

Overview of OHDSI's Vocabulary Tables in OMOP

Timothy Quinn, PhD

Scientific Computing & Data

Icahn School of Medicine at Mount Sinai

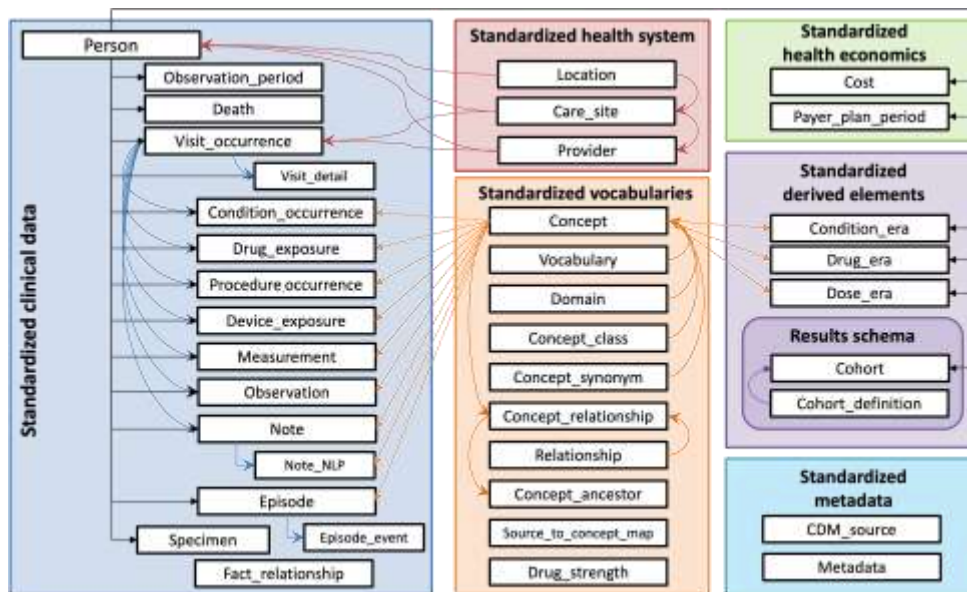
Wednesday, November 2, 2022



Icahn
School of
Medicine at
**Mount
Sinai**

OMOP Common Data Model Requirements

1. Standardize **data structure** via common format



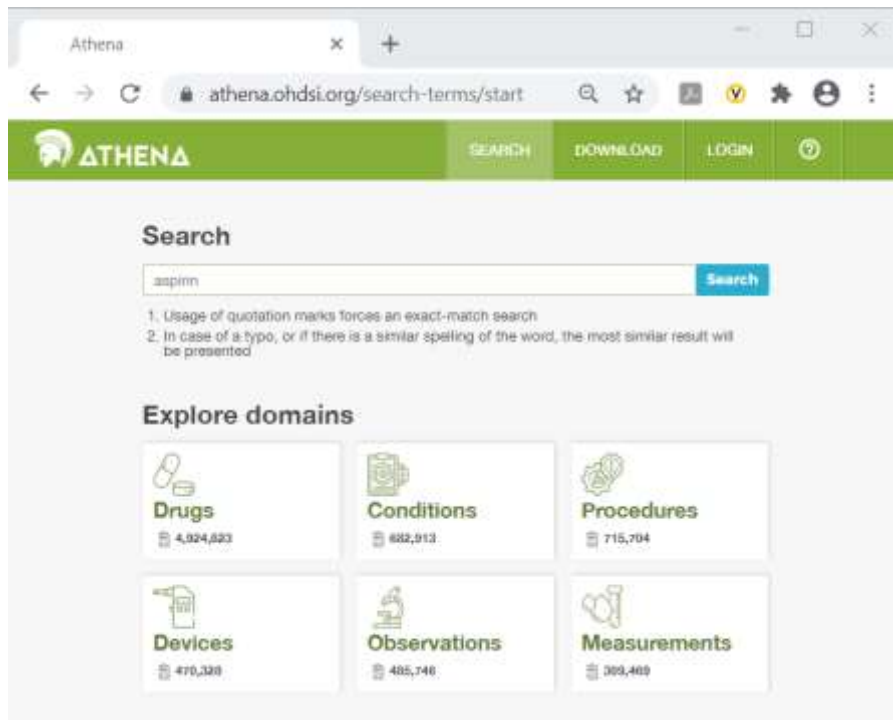
2. Standardize **data content** via mapping EHR codes to standard healthcare vocabularies

OMOP Domain	Standard Vocabularies	Non-standard Vocabularies
Condition	SNOMED-CT	ICD-10-CM, ICD-9-CM
Drug	RxNorm, CVX	ATC, NDC, Multum
Measurement	LOINC	SNOMED-CT, Nebraska Lexicon
Procedure	CPT4, HCPCS, ICD-10-PCS	ICD-9-Proc
Observation	SNOMED-CT, LOINC	ICD-10-CM, ICD-9-CM
Race, Ethnicity	OMOP Race, OMOP Ethnicity	SNOMED-CT, Nebraska Lexicon
Provider (Specialty)	NUCC, Medicare Specialty	SNOMED-CT, Nebraska Lexicon
Route	SNOMED-CT	Nebraska Lexicon
Unit	UCUM	SNOMED-CT, Nebraska Lexicon

