Update on the OHDSI Standardized Vocabularies

03-September-2024

Alexander Davydov
Masha Khitrun
Tanya Skugarevskaya
Anna Ostropolets

Overview

Alexander Davydov

Masha Khitrun

Anna Ostropolets

Tanya Skugarevskaya

Release highlights

Vocabulary maintenance

Improvement highlights

Reporting improvements

Release overview & highlights

The team

Dmitry Dymshyts
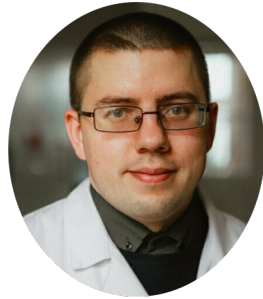
Aliaksei Katyshou

Vlad Korsik

Oleg Zhuk

Maria Rogozhkina

Mikita Salavei

Varvara Savitskaya

Irina Zherka

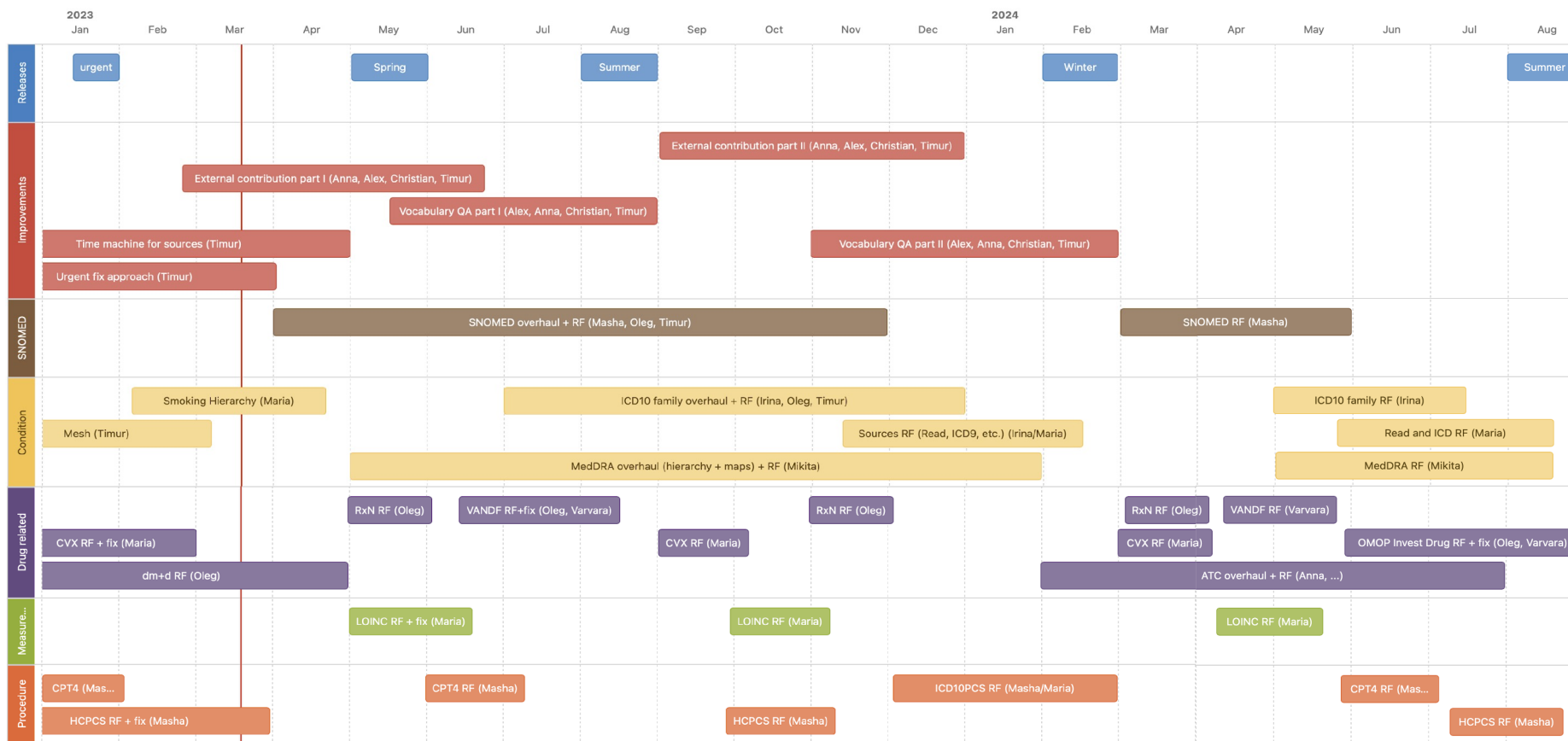Polina Talapova

Tetiana Orlova

Anton Tatur

Janice Cruz

Roadmap and releases

- Summer release **v20240830** has just arrived to Athena

- New roadmap planning in progress, stay tuned!

Vocabulary versioning in Athena

- You can pick a version (release) when downloading

- You can generate delta when downloading

- Memorial Sloan Kettering Cancer Center sponsored solution

Vocabulary versioning in Athena

- E-mail with installation instructions and a download link

Vocabulary versioning in Athena

- SQL script output

```
INSERT INTO concept (concept_name, domain_id, vocabulary_id, concept_class_id, standard_concept, concept_code, valid_start_date,
valid_end_date, invalid_reason, concept_id) VALUES ('Subject of (SNOMED)', 'Metadata', 'Relationship', 'Relationship', NULL, 'OMOP
generated', '1970-01-01', '2099-12-31', NULL, 33105);
UPDATE concept SET  (concept_name, domain_id, vocabulary_id, concept_class_id, standard_concept, concept_code, valid_start_date,
valid_end_date, invalid_reason) = ('CO-CONNECT (University of Nottingham)', 'Metadata', 'Vocabulary', 'Vocabulary', NULL, 'OMOP generated',
'1970-01-01', '2099-12-31', NULL) WHERE concept_id=33091;
UPDATE concept SET  (concept_name, domain_id, vocabulary_id, concept_class_id, standard_concept, concept_code, valid_start_date,
valid_end_date, invalid_reason) = ('CO-CONNECT MIABIS (University of Nottingham)', 'Metadata', 'Vocabulary', 'Vocabulary', NULL, 'OMOP
generated', '1970-01-01', '2099-12-31', NULL) WHERE concept_id=33092;
UPDATE concept SET  (concept_name, domain_id, vocabulary_id, concept_class_id, standard_concept, concept_code, valid_start_date,
valid_end_date, invalid_reason) = ('CO-CONNECT TWINS (University of Nottingham)', 'Metadata', 'Vocabulary', 'Vocabulary', NULL, 'OMOP
generated', '1970-01-01', '2099-12-31', NULL) WHERE concept_id=33093;
```

