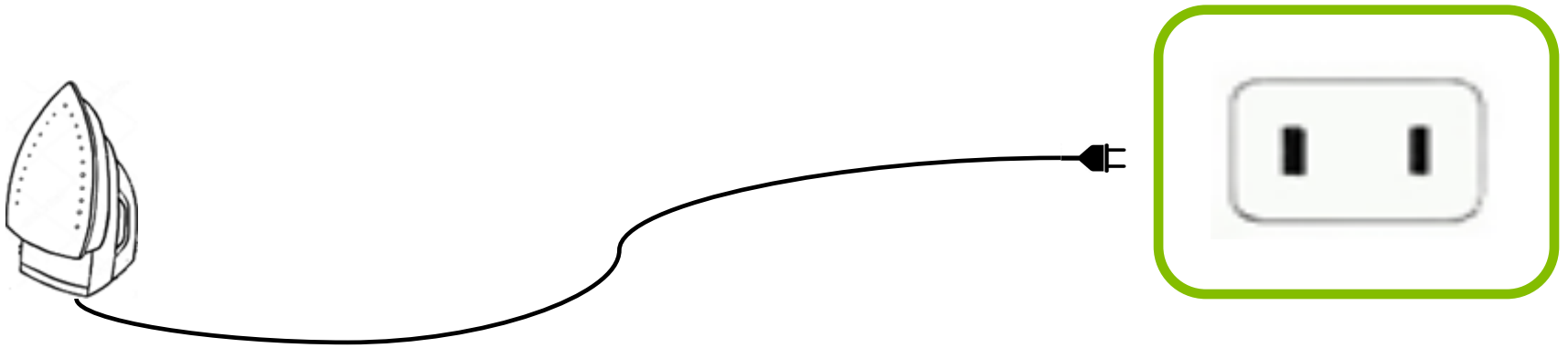
OMOP CDM

Analytics can be remote

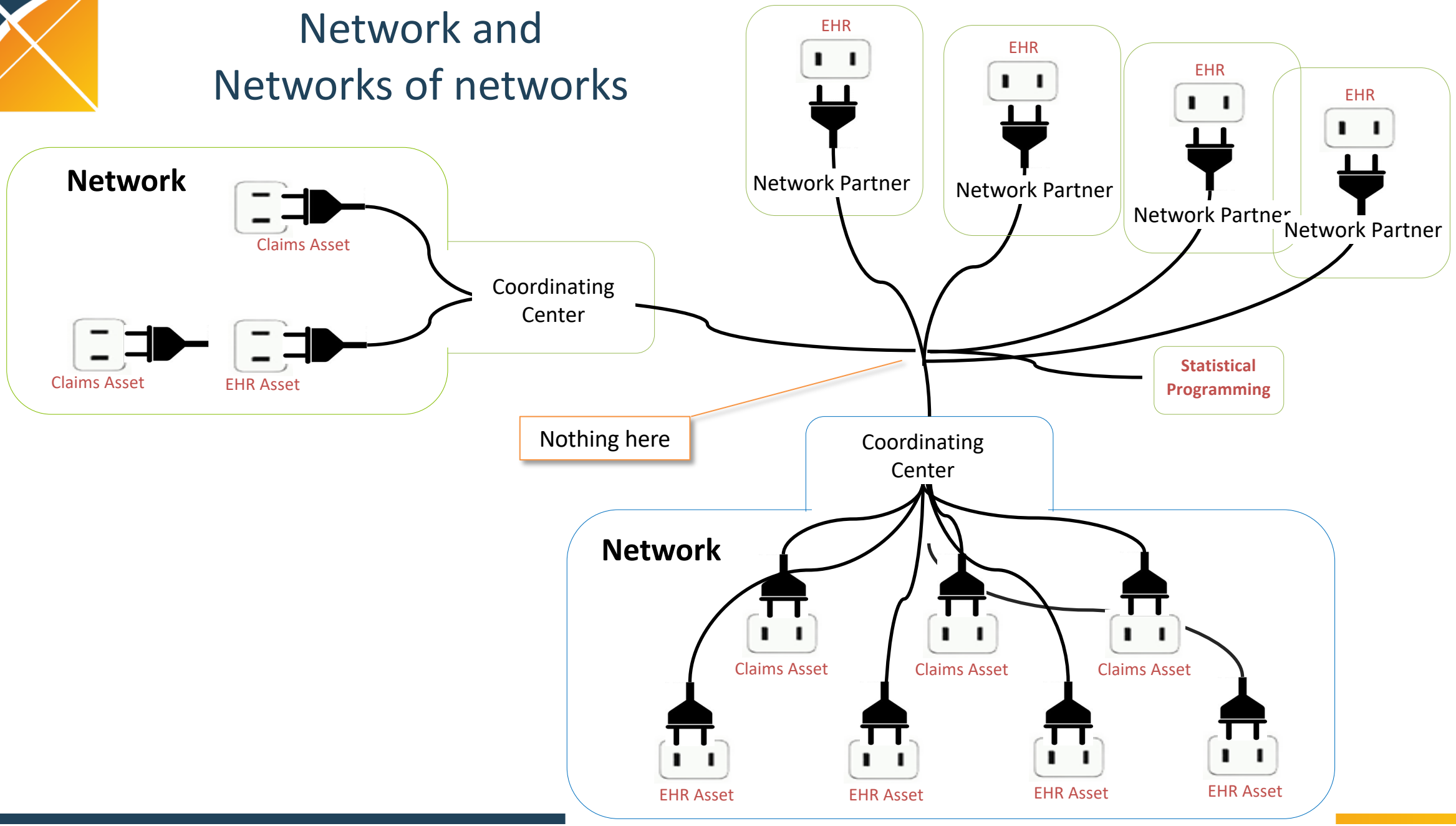




Analytics can run behind firewall



Network and Networks of networks



How do you create data that you query without having access?

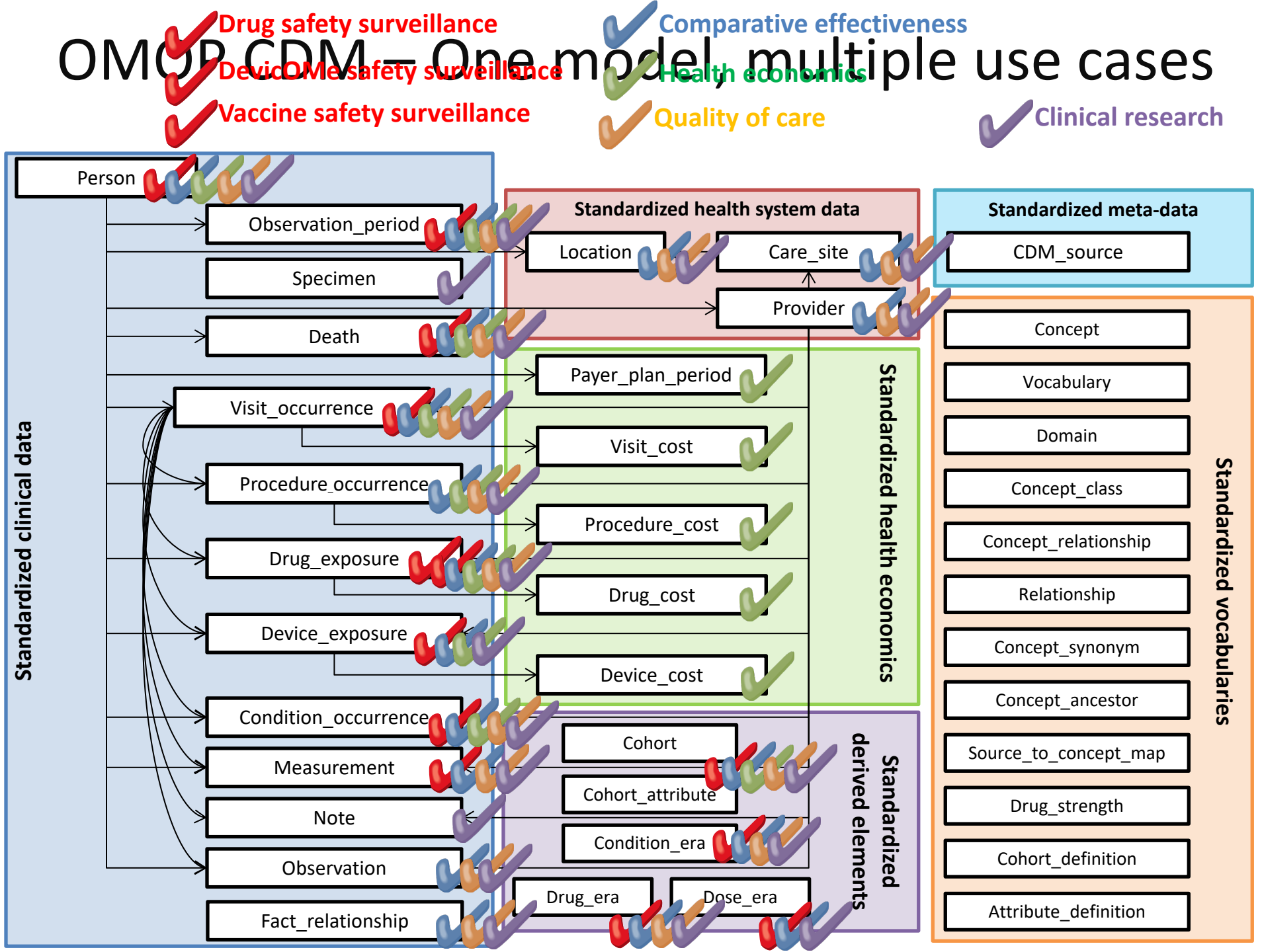
The OHDSI approach: Strict Standardization

