# A peek into the OHDSI vocabulary engine room

Michael Kallfelz, MD

Odysseus Data Services Inc

# Why Standardized Vocabularies?

"The nice thing about standards is that you have so many to choose from." (Andrew S. Tanenbaum)

"Standards are like toothbrushes. Everybody wants one but nobody wants to use anybody else's." (Connie Morella)

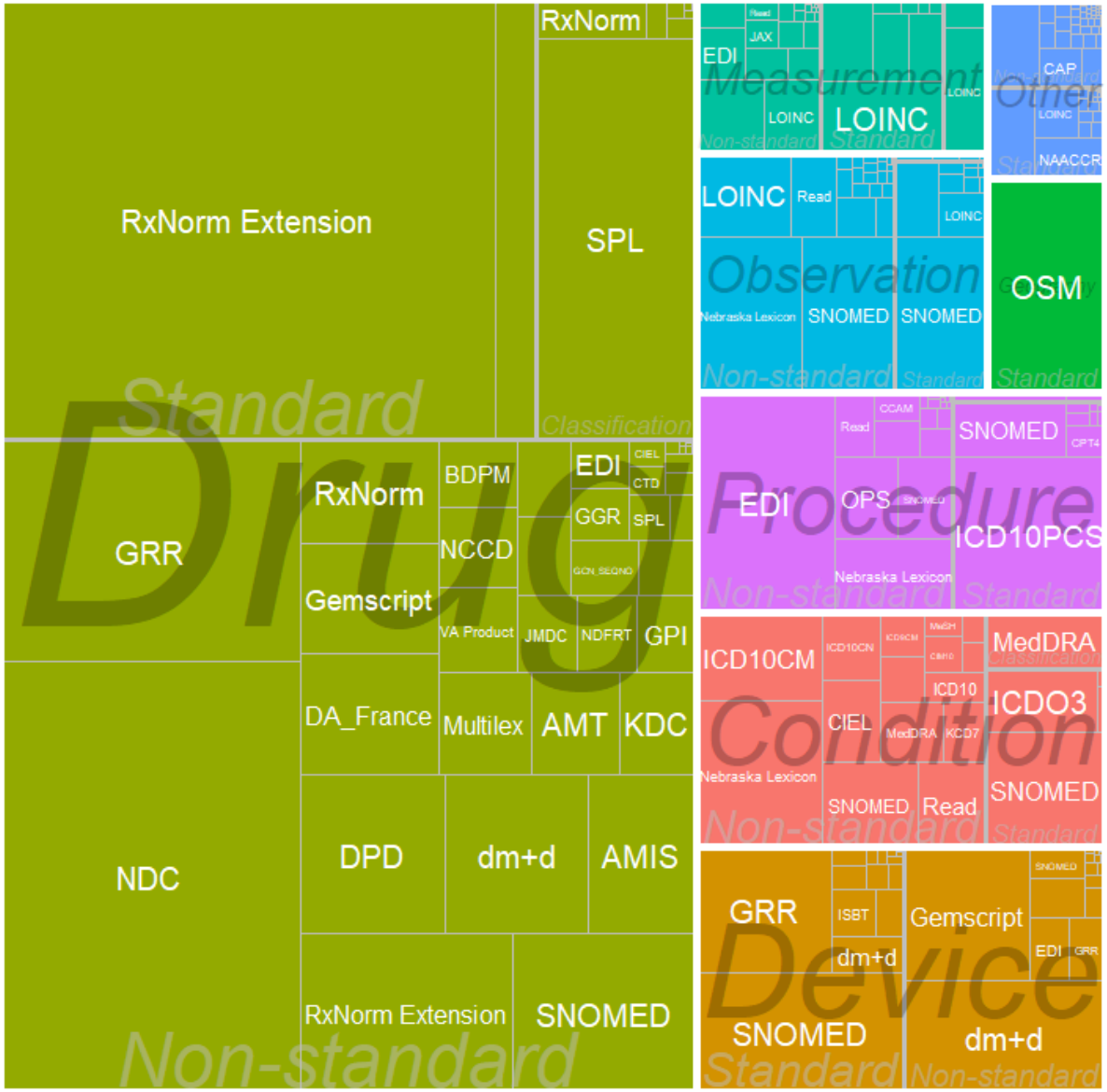
- A myriad of standard terminologies exist for clinical documentation
- Many of them started from a defined use case and have grown since
- In a Common Data Model there ideally is one standard concept
- Backbone terminologies have been chosen and many other terminologies "ingested" to build relationship entries to target STANDARD concepts

# Purpose of the Standardized Vocabularies

1. Give you ideally one *standard* concept for a clinical definition

2. By providing relationships from *non-standard* to *standard* concepts, support the ETL step

3. Support analytical use cases by providing additional information such as hierarchies




OHDSI vocabularies by the numbers
(as of 20 March 2022):

**134** vocabularies

**40** domains

**10,088,289** concepts

**3,533,508** standard concepts

**752,175** classification concepts

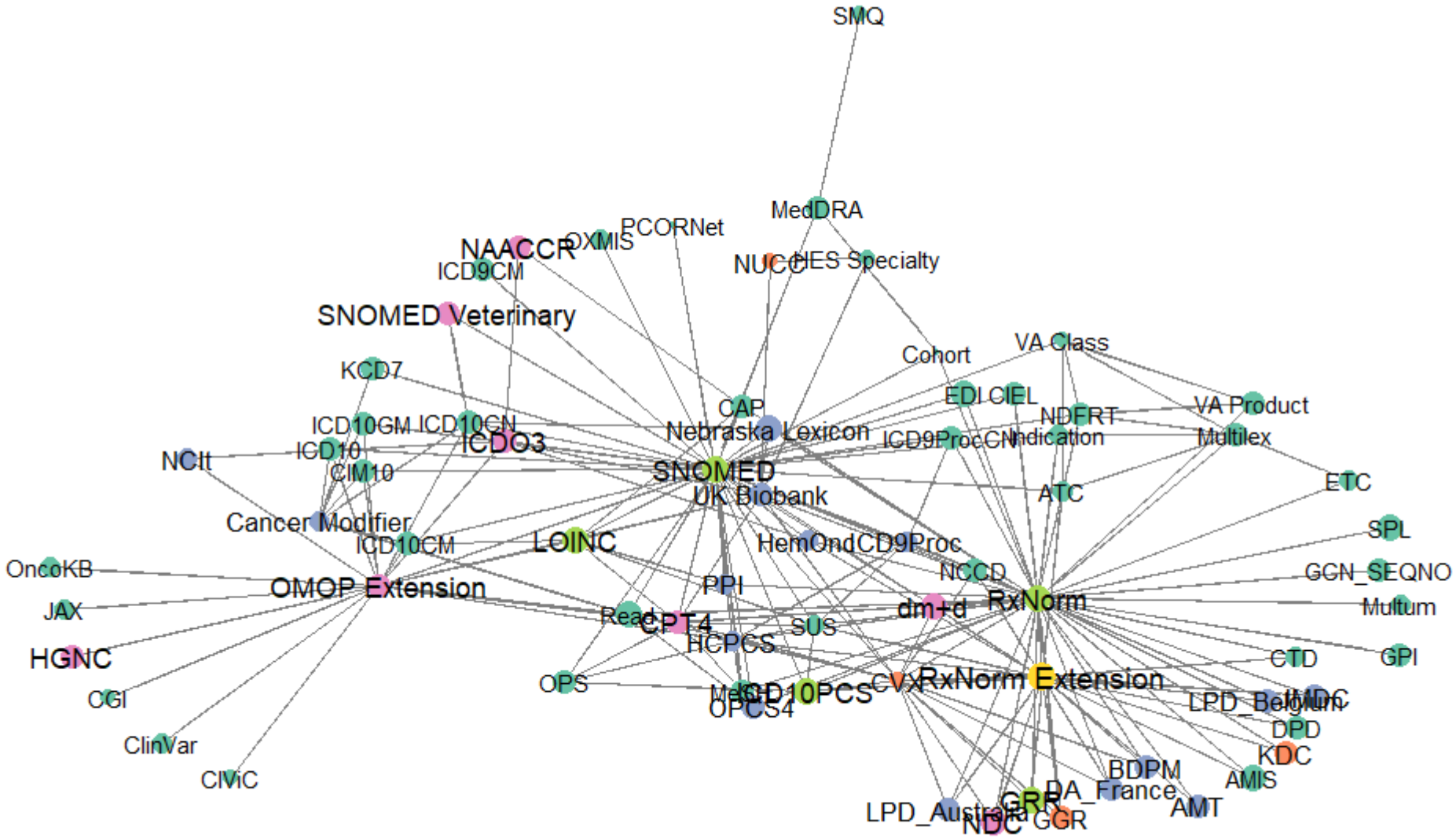
**76,192,476** Concept relationships

**84,898,314** Ancestral relations

**3,016,130** Concept synonyms

**1** shared resource to enable OHDSI

# Vocabulary clusters and relations



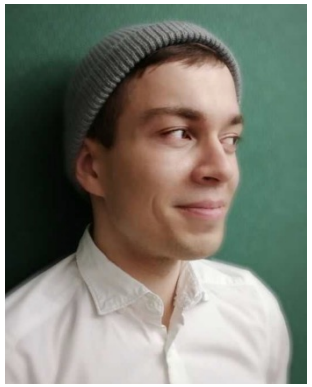
# Vocabularies in detail

- 90 Vocabularies in active maintenance
  - 10 adopted from UMLS including mappings
  - 10 OMOP metadata
  - 3 OMOP internal creation
  - Remaining ones from various sources
- 21 somewhat outdated
- 9 currently inactive
- 14 deprecated

# The team

Timur

Oleg



Vlad

Varvara

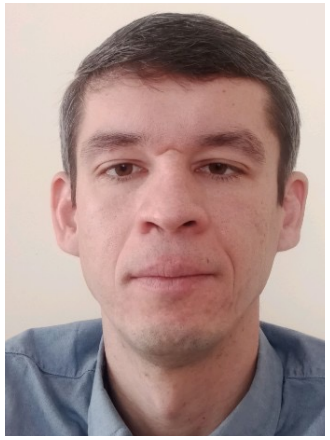
Masha

Mikita

Mikhail

Irina

Maria

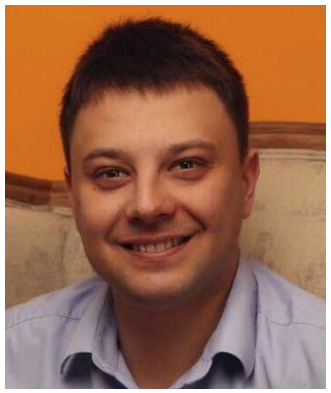


Dmitry

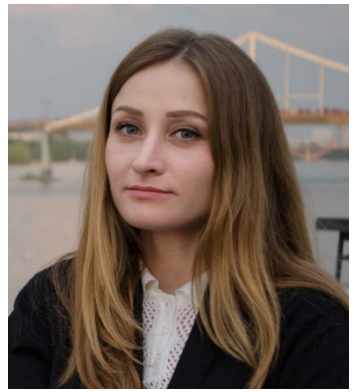
# The team (cont.)