Vocabulary versioning in Athena

- Table format output
- Generated for each table

row_change_type	attribute_modified	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id
U	domain_id	9173	Inactive	Meas Value	SNOMED	Qualifier Value
U	domain_id	9174	Obsolete	Meas Value	SNOMED	Qualifier Value
U	concept_class_id	72245	Chlamydial infection of lower genitourinary tract	Condition	SNOMED	Disorder
U	concept_name					
U	concept_class_id	72332	Serous labyrinthitis	Condition	SNOMED	Disorder
I		33100	Process acts on (SNOMED)	Metadata	Relationship	Relationship
I		33101	Affected by process (SNOMED)	Metadata	Relationship	Relationship
I		33102	Before (SNOMED)	Metadata	Relationship	Relationship
I		33103	After (SNOMED)	Metadata	Relationship	Relationship
I		33104	Towards (SNOMED)	Metadata	Relationship	Relationship
I		33105	Subject of (SNOMED)	Metadata	Relationship	Relationship

New synonym panel

ΛTHENΛ

← 1,1-Dichloroethylene [Mass/volume] in Air

DETAILS

Domain ID	Measurement
Concept Class ID	Lab Test
Vocabulary ID	LOINC ⑦
Concept ID	3048663
Concept code	38735-7
Validity	Valid
Concept	Standard

LANGUAGE	SYNONYM CONCEPT
English	1,1-DCE Air-mCnc
English	1,1-DCE; 1,1-Dichloroethene; Asym-dichloro-ethylene; DCE; DRUG/TOXICOLOGY; Drugs; i; Level; Mass concentration; Point in time; QNT; Quan; Quant; Quantitative; Random; Vinylidene; Vinylidine
Chinese	1,1-二氯乙烯:质量浓度:时间点:空气:定量型