- **Data**
  - **Structure**: tables, fields, data types
  - **Content**: vocabulary to codify clinical domains
  - **Semantics**: conventions about meaning

OMOP CDM
OMOP Standardized Vocabularies

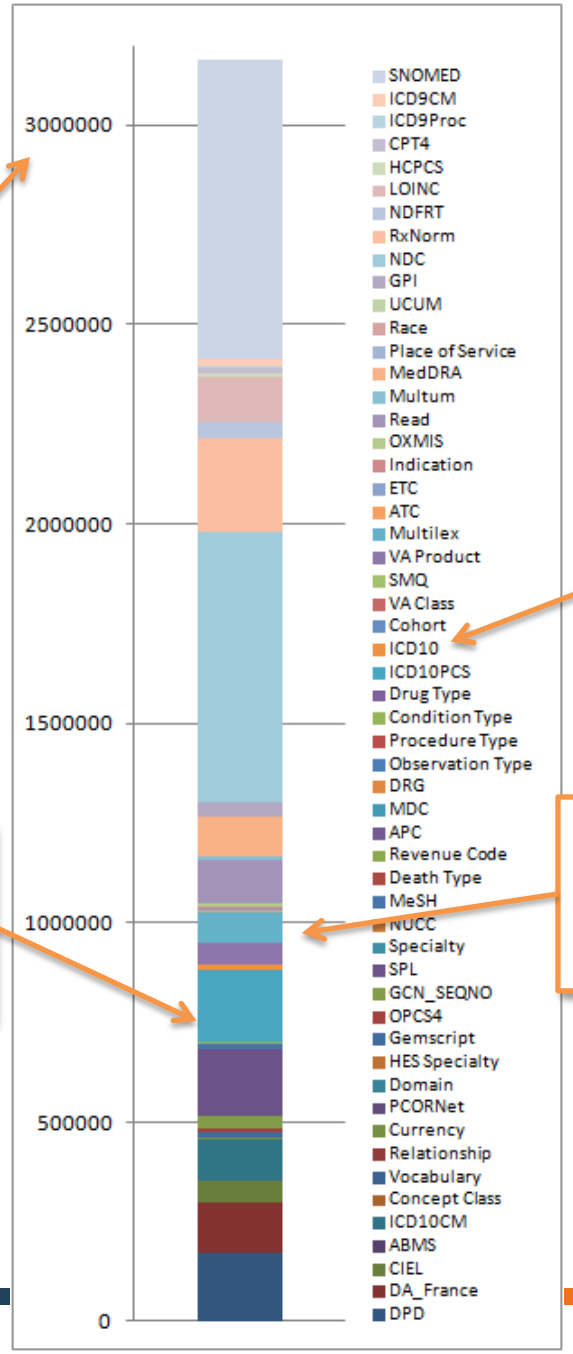
- **Research**
  - **Cohort definition** : algorithms for identifying the set of patients who meet a collection of criteria for a given interval of time
  - **Covariate construction** : logic to define variables available for use in statistical analysis
  - **Analysis** : collection of decisions and procedures required to produce aggregate summary statistics from patient-level data
  - **Results reporting** : series of aggregate summary statistics presented in tabular and graphical form

OMOP CDM – One model, multiple use cases



Single Standardized Reference Table

4,809,472 total



Vocabulary ID

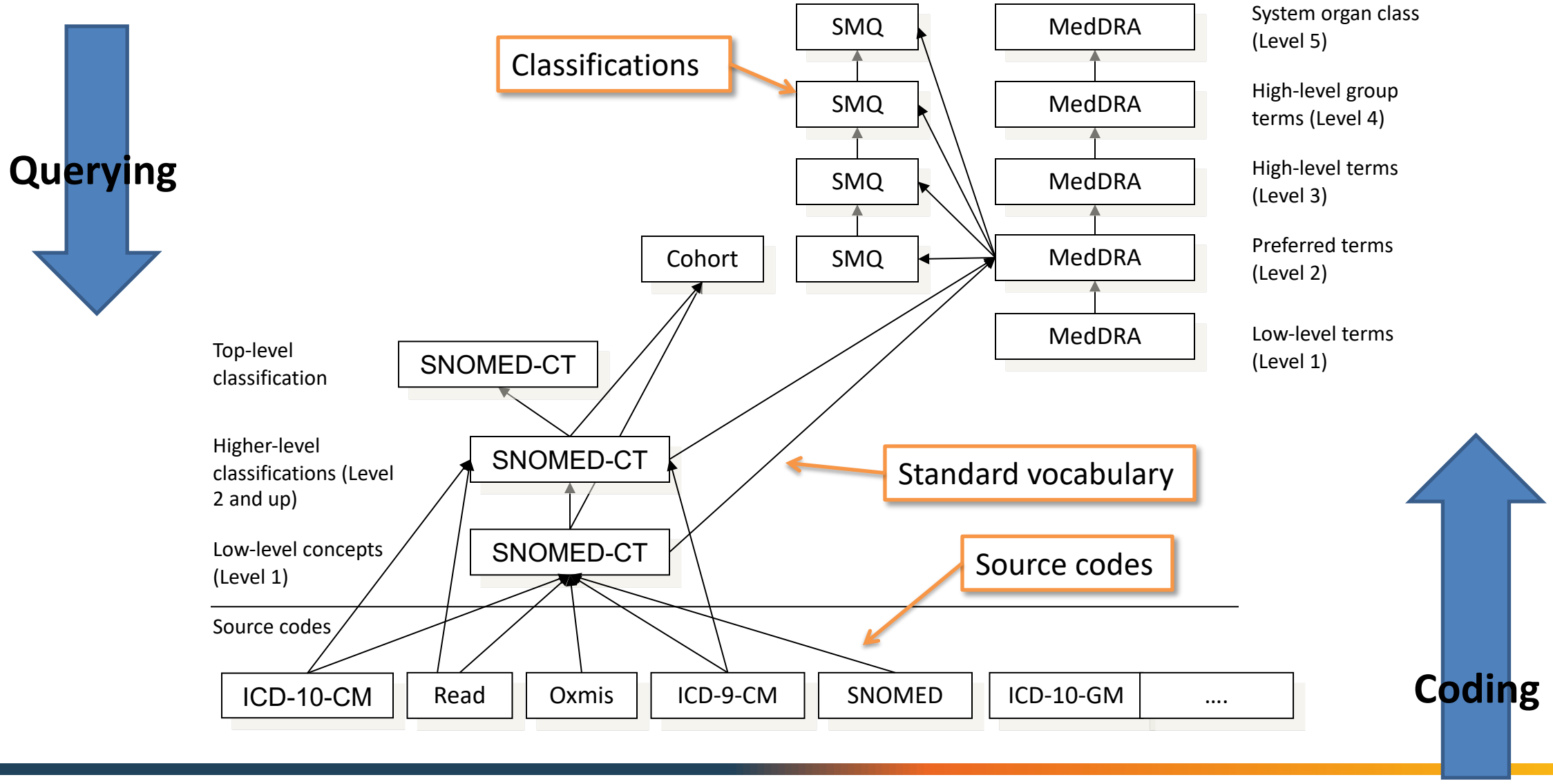
All vocabularies stacked up in one table