Christian

Anna



Dmitry



Polina

Eduard

Denys

Alexander

Michael

# What the team does…

- New Vocabularies
- Vocabulary refreshes
- Maintenance & Support
- Documentation
- Architecture improvements
- Automation
- Innovation

# Other contributors

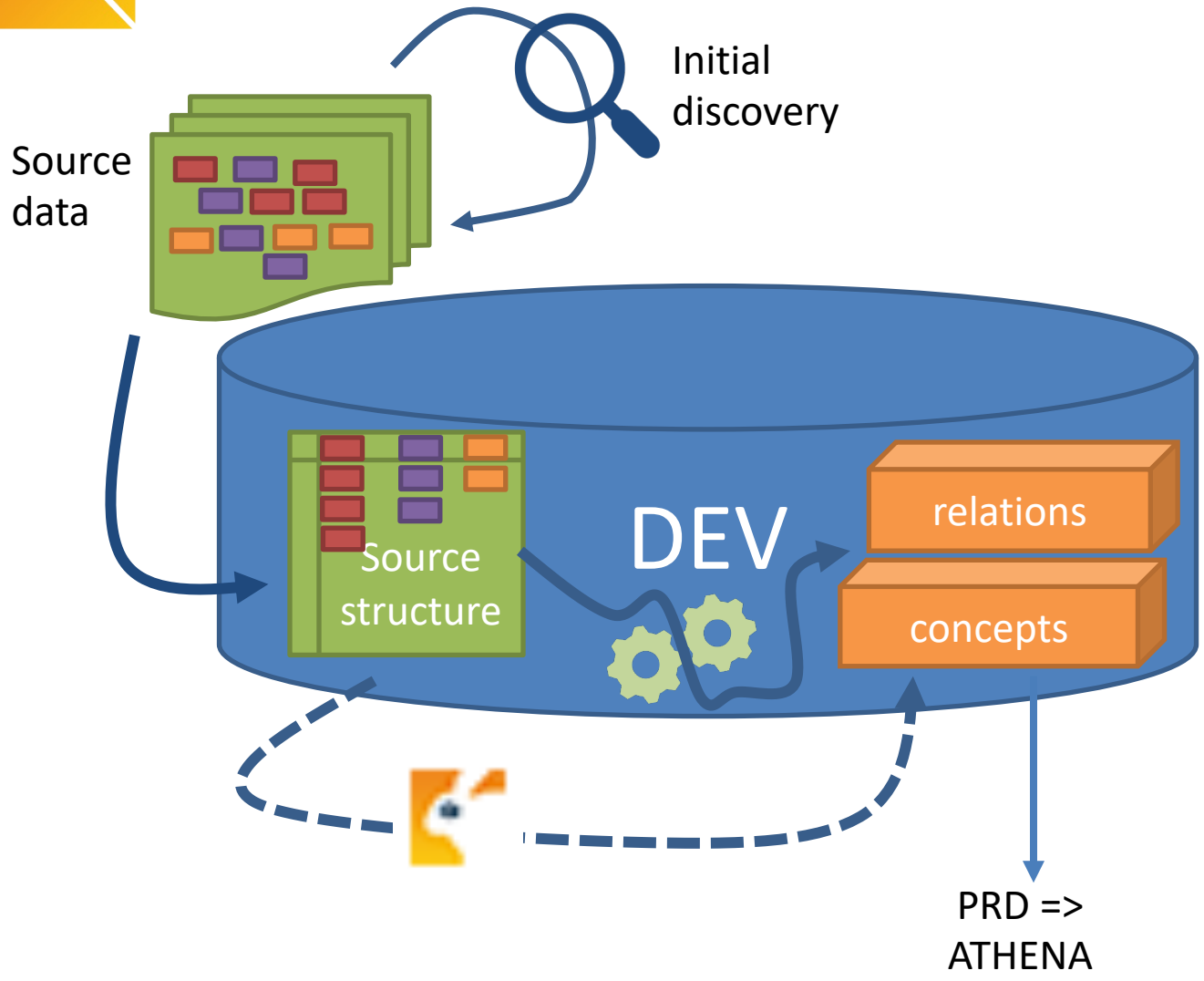
- Vaccine Workgroup
  - Vaccine Ontology
- Korean Community (Hira)
  - Korean Drug Catalog, KCD7, Korean EDI, …
- Chinese Community
  - NCCD, ICD10CN, ICD9ProcCN
- Health Equity Workgroup
  - Race & Ethnicity

# How does an OHDSI vocabulary come into life?

# Creating a vocabulary

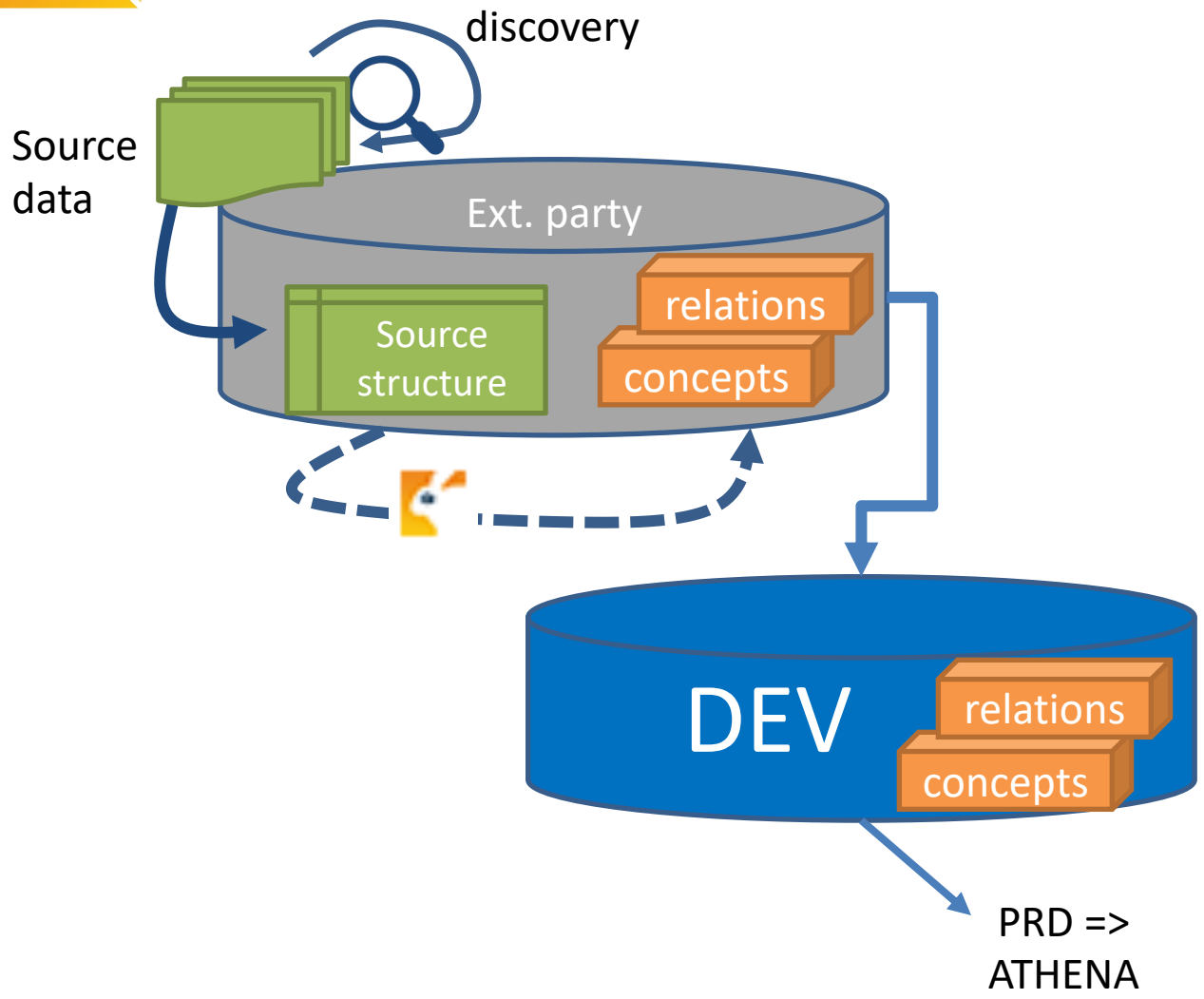
- General approach
  - Get access to source data
  - Initial discovery of data structure and semantics
  - “Import” to Development Database
    - Create digestible extracts
    - Perform automated translation if required, followed by manual review
    - Optionally run Usagi, import results
  - Develop scripts (SQL or Python) to perform structural conversion and automated mapping

