

Release Notes V4.5 01-Nov-2014

Short version

This is a new version of the OMOP Standard Vocabulary in version 4. But the amount of structural changes demanded a new sub-version .5.

Changes (only big-ticket items)

Primary map no longer unique in active concepts

The idea of a unique mapping between a source_code/source_vocabulary and a target_vocabulary was dropped. Hence, the need for a flag to indicated the primary among all active (invalid_reason is null) records in the source_to_concept_map is no longer maintained. There can be now two target concepts, requiring the ETL to write out two (in this case CONDITION_OCCURRENCE) records. For example, ICD-9-CM 070.43 'Hepatitis E with hepatic coma' is both concept 196029 'Viral hepatitis with hepatic coma' as well as 197490 'Acute hepatitis E'.

All SNOMED now Standard vocab

With the exception of drug concepts (to be added later), all of SNOMED-CT (vocabulary_id=1) concepts have a concept_level>0, making them Standard Concepts that can be used in CDM tables. Drugs are still coded in RxNorm (vocabulary_id=8) or Multilex (vocabulary_id=22).

Introduction of domains and relationships to concepts

To prepare for Vocabulary V5.0, new concepts representing the various clinical Domains were introduced. These concepts are vocabulary_id=59. These concepts are linked to the respective Concepts through new relationships (records in the CONCEPT_RELATIONSHIP table), the relationship_id is 359/360. So, in order to find all concepts if the, say, Procedure Domain find all the concepts that are have a relationship to concept_id=10 (Procedure Domain concept).

Note: All active (invalid_reasons is null) Standard Concepts will have one and only one Domain, but this has not yet been achieved quite yet. Currently 642,888 of the total of active 661,217 Standard Concepts have a Domain designation.

In particular, vocabularies that contain concepts of more than one Domain such as SNOMED-CT (vocabulary_id=1), CPT-4 (vocabulary_id=4) and HCPCS (vocabulary_id=5) have been classified:

Note: Relationships to domains are not part of the CONCEPT_ANCESTOR table. Concepts connected ancestry vary in their Domains.

Vocabulary	Domain	Number of concepts
SNOMED-CT	Condition	101,464
SNOMED-CT	Device	4,474
SNOMED-CT	Drug	23,390
SNOMED-CT	Gender	6
SNOMED-CT	Measurement	9,940
SNOMED-CT	Metadata	558
SNOMED-CT	Observation	133,551
SNOMED-CT	Procedure	50,633

SNOMED-CT	Provider	672
SNOMED-CT	Race	281
CPT-4	Measurement	1,492
CPT-4	Metadata	9
CPT-4	Observation	644
CPT-4	Procedure	9,117
HCPCS	Device	2,846
HCPCS	Measurement	82
HCPCS	Observation	1,550
HCPCS	Procedure	1,883

Revision of Mapping Types

In SOURCE_TO_CONCEPT_MAP records, the mapping_type entries have been revised as follows:

- 'CONDITION-OBS' became 'Observation'
- 'INDICATION' became 'Indication'
- 'DRUG' became 'Drug'
- 'PROCEDURE' became 'Procedure'
- 'OBSERVATION' became 'Observation'
- 'ETHNICITY' became 'Ethnicity'
- 'COST' became 'Revenue code'
- 'PROCEDURE DRUG' became 'Drug'
- 'UNIT' became 'Unit'
- 'RACE' became 'Race'
- 'CONDITION' became 'Condition'
- 'CONDITION-MEDDRA' became 'MedDRA'
- 'OTHER' became 'Drug' if it is Oxmis vaccinations
- 'PLACE OF SERVICE' became 'Place of Service'
- 'PROVIDER' became 'Provider'
- 'CONDITION-PROCEDURE' became 'Procedure'
- Maps to 0 have mapping_type='Unmapped'

Note: the mapping from ICD-9-CM to MedDRA will be removed in future versions.

In addition, the mapping_type entries are now identical to the Domain assignments of the Target Concepts. For example, in the record

source_code	source_vocabulary_id	source_code_description	target_concept_id	target_vocabulary_id	mapping_type	primary_map	valid_start_date	valid_end_date
27...00	17	Obstetric examination	4038747	1	Procedure	Y	01-Jan-70	31-Dec-99

links the source_code '27...00' to the Target Concept 4038747 'Obstetric examination'. The Mapping Type is 'Procedure'.

