



Update on OHDSI Standardized Vocabularies

29-Aug-2023

Alexander Davydov
Anna Ostroplets
Oleg Zhuk
Christian Reich
& Vocabulary Team



Overview

1. Recap of the plan for Vocabularies improvement

2. Accomplishments so far

- New process & Vocabulary Committee
- Landscape assessment
- Roadmap and releases
- Community contribution
- August release
- Experience with community contribution
- Vocabulary WG
- Quality management system

3. What is missing, issues faced and ways to improvement



Recap of Vocabularies improvement:
What was planned?



Vocabularies process improvement

1. Landscape assessment of community needs
2. Roadmap and release schedule
3. Documentation & guidelines (developer, end-user, community contribution)
4. Quality framework and QA system
5. Vocabulary versioning
6. Common development infrastructure
7. Better code and automation (better machinery)
8. Vocabulary one-stop-shop and self-service



Vocabulary Committee

Committee and improvement activities



- Oversees the improvement activities
- Output and decisions are communicated to the community



Landscape assessment

Part I: 188 responses from 144 institutions | Part II: Vocabularies use in 60 data sources

FINDINGS

- 87% of the community feels confident about Vocabularies' integrity
- Most commonly used vocabularies: SNOMED, ICD 9/10(US and int versions), MedDRA, ICDO3, ATC, RxNorm/RxE, ICD10PCS, ICD9Proc, CPT4, LOINC, CVX, HCPCS, UCUM, NDC, NAACCR, Cancer Modifier
- Most update data annually or semi-annually

NEEDS

- Transparent release schedule
- Vocabulary changes, versioning
- Transparent QA
- Better coverage and hierarchies
- More documentation and educational materials



Roadmap

OHDSI / Vocabulary-v5.0

Edit Pins Unwatch 47 Fork 67

Code Issues 195 Pull requests 28 Discussions Actions Projects 3 Wiki Security Insights Settings

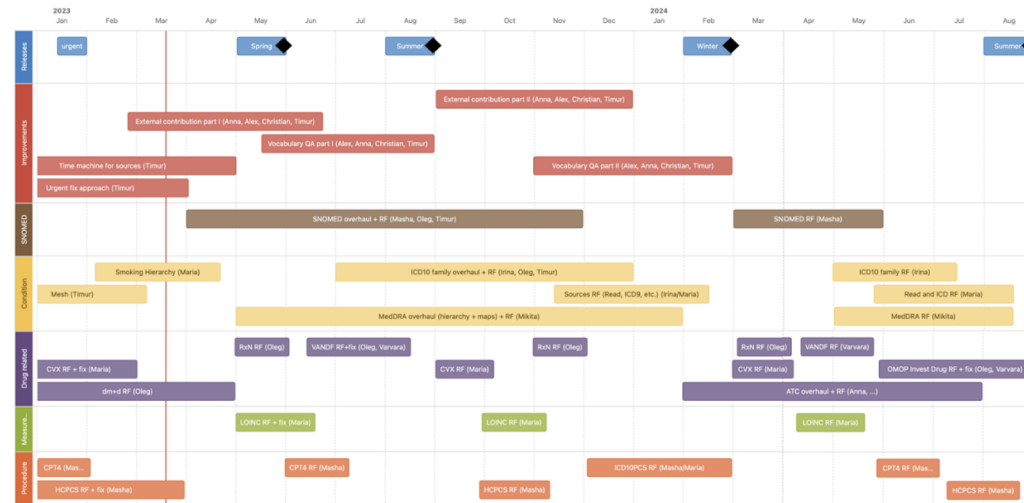
Release planning

Anna Ostropelets edited this page last week · 1 revision

Edit New page

This page provides you with the planned maintenance and improvement activities around the OHDSI Standardized Vocabularies. This is to be treated as a forecast. Below you can find the content of each release and an overview of the planned improvement activities (detailed content to be posted separately).

Roadmap 2023 Q1 - 2024 Q2:



Pages 56

Quick access:

- Home
- Introduction
- The Vocabulary Team
- Roadmap
- Release notes
- General structure and use
- Vocabularies
- Vocabulary Statistics
- Quality Assurance
- Known Issues in Vocabularies
- Maintain your vocabulary
- Glossary
- Articles
- COVID-19 Vocabulary/ETL Instructions

Clone this wiki locally

https://github.com/OHDSI/Vocab



Roadmap: vocabularies

Vocabularies on maintenance: ATC, CPT4, CVX, HCPCS, ICDs (10, 10CM, 10GM, 10PCS, CIM10, KCD7), LOINC, MedDRA, NDC, OMOP Invest Drug, RxNorm, RxNorm Extension, SNOMED, VANDF + administrative/supportive vocabularies (type, status).