# Creating a vocabulary



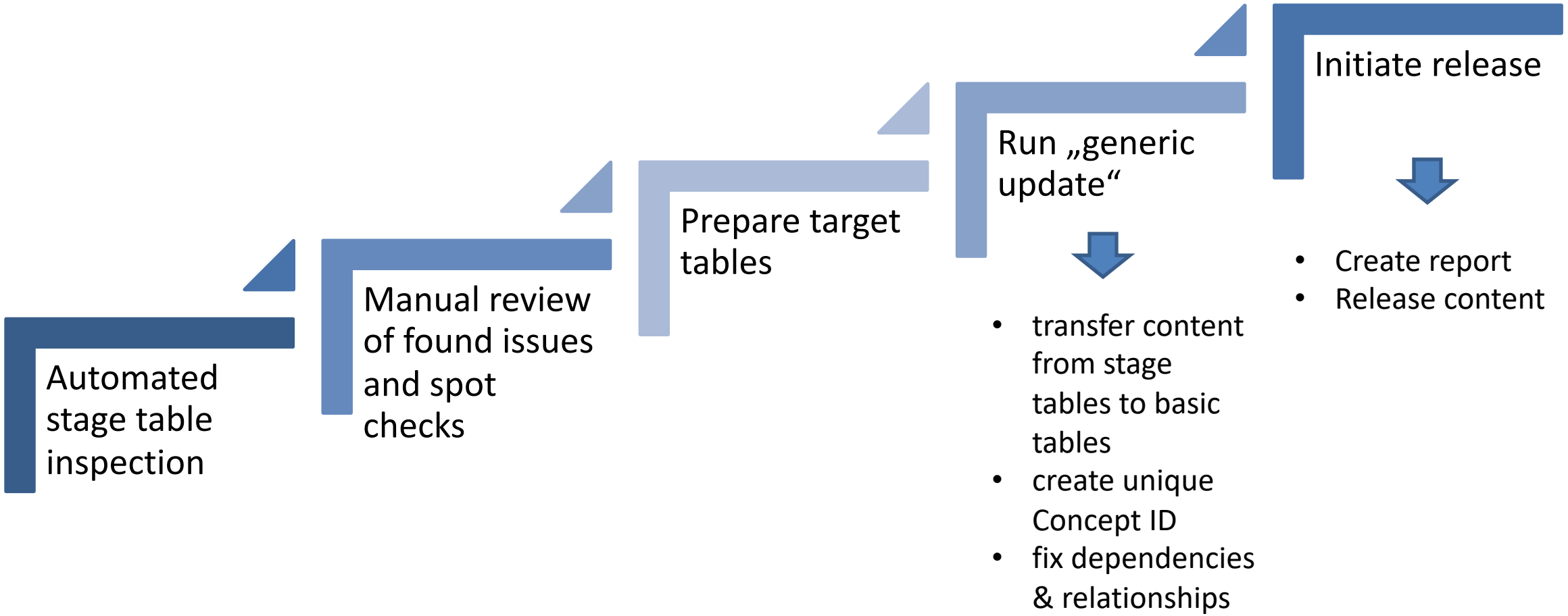

- Import and process
  - Identify source structure
  - Identify possible mapping alternatives (external reference sets…)
  - Translate to English
  - Apply automated mapping
  - Curate results manually
  - Load stage tables
  - Go through QA process

# Creating a vocabulary - alternative

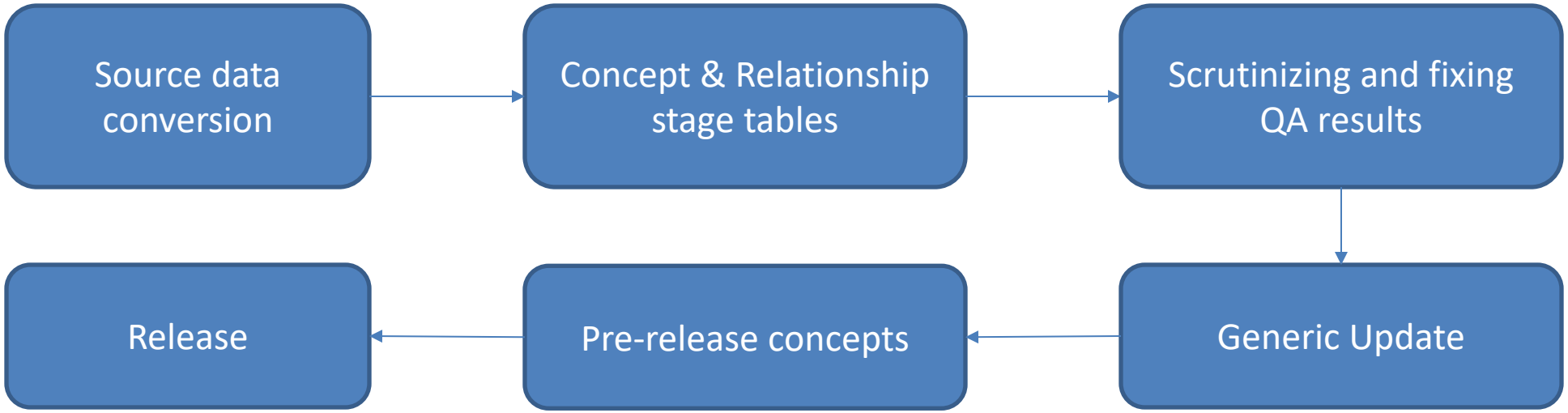



- Import and process
  - Complete preparation process to stage tables
  - Go through QA process
- Export in stage table format
  - Secondary QA process
  - Release to Athena

# Vocabulary QA and release



# Production steps




- Automatic and manual QA on stage table level
- Automatic checks when moving from stage tables to concept
- Automatic checks within release process, creating release notes

# Steps for vocabulary creation

- Condition type
  - Preserve Condition catalogue hierarchy
  - Translation of terms when necessary
  - Process mapping to standard concepts (mostly SNOMED)
  - Conditions are often complex > multiple mappings
  - Provide stage table content
  - Process QA scripts and adjust accordingly

# Steps for vocabulary creation

- Drug type
  - Recover products, clinical drugs, dose forms, ingredients etc.
  - Provide drug data in pre-defined tables
  - Do manual mapping of leftovers / not automatically mapped items
  - Process QA scripts and adjust accordingly
  - Start Drug-“Boiler” for RxNorm mapping or RxE processing

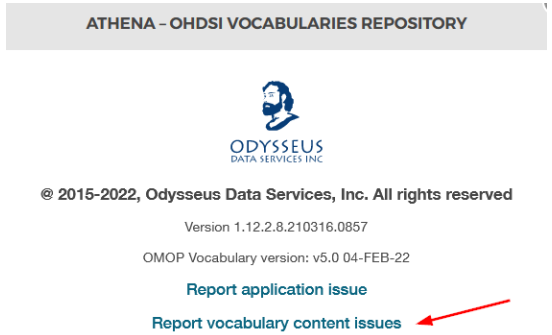
# How does an OHDSI vocabulary stay alive?

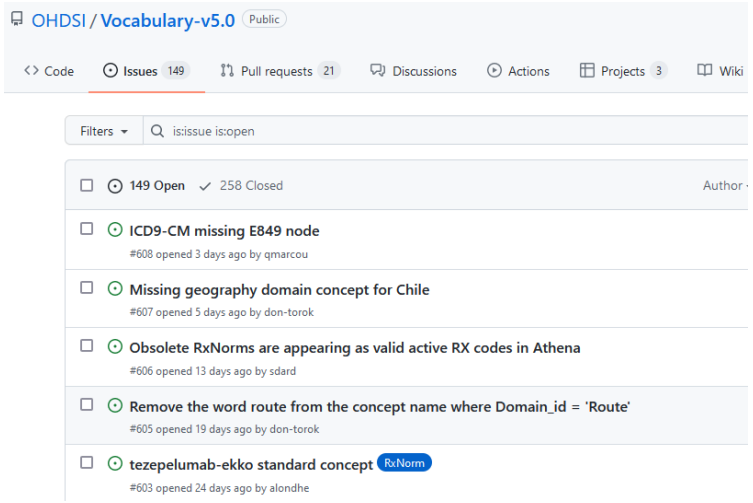
# Refresh & maintenance

- Automated source refresh for stable sources, manual source data refresh on demand
- Manual pull of updated sources into the development environment and subsequent processing
- Extensive review with 4-eyes principle

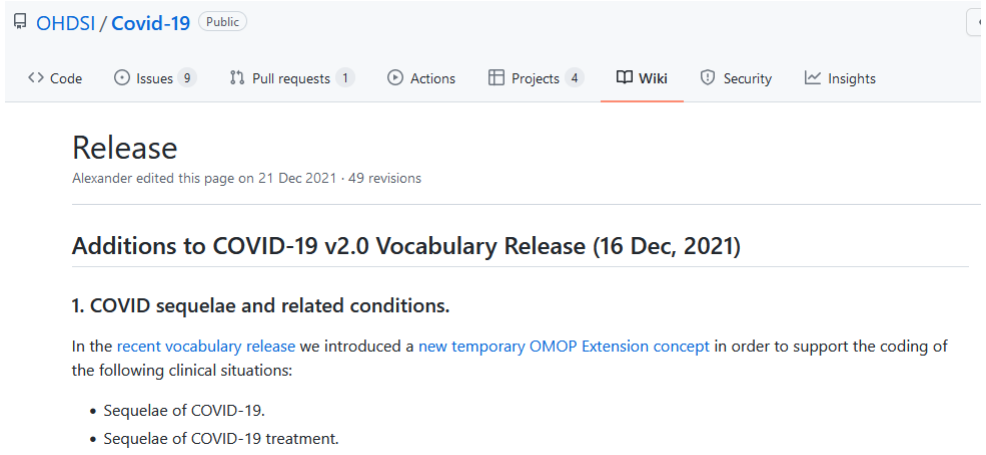
# Fixes & additions

- Github issue triggered repair




- New urgently needed concepts (e.g. CoViD)