Note: The implication for the ETL process is that the entry in the mapping_type will determine, where the target record is going to be written. You can no longer assume a fixed relationship between concepts of a certain vocabulary and a target table in the CDM (HCPCS=procedure).

Currently, the following Mapping Types are assigned to the different vocabularies:

Source Vocabulary		Domain of Target Concept	# maps		Source Vocabulary		Domain of Target Concept	# maps
2	ICD-9-CM	Condition	19077		18	OXMIS	Condition	5447
2	ICD-9-CM	Device	8		18	OXMIS	Device	16
2	ICD-9-CM	Indication	12017		18	OXMIS	Drug	3
2	ICD-9-CM	Measurement	63		18	OXMIS	Measurement	184
2	ICD-9-CM	MedDRA	16794		18	OXMIS	Observation	1011
2	ICD-9-CM	Observation	1573		18	OXMIS	Procedure	1198
2	ICD-9-CM	Procedure	500		18	OXMIS	Provider	3
2	ICD-9-CM	Unmapped	1124		18	OXMIS	Race	4
3	ICD-9-Procedure	Drug	17		18	OXMIS	Unmapped	423
3	ICD-9-Procedure	Procedure	4652		22	Multilex	Drug	51220
4	CPT-4	Condition	32		22	Multilex	Unmapped	41192
4	CPT-4	Drug	209		28	VA Product	Drug	26899
4	CPT-4	Measurement	1614		34	ICD-10-CM	Condition	61450
4	CPT-4	Metadata	9		34	ICD-10-CM	Unmapped	10121
4	CPT-4	Observation	630		35	ICD-10-PCS	Procedure	67413
4	CPT-4	Procedure	10744		35	ICD-10-PCS	Unmapped	4681
4	CPT-4	Unmapped	10		43	Revenue Code	Revenue code	538
5	HCPCS	Condition	123		44	Ethnicity	Ethnicity	43
5	HCPCS	Device	2857		46	NLM Mesh	Condition	4758
5	HCPCS	Drug	1271		46	NLM Mesh	Drug	3951
5	HCPCS	Measurement	82		46	NLM Mesh	Measurement	140
5	HCPCS	Observation	1434		46	NLM Mesh	Observation	94
5	HCPCS	Procedure	2879		46	NLM Mesh	Procedure	745
6	LOINC	Observation	77178		47	NUCC	Provider	829
9	NDC	Drug	604588		48	CMS Special	Provider	111
9	NDC	Unmapped	81296		50	FDA SPL	Drug	66074
10	GPI	Drug	23567		50	FDA SPL	Unmapped	32978
10	GPI	Unmapped	4332		53	FDB Genseq	Drug	15656
11	UCUM	Unit	766		53	FDB Genseq	Unmapped	7185
13	CDC Race	Race	921		56	Gemsript	Drug	15239
14	CMS Place of	Place of Service	49					
16	Multum	Drug	10952					
17	Read	Condition	61057					
17	Read	Device	280					
17	Read	Drug	560					
17	Read	Gender	2					
17	Read	Measurement	4057					
17	Read	Observation	31374					
17	Read	Procedure	29771					
17	Read	Provider	165					
17	Read	Race	108					
17	Read	Unmapped	9737					

Introduction of relationships and ancestry between CPT-4 and HCPCS with Procedures, Observations and Measurements

HCPCS and CPT-4 have now hierarchical relationships to Procedure (CPT-4 had that before), Observation and Measurement concepts (all in SNOMED-CT, vocabulary_id=1). For example, HCPCS 2617208 'Colorectal cancer screening; colonoscopy on individual at high risk' is now a descendant of SNOMED-CT 4152046 'Screening for malignant neoplasm of large intestine'.

Fixed relationship direction between cpt4 and snomed

In previous releases, the direction of relationships and ancestry definitions between CPT-4 and SNOMED was reversed, creating incorrect links between lower-level CPT-4 records to be ancestors of higher-level SNOMED-CT concepts. This has been fixed.

Introduction of a hierarchical system to the Observation and Procedure Domains

This, in conjunction with similar relationships between ICD-9-Procedure (vocabulary_id=3) and CPT-4 (vocabulary_id=4) creates a fully hierarchical system for the Procedure and Observation domains. However, the quality and the coverage of the hierarchy is not nearly as good as for the Drug and Condition Domain hierarchies. For the Measurement Domain, LOINC has to be added to the existing hierarchy.