Concepts related through semantic links, forming hierarchy

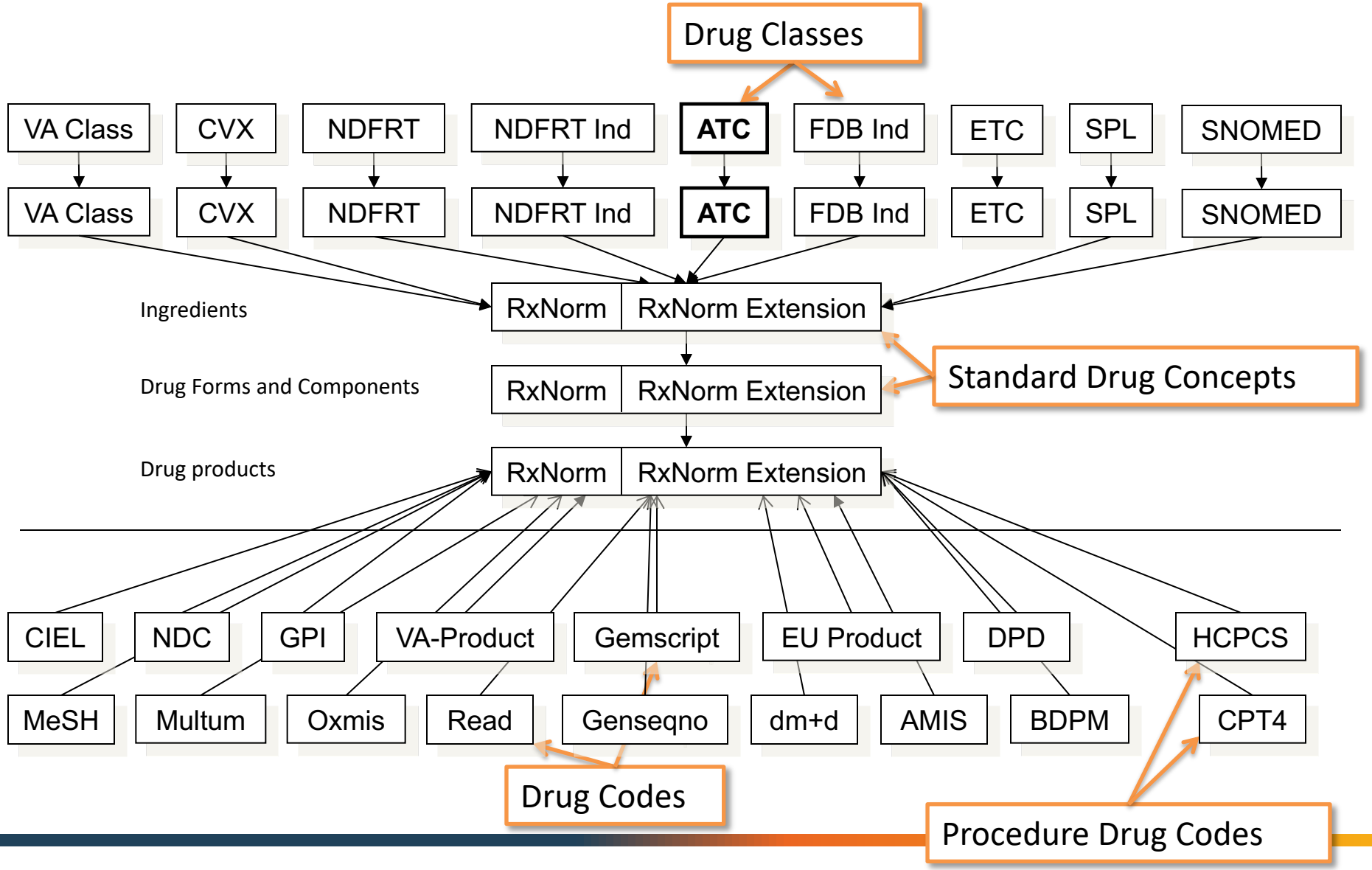
Vocabularies: Goals

- ✓ **Domains:** Every Standard Concept belongs to the right Domain
- **No duplicates:** For every entity exists one Standard Concept
- **Comprehensive:** For every Domain exists a complete finite set of Concepts covering all possible entities in this domain
- **Hierarchy:** All Concepts are connected through a comprehensive hierarchy
- **Mapping:** For every existing code in a vocabulary there is a map to a Standard Concept or a map to 0

Standardized Vocabularies: Conditions

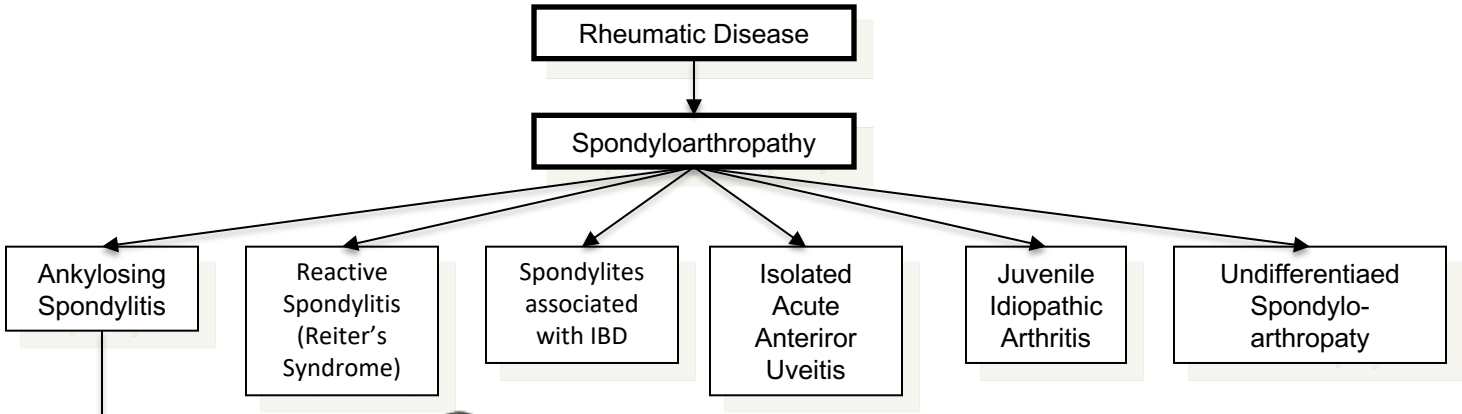


Standardized Vocabularies: Drugs



Why a Hiearchy?

Diseases are not not “parallel“



Why a Drug Hierarchy?

Availabe data might have ingredients, brand names, doses...

5 ML Deferoxamine 100 MG/ML Injectable Solution [Desferal] Box of 4 by Novartis Pharma

Classes / Attributes	Ingredient	Drug Strength	Dose Form	Brand Name	Qty Factor	Box size	Supplier
Ingredient	Deferoxamine						
Clinical Drug Component	Deferoxamine	100 mg/ml					
Clinical Drug	Deferoxamine	100 mg/ml	Injectable Solution				
Branded Drug	Deferoxamine	100 mg/ml	Injectable Solution	Desferal			
Quantified Branded Drug	Deferoxamine	100 mg/ml	Injectable Solution	Desferal	5 ml		
Quantified Branded Box	Deferoxamine	100 mg/ml	Injectable Solution	Desferal	5 ml	Box of 4	
Marketed Product	Deferoxamine	100 mg/ml	Injectable Solution	Desferal	5 ml	Box of 4	Novartis Pharma

RxNorm

RxNorm Extension

Granularity through Hierarchies

- Find (count, create cohort of) patients who ...
  - Had acute ST segment elevation myocardial infarction involving left anterior descending coronary artery
  - Had myocardial infarction
  - Had cardiovascular disease
  - Have family history of malignant disease
  - Took diphtheria and tetanus toxoids, adsorbed for pediatric use
  - Took diphtheria toxoid vaccine, inactivated
  - Took toxoid vaccine
  - Took vaccine
  - Took drug by FDA-approced indication
  - Took off-label drug
  - Underwent hysterectomy
  - Underwent abdominal surgery

  - Had HbA1C > 7 consistently in last 6 months
  - Smoked longer than 10 packyears
  - Received IUD
  - Has follow-up of 2 years after event
  - Had treatment in excess of $100k per year

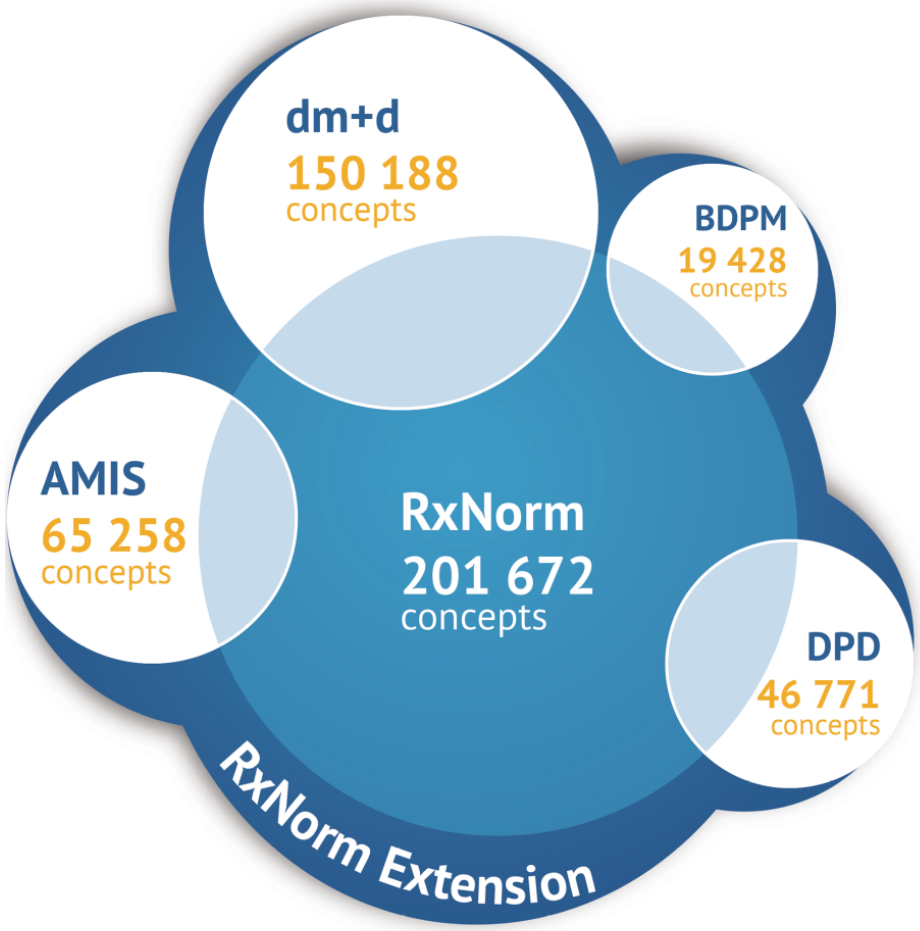
Different European Drug Markets – Different Formulations

Nicotine Chewing Gum [Nicorette]	RxNorm	France, Germany, UK
Nicotine Oral Lozenge [Nicorette]	RxNorm	France, Germany, UK
Nicotine Cartridge [Nicorette]		Germany
Nicotine Dry Powder Inhaler [Nicorette]		France
Nicotine Inhalant Solution [Nicorette]		UK
Nicotine Metered Dose Inhaler [Nicorette]		Germany
Nicotine Nasal Spray [Nicorette]		Germany, UK
Nicotine Oral Solution [Nicorette]		Germany
Nicotine Oral Spray [Nicorette]		France, Germany
Nicotine Sublingual Tablet [Nicorette]		France, Germany, UK
Nicotine Topical Solution [Nicorette]		Germany
Nicotine Transdermal System [Nicorette]		France, Germany, UK