# Where to find more information?

- Vocabulary documentation

OHDSI / Vocabulary-v5.0 Public

Code | Issues 149 | Pull requests 21 | Discussions | Actions | Projects 3 | Wiki | Security

## Standardized Vocabularies

Michael Kallfelz edited this page on 25 Oct 2021 · 8 revisions

== this page is currently under construction ==

The table below reflects the vocabularies available inside ATHENA and provides some additional information for each of them:

(* means the page is not complete)

| Name | Description | Update Cycle | Default |
|---|---|---|---|
| *ABMS | Provider Specialty (American Board of Medical Specialties) | on demand | |
| AMT | Australian Medicines Terminology (NEHTA) | on demand | |
| APC | Ambulatory Payment Classification (CMS) | on demand | |
| ATC | WHO Anatomic Therapeutic Chemical Classification | yearly | X |
| BDPM | Public Database of Medications (Social-Sante) | on demand | |

- Vocabulary Roadmap

OHDSI / Vocabulary-v5.0 Public

Code | Issues 149 | Pull requests 21 | Discussions | Actions | Projects 3 | Wiki | Security | Ins

## Release planning quarterly updated

Michael Kallfelz edited this page 21 days ago · 2 revisions

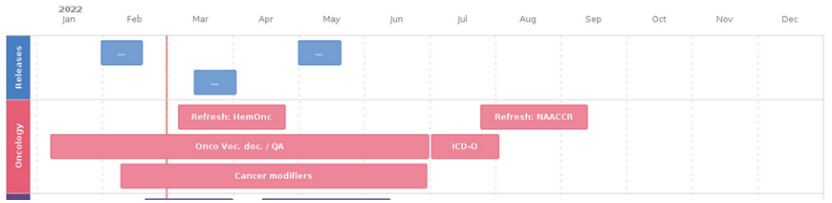
### OHDSI Vocabularies Roadmap

This page provides you with a graphical representation of the planned activities around OHDSI standardized vocabularies.

This is to be treated as a forecast and is subject to changes. The timeline can substantially change due to bandwidth issues or a change in priority per individual tasks. We will update this page at the beginning of each quarter of the year.

Please provide requests for improvement of this page to the vocabulary team as a github issue.

### Roadmap 2022 Q1:



# Work in progress

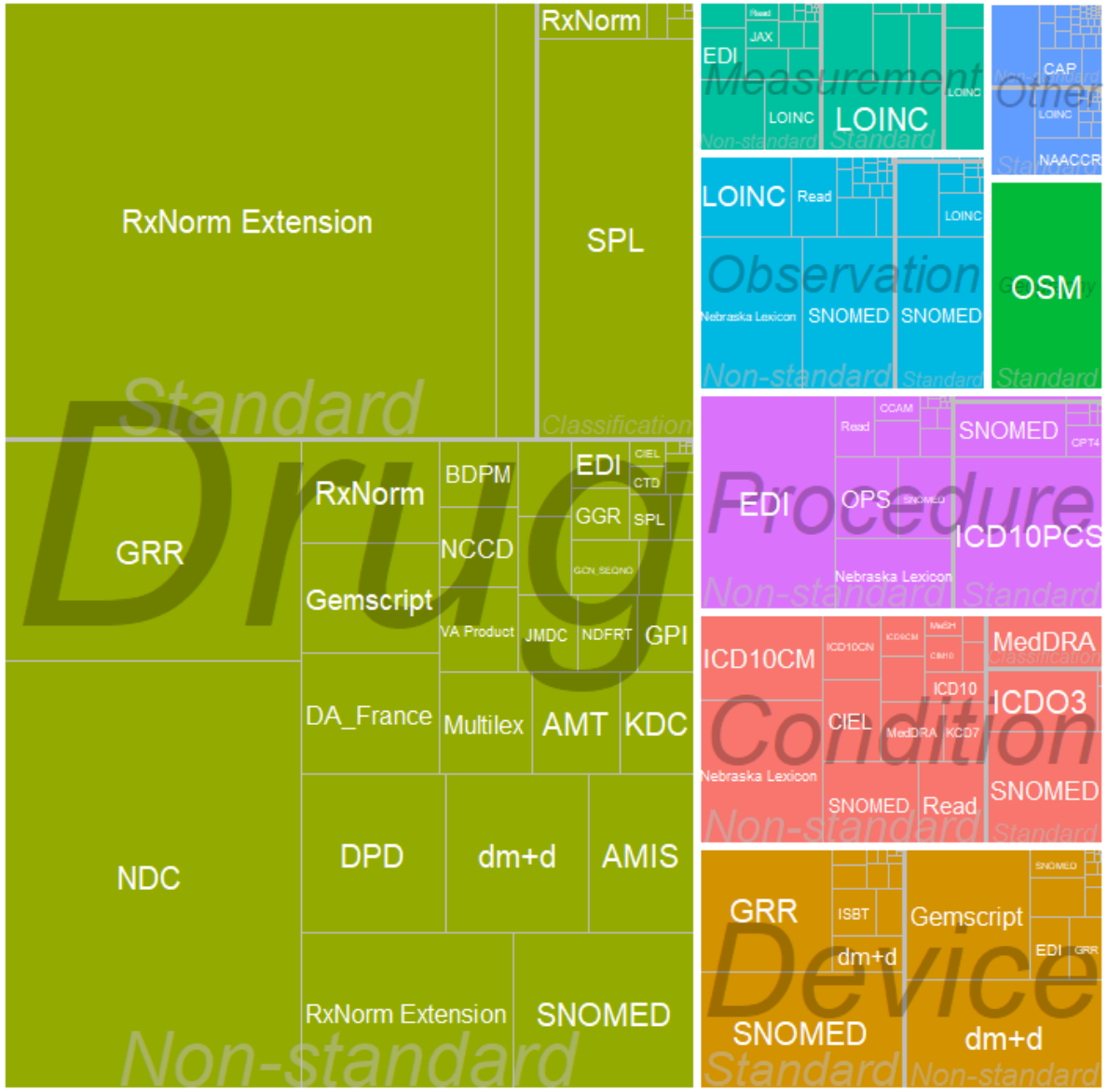
- Vocabulary and Concept Version control
  - Current status: definition of requirements
- Releases
  - Process
  - Documentation
- General Documentation
  - Dashboard
  - Best practices

# Fun things you can learn with the OHDSI Standardized Vocabularies

Patrick Ryan, PhD

Johnson & Johnson Global Epidemiology

Columbia University Irving Medical Center




OHDSI vocabularies by the numbers
(as of 20 March 2022):

