

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

OMOP Standard Vocabulary
Version 4.3 2013-09-15
Release Notes

Christian Reich, Patrick Ryan, Don Torok, Sergei Vereshagin,
Mark Khayter and Emily Welebob

15-September-2013

License

© 2009-2013 Observational Medical Outcomes Partnership (OMOP).

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this document except in compliance with the License. You may obtain a copy of the License at <http://omop.org/publiclicense>.

Unless required by applicable law or agreed to in writing, documentation and software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. Any redistributions of this work or any derivative work or modification based on this work should be accompanied by the following source attribution: "This work is based on work by the Observational Medical Outcomes Partnership (OMOP) and used under license from the RUF at <http://omop.org/publiclicense>".

Any scientific publication that is based on this work should include a reference to <http://omop.org>.

Release

This document reflects all the changes to the Vocabulary introduced since the Version 4.3 of 1-Aug-2013. In addition to executing standard updates (see Appendix for latest content), the following changes were made:

List of changes

1. DEPRECATED AND RESTORED CONCEPTS AND RELATIONSHIPS

Any concept or relationship can have only one life cycle (`valid_start_date` and `valid_end_date`). For concepts that were deprecated and then restored in the source this caused a problem, because no concept could exist twice in the vocabulary and the update script would reject it. Now, the concept will be written out with a `valid_start_date` of the restored date.

2. MAPPING OF ICD-9-CM AND ICD-9-PROC CODES

Some of the existing manual mapping records were manually curated and therefore overwritten by the automatic update scripts and had to be restored. Also, some codes had erroneous cross-mappings from ICD-9-CM to procedure and ICD-9-Proc to Clinical Findings. All these issues are now fixed.

3. AVAILABILITY OF ACTIVE MAPPING RECORDS FOR ALL PERIODS

Due to inconsistencies in the source data, some codes ended up with no valid mapping in the `SOURCE_TO_CONCEPT_MAP` table for a certain time period. In those cases, the latest mapping was promoted to active, such that there is always one active mapping available per code and time period.

4. AVAILABILITY OF PRIMARY MAPPING RECORDS FOR ALL PERIODS

Similarly, some codes had no primary record in the `SOURCE_TO_CONCEPT_MAP` table for some time periods. In those cases the latest added secondary mapping was promoted to primary. In cases where two or more records were primary, all but the latest added mapping were demoted to secondary.

5. MAPPING TO UPDATED CONCEPTS

If a record in the `SOURCE_TO_CONCEPT_MAP` table points to a deprecated concept with an `INVALID_REASON='U'` (updated), and a relationship is available that points to the updated concept (`RELATIONSHIP_IDs` 1, 309, 311, 313), the record is now updated to the valid concept. Sometimes, this can take several such steps in sequence to arrive at a valid target concept.

6. BUILDING OF A COMPLETE ICD-9-CM AND NDC CODE BASE

For the vocabularies ICD-9-CM and NDC, an effort was made to collect all codes used in the various databases, even they are not officially sanctioned, and for some no valid mapping could be derived (ending up with a mapping to concept 0). If you have such codes that are still not available, let us know.

7. IMPROVING THE NDC MAPPING

The NDC mapping has now all three sources included:

- RxNorm (11-digit codes)
- FDB (11 digit codes)
- FDA (8 and 9-digit codes with dashes)

The latter ones were converted to 9-digit codes. These are equivalent to the 11-digit codes RxNorm and FDB provide, but are missing the last two digits. For most products, the last two digits only the package, RxNorm-based concepts don't contain package information, so strictly speaking only the first 9 digits are sufficient for mapping. However, some very common OTC products like NyQuil deviate from that rule, where the manufacturer uses the last two digits to differentiate products

This will improve the coverage of NDC mapping significantly. However it also means that in the SOURCE_TO_CONCEPT_MAP are two types of codes: 11-digit NDC and 9-digit NDC, both mapped to RxNorm-based concepts. This is important for ETL jobs of NDC-based source data. We recommend the following mapping procedure:

1. For each NDC in the source, convert to 9 or 11 digit. If you don't know how to do that visit <http://omop.org/NDC> for more details.
2. Try mapping the full code. If a map is available use it.
3. Otherwise:
 - a. If the source code is 9 digits long, try finding all 11 digit codes in the SOURCE_TO_CONCEPT_MAP that have the first 9 digits match. Test whether these all map to the same drug (TARGET_CONCEPT_ID). If yes, use it. If not, see whether they map to the equivalent clinical and branded drugs. If yes, pick the branded drug. If not, see whether they map to different packs of the same drug. If yes, use the pack. If not, discard the mapping (and let us know).
 - b. If the source code is 11 digits long, truncate the last two digits and try mapping again. If a match is found, use it. If not, try 3a.