Solution: RxNorm Extension

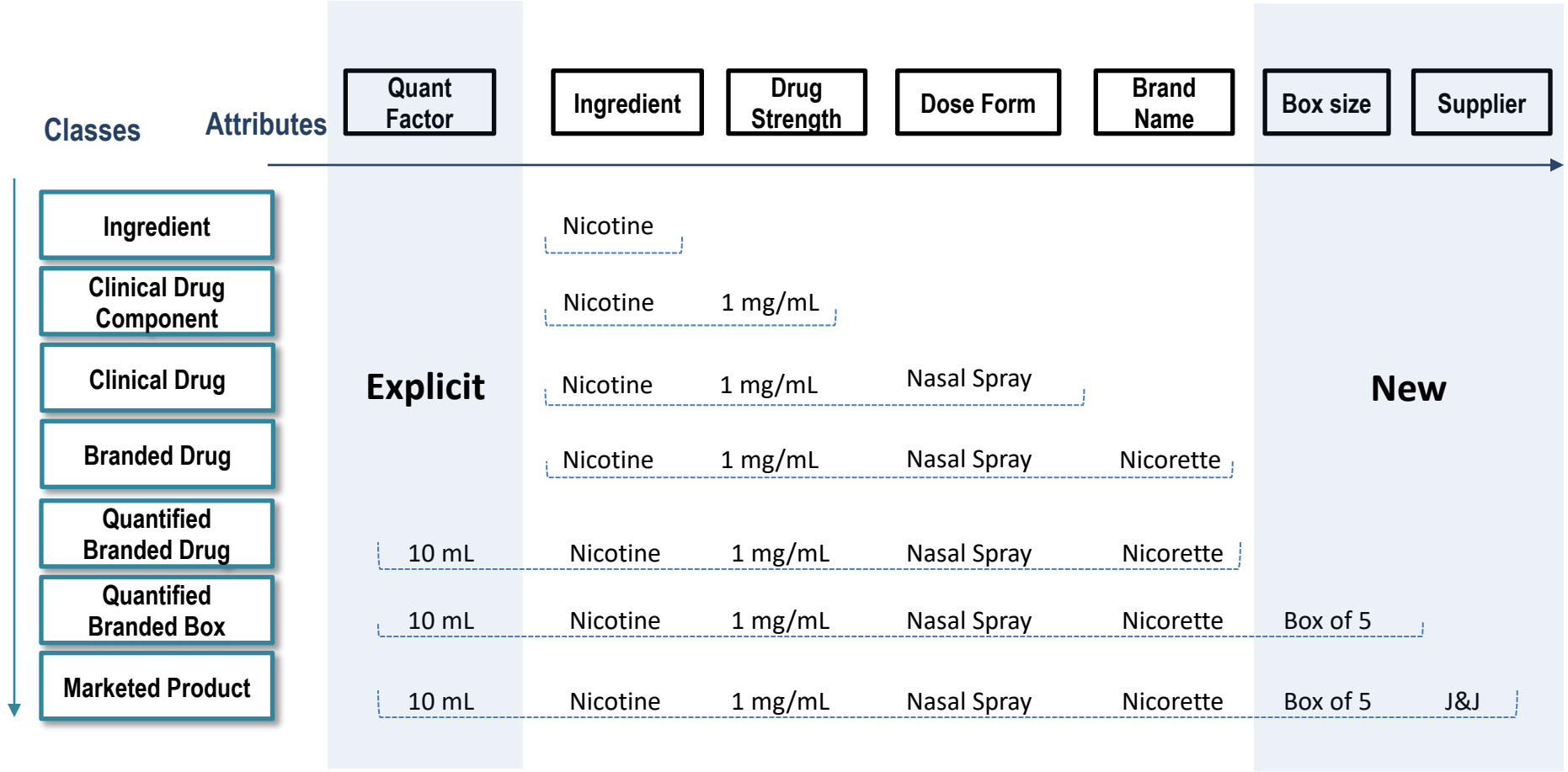
- Make RxNorm for the World
- Build up with relationships to existing RxNorm drugs and attributes
- Same logic, structure and attributes

Drug Markets overlap only partially



RxNorm Extension Hierarchy

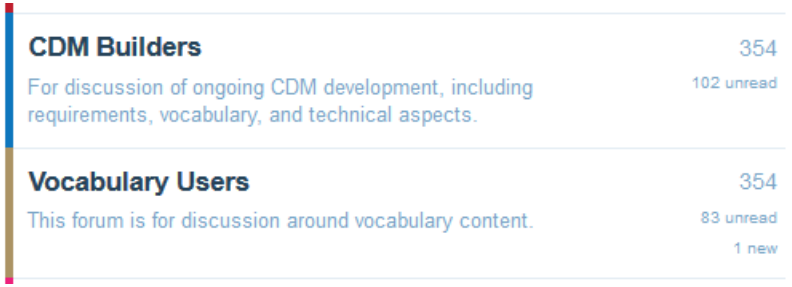
"10 ML Nicotine 1 MG/ML Nasal Spray [Nicorette] Box of 5 by Johnson & Johnson"