Valid start	01-Jan-1970
Valid end	31-Dec-2099

Metadata in Vocabularies
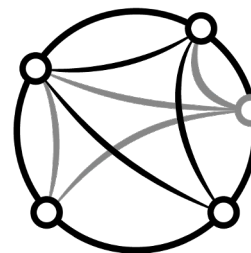
SSSOM
SIMPLE STANDARD FOR SHARING
ONTOLOGY MAPPINGS
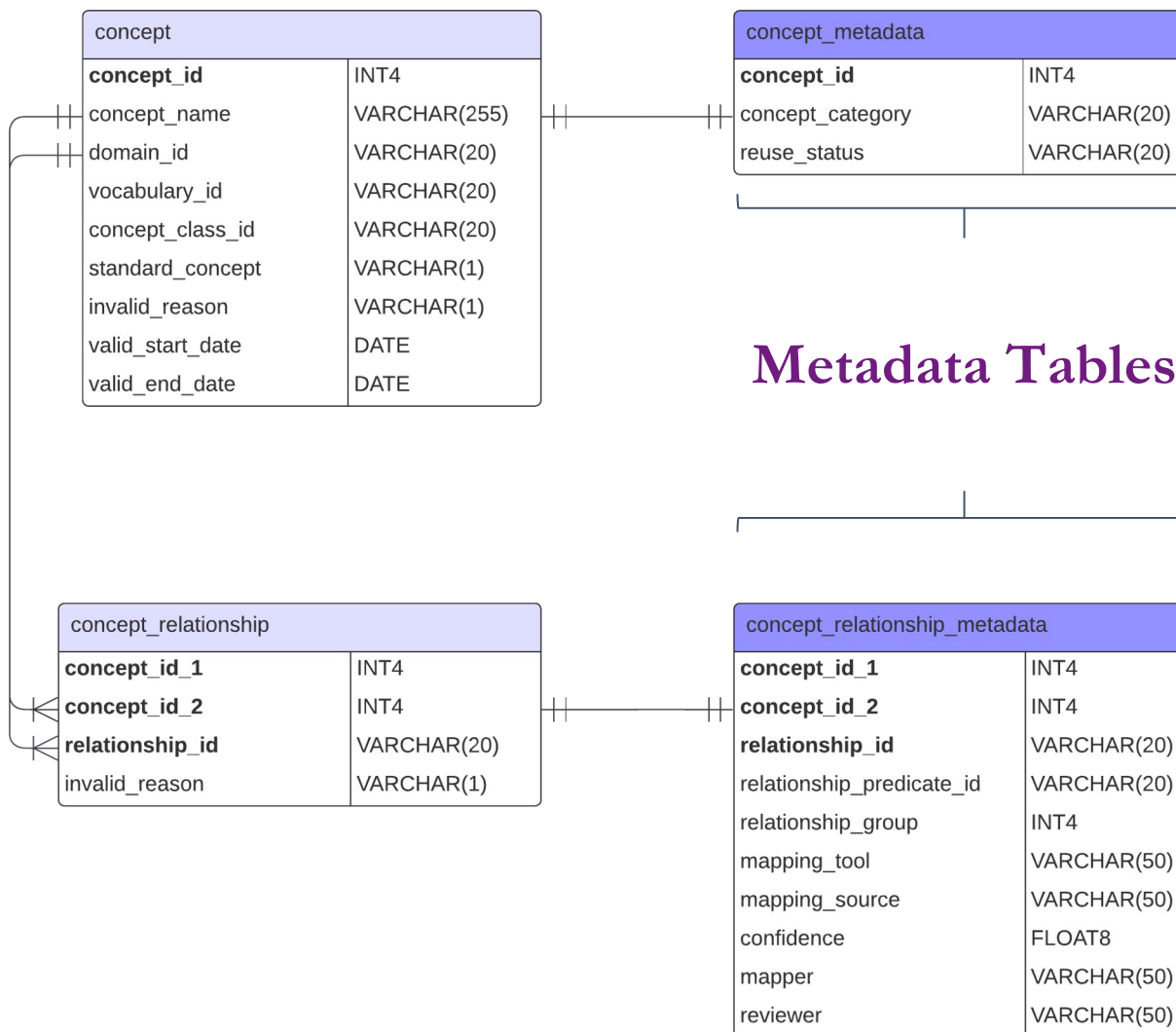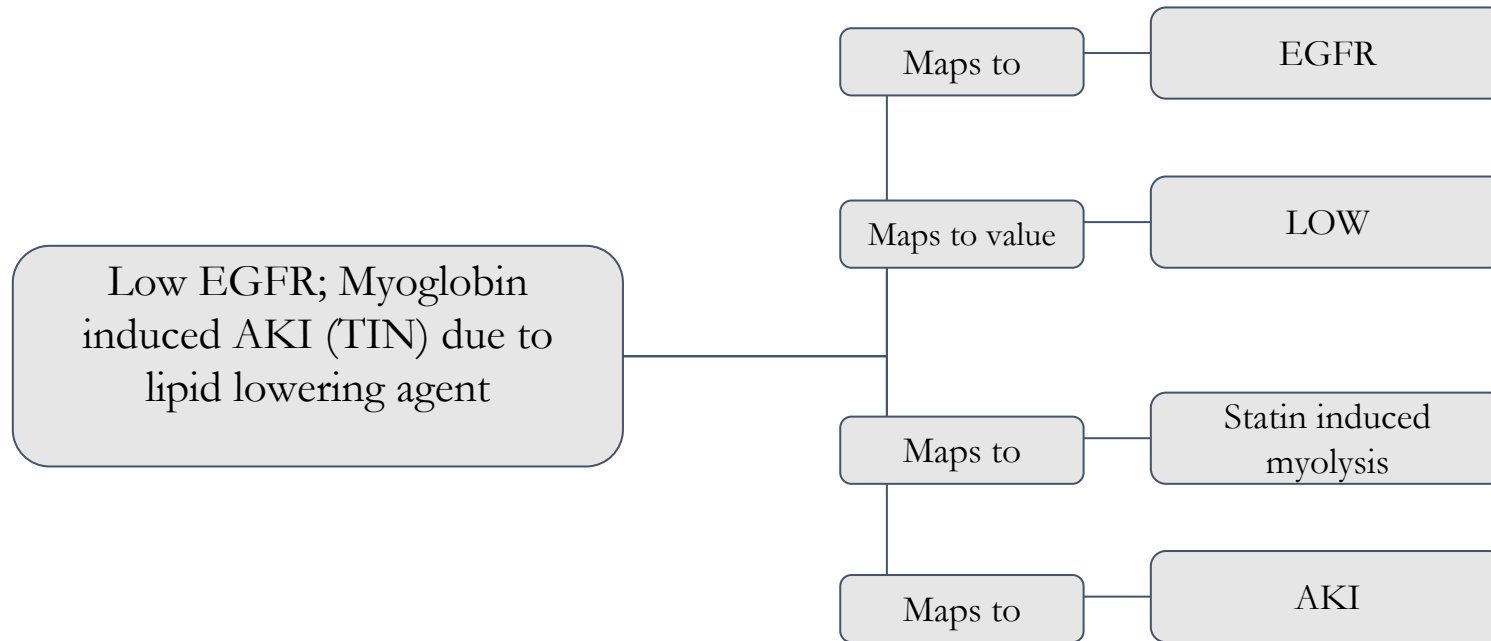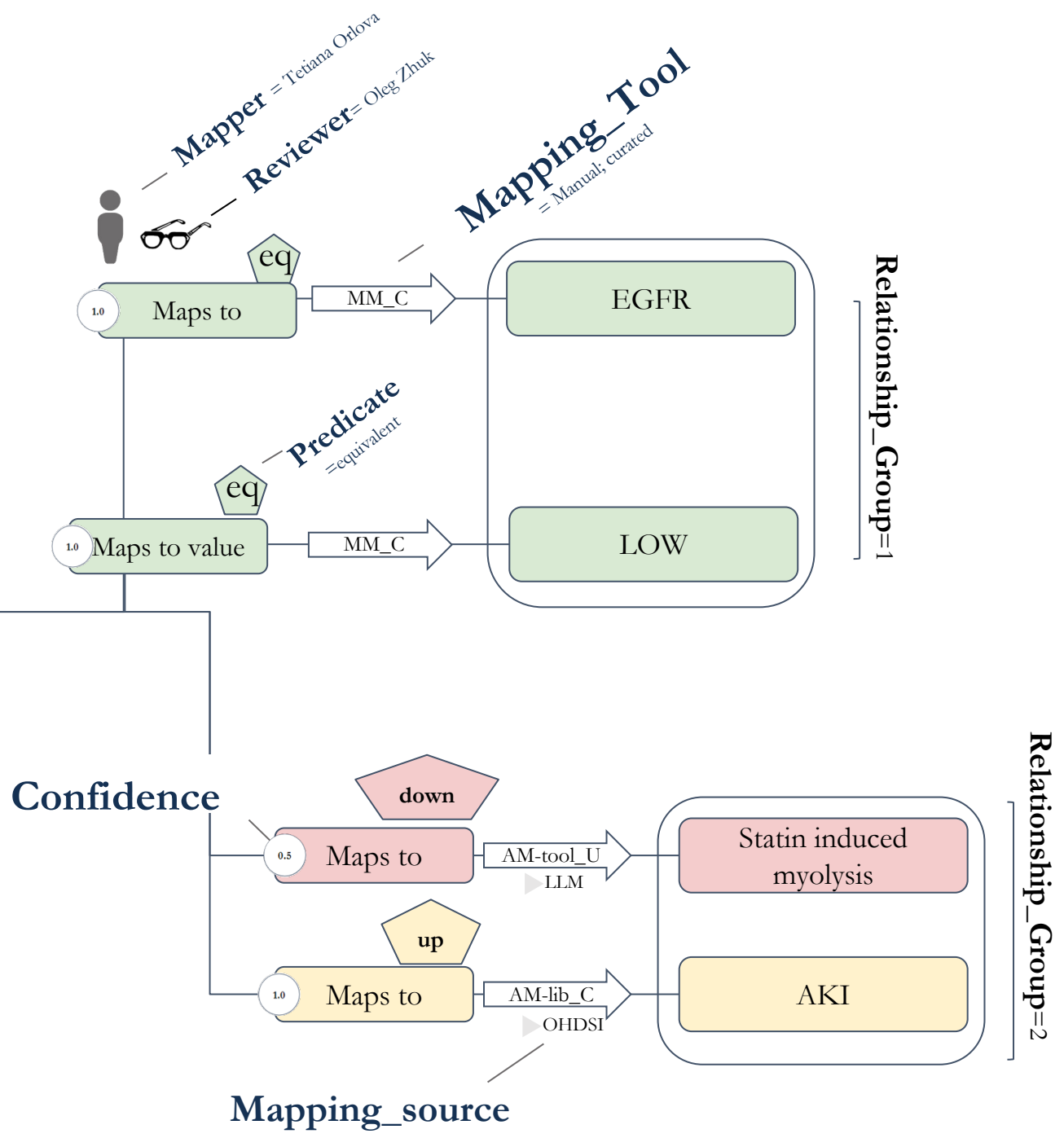
concept

concept_id	INT4
concept_name	VARCHAR(255)
domain_id	VARCHAR(20)
vocabulary_id	VARCHAR(20)
concept_class_id	VARCHAR(20)
standard_concept	VARCHAR(1)
invalid_reason	VARCHAR(1)
valid_start_date	DATE
valid_end_date	DATE

concept_metadata

concept_id	INT4
concept_category	VARCHAR(20)
reuse_status	VARCHAR(20)

Metadata Tables

concept_relationship

concept_id_1	INT4
concept_id_2	INT4
relationship_id	VARCHAR(20)
invalid_reason	VARCHAR(1)

concept_relationship_metadata

concept_id_1	INT4
concept_id_2	INT4
relationship_id	VARCHAR(20)
relationship_predicate_id	VARCHAR(20)
relationship_group	INT4
mapping_tool	VARCHAR(50)
mapping_source	VARCHAR(50)
confidence	FLOAT8
mapper	VARCHAR(50)
reviewer	VARCHAR(50)

https://github.com/OHDSI/Vocabulary-v5.0/wiki/Vocabulary-Metadata

Relationships before metadata

Concept Relationship Metadata

Concept Metadata

devv5.concept_metadata

123	**concept_id**	int4
A-Z	concept_category	varchar(20)
A-Z	reuse_status	varchar(20)

reuse_status

concept_category

- **Indicated the code reuse (serious change of code semantics over time)**

Permissible values are:
- **R - Designates that code was reused;**
- no information available on temporal context of Reuse; state of name in concept table is undetermined;