OMOP Content Standardization via Concept Mapping

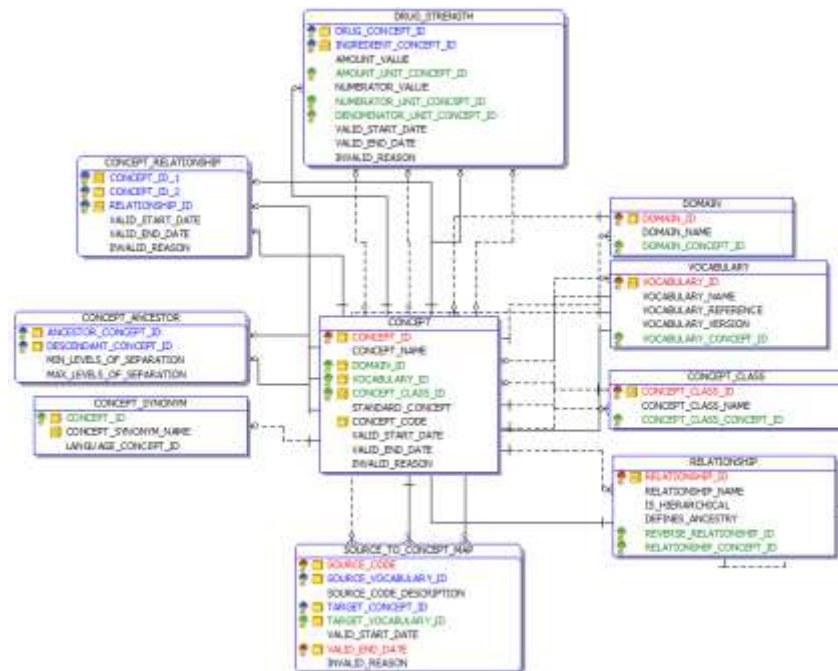
Column	Description	Notes
condition_occurrence_id	Surrogate primary key	
person_id	Surrogate foreign key	One-to-one with Epic's Patient ID
condition_concept_id	OMOP standard diagnosis code	SNOMED-CT from ATHENA
condition_source_concept_id	Epic diagnosis ID	From Epic's EDG master file
condition_source_value	Verbatim Epic diagnosis name	
condition_status_concept_id	OMOP standard diagnosis status code	From Epic's problem list
condition_status_source_value	Verbatim Epic diagnosis status name	
condition_start_date	Required diagnosis date	
condition_end_date	Optional diagnosis resolution date	

OMOP's Nine Vocabulary Tables



The screenshot shows the Athena OHDSI search interface. At the top, there's a browser window with the URL athena.ohdsi.org/search-terms/start. Below the browser window is the Athena logo and navigation buttons: SEARCH, DOWNLOAD, LOGIN, and a help icon. The main section is titled "Search" and contains a search bar with the text "aspirin" and a "Search" button. Below the search bar, there are two instructions: 1. Usage of quotation marks forces an exact-match search. 2. In case of a typo, or if there is a similar spelling of the word, the most similar result will be presented. Below the search section is the "Explore domains" section, which displays six domain cards: Drugs (4,924,823), Conditions (682,913), Procedures (715,704), Devices (470,328), Observations (485,746), and Measurements (309,469). Each card has an icon representing the domain.

<https://athena.ohdsi.org>



Note: SOURCE_TO_CONCEPT_MAP is no longer an “official” vocabulary table since OMOP version 5.3

Three Things to Know about OMOP's Vocabulary Tables

1. “Standard” versus “Non-standard” concepts

- Defined by OHDSI for OMOP
- All *_concept_id columns should contain standard concepts (except observation_concept_id)
- The *_source_concept_id columns can contain standard or non-standard concepts

2. “2 billionaire” concepts

- OHDSI's ATHENA concepts have concept_id < 2,000,000,000
- Custom or source concepts have concept_id > 2,000,000,000
- To avoid ambiguity, we do not use concept_id = 2,000,000,000

3. “Everything is a concept”

- Healthcare codes
- Domains
- Vocabularies
- Concept classes
- Relationships
- Metadata like the OMOP CDM tables and columns

Mount Sinai's Additions to the OMOP Vocabulary Tables

OMOP Table	ATHENA Records	Epic Records	MSDW Records
DOMAIN	48	<i>n/a</i>	0
VOCABULARY	118	116	6
CONCEPT_CLASS	420	0	13
CONCEPT	8,135,987	2,096,839	1,122
RELATIONSHIP	672	<i>n/a</i>	2
CONCEPT_RELATIONSHIP	61,327,540	75,599,221	2,329,653