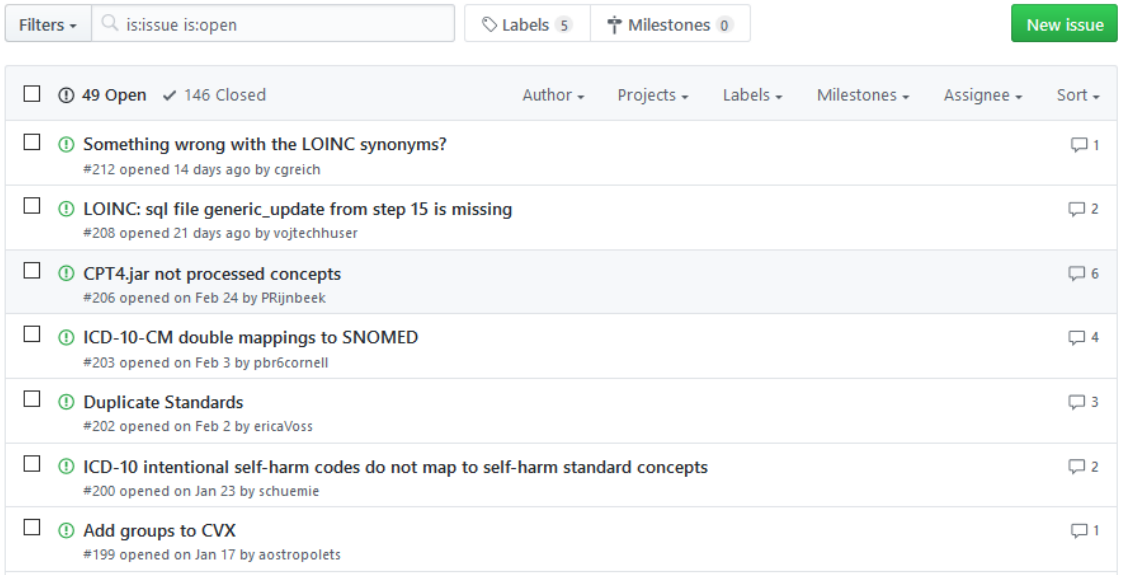


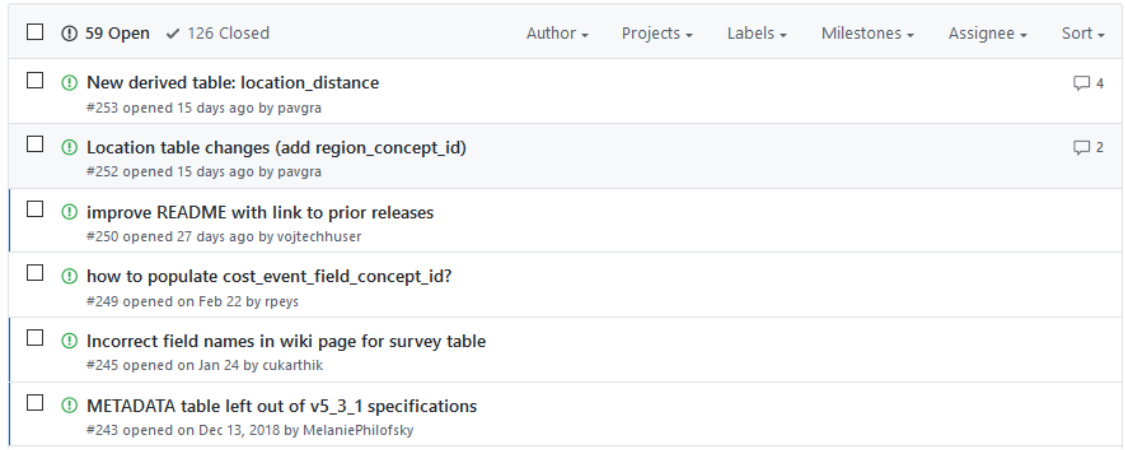
Resources to get help

https://github.com/ohdsi

http://forums.ohdsi.org/



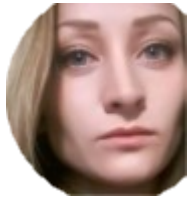




Teams to help you

Vocabulary Team


Alina


Polina


Anna


Oleg


Eduard


Vlad


Alex


Denis


Polina


Eldar


Dima

CDM WG


Clair

Extract, Transform, & Load (ETL) & OHDSI ETL Tools

ETL Process

Data experts and CDM experts together design the ETL

People with medical knowledge create the code mappings

ETL Documentation

A technical person implements the ETL

All are involved in quality control

OHDSI Tools

White Rabbit

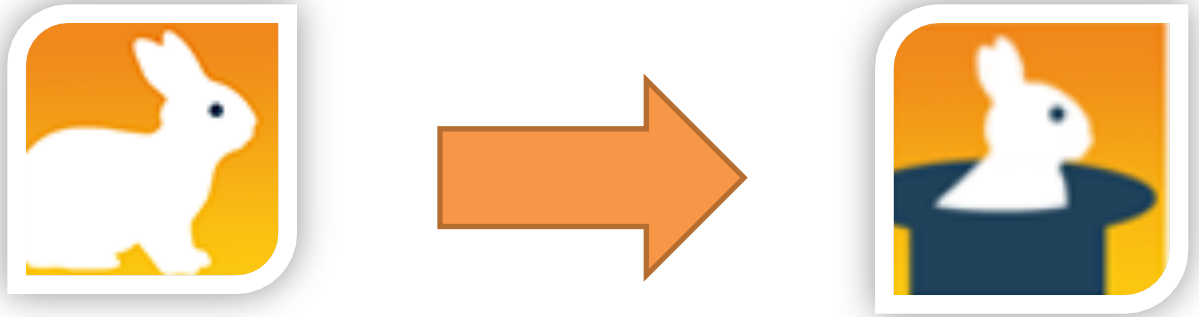
Rabbit In a Hat

Usagi

White Rabbit

ACHILLES

Designing an ETL



White Rabbit

- performs a scan of the source data, providing detailed information on the tables, fields, and values that appear in a field

Rabbit In a Hat

- Uses White Rabbit scan to provide a graphical user interface to help build an ETL document
- Does not generate code

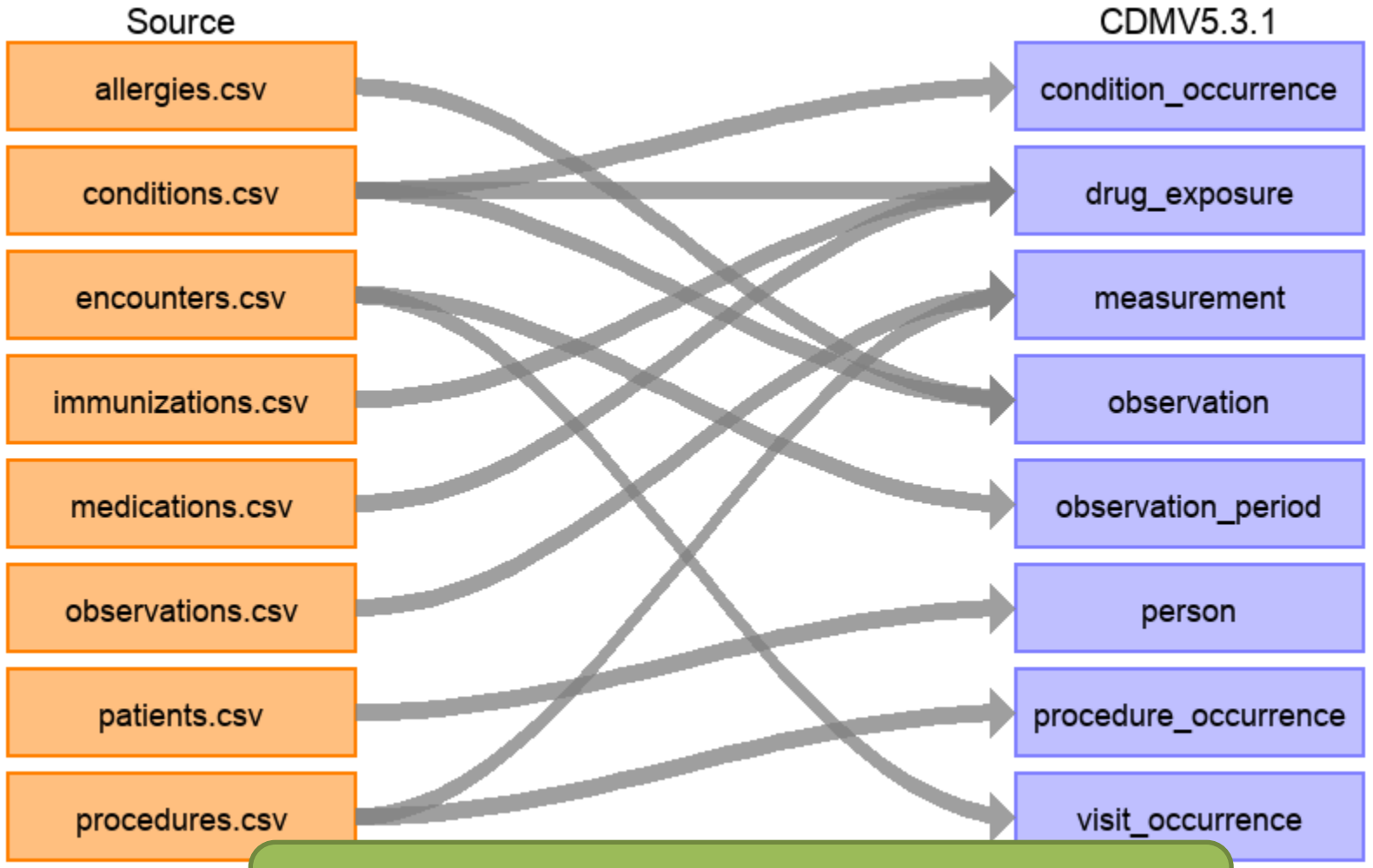
White Rabbit

W	X	Y	Z	AA	AB	AC	AD	AE	AF
marital	Frequency	race	Frequency	ethnicity	Frequency	gender	Frequency	birthplace	Frequenc
M	610	white	835	irish	224	M	562	Boston	130
	344	hispanic	112	italian	145	F	557	Springfiel	30
S	166	black	82	english	102	U	1	Worceste	28
		asian	70	puerto_ri	72			Lowell	22
		native	20	french	72			Brockton	21
		List trunca		german	64			Cambridg	18
				chinese	51			Methuen	18
				polish	49			Newton	17
				american	39			Quincy	16
				portugues	37			Framingh	16
				french_ca	35			Lynn	12
				african	33			Arlington	12
				west_indi	28			Weymout	12
				dominicar	21			New Bedf	12
				russian	20			Lawrence	11
				american_	20			Haverhill	11
				asian_indi	19			Fitchburg	11
				scottish	19			Marshfiel	10
				mexican	18			Somervill	10
				swedish	17			Barnstabl	10
									9
									9