More details for all this complexity will be published at <http://omop.org/NDC>.

8. IMPROVING THE MULTILEX MAPPING

Several improvements in the Multilex mapping have been introduced:

- Mapping from Multilex is now always to active valid concepts, whether RxNorm (preferred) or Multilex (if no RxNorm exists for the product).
- The CONCEPT_LEVEL is set to 0 for all Multilex codes that have RxNorm equivalents, and to 1 and 2 if no equivalent RxNorm drug exists.
- The CONCEPT_ANCESTOR table is including these non-0 concepts, providing relationships to relevant drug classes and between drug products and ingredients very much like it works for RxNorm-based concepts.

For databases that are coded in Multilex (like CPRD and THIN), the Standard Vocabulary for drugs is a mixture of RxNorm and Multilex-based concepts with non-0 CONCEPT_LEVELS. See <http://omop.org/multilex> for more details.

9. INCORRECT RECORDS IN VOCABULARY_ANCESTOR TABLE FOR THE HOI VOCABULARY

Health Outcome of Interest concepts (VOCABULARY_ID=33) were wrongly assigned as descendants of RxNorm Ingredient concepts, instead of the other way around. This has been fixed.

10. CORRECT DEPRECATION OF MEDDRA CONCEPTS

MedDRA concepts were not correctly deprecated. In particular, the field LLT_CURRENCY in the source files was not taken into account. This is now fixed.

11. IMPROVING THE CREATION OF THE VA CLASS AND VA PRODUCT CONCEPTS

The processing of these vocabularies (VOCABULARY_IDs 32 and 28) has been improved, fixing issues with the ancestry builder.

12. INTRODUCING THE INTERNAL HIERARCHY INTO THE ICD-9-PROC CONCEPTS

These concepts have now concept_levels representing their position within the ICD-9-Proc hierarchy. This hierarchy is conceptually not very strong, but now correctly represented:

- 2-digit codes: CONCEPT_LEVEL=3 (top level)
- 3-digit codes: CONCEPT_LEVEL=2
- 4-digit codes: CONCEPT_LEVEL=1 (bottom level)

13. INCORRECT SET IDS OF THE SPL MAPPING

The Structured Product Labels (SPL, VOCABULARY_ID 50) are referred to by a 37-digit Set ID. In the previous release, the 7 and 8-digit NDC was erroneously used for source code. This is now fixed.

14. EXTENSION OF THE SOURCE_CODE FIELD TO 40 CHARACTERS

Because of the above, the size of the field SOURCE_CODE in the SOURCE_TO_CONCEPT_MAP had to be extended from 20 to 40 characters.

15. LOW-QUALITY READ MAPPING

A new mapping from Read codes (VOCABULARY_ID 17) has been provided, with increased coverage and improved MAPPING_TYPE, which now more correctly reflects whether a Read code reflects a condition, an observation or a procedure.

Appendix A: Statistic about Concepts, Vocabularies, Classes and Levels

This table is no longer provided in the release notes. But it can be produced by a standardized query. Please visit <https://sites.google.com/site/omopqueries/general-queries/g15>.

Appendix B: Statistics about Relationships between Concepts

This table is no longer provided in the release notes. But it can be produced by a standardized query. Please visit <https://sites.google.com/site/omopqueries/general-queries/g14>.

Appendix C: Updated Statistics about Mapping of Source Vocabularies

This table is no longer provided in the release notes. But it can be produced by a standardized query. Please visit <https://sites.google.com/site/omopqueries/general-queries/g16> and

<https://sites.google.com/site/omopqueries/general-queries/g17>

Appendix D: Source Vocabulary Versions and Sources

The following table contains the location for all external source vocabularies used for the OMOP Standard Vocabularies

