

**Observational Health Data
Sciences and Informatics (OHDSI)**

**Natural Language Processing (NLP)
Working Group**

09/12/2018

AGENDA

- ❖ Introduction of new members
- ❖ Presentation – Mapping of CUIs to standard terminology – Juan Banda
- ❖ Updates on ongoing projects
 - ❖ Rules for defining term_exists – led by Stephane Meystre
 - ❖ Mapping of CUIs to standard terminology – Juan Banda
 - ❖ Mapping of Note Types to LOINC/standard vocabulary –Karthik Natarajan, Ruth Reeves, Jon Duke and Hua Xu– Report type list discussion
 - ❖ Landscape Analysis of section identifier systems and proposal of a standard terminology for use – Hua Xu, Karthik Natarajan
 - ❖ Examples and rules for term_temporal – led by George Hripsack (Sunny)
 - ❖ Standardization of term_modifiers and values – Hua Xu

Mapping from UMLS CUI into the OHDSI Vocabulary Release 2.0

Juan M. Banda

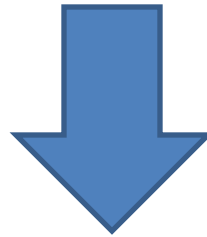




What are we trying to do?



Concept Unique Identifiers (CUI)



OHDSI concept_id's



ATHENA Standardized Vocabularies



Challenges

- UMLS (2018AA) has 140 English vocabularies
- OHDSI Vocabulary (Jan-2018) has 76 vocabularies
- Some versioning issues



The Real Challenge

- OHDIS concept_code which is the source code (in RxNorm, SNOMED) can be overlapping between different vocabularies
- Filtering on a per-vocabulary basis is needed



Side note

- This work was accepted for a poster at the OHDSI 2018 symposium



New name

- Keeping with the OHDSI tradition, this tool will be named:

- Ananke



personification of inevitability, compulsion and necessity



Our approach

- 1) Identify matching vocabularies on both sides (HPCS, ICD9, ICD10, SNOMED are easy)

UMLS2018AA – MRSAB table

RSAB	SON	SVER	SLC	SSN	SCIT
HCPT	HCPCS Version of Current Procedural Terminology (CPT),...	2017	Cynthia Hake;CMS HCPCS Workgroup Chair;Centers for ...	CPT in HCPCS	;;;Version of Physicians' Current Procedural Terminology...
HGNC	HUGO Gene Nomenclature Committee, 2017_05	2017_05	Elsbeth Bruford, PhD;Group Co-ordinator;HUGO Gene No...	HUGO Gene Nomenclature Committee	Gray KA, Yates B, Seal RL, Wright MW, Bruford EA;;HGN...
HL7V2.5	HL7 Vocabulary Version 2.5, 2003_08_30	2003_08_30	Health Level Seven;;;3300 Washtenaw Avenue.;Suite 22...	HL7 Version 2.5	Mark McDougall, Executive Director, Health Level Seven; ...
HL7V3.0	HL7 Vocabulary Version 3.0, 2016_07	2016_07	Health Level Seven International;;;3300 Washtenaw Ave...	HL7 Version 3.0	;;;Health Level Seven (HL7) Vocabulary;;;Ann Arbor, M...
HPO	Human Phenotype Ontology, 2017_06_28	2017_06_28	Dr. Peter N. Robinson;Professor of Computational Biolog...	HPO	;;;The Human Phenotype Ontology (HPO);;;The Huma...
ICD10	ICD10, 1998	1998	;;Office of Publications, World Health Organization;1211 ...	ICD-10	;;;International Statistical Classification of Diseases and ...
ICD10AE	ICD10, American English Equivalents, 1998	1998	;;Office of Publications, World Health Organization;1211 ...	ICD-10 Am Engl	;;;International Statistical Classification of Diseases and ...
ICD10AM	International Statistical Classification of Diseases and Rel...	2000	;;National Centre for Classification in Health, University o...	ICD-10 Austral Mod	;;National Centre for Classification in Health, University o...
ICD10AMAE	International Statistical Classification of Diseases and Rel...	2000	;;National Centre for Classification in Health University of...	ICD-10 Austral Mod Am Engl	;;;International Statistical Classification of Diseases and ...
ICD10CM	International Classification of Diseases, 10th Edition, Clini...	2018	Donna Pickett;Medical Classification Administrator;Nation...	ICD-10-CM	;;National Center for Health Statistics (NCHS), under aut...
ICD10PCS	ICD-10-PCS, 2018	2018	Pat Brooks;Senior Technical Advisor;Center for Medicare ...	ICD-10-PCS	;;Centers for Medicare and Medicaid Services;;Internatio...
ICD9CM	International Classification of Diseases, Ninth Revision, C...	2014	Patricia Brooks;Contact for Procedures;Health Care Fin...	ICD-9-CM	;;National Center for Health Statistics (NCHS);;ICD-9-CM...