izations | m

Scan your Data

Rabbit in a Hat



Map Source Tables to CDM Tables

Rabbit in a Hat

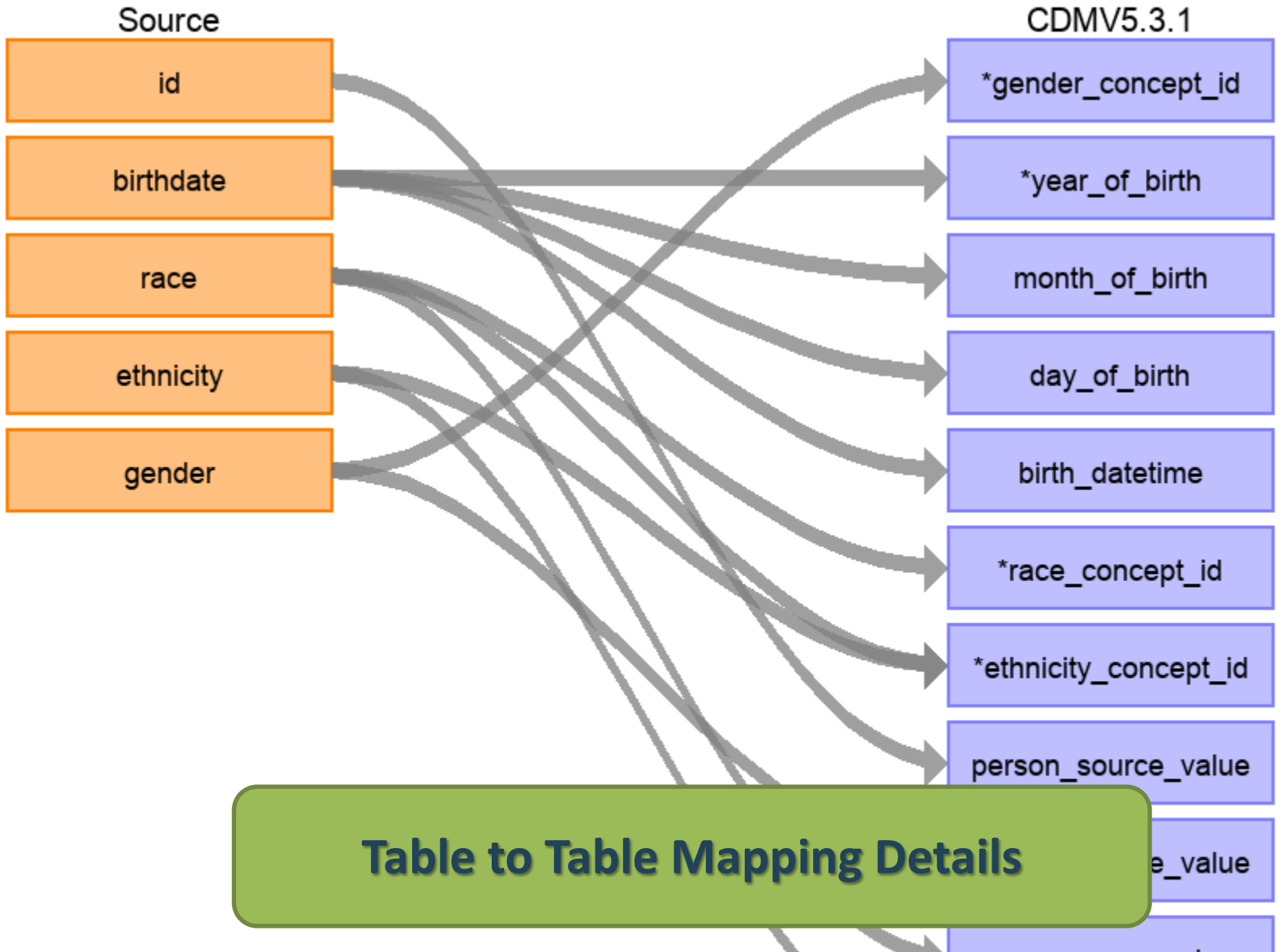
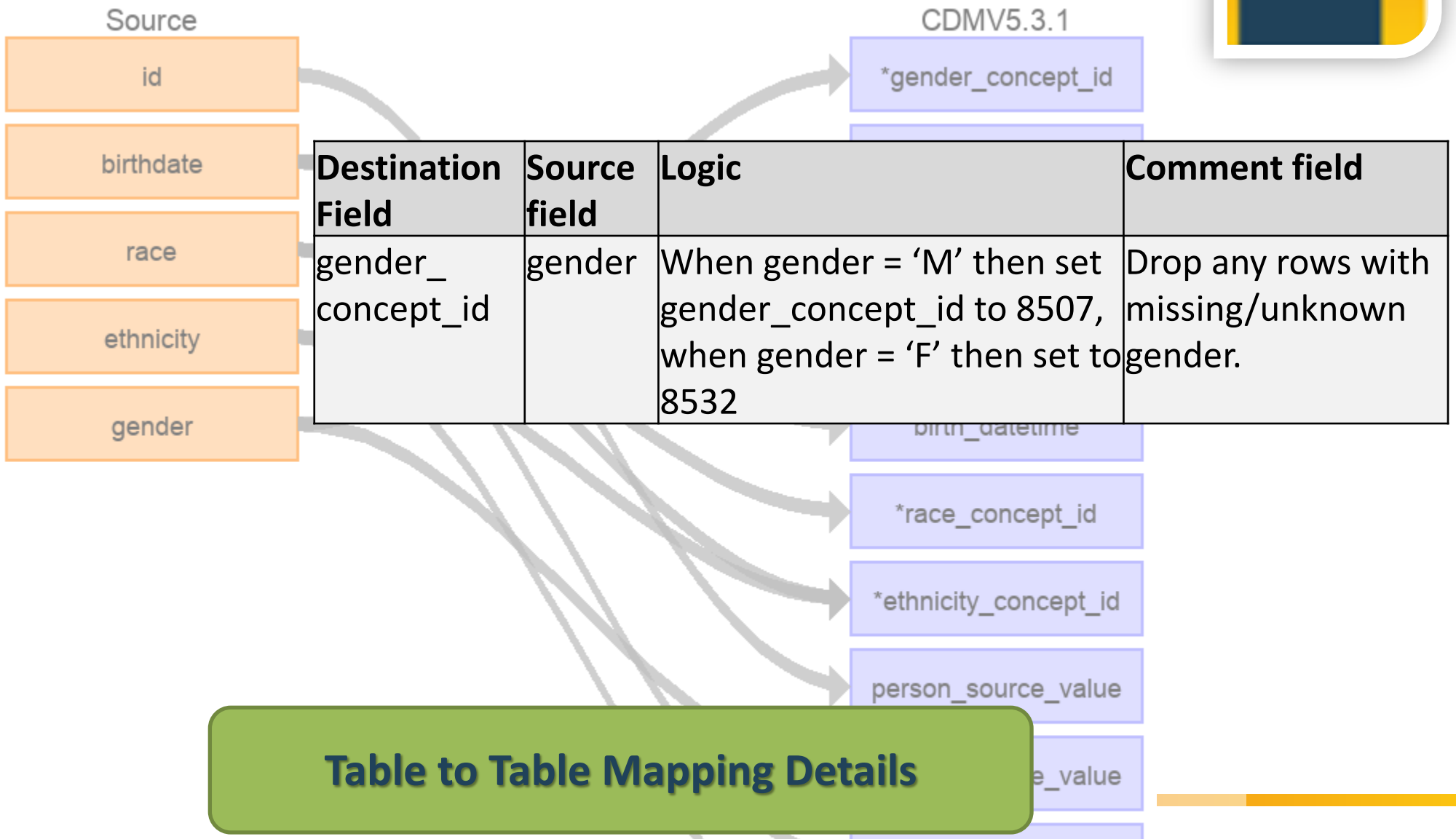


Table to Table Mapping Details

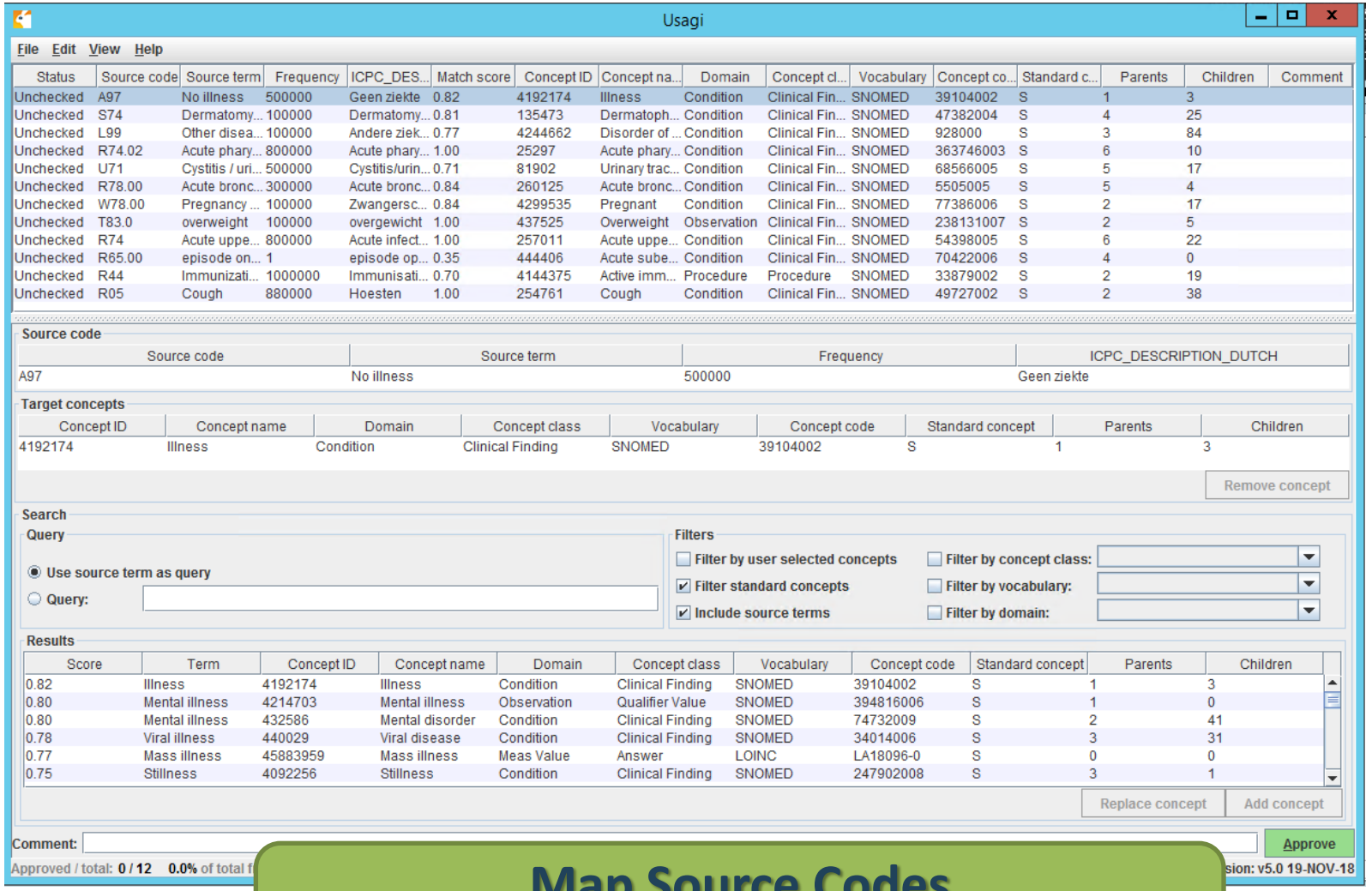
Rabbit in a Hat



Destination Field	Source field	Logic	Comment field
gender_ concept_id	gender	When gender = ‘M’ then set gender_concept_id to 8507, when gender = ‘F’ then set to 8532	Drop any rows with missing/unknown gender.