**10,088,289** concepts

**3,533,508** standard concepts

**752,175** classification concepts

**134** vocabularies

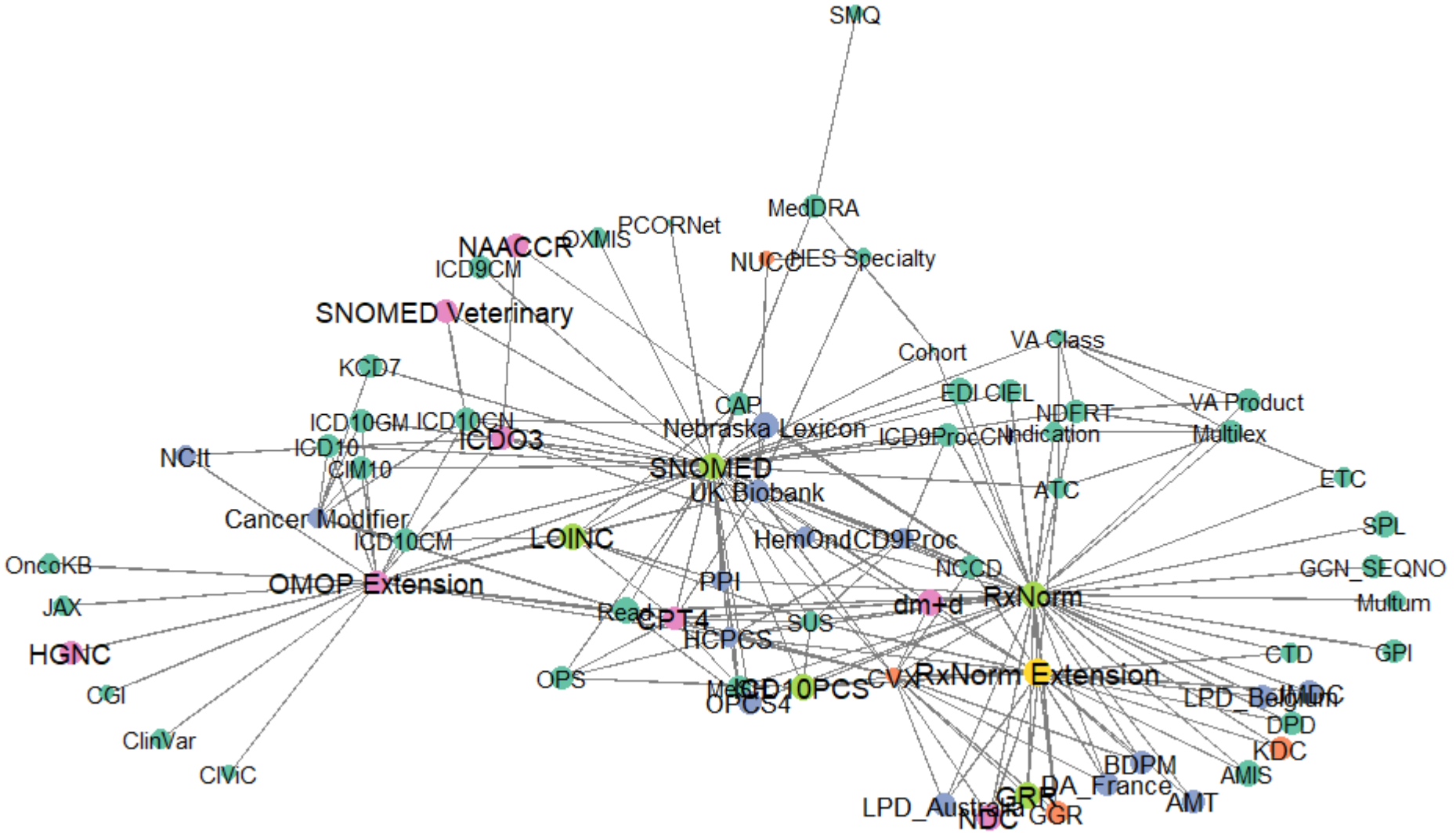
**40** domains

**76,192,476** Concept relationships

**84,898,314** Ancestral relations

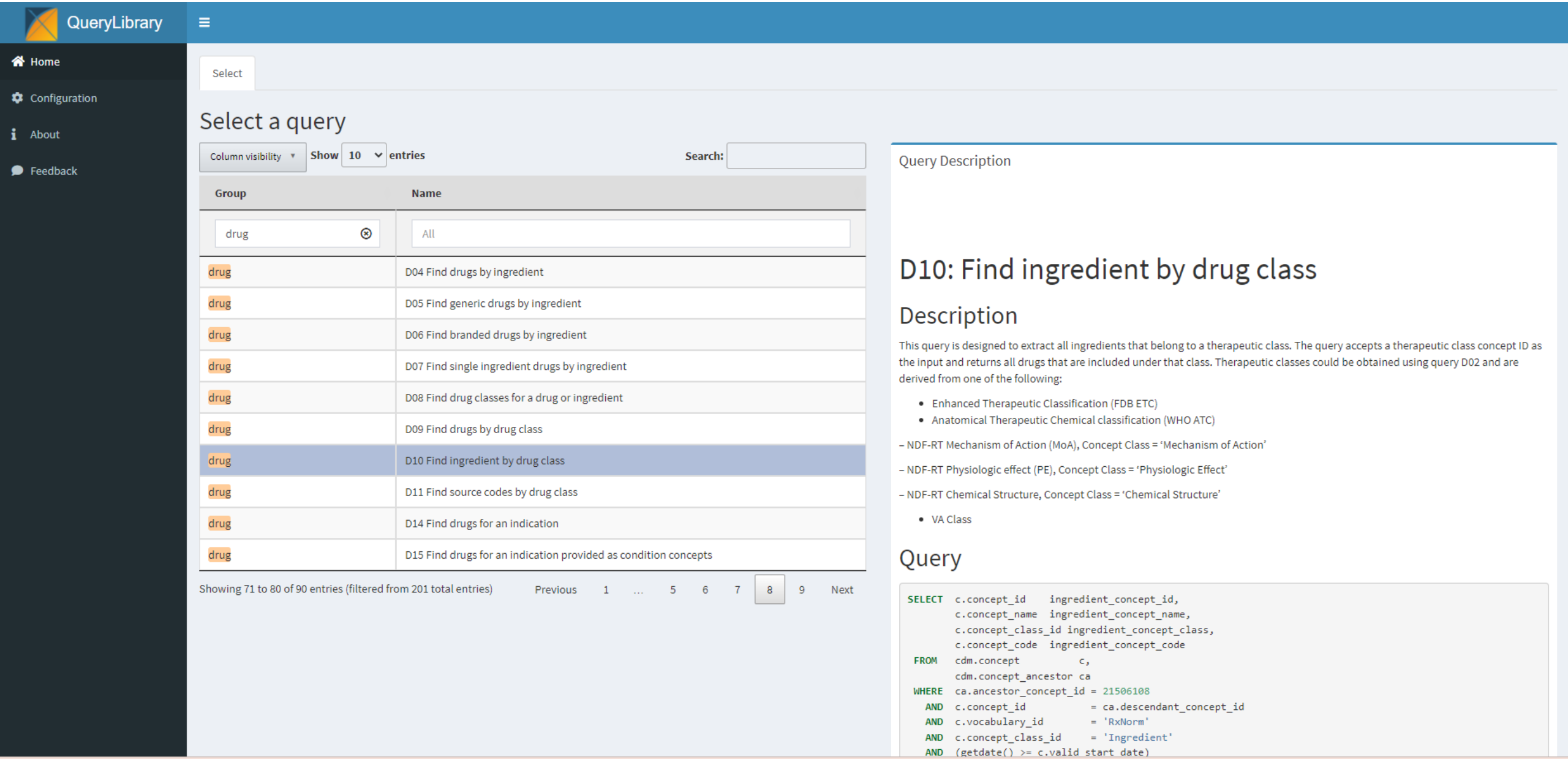
**3,016,130** Concept synonyms

**1** shared resource to enable OHDSI


SMQ
MedDRA
PCORNet
OXMIS
NAACCR
ICD9CM
NUCC
HES Specialty
SNOMED Veterinary
KCD7
Cohort
VA Class
CAP
EDI CIEL
VA Product
ICD10GM
ICD10CN
Nebraska Lexicon
NDFRT
ICD9ProcCN
Indication
Multilex
ICDO3
NCIt
ICD10
CIM10
SNOMED
UK Biobank
ATC
ETC
Cancer Modifier
ICD10CM
LOINC
HemOncCD9Proc
SPL
OncoKB
NCCD
GCN_SEQNO
OMOP Extension
PPI
RxNorm
JAX
dm+d
Multum
Read
CPT4
SUS
HCPCS
HGNC
CTD
GPI
OPS
CVX
RxNorm Extension
ICD10PCS
LPD_Belgium
CGI
OPCS4
DPD
KDC
ClinVar
BDPM
AMIS
DA_France
CIViC
GRR
AMT
LPD_Australia
NDC
GGR

OHDSI Standardized Vocabularies are NOT just a tool for ETL to standardize real-world data

OHDSI Standardized Vocabularies are a tool for ANALYSIS TO GENERATE RELIABLE EVIDENCE




https://data.ohdsi.org/QueryLibrary/

# Questions that the vocabulary can answer

Drug:

- Given drug class, find all ingredients
- Given drug, find an indication
- Given drug, find a comparator drug (drug with same indication or same class)

Condition:

- Given a list of source codes, find all standard concepts
- Given a condition, find all related conditions

…and the list goes on and on…

# Given a name of drug class, find me the ingredients
# Step 1: Find classes for 'TNFs'

```
SELECT *
FROM concept
WHERE vocabulary_id IN ('NDFRT', 'ETC', 'ATC', 'VA Class')
AND LOWER(concept_name) LIKE '%tumor necrosis factor%'
```

| concept_id | concept_name | domain_id | vocabulary_id ▲ | concept_class_id | standard_concept |
|---|---|---|---|---|---|
| 21603907 | Tumor necrosis factor alpha (TNF-a) inhibitors | Drug | ATC | ATC 4th | C |
| 21505707 | Anti-inflammatory Tumor Necrosis Factor Inhibiting Agnts,Non-SeIective | Drug | ETC | ETC | C |
| 21500480 | Inflammatory Bowel Agent - Tumor Necrosis Factor Alpha Blockers | Drug | ETC | ETC | C |
| 21505708 | Anti-inflammatory Tumor Necrosis Factor Inhibiting Agnts,TNF-alpha Sel | Drug | ETC | ETC | C |
| 21500601 | DMARD - Anti-inflammatory Tumor Necrosis Factor Inhibiting Agents | Drug | ETC | ETC | C |
| 4260302 | Receptors, Tumor Necrosis Factor, Member 25 | Drug | NDFRT | Chemical Structure | |
| 4317589 | Receptors, Tumor Necrosis Factor, Member 10c | Drug | NDFRT | Chemical Structure | |
| 4319368 | Receptors, Tumor Necrosis Factor, Member 6b | Drug | NDFRT | Chemical Structure | |
| 4256089 | Tumor Necrosis Factor-alpha | Drug | NDFRT | Chemical Structure | |
| 4257024 | Receptors, Tumor Necrosis Factor, Type II | Drug | NDFRT | Chemical Structure | |
| 4258863 | Receptors, Tumor Necrosis Factor, Type I | Drug | NDFRT | Chemical Structure | |
| 4334083 | Tumor Necrosis Factor Blocker | Drug | NDFRT | Pharmacologic Class | |
| 4317874 | Tumor Necrosis Factor Decoy Receptors | Drug | NDFRT | Chemical Structure | |
| 4318304 | Tumor Necrosis Factor beta Receptor Blocking Activity | Drug | NDFRT | Mechanism of Action | |
| 4317871 | Receptors, Tumor Necrosis Factor, Member 14 | Drug | NDFRT | Chemical Structure | |
| 718459 | Tumor Necrosis Factor alpha-Induced Protein 3 | Drug | NDFRT | Chemical Structure | |
| 4256088 | Tumor Necrosis Factors | Drug | NDFRT | Chemical Structure | |
| 4318306 | Tumor Necrosis Factor Receptor Blocking Activity | Drug | NDFRT | Mechanism of Action | |
| 4319941 | Tumor Necrosis Factor alpha Receptor Blocking Activity | Drug | NDFRT | Mechanism of Action | |
| 4332483 | Receptors, Tumor Necrosis Factor | Drug | NDFRT | Chemical Structure | |
| 4317604 | Tumor Necrosis Factor Ligand Superfamily Member 13 | Drug | NDFRT | Chemical Structure | |
| 4318158 | Tumor Necrosis Factor Ligand Superfamily Member 14 | Drug | NDFRT | Chemical Structure | |
| 4256093 | Tumor Necrosis Factor Receptor-Associated Peptides and Proteins | Drug | NDFRT | Chemical Structure | |
| 4331352 | Tumor Necrosis Factor Receptor Interactions | Drug | NDFRT | Mechanism of Action | |
| 4318159 | Tumor Necrosis Factor Ligand Superfamily Member 15 | Drug | NDFRT | Chemical Structure | |
| 4334516 | TUMOR NECROSIS FACTOR BLOCKER | Drug | VA Class | VA Class | |