OHDSI Vocabulary – Vocabulary table

HCPCS	Healthcare Common Procedure Coding System (CMS)	2017 Alpha Numeric HCPCS File
HES Specialty	Hospital Episode Statistics Specialty (NHS)	Not implemented yet
ICD10	International Classification of Diseases, Tenth Revision (...)	2016 Release
ICD10CM	International Classification of Diseases, Tenth Revision, ...	ICD10CM FY2018 code descriptions
ICD10PCS	ICD-10 Procedure Coding System (CMS)	ICD10PCS 20160518
ICD9CM	International Classification of Diseases, Ninth Revision, C...	ICD9CM v32 master descriptions
ICD9Proc	International Classification of Diseases, Ninth Revision, C...	ICD9CM v32 master descriptions
ICD03	ICD-O-3	ICD03 SEER Site/Histology Released 09/18/2015
ISBT	ISBT 128 Product Description Code Database	7.9.0
ISBT Attribute	ISBT 128 Product Description Code Database	7.9.0
LOINC	Logical Observation Identifiers Names and Codes (Regen...	LOINC 2.61
MDC	Major Diagnostic Categories (CMS)	2013-01-06
Meas Type	OMOP Measurement Type	
MedDRA	Medical Dictionary for Regulatory Activities (MSSO)	MedDRA version 19.1
MeSH	Medical Subject Headings (NLM)	2016 Release



Our approach (2)

2) Obtain CUI via concept_code (OHDSI) and code (UMLS)

CUI	concept_id	vocabulary_id
C0000039	1592753	RxNorm
C0000266	19041783	RxNorm
C0000294	1354698	RxNorm
C0000378	44814542	RxNorm
C0000392	19123344	RxNorm
C0000464	19016670	RxNorm
C0000473	19018384	RxNorm

UMLS via CUI

CUI	LAT	TS	LUI	STT	SUI	ISPREF	AUI	SAUI	SCUI	SDUI	SAB	TTY	CODE	STR	SRL	SUPPRESS	CVF
C0000039	ENG	P	L0000039	PF	S17175117	N	A28315139	9194921	1926948	(NULL)	RXNORM	IN	1926948	1,2-dipalmitoylphosphatidylcholine	0	N	(NULL)

OHDSI Vocab via concept_id

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
1592753	1,2-dipalmitoylphosphatidylcholine	Drug	RxNorm	Ingredient	S	1926948	2017-08-07	2099-12-31	



Version 2 release

- New release of mappings include:

- **Standard vocabularies (v1.0)**

- ICD9CM
- ICD10
- ICD10PCS
- ICD10CM
- HCPCS
- SNOMED
- RxNorm
- CPT
- MedDRA
- NDFRT
- LOINC

Added in v2.0

- ICD9Proc
- WHO Anatomic Therapeutic Chemical Classification (ATC)
- CDC Vaccine Administered (CVX)
- Medical Subject Headings (MeSH)
- National Uniform Claim Committee Health Care Provider Taxonomy Code Set (NUCC)
- Veterans Health Administration National Drug File