Table to Table Mapping Details

Usagi



Status	Source code	Source term	Frequency	ICPC_DES...	Match score	Concept ID	Concept na...	Domain	Concept cl...	Vocabulary	Concept co...	Standard c...	Parents	Children	Comment
Unchecked	A97	No illness	500000	Geen ziekte	0.82	4192174	Illness	Condition	Clinical Fin...	SNOMED	39104002	S	1	3	
Unchecked	S74	Dermatomy...	100000	Dermatomy...	0.81	135473	Dermatoph...	Condition	Clinical Fin...	SNOMED	47382004	S	4	25	
Unchecked	L99	Other disea...	100000	Andere ziek...	0.77	4244662	Disorder of ...	Condition	Clinical Fin...	SNOMED	928000	S	3	84	
Unchecked	R74.02	Acute phary...	800000	Acute phary...	1.00	25297	Acute phary...	Condition	Clinical Fin...	SNOMED	363746003	S	6	10	
Unchecked	U71	Cystitis / uri...	500000	Cystitis/urin...	0.71	81902	Urinary trac...	Condition	Clinical Fin...	SNOMED	68566005	S	5	17	
Unchecked	R78.00	Acute bronc...	300000	Acute bronc...	0.84	260125	Acute bronc...	Condition	Clinical Fin...	SNOMED	5505005	S	5	4	
Unchecked	W78.00	Pregnancy ...	100000	Zwangersc...	0.84	4299535	Pregnant	Condition	Clinical Fin...	SNOMED	77386006	S	2	17	
Unchecked	T83.0	overweight	100000	overgewicht	1.00	437525	Overweight	Observation	Clinical Fin...	SNOMED	238131007	S	2	5	
Unchecked	R74	Acute uppe...	800000	Acute infect...	1.00	257011	Acute uppe...	Condition	Clinical Fin...	SNOMED	54398005	S	6	22	
Unchecked	R65.00	episode on...	1	episode op...	0.35	444406	Acute sube...	Condition	Clinical Fin...	SNOMED	70422006	S	4	0	
Unchecked	R44	Immunizati...	1000000	Immunisati...	0.70	4144375	Active imm...	Procedure	Procedure	SNOMED	33879002	S	2	19	
Unchecked	R05	Cough	880000	Hoesten	1.00	254761	Cough	Condition	Clinical Fin...	SNOMED	49727002	S	2	38	

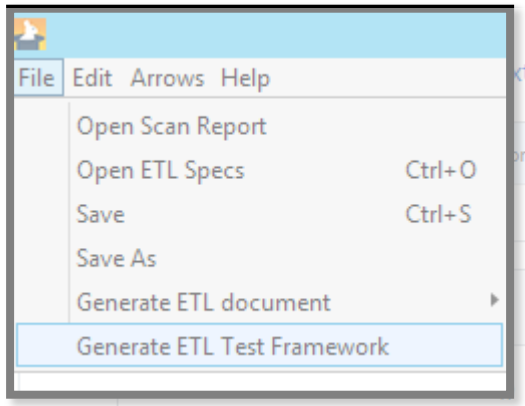
Source code	Source term	Frequency	ICPC_DESCRIPTION_DUTCH
A97	No illness	500000	Geen ziekte

Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard concept	Parents	Children
4192174	Illness	Condition	Clinical Finding	SNOMED	39104002	S	1	3

Score	Term	Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard concept	Parents	Children
0.82	Illness	4192174	Illness	Condition	Clinical Finding	SNOMED	39104002	S	1	3
0.80	Mental illness	4214703	Mental illness	Observation	Qualifier Value	SNOMED	394816006	S	1	0
0.80	Mental illness	432586	Mental disorder	Condition	Clinical Finding	SNOMED	74732009	S	2	41
0.78	Viral illness	440029	Viral disease	Condition	Clinical Finding	SNOMED	34014006	S	3	31
0.77	Mass illness	45883959	Mass illness	Meas Value	Answer	LOINC	LA18096-0	S	0	0
0.75	Stillness	4092256	Stillness	Condition	Clinical Finding	SNOMED	247902008	S	3	1

Map Source Codes to Standard Concepts

Rabbit in a Hat



xtras / SyntheaTestFramework.r

Find file | Copy path

orrect location c380603 Mar 15, 2019

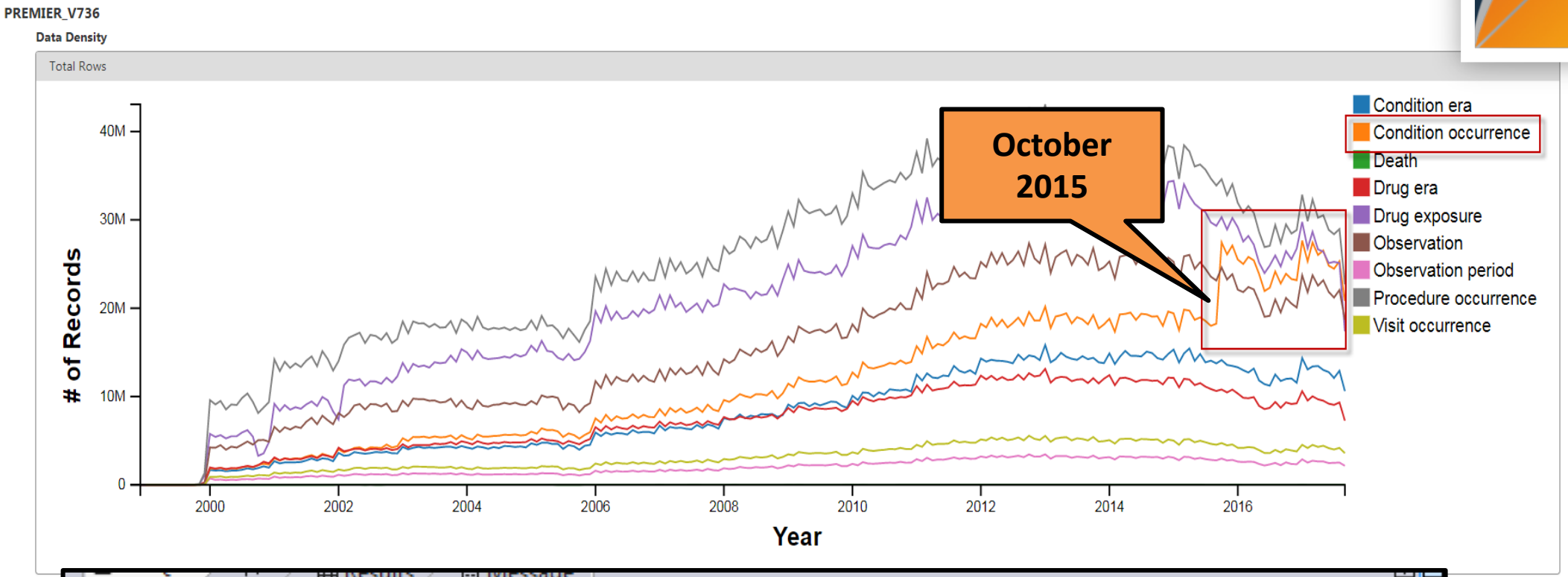
Raw | Blame | History

```
{
  frameworkContext <- new.env(parent = globalenv())
  assign('frameworkContext', frameworkContext, envir = globalenv())
  frameworkContext$inserts <- list()
  frameworkContext$expects <- list()
  frameworkContext$testId <- -1
  frameworkContext$testDescription <- ""
  frameworkContext

  defaults <- list
  defaults$start <
  defaults$patient <-
  defaults$encounter <- 'c430e49d-9612-4ba0-92ec-70ab4afdcb5f'
  defaults$code <- '419474003'
  defaults$description <- 'Allergy to mould'
  assign('allergies', defaults, envir = frameworkContext$defaultValues)

  defaults <- list()
  defaults$id <- '90
  defaults$start <-
```

```
patient <- createPatient()
declareTest(id = patient$id, description = "Patient with strange birth date, id is PERSON_SOURCE_VALUE")
add_patients(id = patient$id, birthdate = "2099-01-01")
```

Helps Build Test Cases

ACHILLES



	SOURCE_CODE	SOURCE_CONCEPT_ID	SOURCE_CODE_DESCRIPTION	SOURCE_VOCABULARY_ID	SOURCE_DOMAIN_ID
1	A41.9	35205553	Sepsis, unspecified organism	ICD10CM	Condition
2	A41.9	45585911	Sepsis, unspecified	ICD10	Condition