# Given a name of drug class, find me the ingredients
# Step 2: Follow ancestry to find drugs

```
SELECT DISTINCT c1.concept_id, c1.concept_name
FROM
(
SELECT *
FROM concept
WHERE vocabulary_id IN ('NDFRT', 'ETC', 'ATC', 'VA Class')
AND LOWER(concept_name) LIKE '%tumor necrosis factor%'
) classes
INNER JOIN concept_ancestor ca1
ON classes.concept_id = ca1.ancestor_concept_id
INNER JOIN concept c1
ON ca1.descendant_concept_id = c1.concept_id
AND c1.concept_class_id = 'Ingredient'
;
```

| concept_id | concept_name |
|---|---|
| 912263 | certolizumab pegol |
| 1119119 | adalimumab |
| 937368 | infliximab |
| 1151789 | etanercept |
| 19041065 | golimumab |

# Given a name of drug class, find me the ingredients
# Example 2: ‘drugs used for diabetes’

```
SELECT DISTINCT c1.concept_id, c1.concept_name
FROM
(
SELECT *
FROM concept
WHERE vocabulary_id IN ('NDFRT', 'ETC', 'ATC', 'VA Class')
AND LOWER(concept_name) LIKE '%drugs used in diabetes%'
) classes
INNER JOIN concept_ancestor ca1
ON classes.concept_id = ca1.ancestor_concept_id
INNER JOIN concept c1
ON ca1.descendant_concept_id = c1.concept_id
AND c1.concept_class_id = 'Ingredient'
;
```

| concept_id | concept_name ▲ |
|---|---|
| 19035533 | benfluorex |
| 40798673 | Buformin |
| 43526465 | canagliflozin |
| 19033498 | carbutamide |
| 1594973 | chlorpropamide |
| 44785829 | dapagliflozin |
| 45774435 | dulaglutide |
| 45774751 | empagliflozin |
| 793293 | ertugliflozin |
| 43009051 | evogliptin |
| 1583722 | exenatide |
| 43009089 | gemigliptin |
| 19001409 | glibornuride |
| 19059796 | gliclazide |
| 1597756 | glimepiride |
| 1560171 | glipizide |
| 19097821 | gliquidone |
| 40798860 | Glisoxepide |
| 1559684 | glyburide |
| 19001441 | glymidine |
| 1000979 | guar gum |
| 35198096 | insulin aspart |
| 1531601 | insulin aspart protamine, human |
| 1567198 | insulin aspart, human |
| 35602717 | insulin degludec |
| 1516976 | insulin detemir |
| 1502905 | insulin glargine |
| 1544838 | insulin glulisine, human |
| 1588986 | insulin human, rDNA origin |

77 distinct ingredients!

Many are used in only select geographies, and may not be listed in country-specific guidelines!

# Given drug, find an indication

Step 1: find ingredient concept

```
(SELECT *
FROM concept
WHERE concept_class_id = 'Ingredient' AND standard_concept = 'S'
AND concept_name = 'adalimumab') ingredient
```

| concept_id | concept_name | domain_id | vocabulary_id | concept_class_id | standard_concept | concept_code | valid_start_date | valid_end_date | invalid_reason |
|---|---|---|---|---|---|---|---|---|---|
| 1119119 | adalimumab | Drug | RxNorm | Ingredient | S | 327361 | 1970-01-01 | 2099-12-31 | |

Step 2: exploit vocabulary relationships: ingredient → clinical drug → indication

```
SELECT DISTINCT ingredient.concept_name as ingredient_concept_name, c1.concept_name as indication_concept_name
FROM
  (SELECT *
  FROM concept
  WHERE concept_class_id = 'Ingredient' AND standard_concept = 'S'
  AND concept_name = 'adalimumab') ingredient
INNER JOIN concept_ancestor ca1
ON ingredient.concept_id = ca1.ancestor_concept_id
INNER JOIN concept_relationship cr1
ON ca1.descendant_concept_id = cr1.concept_id_1
AND cr1.relationship_id IN ('Has FDA-appr ind')
INNER JOIN concept c1
ON cr1.concept_id_2 = c1.concept_id
;
```

| ingredient_concept_name | indication_concept_name ▲ |
|---|---|
| adalimumab | Ankylosing Spondylitis |
| adalimumab | Crohn's Disease |
| adalimumab | Hidradenitis Suppurativa |
| adalimumab | Moderate to Severe Plaque Psoriasis |
| adalimumab | Polyarticular Juvenile Idiopathic Arthritis |
| adalimumab | Psoriatic Arthritis |
| adalimumab | Rheumatoid Arthritis |
| adalimumab | Ulcerative Colitis |

# Given drug, find a comparator drug with the same indication

Exploit vocabulary relationships: ingredient → clinical drug → indication → clinical drug → ingredient

```
SELECT DISTINCT ingredient.concept_name as ingredient_concept_name, c1.concept_name as indication_concept_name, c2.concept_name as comparator_concept_name
FROM
  (SELECT *
  FROM concept
  WHERE concept_class_id = 'Ingredient' AND standard_concept = 'S'
  AND concept_name = 'adalimumab') ingredient
INNER JOIN concept_ancestor ca1
ON ingredient.concept_id = ca1.ancestor_concept_id
INNER JOIN concept_relationship cr1
ON ca1.descendant_concept_id = cr1.concept_id_1
AND cr1.relationship_id IN ('Has FDA-appr ind')
INNER JOIN concept c1
ON cr1.concept_id_2 = c1.concept_id
INNER JOIN concept_relationship cr2
ON cr2.concept_id_2 = c1.concept_id
AND cr2.relationship_id IN ('Has FDA-appr ind')
INNER JOIN concept_ancestor ca2
ON ca2.descendant_concept_id = cr2.concept_id_1
INNER JOIN concept c2
ON ca2.ancestor_concept_id = c2.concept_id
and c2.concept_class_id = 'Ingredient'
;
```

| ingredient_concept_name | indication_concept_name | comparator_concept_name |
|---|---|---|
| adalimumab | Ankylosing Spondylitis | adalimumab |
| adalimumab | Ankylosing Spondylitis | betamethasone |
| adalimumab | Ankylosing Spondylitis | celecoxib |
| adalimumab | Ankylosing Spondylitis | certolizumab pegol |
| adalimumab | Ankylosing Spondylitis | corticotropin |
| adalimumab | Ankylosing Spondylitis | cortisone |
| adalimumab | Ankylosing Spondylitis | dexamethasone |
| adalimumab | Ankylosing Spondylitis | diclofenac |
| adalimumab | Ankylosing Spondylitis | etanercept |
| adalimumab | Ankylosing Spondylitis | golimumab |
| adalimumab | Ankylosing Spondylitis | hydrocortisone |
| adalimumab | Ankylosing Spondylitis | indomethacin |
| adalimumab | Ankylosing Spondylitis | infliximab |
| adalimumab | Ankylosing Spondylitis | lansoprazole |
| adalimumab | Ankylosing Spondylitis | methylprednisol |
| adalimumab | Ankylosing Spondylitis | naproxen |
| adalimumab | Ankylosing Spondylitis | prednisolone |
| adalimumab | Ankylosing Spondylitis | prednisone |
| adalimumab | Ankylosing Spondylitis | sulindac |
| adalimumab | Ankylosing Spondylitis | triamcinolone |
| adalimumab | Crohn's Disease | adalimumab |
| adalimumab | Crohn's Disease | betamethasone |
| adalimumab | Crohn's Disease | budesonide |
| adalimumab | Crohn's Disease | certolizumab pe |
| adalimumab | Crohn's Disease | cortisone |
| adalimumab | Crohn's Disease | dexamethasone |
| adalimumab | Crohn's Disease | hydrocortisone |
| adalimumab | Crohn's Disease | infliximab |
| adalimumab | Crohn's Disease | methylprednisol |
| adalimumab | Crohn's Disease | natalizumab |
| adalimumab | Crohn's Disease | prednisolone |
| adalimumab | Crohn's Disease | prednisone |
| adalimumab | Crohn's Disease | triamcinolone |
| adalimumab | Crohn's Disease | vedolizumab |
| adalimumab | Hidradenitis Suppurativa | adalimumab |
| adalimumab | Moderate to Severe Plaque Psoriasis | adalimumab |

Each indication has different comparators!

**19** for Ankylosing Spondylitis
**13** for Crohn's disease
**15** for Ulcerative colitis
**51** for Rheumatoid arthritis
0 for Hidradenitis suppurativa

Comparator selection can be hard, and expert opinion can be often insufficient.
**let the vocabulary help you!**

# Given source codes, find the standard concepts

Research

JAMA | Original Investigation

## Association of Rivaroxaban vs Apixaban With Major Ischemic or Hemorrhagic Events in Patients With Atrial Fibrillation

Wayne A. Ray, PhD; Cecilia P. Chung, MD, MPH; C. Michael Stein, MB, ChB; Walter Smalley, MD, MPH; Eli Zimmerman, MD; William D. Dupont, PhD; Adriana M. Hung, MD, MPH; James R. Daugherty, MS; Alyson Dickson, MA; Katherine T. Murray, MD

*e. Atrial fibrillation.*

| ICD-9 | Rubric | ICD-10 | Rub |
|---|---|---|---|
| 427.31 | Atrial fibrillation | I48.0 | Par |
| 427.32 | Atrial flutter | I48.1 | Per |
| | | I48.2 | Chr |
| | | I48.3 | Typ |
| | | I48.4 | Aty |
| | | I48.91 | Uns |
| | | I48.92 | Uns |

OHDSI Home | Forums | Wiki | Github

OHDSI

### Phenotype Phebruary Day 3 - Atrial Fibrillation

General

Patrick_Ryan Feb 4

Today, I'd like to a little side step from the problem of creating phenotypes de novo, and talk about how to try to implement a phenotype algorithm using OHDSI tools based on an existing description from some external material, just as a publication. I'll use the phenotype of Atrial Fibrillation to demonstrate some tips and tricks, but I hope you'll see that the steps I'll follow here are completely transportable to whatever phenotype you may want to be working on.