Obsolete vocabularies: CCS, Cohort, Cohort Type, Condition Type, Cost Type, Death Type, Device Type, Drug Type, Episode Type, Meas Type, Multilex, NDFRT, Note Type, Obs Period Type, Observation Type, OXMIS, Procedure Type, SMQ, Specimen Type, Visit Type.

Vocabularies no longer maintained:

- **Oncology:** Cancer Modifier, CAP, HemOnc, ICDO3, NAACCR, Nebraska Lexicon, OncoTree, OMOP Genomic group (OncoKB, CGI, CIViC, ClinVar, JAX, HGNC, NCI).
- **Source:** AMIS, AMT, BDPM, CCAM, CIEL, CTD, dm+d*, DPD, EDI, EU Product, GCN_SEQNO, Gemsript, GGR, GPI, JMDC, KDC, KNHIS, MeSH*, Multum, NCCD, OPS, PCORNet, SUS, UK Biobank, VA Class.
- **Standard/Classification:** APC, CDT, DRG, EphMRA ATC, ETC, Indication, MDC, NFC, OPCS4, OSM, US Census, SNOMED Veterinary.



Maintenance: Aug 2023 and February 2024 releases

August 2023
CPT4 refresh
LOINC refresh
NDC refresh
RxNorm refresh
RxNorm Extension update
SPL refresh
VANDF refresh

February 2024	
CVX refresh	<i>SNOMED overhaul</i> and refresh (UK, US, Int)
LOINC refresh	
NDC refresh	<i>ICD10 improvement</i> + refresh (ICD10CM, ICD10, ICD10CN, ICD10GM, CIM10)
RxNorm refresh	
RxNorm Extension update	MedDRA <i>improvement</i> + refresh
SPL refresh	Read mapping refresh
HCPCS refresh	ICD10PCS refresh

August 2024: refreshes of vocabularies above + *ATC overhaul*



What have you got so far?



Community contribution Part I: currently supported use cases

ADD NEW CONTENT

Does not impact the rest of the community
Can be incorporated quickly (little review)

- Adding **non-standard concepts** to existing vocabularies
- Adding **synonyms** to existing concepts
- Adding **new mappings**
- Adding non-standard **vocabularies**

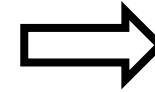
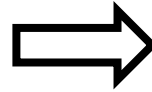
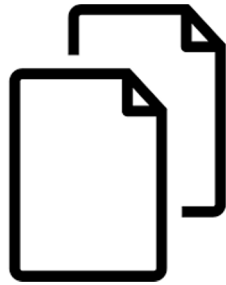
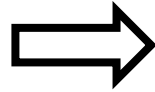
MODIFY EXISTING CONTENT

Requires review of the Vocab Team & other WGs/broader community

- Modifying existing **mappings**
- Modifying **concept fields** (e.g., description, domain, but not code)
- **Promoting non-standard** concepts to standard



Community contribution: pipeline



Issue on GitHub

Template

- Content
- Meta-data
- Checklist (QA)

Submission

Review

Release



August release



August 2023 release: vocabularies on maintenance

Vocabulary	Version planed	Version released (if different)	Author
CPT4	Spring 2023		Masha Khitrun
LOINC	2.74	2.75 version	Maria Rogozhkina
NDC	August 2023		Oleg Zhuk, Timur Vakhitov
RxNorm	August 2023	July 2023	Oleg Zhuk, Timur Vakhitov
RxNorm Extension	August 2023		Anna Ostropolets, Timur Vakhitov
SPL	August 2023		Oleg Zhuk, Timur Vakhitov
VANDF/ VA class	20230306	20230807	Varvara Savitskaya



LOINC

FRESH DELIVERY



LOINC 2.75

- One of the core vocabularies
- Released 2023-08-15 by Regenstrief Institute - only two weeks before the end of August
- LOINC-SNOMED hierarchy improved
- One-to-many mappings fixed
- SARS-2 related lab tests incorporated into the existing hierarchy



LOINC

- **New concepts**

vocabulary_id	domain_id	count
LOINC	Measurement	2453
LOINC	Observation	1217
LOINC	Procedure	402
LOINC	Meas Value	215
LOINC	Note	114

- **New relationships**

vocabulary_id_1	vocabulary_id_2	relationship_id	count	percentage
LOINC	SNOMED	Is a	76	0,432%
LOINC	SNOMED	Maps to	-1	-0,19%
LOINC	OMOP Extension	Is a	7	5,512%



VANDF / VA Class

- Veterans Health Administration (VHA) produces the National Drug File (NDF)
- Updated overdue: v 2021-10-04 currently in Athena
- Huge changes to the algorithm of domain assignments. Previously, all the concepts from VANDF vocabulary were **Drugs**. Now we are separating **Devices** using internal hierarchy