- RF - Designates that code is Reused in Future; current state of concept in concept table displays historically valid attributes (name, validity dates) provided by source

- RP - Designates that code is Reused in Past; current state of concept in concept table displays actual valid attributes (name, validity dates) provided by source

Descriptor of expected behaviour of code within OMOP CDM

Permissible values are:

- **J - J**unk codes that either not OHDSI usecase or flavor of NULL
- **M** - **M**etadata concept
- **A - A**ttributive concept not designed to crete any Event
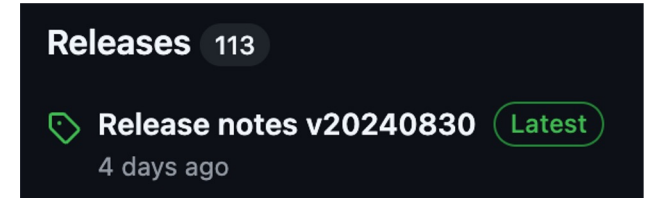- *NULL* - regular concept

Where to find this content

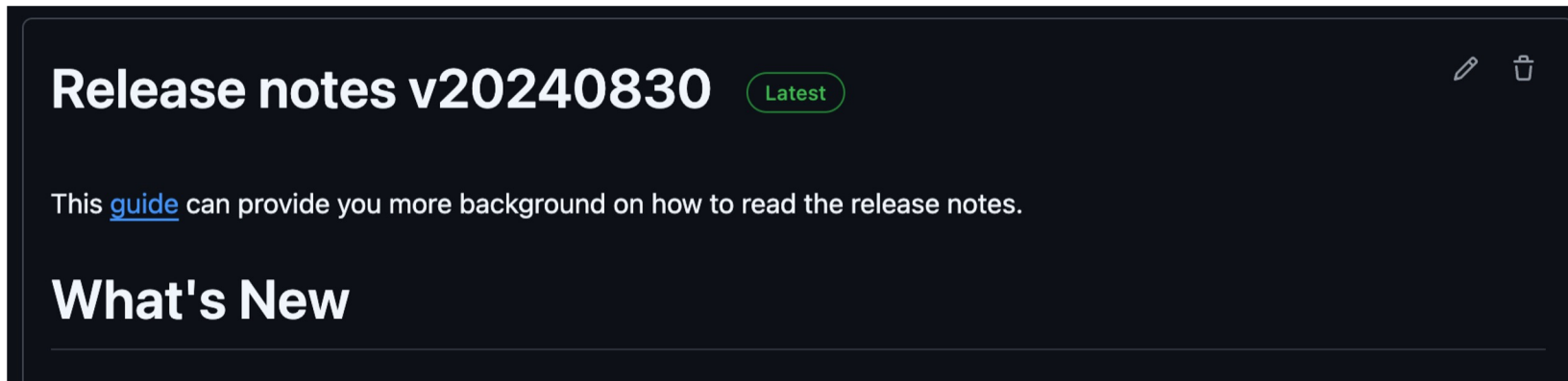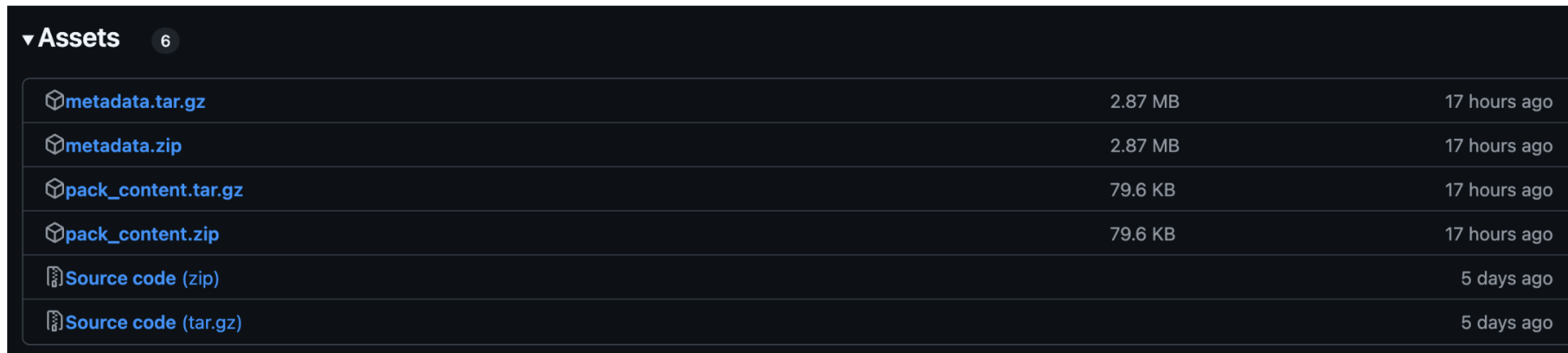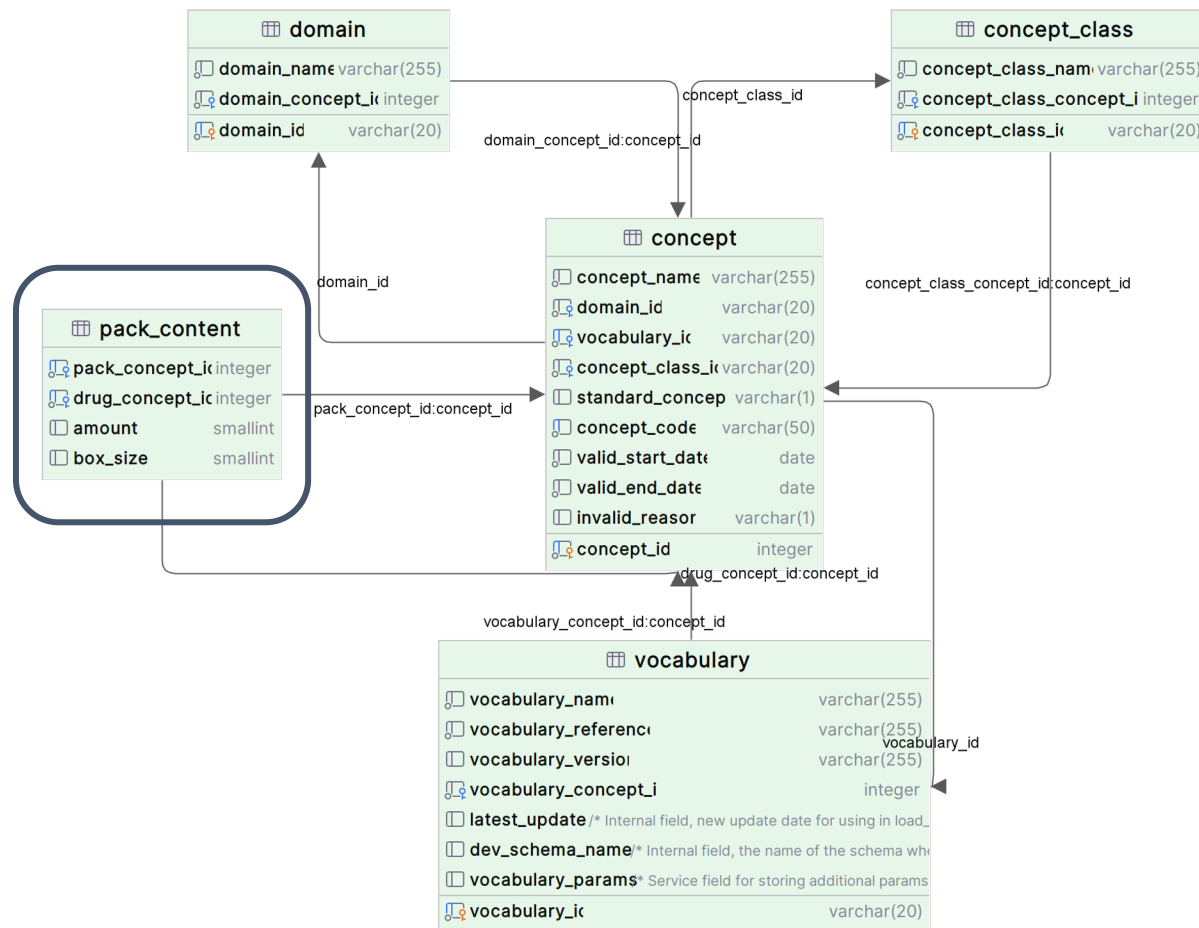1. **OHDSI**