Results of V2 mapping

- Total number of concepts mapped: 1,321,985
(~200K concepts more than on v1)

V1

vocabulary_id	number_of_mapped_concepts
CPT4	39,947
HCPCS	6,376
ICD10CM	101,725
ICD10PCS	180,450
ICD9CM	16,376
LOINC	124,813
MedDRA	52,953
NDFRT	23,345
RxNorm	202,627
SNOMED	371,013

vocabulary_id	totalC
SNOMED	445,927
RxNorm	202,192
ICD10PCS	180,451
LOINC	125,841
ICD10CM	117,564
MedDRA	67,583
CPT4	40,089
VA Product	37,966
MeSH	30,391
NDFRT	26,389
ICD9CM	17,496
ICD10	12,034
HCPCS	6,445
ATC	5,389
ICD9Proc	4,477
NUCC	828
VA Class	513
CVX	410



Release Notes:

- All overlapping vocabularies are now mapped
- Issues remain with Deprecated concepts



In SNOMED

- OHDSI has 801,303 unique concept_codes
- UMLS only has 449,474 unique CODES

Result #1 (28x358,421)

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason	CUI	LAT	TS	LUI	STT
37392117	Urine tryptophan:creatinine ratio	Observation	SNOMED	Observable Entity	S	1000001000000103	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46139744	Nu-Hope belt left small 6460C 70cm-78cm length, 225mm...	Drug	SNOMED	Pharma/Biol Product		10000011000001109	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37398455	Urine threonine:creatinine ratio	Observation	SNOMED	Observable Entity	S	1000011000000101	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46004596	Nu-Hope belt right large 6457A 90cm-100cm length, 200...	Drug	SNOMED	Pharma/Biol Product		10000111000001105	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37392118	Urine taurine:creatinine ratio	Observation	SNOMED	Observable Entity	S	1000021000000107	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46004597	Nu-Hope belt left small 6460DC 70cm-78cm length, 225m...	Drug	SNOMED	Pharma/Biol Product		10000211000001104	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37392119	Urine phenylalanine:creatinine ratio	Observation	SNOMED	Observable Entity	S	1000031000000109	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46139745	Nu-Hope belt left small 6460DC 70cm-78cm length, 225m...	Drug	SNOMED	Pharma/Biol Product		10000311000001107	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37392120	Urine ornithine:creatinine ratio	Observation	SNOMED	Observable Entity	S	1000041000000100	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46139746	Nu-Hope belt right large 6457A 90cm-100cm length, 200...	Drug	SNOMED	Pharma/Biol Product		10000411000001100	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37392121	Urine methionine:creatinine ratio	Observation	SNOMED	Observable Entity	S	1000051000000102	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46004598	Nu-Hope belt left small 6460E 70cm-78cm length, 225mm ...	Drug	SNOMED	Pharma/Biol Product		10000511000001101	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37392122	Urine lysine:creatinine ratio	Observation	SNOMED	Observable Entity	S	1000061000000104	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)

V2 maps 445,927 of them – considerably better than on V1 (80K + better).

All remaining concepts are not found in UMLS because they have been removed over the years



How do we fix this?

- Use older versions or UMLS
 - I have from 2014 to 2018 to test this.
- Might introduce CUI collisions (yikes!!) as retired CUIs might surface again later in time



V2 availability

- Three flavors

- Mapping table

- <https://github.com/jmbanda/Ananke/blob/master/AnankeV2.zip>

- Mapping Script

- https://github.com/jmbanda/Ananke/blob/master/cui2ohdsi_concept_idv1.1.sql

- R Package

- Ready by Symposium



What is next ?

- Fully customizable R package
 - Will be ready for symposium



Question?

Thank you for your time!

jbanda@gsu.edu