Before

BANDAGE TUBULAR ELASTIC SZ C	Drug
TRACHEOSTOMY TUBE HOLDER WELCON #2740	Drug
DRESSING,SORBACELL FOAM 4IN X 4IN #84344	Drug

After

BANDAGE TUBULAR ELASTIC SZ C	Device
TRACHEOSTOMY TUBE HOLDER WELCON #2740	Device
DRESSING,SORBACELL FOAM 4IN X 4IN #84344	Device



VANDF / VA Class

Note: mapping provided by the source can be inaccurate in OMOP sense

VANDF code	VANDF name	relationship_id	RxNorm code	RxNorm name
4003882	CEPHAPIRIN NA 20GM/VIL INJ	Maps to	313899	cephapirin 417 MG/ML Injectable Solution
4004136	GENTAMICIN SO4 10MG/ML (PF) INJ	Maps to	1870676	2 ML gentamicin 10 MG/ML Injection
4004199	MEPIVACAINE HCL 3% INJ	Maps to	1305263	1.7 ML mepivacaine hydrochloride 30 MG/ML Cartridge
4004573	MEZLOCILLIN NA 1GM/VIL INJ	Maps to	311689	mezlocillin 250 MG/ML Injectable Solution
4005823	DOPAMINE HCL 80MG/ML INJ	Maps to	1743953	5 ML dopamine hydrochloride 80 MG/ML Injection
4006008	MANNITOL 100MG/ML INJ	Maps to	1791395	500 ML mannitol 100 MG/ML Injection
4007046	RANITIDINE HCL 150MG GRANULES,EFFERVSC	Maps to	827179	ranitidine 1 MG/ML Oral Solution



CPT4

- Updated to UMLS 2023AA version
- Visit concepts have been back rolled back to their original domains (after debates with the CDM WG)
- Deduplication of semantic counterparts between CPT4 categories and SNOMED
- Relationships between CPT and SNOMED were enriched by mappings and hierarchical 'Is a'
- SNOMED-CPT crosswalks are now extracted directly from the source



CPT4

- **New concepts**

vocabulary_id	domain_id	count
CPT4	Procedure	142
CPT4	Measurement	83
CPT4	Observation	23
CPT4	Drug	1

- **Domain change**

vocabulary_id	old_domain_id	new_domain_id	count
CPT4	Visit	Observation	175
CPT4	Visit	Procedure	8
CPT4	Visit	Measurement	3

- **New relationships**

vocabulary_id_1	vocabulary_id_2	relationship_id	count	percentage
CPT4	SNOMED	Is a	1344	212,995%
CPT4	SNOMED	Maps to	1562	159,877%
CPT4	SNOMED	CPT4 - SNOMED cat	86	0,516%
CPT4	SNOMED	CPT4 - SNOMED eq	-105	-0,644%



Other changes

- RxNorm is updated to July 2023 version.
- NDC/SPL are updated with August 2023 version
- Refined Smoking-related concept mappings: HCPCS, Read, UK Biobank, ICD10CM
- Family of previously proprietary CO-CONNECT vocabularies is now distributed under an open-source license - thank you for your contribution!
- Minor bug fixes



Overview

1. Recap of the plan for Vocabularies improvement

2. Accomplishments so far

- New process & Vocabulary Committee
- Landscape assessment
- Roadmap and releases
- Community contribution
- August release
- Experience with community contribution
- Vocabulary WG
- Quality management system

3. **What is missing**, issues faced and ways to improvement



Community contribution in more details



Community contribution part I

- **11** contributions submitted so far:
 - Addition of content: 2 new non-standard vocabulaires
 - Modification of content: 1 with domain change, 3 with mapping changes
 - Complex: 2 with deduplication, 1 with hierarchy, 2 with new standard terms
- **4** contributions submitted and approved *before* the end of June and are *incorporated* into the August release
- **4** contributions submitted *after* the end of July and will be incorporated into the February release
- **3** contributions are not fully baked yet



Community contributions in the August release

Adding NHS Ethnicity codes and Admission and Discharge codes to the OMOP vocabulary community contribution 13
 #843 opened on Jun 14 by solmazeradat

ICD10CM: Limb vs. Lip in Carcinomas mapping bug community contribution ICD10CM 2
 #845 opened on Jun 21 by p-talapova

Adding New RxE Drugs community contribution
 #853 opened on Jul 1 by aostroplets



- 2 new non-standard vocabularies added with mappings (NHS Ethnic Category and NHS Place of Service)
- 4 mapping bugs fixed in existing vocabularies
- RxNorm Extension drug concepts added to support the local source drug system (Z-index)