2. **Vocabulary-v5.0** Public

3. Releases 113
   Release notes v20240830 Latest
   4 days ago

4. **Release notes v20240830** Latest

   This guide can provide you more background on how to read the release notes.

   What's New

5. ▼Assets 6

metadata.tar.gz	2.87 MB	17 hours ago
metadata.zip	2.87 MB	17 hours ago
pack_content.tar.gz	79.6 KB	17 hours ago
pack_content.zip	79.6 KB	17 hours ago
Source code (zip)		5 days ago
Source code (tar.gz)		5 days ago

https://github.com/OHDSI/Vocabulary-v5.0/releases

GitHub Drug Tables release

Vocabularies maintenance

Release scope

Vocabularies

CDISC

ICD10-family	SNOMED
Read	MedDRA
CPT4	LOINC
HCPCS	Rx/RxE
VANDF	CVX
OMOP InvDrug	ATC
NDC	OMOP Ext
+ [SOURCES]	CMS Place of Service

Community Contribution

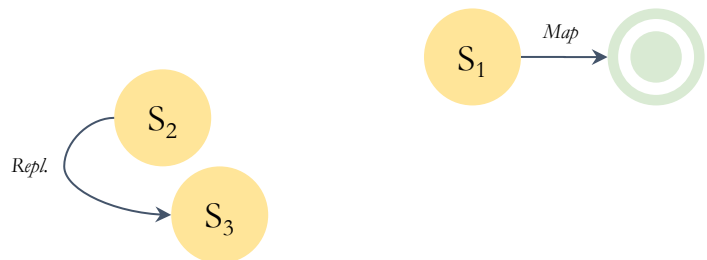| EDI	ICD10PCS (1)	SNOMED (1)
HCPCS (1)	ICD10CM (2)	PPI (3)

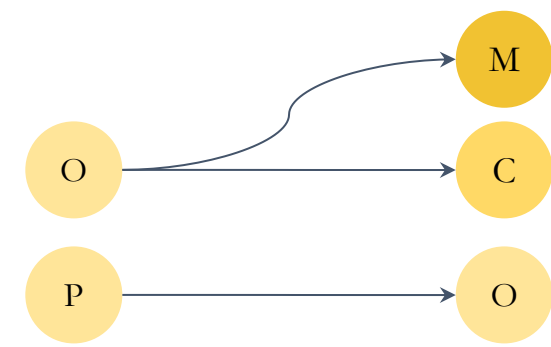# SNOMED. Refresh + adjustments
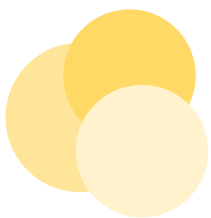
Int.: 2024-02-01
US: 2024-03-01
UK: 2024-04-10

+137 curated mappings
+1,012 replacements by SNOMED
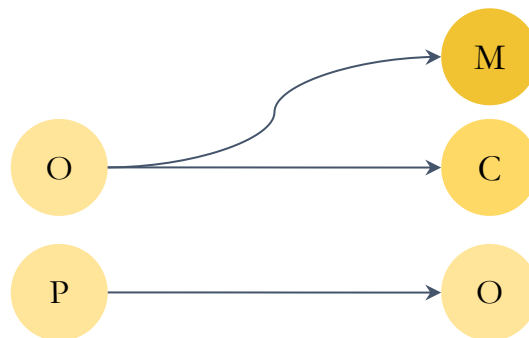-576 Standard codes, no mappings
provided by SNOMED

406 standard codes
change domain

According to the discussion around the post-coordination of pre-coordinated SNOMED measurements, **we do not perform any further post-coordination**.

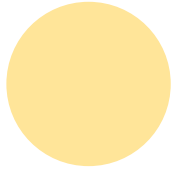# SNOMED. Refresh + adjustments

406 standard codes
change domain

- Anaphylaxis reactions
- Other allergic reactions

- Body measurements
- Measured quantity of intake

- Consultations

concept_name	old_domain_id	new_domain_id
Allergic reaction caused by food additive	Observation	Condition
Anaphylaxis caused by mollusk	Observation	Condition
Measured quantity of intake of amino acids in 24 hours	Observation	Measurement
Lactate in umbilical cord blood	Observation	Measurement
Midwife consultation	Procedure	Observation

ICD family. Refresh + adjustments
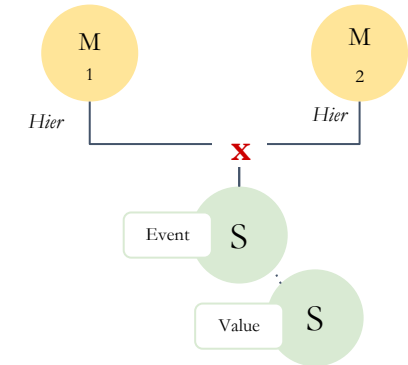
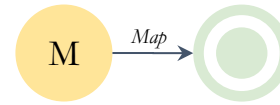ICD10CM version FY2025
(October 01, 2024)