Vocabulary ID	Vocabulary Name	Source	Date used for Vocabulary
1	SNOMED-CT	UMLS distribution, http://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html	31-Jul-2013
2	ICD-9-CM	Diagnosis and Procedure Codes: Abbreviated and Full Code Titles, Version 30, http://www.cms.gov/Medicare/Coding/ICD9ProviderDiagnosticCodes/codes.html	4-Dec-2012
3	ICD-9-Procedure		
4	CPT-4	UMLS distribution, http://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html	8-May-2013
5	HCPCS	UMLS distribution, http://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html	8-May-2013
6	LOINC	LOINC distribution, http://loinc.org/downloads/loinc	18-Jun-2013
7	NDF-RT	RxNorm distribution, http://www.nlm.nih.gov/research/umls/rxnorm/docs/rxnormfiles.html	5-Aug-2013
8	RxNorm	RxNorm distribution, http://www.nlm.nih.gov/research/umls/rxnorm/docs/rxnormfiles.html	5-Aug-2013
9	NDC	RxNorm distribution, http://www.nlm.nih.gov/research/umls/rxnorm/docs/rxnormfiles.html	5-Aug-2013
		FDB distribution package	20-Jun-2013
		FDA NDC distribution, http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/UCM070838.zip	6-May-2013
10	GPI	Medi-Span distribution package	1-May-2011
11	UCUM	The Unified Code for Units of Measure, Version 1.8.2, http://aurora.regenstrief.org/~ucum/ucum.html#section-Alphabetic-Index	5-Apr-2010
12	HL7 Administrative Sex	HL7, Appendix A: Data Definition Tables, http://www.hl7.org/special/committees/vocab/v26_appendix_a.pdf	27-Nov-2007
13	CDC Race	http://www.cdc.gov/nchs/data/dvs/Race_Ethnicity_CodeSet.pdf	27-Jun-2000
14	CMS Place of Service	Place of Service Codes for Professional Claims Database, http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/downloads/Website_POS_database.pdf	1-Nov-2009
15	MedDRA	MedDRA Version 16.0, http://www.meddramsso.com/subscriber_download.asp	14-Mar-2013
16	Multum	RxNorm distribution,	5-Aug-2013

OMOP STANDARD VOCABULARY 4.3 2013-09-15

Vocabulary ID	Vocabulary Name	Source	Date used for Vocabulary
		http://www.nlm.nih.gov/research/umls/rxnorm/docs/rxnormfiles.html	
17	Read	UMLS distribution, http://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledge/sources.html	2-May-2011
18	OXMIS	Codes extracted from GPRD database, courtesy of GSK	1-Mar-2009
19	FDB Indication	FDB US distribution package	20-Jun-2013
20	FDB ETC	FDB US distribution package	20-Jun-2013
21	WHO ATC	FDB US distribution package	20-Jun-2013
22	Multilex	FDB UK distribution package	01-Jun-2013
28	VA Product	RxNorm distribution, http://www.nlm.nih.gov/research/umls/rxnorm/docs/rxnormfiles.html	05-Aug-2013
31	SMQ	Standardised MedDRA Queries (SMQs) http://www.meddrasso.com/secure/smq/SMQ_Spreadsheet_14_0_English_update.xlsx	23-Jan-2013
32	VA Class	RxNorm distribution, http://www.nlm.nih.gov/research/umls/rxnorm/docs/rxnormfiles.html	05-Aug-2013
34	ICD-10-CM	2011 ICD-10-PCS and GEMs, Code Descriptions, http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html	21-Nov-2011
35	ICD-10-PCS	2012 ICD-10-PCS and GEMs, Code Descriptions, http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-PCS.html	8-Jun-2011
40	DRG	List of Diagnosis Related Groups (DRGs), FY 2005, http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareFeeforSvcPartsAB/downloads//DRGDesc05.pdf	5-Dec-2006
	MS-DRG	Acute Inpatient - Files for Download , https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Acute-Inpatient-Files-for-Download.html , Files for FY 20XX Final Rule and Correction Notice, Table 5	18-Feb-2011
41	MDC	MDC Description File, http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Acute-Inpatient-Files-for-Download-Items/CMS1247844.html	1-Jun-2013
42	APC	Hospital Outpatient PPS, Addendum A and Addendum B Updates, http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalOutpatientPPS/Addendum-A-and-Addendum-B-Updates.html	28-Sep-2011
47	NUCC	http://www.nucc.org/images/stories/CSV/nucc_taxonomy_131.csv , Version 13.1	7-Jan-2013
48	CMS Specialty	http://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/MedicareProviderSupEnroll/downloads/JSMTDL-08515MedicarProviderTypetoHCPTaxonomy.pdf	22-Sep-2008
50	FDA SPL	RxNorm distribution, http://www.nlm.nih.gov/research/umls/rxnorm/docs/rxnormfiles.html	05-Aug-2013
53	FDB Genseqno	FDB US distribution package	20-Jun-2013