Community contributions going into February release

Issues 850, 856: mapping changes



remap coma codes community contribution ICD10CM   1
#850 opened on Jun 29 by dimshitc

To fix that ICD10CM "substance use with other disorders" mapped to Mood disorder community contribution ICD10CM   3
#856 opened on Jul 10 by dimshitc

Issue 872: concept attribute change

Sexual activity concepts in the incorrect domain of Condition.Should be Observation community contribution SNOMED
#872 opened 3 weeks ago by MelaniePhilofsky



Issue 795: most of the concepts are new SNOMEDs absent in vocabularies => postponed till February

Neuropsychiatric Scores and Scales - Deduplication request [Psychiatry WG] community contribution OMOP Extension SNOMED   7
#795 opened on Apr 6 by p-talapova




Not fully baked community contribution

Issue **846**: ongoing discussion in Oncology WG

- Intersection between Histopattern and Staging/Grading concept classes **Cancer Modifier** **community contribution**   4
#846 opened on Jun 28 by dimshitc

Issue **860**: complex case (new hierarchy) with more discussion needed

- suggest new tidier route hierarchy through linkage to dose form **community contribution**  9
#860 opened on Jul 25 by tiozab

Issue **851**: ongoing discussion

- Derived Value Standard Concept **community contribution**  8
#851 opened on Jun 29 by cukarthik

Issue **871**: to be reviewed and discussed

- Feature Request: Vocabulary Contribution Linter **community contribution**  4
#871 opened 3 weeks ago by haydenbspence



How first run of community contribution went

1. We incorporated all contributions submitted and approved by the end of June
(many thanks to our contributors!)
1. Contributions can be made through templates but require involvement on both sides (and we advise to create an issue first!)
1. Quality assurance on the side of contributors and resolvment of inconsistencies on the Vocabularies development side are the areas for improvement



Examples of issues to watch out for in the checklist

Wrong dates

relationship_valid_end_date
for old mappings - current date
new mappings - <u>12/31/2099</u>
date
8/1/2023
8/1/2023
8/1/2023
8/1/2023
12/30/2099
12/30/2099
12/30/2099
12/30/2099

Different vocabulary_ids in different files, referring to the same concepts

vocabulary_id
short name (id) of your vocabulary
max 20 characters
Vocabulary name
Vocabulary name
Vocabulary name

vocabulary_id_1
short name (id) of your vocabulary
max 20 characters
Different Vocabulary name
Different Vocabulary name
Different Vocabulary name

Violating constraints: vocabulary_id longer than varchar(20)

vocabulary_id
short name (id) of your vocabulary
max 20 characters
Too long vocabulary name
Too long vocabulary name
Too long vocabulary name

Wrong dates

relationship_valid_start_date	relationship_valid_end_date
usually current date	12/31/2099
date	date
12/31/2099	12/31/2099
12/31/2099	12/31/2099
12/31/2099	12/31/2099
12/31/2099	12/31/2099
12/31/2099	12/31/2099
12/31/2099	12/31/2099

...and it gets harder for more complex cases (new drugs, new hierarchies)



Community contribution: summary

1. Community is getting increasingly involved in collaborative Vocabularies development
1. Current template approach supports advertised use cases well but requires time
1. Further steps include more education (e.g., demos, Vocabulary WG) and building automated pipelines



Current QA-System

- Look in Github Vocabulary-v5.0 repo under Wiki
- QA consists of:
 1. Quality issue tracking and resolution in Forum or as Github Issue
 2. Scripts
 - Quality control scripts, mostly conformance and logic checks, a few semantic checks
 - Quality enforcing scripts, overwriting deviating content
 3. Manual content checks



Future QA-System: Quality Management System

- QMS is a formalised approach with documented objectives, processes, procedures, responsibilities
- Consists of quality (i) planning, (ii) assurance, (iii) control, (iv) improvement
- We need:
 - Organization
 - Design
 - Documentation
 - Complaint handling and defect tracking
 - Corrective and preventive actions (CAPA)
 - Training



Vocabulary WG

- Changing focus to familiarize community with ongoing efforts and create a collaborative environment to share and discuss needs and ideas and to support other WGs
- Increased attendance (up to 40 people)
- Topics covered so far: community contribution, tracking vocabulary changes, SNOMED-LOINC hierarchy (check out our Teams subchannel!)
- Please come to raise questions and discuss your needs



Summary and next steps

1. We have implemented landscape assessment, established Vocabulary Committee, created roadmap, improved on documentation (incl. QA) and ran community contribution part I
1. More work is needed to establish working pipelines for community contribution, including common development environment
1. More engagement from community is needed (message Vocab Team, come to Vocab WG calls!)
1. Quality Management System is a crucial element and is currently in progress