**Clinical description:**

Atrial fibrillation (AFib) is an abnormal heart rhythm, often with irregular beats in the atrial chamber.

Feb 4

1 / 5
Feb 4

# Given source codes, find the standard concepts

*e. Atrial fibrillation.*

From Ray JAMA 2022

| ICD-9 | Rubric | ICD-10 | Rubric |
|---|---|---|---|
| 427.31 | Atrial fibrillation | I48.0 | Paroxysmal atrial fibrillation |
| 427.32 | Atrial flutter | I48.1 | Persistent atrial fibrillation |
| | | I48.2 | Chronic atrial fibrillation. Note code expanded to I48.20, I48.21 in FY20 |
| | | I48.3 | Typical atrial flutter, Type I atrial flutter |
| | | I48.4 | Atypical atrial flutter, Type II atrial flutter |
| | | I48.91 | Unspecified atrial fibrillation |
| | | I48.92 | Unspecified atrial flutter |

```
SELECT *
FROM concept
WHERE (vocabulary_id = 'ICD9CM' AND concept_code IN ('427.31','427.32'))
OR (vocabulary_id = 'ICD10CM' AND concept_code IN ('I48.0','I48.1','I48.2','I48.3','I48.4','I48.91','I48.92'))
```

| 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 |
| 44820868 | Atrial flutter | Condition | ICD9CM | 5-dig billing code | | 427.32 |
| 35207784 | Paroxysmal atrial fibrillation | Condition | ICD10CM | 4-char billing code | | I48.0 |
| 35207785 | Persistent atrial fibrillation | Condition | ICD10CM | 4-char nonbill code | | I48.1 |
| 1569171 | Chronic atrial fibrillation | Condition | ICD10CM | 4-char nonbill code | | I48.2 |
| 1569172 | Typical atrial flutter | Condition | ICD10CM | 4-char billing code | | I48.3 |
| 1569173 | Atypical atrial flutter | Condition | ICD10CM | 4-char billing code | | I48.4 |
| 45576876 | Unspecified atrial fibrillation | Condition | ICD10CM | 5-char billing code | | I48.91 |
| 45572094 | Unspecified atrial flutter | Condition | ICD10CM | 5-char billing code | | I48.92 |

# Given source codes, find the standard concepts

```
SELECT DISTINCT c1.*
FROM
(
SELECT *
FROM concept
WHERE (vocabulary_id = 'ICD9CM' AND concept_code IN ('427.31','427.32'))
OR (vocabulary_id = 'ICD10CM' AND concept_code IN ('I48.0','I48.1','I48.2','I48.3','I48.4','I48.91','I48.92'))
) source_codes
INNER JOIN concept_relationship cr1
ON source_codes.concept_id = cr1.concept_id_1
AND cr1.relationship_id = 'Maps to'
INNER JOIN concept c1
ON cr1.concept_id_2 = c1.concept_id
AND c1.standard_concept = 'S'
;
```

| concept_id | concept_name | domain_id | vocabulary_id | concept_class_id | standard_concept | concept_code | valid_start_date | valid_end_date |
|---|---|---|---|---|---|---|---|---|
| 313217 | Atrial fibrillation | Condition | SNOMED | Clinical Finding | S | 49436004 | 2002-01-31 | 2099-12-31 |
| 314665 | Atrial flutter | Condition | SNOMED | Clinical Finding | S | 5370000 | 2002-01-31 | 2099-12-31 |
| 4141360 | Chronic atrial fibrillation | Condition | SNOMED | Clinical Finding | S | 426749004 | 2007-07-31 | 2099-12-31 |
| 4232697 | Persistent atrial fibrillation | Condition | SNOMED | Clinical Finding | S | 440059007 | 2009-01-31 | 2099-12-31 |
| 4154290 | Paroxysmal atrial fibrillation | Condition | SNOMED | Clinical Finding | S | 282825002 | 2002-01-31 | 2099-12-31 |
| 36712986 | Atypical atrial flutter | Condition | SNOMED | Clinical Finding | S | 15964901000119107 | 2017-01-31 | 2099-12-31 |
| 36714994 | Typical atrial flutter | Condition | SNOMED | Clinical Finding | S | 720448006 | 2017-01-31 | 2099-12-31 |

OHDSI Standardized Vocabularies are NOT just a tool for ETL to standardize real-world data

OHDSI Standardized Vocabularies are a tool for ANALYSIS TO GENERATE RELIABLE EVIDENCE

# Time for reflection

## Where are we?
## Where should we be?

Christian Reich

IQVIA

OHDSI Center at Roux Institute, Northeastern University

# Reflection on

1. The content
2. The service
3. The technology

# Report Card for Domains: Content

| Domain | # | standard class source | Fitness | Complete-ness | Hierarchy | Duplication | Mapping | Overall |
|---|---|---|---|---|---|---|---|---|
| Drug | 6,392,739 | | | | | | | |
| Device | 834,979 | | | | | | | |
| Condition | 784,144 | | | | | | | |
| Procedure | 745,866 | | | | | | | |
| Observation | 572,093 | | | | | | | |
| Measurement | 383,227 | | | | | | | |
| Geography | 203,352 | | | | | | | |
| Spec Anatomic Site | 79,424 | | | | | | | |
| Meas Value | 65,852 | | | | | | | |
| Type Concept | 4,574 | | | | | | | |
| Unit | 3,706 | | | | | | | |
| Provider | 2,313 | | | | | | | |
| Regimen | 2,032 | | | | | | | |
| Race | 1,313 | | | | | | | |
| Visit | 1,269 | | | | | | | |
| Payer | 171 | | | | | | | |
| Cost | 51 | | | | | | | |
| Episode | 26 | | | | | | | |
| Condition Status | 22 | | | | | | | |
| Plan | 11 | | | | | | | |
| Sponsor | 6 | | | | | | | |
| Ethnicity | 2 | | | | | | | |

# Report Card for Domains: Content

| Domain | # | standard | class | source | Fitness | Complete-ness | Hierarchy | Duplication | Mapping | Overall |
|---|---|---|---|---|---|---|---|---|---|---|
| Drug | 6,392,739 | | | | | | | | | |
| Device | 834,979 | | | | | | | | | |
| Condition | 784,144 | | | | | | | | | |
| Procedure | 745,866 | | | | | | | | | |
| Observation | 572,093 | | | | | | | | | |
| Measurement | 383,227 | | | | | | | | | |
| Geography | 203,352 | | | | | | | | | |
| Spec Anatomic Site | 79,424 | | | | | | | | | |
| Meas Value | 65,852 | | | | | | | | | |
| Type Concept | 4,574 | | | | | | | | | |
| Unit | 3,706 | | | | | | | | | |
| Provider | 2,313 | | | | | | | | | |
| Regimen | 2,032 | | | | | | | | | |
| Race | 1,313 | | | | | | | | | |
| Visit | 1,269 | | | | | | | | | |
| Payer | 171 | | | | | | | | | |
| Cost | 51 | | | | | | | | | |
| Episode | 26 | | | | | | | | | |
| Condition Status | 22 | | | | | | | | | |
| Plan | 11 | | | | | | | | | |
| Sponsor | 6 | | | | | | | | | |
| Ethnicity | 2 | | | | | | | | | |

Does it work for our use cases?

# Report Card for Domains: Content

| Domain | # | standard | class | source | Fitness | Complete-ness | Hierarchy | Duplication | Mapping | Overall |
|---|---|---|---|---|---|---|---|---|---|---|
| Drug | 6,392,739 | | | | | | | | | |
| Device | 834,979 | | | | | | | | | |
| Condition | 784,144 | | | | | | | | | |
| Procedure | 745,866 | | | | | | | | | |
| Observation | 572,093 | | | | | | | | | |
| Measurement | 383,227 | | | | | | | | | |
| Geography | 203,352 | | | | | | | | | |
| Spec Anatomic Site | 79,424 | | | | | | | | | |
| Meas Value | 65,852 | | | | | | | | | |
| Type Concept | 4,574 | | | | | | | | | |
| Unit | 3,706 | | | | | | | | | |
| Provider | 2,313 | | | | | | | | | |
| Regimen | 2,032 | | | | | | | | | |
| Race | 1,313 | | | | | | | | | |
| Visit | 1,269 | | | | | | | | | |
| Payer | 171 | | | | | | | | | |
| Cost | 51 | | | | | | | | | |
| Episode | 26 | | | | | | | | | |
| Condition Status | 22 | | | | | | | | | |
| Plan | 11 | | | | | | | | | |
| Sponsor | 6 | | | | | | | | | |
| Ethnicity | 2 | | | | | | | | | |

Are we covering the space?

# Report Card for Domains: Content

| Domain | # | standard class source | Fitness | Complete-ness | Hierarchy | Duplication | Mapping | Overall |
|---|---|---|---|---|---|---|---|---|
| Drug | 6,392,739 | | | | | | | |
| Device | 834,979 | | | | | | | |
| Condition | 784,144 | | | | | | | |
| Procedure | 745,866 | | | | | | | |
| Observation | 572,093 | | | | | | | |
| Measurement | 383,227 | | | | | | | |
| Geography | 203,352 | | | | | | | |
| Spec Anatomic Site | 79,424 | | | | | | | |
| Meas Value | 65,852 | | | | | | | |
| Type Concept | 4,574 | | | | | | | |
| Unit | 3,706 | | | | | | | |
| Provider | 2,313 | | | | | | | |
| Regimen | 2,032 | | | | | | | |
| Race | 1,313 | | | | | | | |
| Visit | 1,269 | | | | | | | |
| Payer | 171 | | | | | | | |
| Cost | 51 | | | | | | | |
| Episode | 26 | | | | | | | |
| Condition Status | 22 | | | | | | | |
| Plan | 11 | | | | | | | |
| Sponsor | 6 | | | | | | | |
| Ethnicity | 2 | | | | | | | |

Is the hierarchy robust and reliable?