Introduction of mappings between HCPCS and Conditions and Drugs

New and improved mapping records in the SOURCE_TO_CONCEPT_TABLE have been introduced between HCPCS codes and RxNorm Drug concepts for procedure drugs. New mapping records were established between HCPCS and SNOMED-CT Condition concepts. For example, HCPCS G8575 'Developed postoperative renal insufficiency or required dialysis' is now mapped to SNOMED-CT 192359 'Renal failure syndrome'

ATC now sourced directly from WHO

ATC codes are now directly sourced from the WHO Collaborating Center for Drug Statistics Methodology, instead of through FDB. They are complete and more up-to-date.

Removed redundant relationships between ATC and RxNorm

In previous releases, RxNorm Ingredients had relationships not only to the equivalent ATC ingredient, but to every concept that is in the hierarchy above them. That didn't create problems with the CONCEPT_ANCESTOR records, which would have established the same links as well, but it caused problems with hierarchical browsing within the ATC hierarchy.

Removed incorrect relationships between MedDRA LLT and SNOMED-CT

In previous releases, relationships between SNOMED-CT and MedDRA Preferred Terms were also extended to the MedDRA Low Level Terms. These are supposed to be synonymous to the PT, but in reality are often hierarchically below the PT. Therefore, these additional relationships have been removed. That means, that not all PT will have a SNOMED-CT descendant in the CONCEPT_ANCESTOR table.

Introduction of escape characters to comma-separated distribution files

In previous releases, comma-separated files ran into problems when quotes, double quotes or commas were part of the name fields. Now, everything is comma separated. Commas within the fields are escaped, i.e. have a backslash character in front of them.

HCPCS and SNOMED Concept Names fixed

In previous releases, the entries in the concept_name fields of the SNOMED-CT (vocabulary_id=1) and HCPCS (vocabulary_id=5) records in the CONCEPT table were not filtered for obsolete content. In some cases, this led to mistakes, such as SNOMED-CT 4303531 'Eosinopenia' was falsely labeled as 'Eosinophilia'.

Introduction of previously missed update relationship in SNOMED

In previous releases, relationships between obsolete and active concepts were established from the 'replaced_by' relationships in the SNOMED-CT distribution files. Now, the relationships 'same as', 'alternative_to', 'possibly_equivalent_to' and 'was a' relationships were added, bringing the number of upgrades from 5,047 to 77,733.

Introduction of Currency concepts

All currencies are now available as concepts in preparation for CDM V5.0. These concepts have vocabulary_id=65.

RELATIONSHIP table fixed

In previous releases, relationship table contained bugs (missing reverse_relationship, incorrect flags, wrong direction), which are now fixed. As a consequence, the CONCEPT_ANCESTOR table can now be constructed automatically out of the individual relationships.

Removed null flavors from gender and race

All but legitimate gender and race concepts were deprecated. All unknown or undefined gender and race data are encoded through concept_id=0.

Read mapping revised and automated

Read codes (vocabulary_id=17) are now derived from the NHS mapping tables and fully automated, including the assignments of the correct mapping_type.

Removed mapping to gender, ethnicity

No Source Codes are available that are in standard use.

Revision of ICD-9-CM mappings

The ICD-9-CM mappings have now a consistent high quality. The only remaining improvement planned is the addition of special "Include children" and "Exclude children" maps for cases when a near-perfect mapping is desired. However, these don't fit into the SOURCE_TO_CONCEPT_MAP model and will therefore be implemented only in V5.0.

Revision of ICD-10-CM mappings

The ICD-10-CM mapping was revised based on the available SNOMED to ICD-10 mappings released by the NLM and the NHS, as well as through triangulation through ICD-9. However, there are a total of 60k ICD-10 codes. Currently, only those codes for which we have actually data were manually curated and revised. These are listed in the file ICD-10-manually-checked.txt.

Todo:

The current test release requires the following additional fixes:

- Fix VA Class
- Fix NDF-RT
- Fix ATC combination, finish concept assignments.
- Fix Gemscrip
- Fix missing UK SNOMED Concepts Relationship
- Modifier and Specimen type not defined yet.
- Resolve with specimen disease status which collides with Measurement values
- Define Generic
- Vet Device Type and Measurement Type
- Resolve mappings to Indications, MedDRA
- Resolve fate of Revenue Codes and DRG
- Create Specimen Type concepts (domain_concept_id=37)
- Resolve overlap of Modifiers with Observations. (domain_concept_id=12)
- Create Specimen disease status concepts (vocabulary_id=39)
- Revise and automate Gemscrip
- Add class and vocabulary concepts