Maps to

I
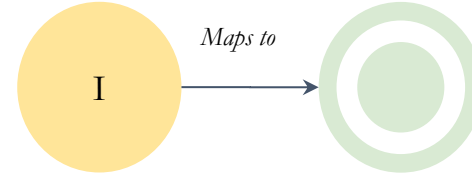
- ○ ICD10
- ○ ICD10CM
- ○ ICD9CM
- ○ ICD10GM
- ○ ICD10CN
- ○ CIM10
- ○ KCD7

We automatically updated the mappings and curated them
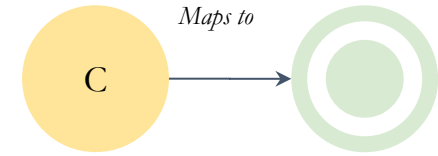
MedDRA. Refresh + adjustments
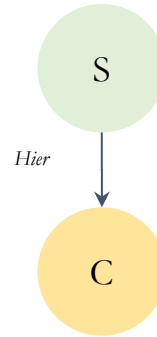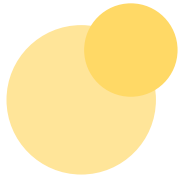

Version 27 (March 2024)
+ 754 new codes

+~2,891 PT mappings
+~215 LLT mappings

Fixed hierarchy bugs

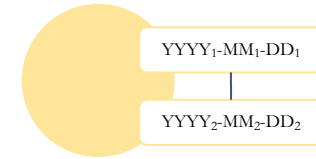# CPT4. Refresh + adjustments

2024AA UMLS
+ 281 codes
+ AMA immunization
codes

+ 1,164 OHDSI-curated
hierarchical relationships
(incl. all **Anesthesia**
codes)

+ 25 curated mappings
of new concepts

HCPCS. Refresh + adjustments

July 2024 version
+ 131 concepts

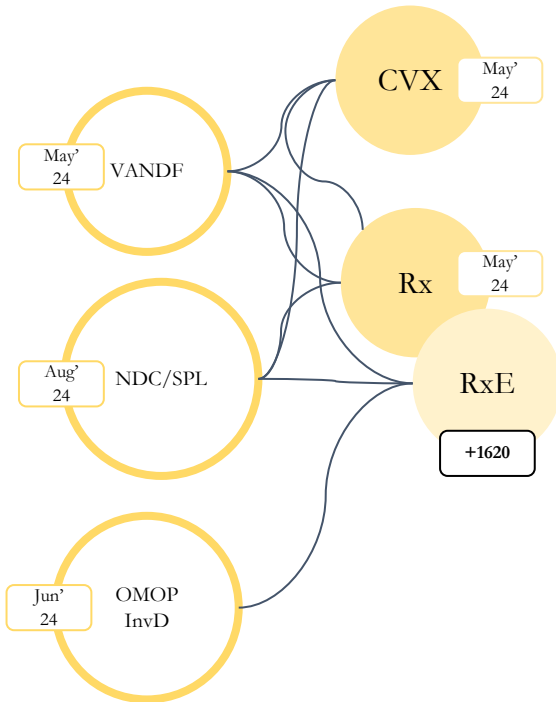+ 70 OHDSI-curated hierarchical rels
+ 62 mappings
+ improved mappings for 27 drugs

Validity dates manual adjustments for reused codes

Drug vocabularies. Refresh

- RxNorm - May 2024 version

- CVX - May 2024 version

- OMOP Invest Drug   +1,640 new concepts

- RxNorm Extension +1,620 new Ingredients from OMOP Invest Drug

- VANDF - June 2024 version

- NDC and SPL - Aug 2024 version (automation)

Other vocabularies

- **LOINC:** refresh, version 2.77; 230 concepts added to LOINC-SNOMED hierarchy

- **Read:** mapping refresh; 600 concepts got lost "Maps to value" restored

- **OMOP Extension**: 13 concepts from different clinical areas added

- **Visit** and **CMS Place of Service** refresh

- **UCUM:** added new units

Read release notes

CDISC* (a SDTM dictionary) introduction

General Data Flow Diagram for CDISC to OMOP
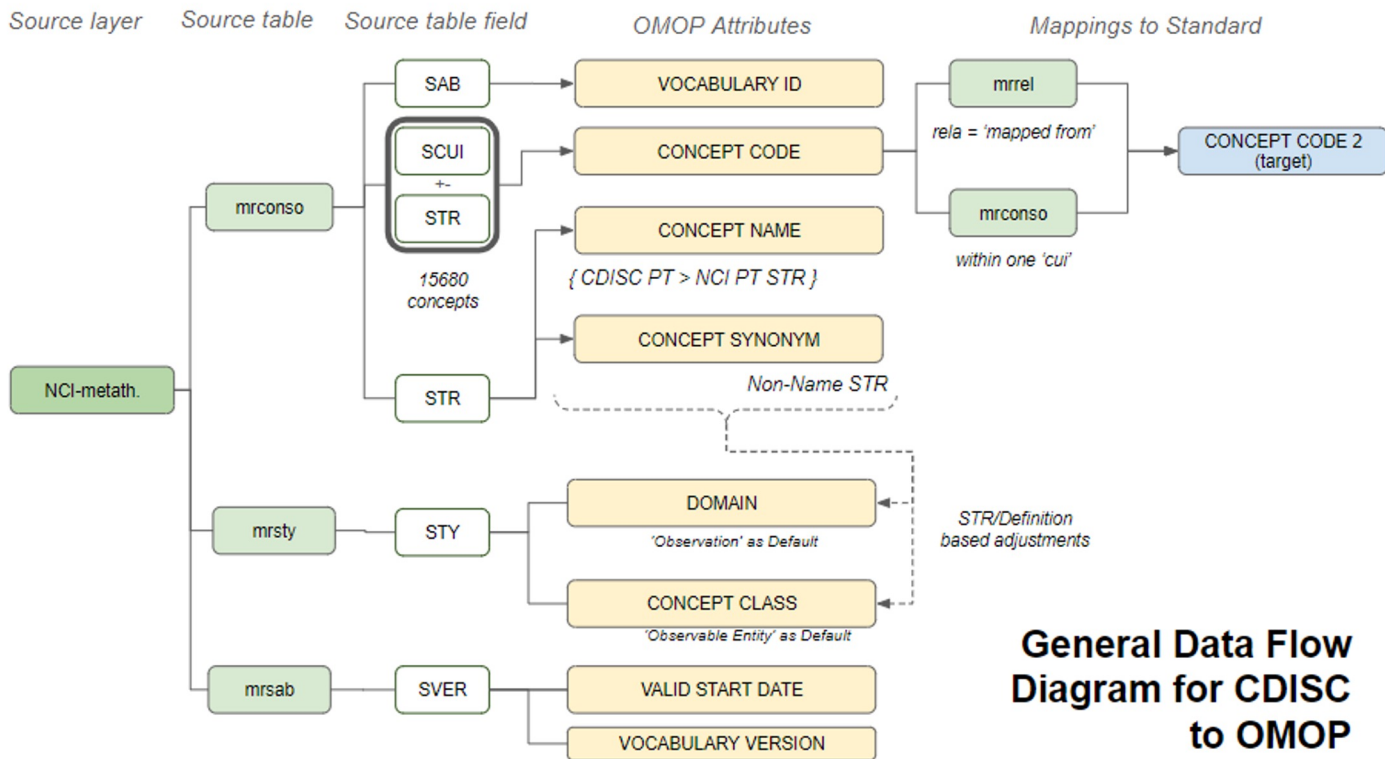
- + 29k of new concepts
- 5k (17%) of them mapped