# Report Card for Domains: Content

| Domain | # | standard class source | Fitness | Complete-ness | Hierarchy | Duplication | Mapping | Overall |
|---|---|---|---|---|---|---|---|---|
| Drug | 6,392,739 | | | | | | | |
| Device | 834,979 | | | | | | | |
| Condition | 784,144 | | | | | | | |
| Procedure | 745,866 | | | | | | | |
| Observation | 572,093 | | | | | | | |
| Measurement | 383,227 | | | | | | | |
| Geography | 203,352 | | | | | | | |
| Spec Anatomic Site | 79,424 | | | | | | | |
| Meas Value | 65,852 | | | | | | | |
| Type Concept | 4,574 | | | | | | | |
| Unit | 3,706 | | | | | | | |
| Provider | 2,313 | | | | | | | |
| Regimen | 2,032 | | | | | | | |
| Race | 1,313 | | | | | | | |
| Visit | 1,269 | | | | | | | |
| Payer | 171 | | | | | | | |
| Cost | 51 | | | | | | | |
| Episode | 26 | | | | | | | |
| Condition Status | 22 | | | | | | | |
| Plan | 11 | | | | | | | |
| Sponsor | 6 | | | | | | | |
| Ethnicity | 2 | | | | | | | |

Do we have one standard concept per entity?

# Report Card for Domains: Content

| Domain | # | standard class source | Fitness | Complete-ness | Hierarchy | Duplication | Mapping | Overall |
|---|---|---|---|---|---|---|---|---|
| Drug | 6,392,739 | | | | | | | |
| Device | 834,979 | | | | | | | |
| Condition | 784,144 | | | | | | | |
| Procedure | 745,866 | | | | | | | |
| Observation | 572,093 | | | | | | | |
| Measurement | 383,227 | | | | | | | |
| Geography | 203,352 | | | | | | | |
| Spec Anatomic Site | 79,424 | | | | | | | |
| Meas Value | 65,852 | | | | | | | |
| Type Concept | 4,574 | | | | | | | |
| Unit | 3,706 | | | | | | | |
| Provider | 2,313 | | | | | | | |
| Regimen | 2,032 | | | | | | | |
| Race | 1,313 | | | | | | | |
| Visit | 1,269 | | | | | | | |
| Payer | 171 | | | | | | | |
| Cost | 51 | | | | | | | |
| Episode | 26 | | | | | | | |
| Condition Status | 22 | | | | | | | |
| Plan | 11 | | | | | | | |
| Sponsor | 6 | | | | | | | |
| Ethnicity | 2 | | | | | | | |

Do we reliably map from source to standard?

# Report Card for Domains: Service

| Domain | # | standard | class | source | Urgency | Automa-tion | Crowd sourced | Documen-tation | Support | Overall |
|---|---|---|---|---|---|---|---|---|---|---|
| Drug | 6,392,739 | | | | | | | | | |
| Device | 834,979 | | | | | | | | | |
| Condition | 784,144 | | | | | | | | | |
| Procedure | 745,866 | | | | | | | | | |
| Observation | 572,093 | | | | | | | | | |
| Measurement | 383,227 | | | | | | | | | |
| Geography | 203,352 | | | | | | | | | |
| Spec Anatomic Site | 79,424 | | | | | | | | | |
| Meas Value | 65,852 | | | | | | | | | |
| Type Concept | 4,574 | | | | | | | | | |
| Unit | 3,706 | | | | | | | | | |
| Provider | 2,313 | | | | | | | | | |
| Regimen | 2,032 | | | | | | | | | |
| Race | 1,313 | | | | | | | | | |
| Visit | 1,269 | | | | | | | | | |
| Payer | 171 | | | | | | | | | |
| Cost | 51 | | | | | | | | | |
| Episode | 26 | | | | | | | | | |
| Condition Status | 22 | | | | | | | | | |
| Plan | 11 | | | | | | | | | |
| Sponsor | 6 | | | | | | | | | |
| Ethnicity | 2 | | | | | | | | | |

Do the use cases demand we fix our problems?

# Report Card for Domains: Service

| Domain | # | standard | class | source | Urgency | Automa-tion | Crowd sourced | Documen-tation | Support | Overall |
|---|---|---|---|---|---|---|---|---|---|---|
| Drug | 6,392,739 | | | | | | | | | |
| Device | 834,979 | | | | | | | | | |
| Condition | 784,144 | | | | | | | | | |
| Procedure | 745,866 | | | | | | | | | |
| Observation | 572,093 | | | | | | | | | |
| Measurement | 383,227 | | | | | | | | | |
| Geography | 203,352 | | | | | | | | | |
| Spec Anatomic Site | 79,424 | | | | | | | | | |
| Meas Value | 65,852 | | | | | | | | | |
| Type Concept | 4,574 | | | | | | | | | |
| Unit | 3,706 | | | | | | | | | |
| Provider | 2,313 | | | | | | | | | |
| Regimen | 2,032 | | | | | | | | | |
| Race | 1,313 | | | | | | | | | |
| Visit | 1,269 | | | | | | | | | |
| Payer | 171 | | | | | | | | | |
| Cost | 51 | | | | | | | | | |
| Episode | 26 | | | | | | | | | |
| Condition Status | 22 | | | | | | | | | |
| Plan | 11 | | | | | | | | | |
| Sponsor | 6 | | | | | | | | | |
| Ethnicity | 2 | | | | | | | | | |

How well can we bring in vocabularies and refresh?

# Report Card for Domains: Service

| Domain | # | standard | class | source | Urgency | Automa-tion | Crowd sourced | Documen-tation | Support | Overall |
|---|---|---|---|---|---|---|---|---|---|---|
| Drug | 6,392,739 | | | | | | | | | |
| Device | 834,979 | | | | | | | | | |
| Condition | 784,144 | | | | | | | | | |
| Procedure | 745,866 | | | | | | | | | |
| Observation | 572,093 | | | | | | | | | |
| Measurement | 383,227 | | | | | | | | | |
| Geography | 203,352 | | | | | | | | | |
| Spec Anatomic Site | 79,424 | | | | | | | | | |
| Meas Value | 65,852 | | | | | | | | | |
| Type Concept | 4,574 | | | | | | | | | |
| Unit | 3,706 | | | | | | | | | |
| Provider | 2,313 | | | | | | | | | |
| Regimen | 2,032 | | | | | | | | | |
| Race | 1,313 | | | | | | | | | |
| Visit | 1,269 | | | | | | | | | |
| Payer | 171 | | | | | | | | | |
| Cost | 51 | | | | | | | | | |
| Episode | 26 | | | | | | | | | |
| Condition Status | 22 | | | | | | | | | |
| Plan | 11 | | | | | | | | | |
| Sponsor | 6 | | | | | | | | | |
| Ethnicity | 2 | | | | | | | | | |

How much is the community involved?

# Report Card for Domains: Service

| Domain | # | standard | class | source | Urgency | Automa-tion | Crowd sourced | Documen-tation | Support | Overall |
|---|---|---|---|---|---|---|---|---|---|---|
| Drug | 6,392,739 | | | | | | | | | |
| Device | 834,979 | | | | | | | | | |
| Condition | 784,144 | | | | | | | | | |
| Procedure | 745,866 | | | | | | | | | |
| Observation | 572,093 | | | | | | | | | |
| Measurement | 383,227 | | | | | | | | | |
| Geography | 203,352 | | | | | | | | | |
| Spec Anatomic Site | 79,424 | | | | | | | | | |
| Meas Value | 65,852 | | | | | | | | | |
| Type Concept | 4,574 | | | | | | | | | |
| Unit | 3,706 | | | | | | | | | |
| Provider | 2,313 | | | | | | | | | |
| Regimen | 2,032 | | | | | | | | | |
| Race | 1,313 | | | | | | | | | |
| Visit | 1,269 | | | | | | | | | |
| Payer | 171 | | | | | | | | | |
| Cost | 51 | | | | | | | | | |
| Episode | 26 | | | | | | | | | |
| Condition Status | 22 | | | | | | | | | |
| Plan | 11 | | | | | | | | | |
| Sponsor | 6 | | | | | | | | | |
| Ethnicity | 2 | | | | | | | | | |

How good is it for ETLers and Analysts?

# Report Card for Domains: Service

| Domain | # | standard class source | Urgency | Automa-tion | Crowd sourced | Documen-tation | Support | Overall |
|---|---|---|---|---|---|---|---|---|
| Drug | 6,392,739 | | | | | | | |
| Device | 834,979 | | | | | | | |
| Condition | 784,144 | | | | | | | |
| Procedure | 745,866 | | | | | | | |
| Observation | 572,093 | | | | | | | |
| Measurement | 383,227 | | | | | | | |
| Geography | 203,352 | | | | | | | |
| Spec Anatomic Site | 79,424 | | | | | | | |
| Meas Value | 65,852 | | | | | | | |
| Type Concept | 4,574 | | | | | | | |
| Unit | 3,706 | | | | | | | |
| Provider | 2,313 | | | | | | | |
| Regimen | 2,032 | | | | | | | |
| Race | 1,313 | | | | | | | |
| Visit | 1,269 | | | | | | | |
| Payer | 171 | | | | | | | |
| Cost | 51 | | | | | | | |
| Episode | 26 | | | | | | | |
| Condition Status | 22 | | | | | | | |
| Plan | 11 | | | | | | | |
| Sponsor | 6 | | | | | | | |
| Ethnicity | 2 | | | | | | | |

How easy is it to get them ETLed/mapped?

# Report Card for Overall Technology

| | | |
|---|---|---|
| Release Management | | 3 |
| Build Process | | 4 |
| Documentation | | 2 |
| Dashboard | | 2 |
| Cookbook | | 2 |
| Access Automation | | 2 |

# OHDSI Standardized Vocabularies play a key role in our mission

A lot more has to be done