Review your Data with High Level Reporting

Tools Demo

- Visit Clair during the Collaborator Showcase

Rabbit in a Hat

Usagi


OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

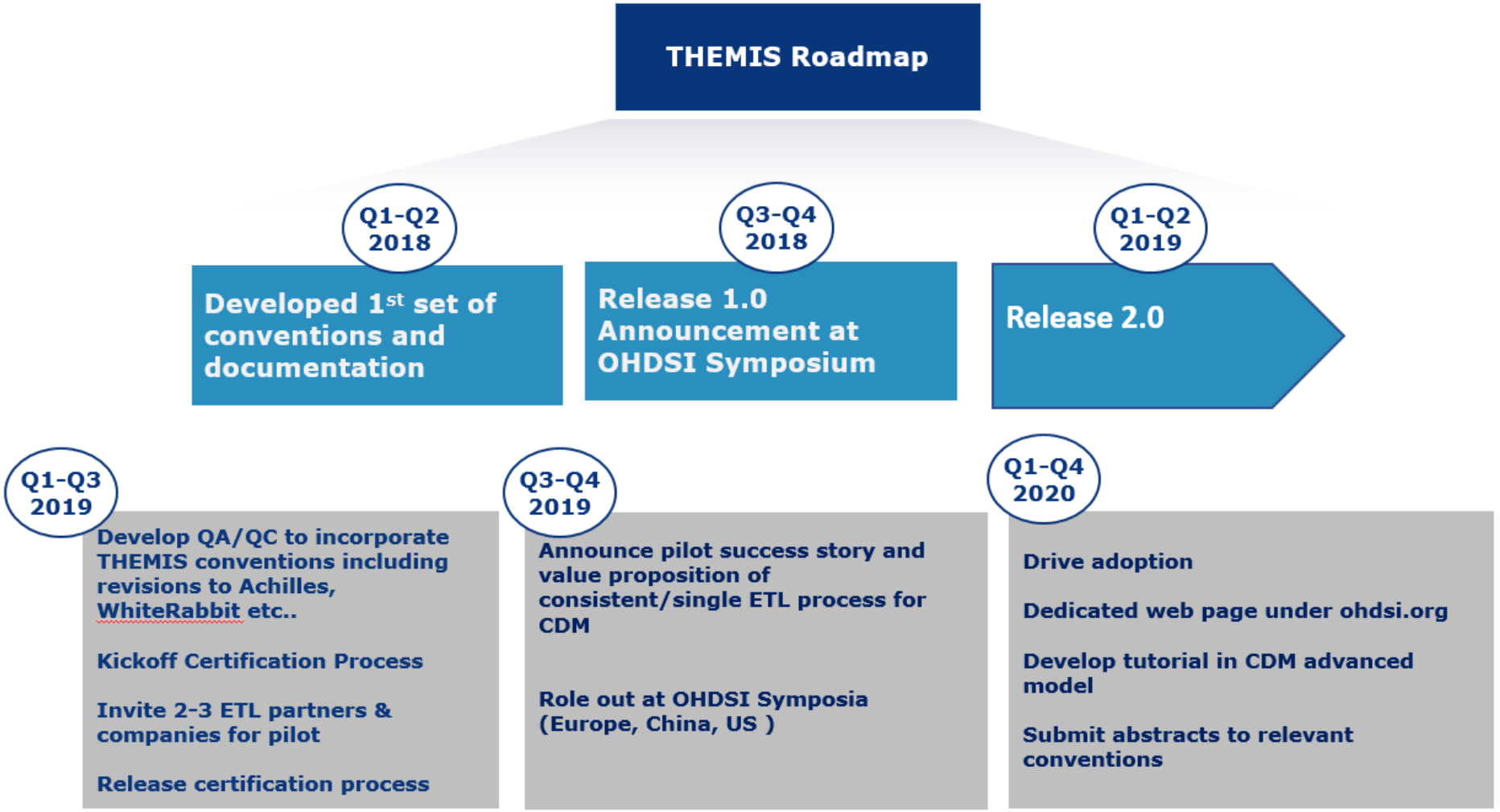
THEMIS


THEMIS

Purpose of THEMIS

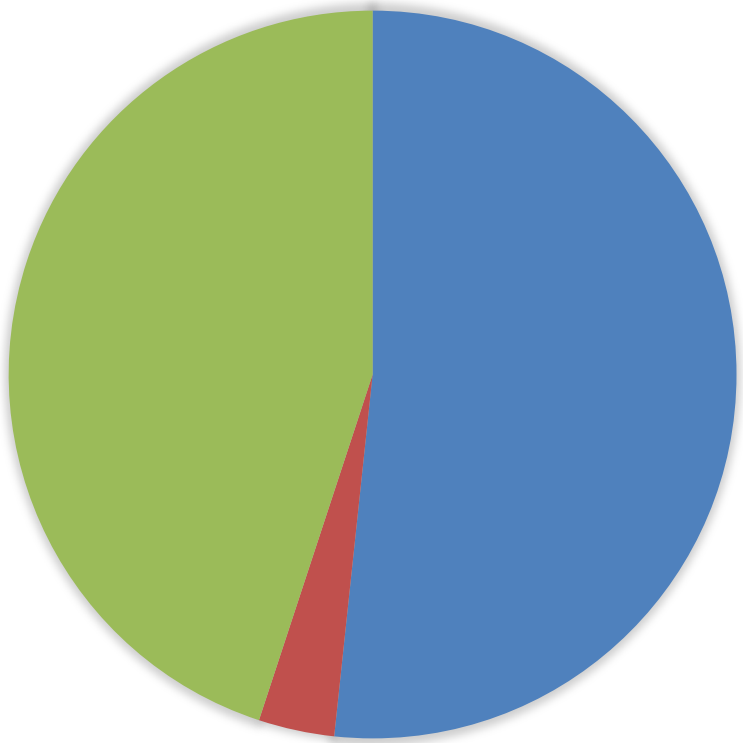
Developing rules and regulations to combat the inconsistent representations of the same data sources across Observational Health Data Sciences and Informatics' (OHDSI) and inability to gather reliable and scalable evidence.

THEMIS Roadmap



March 2018 Hackathon

THEMIS ISSUES



Duplicate Drugs on Same Day

What	What should someone do if duplicate drugs are reported on the same day? Most likely an idea that can be expanded to procedures.
Notes	http://forums.ohdsi.org/t/duplicate-drugs-themis-wg3/4101 **RECOMMENDATION:** `If a patient has multiple records on the same day for the same drug or procedures the ETL should not dedupe them unless there is probable reason to believe the item is a true data duplicate.`
Action	Work with Clair to have posted on the DRUG_EXPOSURE & PROCEDURE_OCCURRENCE page under the CDM Wiki. https://github.com/OHDSI/CommonDataModel/wiki/

Missing Visit End Dates/Pharmacy considered visits

THEMIS Topic	Resolution
Missing Visit End Dates	Visit end dates should be required. Use the best information to infer a visit end date. Examples include: • A claims billing end date instead of a discharge date • An encounter end date instead of a visit end date
Prescription events have visits	concept_ids have been proposed by Gowtham (http://forums.ohdsi.org/t/pharmacy-claims-drug-exposure-how-to-identify-the-dispensing-billing-pharmacy/3795/13) and will be allowed in the CDM.

THEMIS Releases

October Symposium 2018

Release V1.0.0

Past due by 9 months 100% complete

Set of items reviewed and ratified by the THEMIS working groups.

0 Open ✓ 19 Closed

- Source Concepts for Local Variables Incorporated! Ratified! WG#3
  #22 by ericaVoss was cl
- Masking of Item
  #21 by ericaVoss was
- Overlapping Pay
  #18 by aostropolets w
- Missing Visit En
  #13 by aostropolets w
- Person Inclusion
  #9 by ericaVoss was cl
- CONCEPT_ID (to NULL or not to NULL)
  #11 by ericaVoss was closed on Oct 24, 2018
- Multiple deaths on the same day Incorporated! Ratified! WG#4
  #5 by aostropolets was closed on Oct 2, 2018
- Multiple deaths on different days Incorporated! Ratified! WG#4
  #6 by aostropolets was closed on Oct 2, 2018
- Default Time Incorporated! Ratified! WG#3
  #10 by ericaVoss was closed on Oct 2, 2018
- Duplicate Drugs on Same Day Incorporated! Ratified! WG#3
  #14 by ericaVoss was closed on Oct 2, 2018

Release V1.0.0 Highlights:

- Establishing source concepts for local variables
- Handling overlapping Payer Plan periods
- Establishing conventions for:
  - Multiple death codes
  - Person inclusion
  - Masking items related to a person

May 2019

Next Release

No due date 15% complete

17 Open ✓ 3 Closed

- Specialty & Add Clinical Title PRIORITY HIGH
  #49 opened on Nov 27, 2018 by ericaVoss
- Multiple Addres
  #48 opened on Nov
- Place_of_Servic PRIORITY HIGH
  #47 opened on Nov
- Generate Heel PRIORITY HIGH
  #4 opened on Apr 1
- Observation –
  #29 opened on May
- Patient Name Changes PRIORITY LOW
  #39 opened on Jun 19, 2018 by ericaVoss
- Event type concept when vocabulary maps to a different domain PRIORITY HIGH
  #40 opened on Jun 29, 2018 by don-torok
- Where should we store DRG codes? PRIORITY HIGH
  #19 opened on May 1, 2018 by aostropolets
- Payer plan enhancement and clean-up PRIORITY MEDIUM
  #20 opened on May 1, 2018 by aostropolets
- Patient Reported Drugs/Conditions PRIORITY MEDIUM
  #25 opened on May 7, 2018 by ericaVoss

Next Release Activities:

- Cleaning up Place_of_Service
- Handling provider specialty, clinical titles, multiple addresses, multiple providers and 1-to-many NPIs
- Mother to baby and father to baby linkages
- Achilles Heel error for spotting non-sensical units

May 2019 Hackathon

Purpose: To create a certification program, based off the rules and regulations THEMIS has decided on

Who Should Attend: SQL Developers, JAVA Developers, Software Developers

Day 1:

- What are the gaps between THEMIS rules and Achilles
- Create a program to check for these rules
  - Adjust Achilles to take these rules into consideration
  - Take code of Achilles and create new program
  - Start from Scratch

Day 2:

- Code it
- Mostly developers at work
- Analysts/ non technical team work to organize
- Whatever work does not get done strategize the next steps to keep the ball rolling

Where/When: May 14th & 15th at the University of Colorado (Aurora)

Join Us

To learn more about any of these topics, come see us at the tutorials:

- Saturday – OMOP Vocabulary/CDM
- Sunday – OMOP CDM ETL

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Q&A Session

All speakers

Moderator: Patrick Ryan

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

13:00 OHDSI Collaborator Showcase: Demos and Posters

14:00 Lightning Talks

Second Annual

EUROPEAN OHDSI SYMPOSIUM

March 29th 2019

Tutorials 30th and 31st

The Journey from Data to Evidence

Erasmus MC Rotterdam The Netherlands

www.ohdsi-europe.org