* - Clinical Data Interchange Standards Consortium

Community contributions across the releases

Aug 2023

2 simple

1 complex

Feb 2023

8 simple

1 complex

Aug 2024

6 simple

1 complex

Community contribution in Aug release: complex contribution

Community contribution in Aug release: simple contributions

#	GitHub	What
1	944	Update ICD-10 CM G56.0x mapping to SNOMED
2	978	Injection, cabotegravir and rilpivirine, 2mg/3mg fixing mapping
3	979	C83.0*, C83.1*, C91.1*, D03.*, F03.9*, O99.512 ICD10CM and other concepts mapping improvement
4	980	XW03*, XW04* ICD10PCS mapping improvement
8	989	Typo in vocabulary: Concept code 1585652
5	999	Incorrectly labeled codes (PPI)
6	1001	Concept Class ID correction (PPI)

Thank you to all contributors!

Vocabularies improvements

Vocabularies improvements

- ATC

- Quality assessment and reports

Old ATC vs New ATC

Old ATC hierarchy
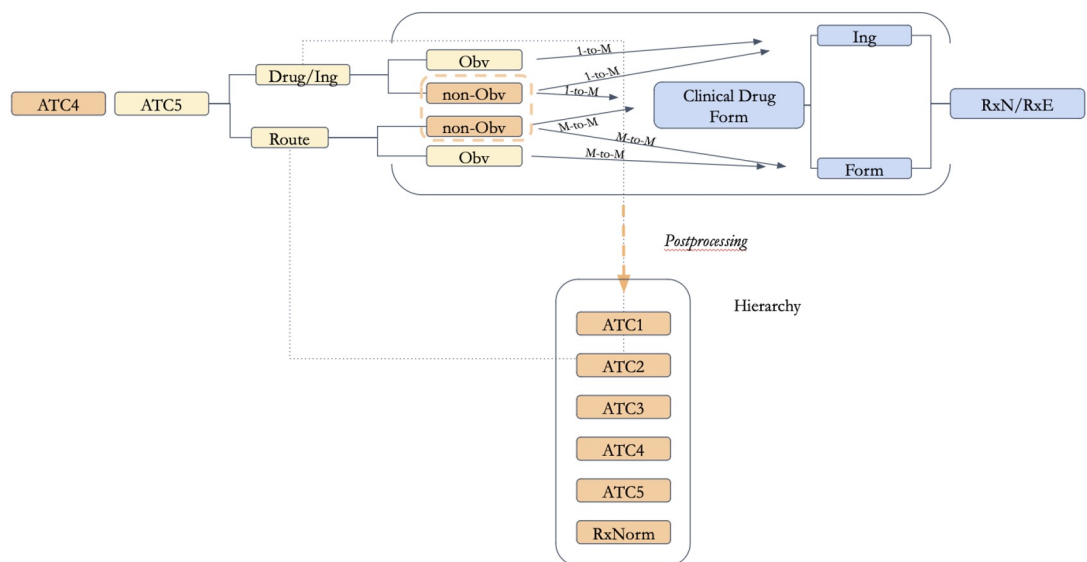
New ATC hierarchy
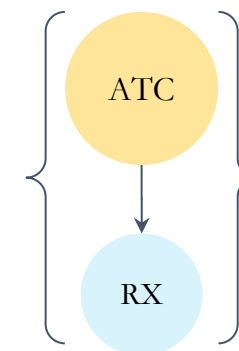
New ATC: how we got there

1. New Source Scraper
   - Weekly content ingestion
   - Automated assign of lifecycle attributes
2. Incorporation of new data sources
3. ETL of Norwegian Drug Database
4. Curation of links based on prevalence in the network

New ATC: what we need

1. More international data sources

1. Your feedback

Vocabularies improvements

- ATC

- Quality assessment and reports

Vocabulary principles and reporting

JOURNAL ARTICLE

OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization 🔓

Christian Reich, MD ✉, Anna Ostropolets, PhD, Patrick Ryan, PhD, Peter Rijnbeek, PhD, Martijn Schuemie, PhD, Alexander Davydov, MD, Dmitry Dymshyts, MD, George Hripcsak, MD

📄 PDF   ▊▊ Split View   66 Cite   🔑 Permissions   ◄ Share ▾

https://github.com/OHDSI/Vocabulary-v5.0/wiki/Introduction

Principles

The Standardized Vocabularies are constructed with a few principles in mind. Not every principle has been executed to perfection, but it represents a general motivation and direction of the ongoing improvement and development process:

1. **OHDSI Use Cases Focused**:
   OHDSI Vocabularies serve OMOP CDM and are driven by the use cases of storing and analyzing patient-level data for evidence generation. Storing data for other purposes such as keeping track of clinical transactions is outside of scope.
2. **Persistency**:
   Nothing gets deleted from the Vocabularies tables.
3. **Openness**:
   OHDSI Vocabularies are maintained as open-source and support transparency in generation processes.
4. **Unique Standard Concepts**:
   For each Clinical Entity there is only one concept representing it, called the Standard Concept. Other equivalent or similar concepts are designated non-Standard and mapped to the Standard ones.
5. **Pre-defined Domains**:
   Each concept is assigned one of a list of predefined domains (defined by OHDSI). "Dirty", i.e., not well-defined concepts, which are mostly non-Standard, can also belong to more than one domain. The assigned domain also defines in which CDM table a clinical entity should be placed into or looked up in at query time.
6. **Comprehensive Coverage**:
   Every event in a patient's healthcare experience (e.g. Conditions, Procedures, Exposures to Drug, etc.) and some of the administrative artifacts of the healthcare system (e.g. Visits, Care Sites, etc.) are covered within the domain. This principle includes concept coverage and mapping coverage (mappings from non-standard concepts to standard concepts).
7. **Polyhierarchy**:

https://github.com/OHDSI/Vocabulary-v5.0/wiki/Quality-assurance-and-control

Four key vocabulary principles

- Unique standard concept

- Comprehensive coverage

- Pre-defined domains

- Polyhierarchy

Vocabulary principles reporting

List of the principles

outlining the
principles we follow

General report

on the main principles and
their state in Feb-2024 and
Aug-2024 releases

Detailed reports

for separate principles

Code

for generating reports
is on GitHub
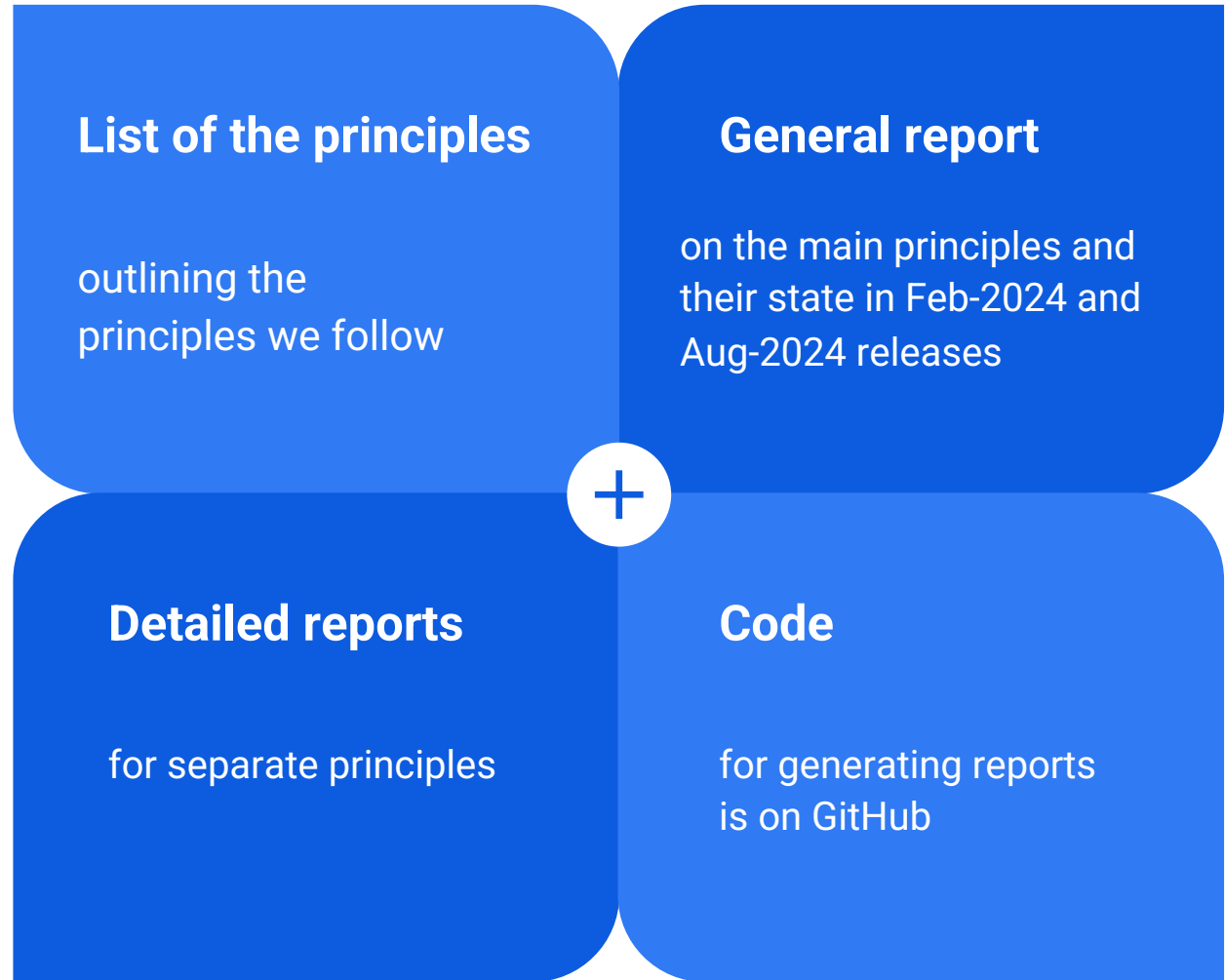
Vocabulary principles reporting example

Upcoming changes and release notes

- **Upcoming changes (before the release)**

How it will impact you + more detailed view on the github

August 2024 release upcoming changes

◼ Vocabulary Users

aostropolets Anna Ostropolets   1 ✎ 17d

Please check out GitHub Upcoming Changes 10 section to learn about what will be implemented in the upcoming August release.

TLDR:

The release will provide you with a new vocabulary for clinical trial data, more new terms, more mappings and more informations about the mappings (meta-data table). No significant impact on ETL/phenotyping processes is expected.

A longer version:

- A brand-new table capturing concept and relationship meta-data will be distributed;
- pack_content (table that stores drug pack info) will be distributed through Github;
- A new vocabulary (CDISC) will be added;
- Several source vocabularies will be refreshed (GitHub provides specific versions): ATC, LOINC, SNOMED, MedDRA, HCPCS, CPT4, CMS Place of Service, ICD10CM and EDI
- Several adjustments were made to SNOMED concepts' domain;
- More mappings were added for MedDRA, HCPCS, CPT4, ICD10PCS, ATC and ICD condition

nships as we changed our approach for ATC-
m US and international drug vocabularies and data

Upcoming changes

Oleg Zhuk edited this page 3 weeks ago · 10 revisions

2024 August Release

1. OMOP Structural changes overview

Vocabulary modeling overview

Pre-/Post-coordination migration

According to the discussion around the post-coordination of pre-coordinated SNOMED measurements, the Vocabulary Team **does not perform** any further post-coordination work.

Canonical vocabulary tables

Pack_content table, unavailable in Athena, will be available via the Vocabulary-v5.0 github repository.

New tables

Upcoming changes and release notes

- **Upcoming changes (before the release)**

How it will impact you + more detailed view on the github

- **Release Notes (with the release)**

Detailed description and statistics

Release notes v20240830 Latest

This guide can provide you more background on how to read the release notes.

What's New

1. Structural changes

- *Pack_content* table is now available via the Vocabulary-v5.0 GitHub repository (link will be added soon).
- We deliver the SSSOM-compatible metadata for certain relationships (link will be added soon). The metadata mimics the one we collect during the community contribution process and contains the following information:
  - Relationship metadata
  - Concept metadata

d, and the first release of

Standard concept changes

vocabulary_id	old_standard_concept	new_standard_concept	count
ATC	Classification with mapping	Non-standard with mapping	78
ATC	Classification without mapping	Non-standard with mapping	7
ATC	Classification without mapping	Non-standard without mapping	2
CMS Place of Service	Standard	Non-standard with mapping	1
CPT4	Standard	Non-standard with mapping	160
HCPCS	Non-standard without mapping	Standard	5
ICD10PCS	Standard	Non-standard with mapping	10
LOINC	Standard	Non-standard with mapping	11
LOINC	Classification without mapping	Non-standard without mapping	6
LOINC	Non-standard without mapping	Classification without mapping	1
MedDRA	Non-standard without mapping	Classification with mapping	3103
MedDRA	Classification with mapping	Non-standard with mapping	1
RxNorm	Standard	Non-standard without mapping	338

Upcoming changes and release notes

- **Upcoming changes (before the release)**

How it will impact you + more detailed view on the github

- **Release Notes (with the release)**

Detailed description and statistics

- **Quality reporting (shortly after the release)**

Adherence to principles

Why you should download this vocabulary release

- **More concepts**:
  - Refresh of SNOMED, CPT4, HCPCS, MedDRA, ICD10CM, LOINC, CVX, RxNorm and more

- **Better hierarchies**:
  - ATC, SNOMED, HCPCS, CPT4,LOINC, MedDRA

- **More good mappings**:
  - ICD family refresh
  - community contributions & bug fixing

- **Totally new features:**
  - Meta-data tables
  - Pack_content table