- ▶ We do not load all vocabularies from ATHENA, only the ones we use or might use
- ▶ We continue to add Epic vocabularies as we integrate more data
- ▶ We will add vocabularies for other sources/systems as we integrate more data
- ▶ MSDW-specific records are almost all metadata

DOMAIN

The Domain Catalog (vocabulary_id = 'Domain')

```
SELECT
  d.domain_id, d.domain_name, d.domain_concept_id
,c.concept_code, c.concept_name, c.domain_id, c.vocabulary_id, c.concept_class_id
FROM cdm_phi.domain d
JOIN cdm_phi.concept c ON d.domain_concept_id = c.concept_id
ORDER BY d.domain_id
```

	domain_id	domain_name	domain_concept_id	concept_code	concept_name	domain_id	vocabulary_id	concept_class_id
1	Condition	Condition	19	OMOP generated	Condition	Metadata	Domain	Domain
2	Condition Status	OMOP Condition Status	32889	OMOP generated	OMOP Condition Status	Metadata	Domain	Domain
3	Condition/Device	Condition/Device	235	OMOP generated	Condition/Device	Metadata	Domain	Domain
4	Condition/Drug	Condition/Drug	53	OMOP generated	Condition/Drug	Metadata	Domain	Domain
5	Condition/Meas	Condition/Measurement	47	OMOP generated	Condition/Measurement	Metadata	Domain	Domain
6	Condition/Obs	Condition/Observation	48	OMOP generated	Condition/Observation	Metadata	Domain	Domain
7	Condition/Procedure	Condition/Procedure	49	OMOP generated	Condition/Procedure	Metadata	Domain	Domain
8	Cost	Cost	581456	OMOP generated	Cost	Metadata	Domain	Domain
9	Currency	Currency	34	OMOP generated	Currency	Metadata	Domain	Domain
10	Device	Device	17	OMOP generated	Device	Metadata	Domain	Domain
11	Device/Drug	Device/Drug	50	OMOP generated	Device/Drug	Metadata	Domain	Domain
12	Device/Obs	Device/Observation	45	OMOP generated	Device/Observation	Metadata	Domain	Domain
13	Device/Procedure	Device/Procedure	41	OMOP generated	Device/Procedure	Metadata	Domain	Domain
14	Drug	Drug	13	OMOP generated	Drug	Metadata	Domain	Domain
15	Drug/Measurement	Drug/Measurement	51	OMOP generated	Drug/Measurement	Metadata	Domain	Domain
16	Drug/Obs	Drug/Observation	52	OMOP generated	Drug/Observation	Metadata	Domain	Domain
17	Drug/Procedure	Drug/Procedure	54	OMOP generated	Drug/Procedure	Metadata	Domain	Domain
18	Episode	Episode	32527	OMOP generated	Episode	Metadata	Domain	Domain
19	Ethnicity	Ethnicity	4	OMOP generated	Ethnicity	Metadata	Domain	Domain
20	Gender	Gender	2	OMOP generated	Gender	Metadata	Domain	Domain

Domains Indicate Standardized Data in OMOP

domain_id	OMOP Column	OMOP Table
Condition	condition_concept_id	condition_occurrence
Condition Status	Condition_status_concept_id	condition_occurrence
Drug	drug_concept_id	drug_exposure
Ethnicity	ethnicity_concept_id	person
Gender	gender_concept_id	person, provider
Measurement	measurement_concept_id	measurement
Observation	observation_concept_id	observation
Procedure	procedure_concept_id	procedure_occurrence
Provider	specialty_concept_id	provider
Race	race_concept_id	person
Route	route_concept_id	drug_exposure
Type Concept	*_type_concept_id	<all clinical data tables>
Unit	unit_concept_id	measurement, observation, specimen

VOCABULARY

The Vocabulary Catalog (vocabulary_id = 'Vocabulary')

```
SELECT
  v.vocabulary_id, v.vocabulary_name, v.vocabulary_concept_id
,c.concept_code, c.concept_name, c.domain_id, c.vocabulary_id, c.concept_class_id
FROM cdm_phi.vocabulary v
JOIN cdm_phi.concept c ON v.vocabulary_concept_id = c.concept_id
ORDER BY c.concept_class_id DESC, v.vocabulary_id
```

	vocabulary_id	vocabulary_name	vocabulary_concept_id	concept_code	concept_name	domain_id	vocabulary_id	concept_class_id
111	UK Biobank	UK Biobank	32976	OMOP generated	UK Biobank	Metadata	Vocabulary	Vocabulary
112	US Census	United States Census Bureau	32570	OMOP generated	United States Census Bureau	Metadata	Vocabulary	Vocabulary
113	VA Class	VA National Drug File Class (VA)	44819122	OMOP generated	VA National Drug File Class (VA)	Metadata	Vocabulary	Vocabulary
114	VA Product	VA National Drug File Product (VA)	44819120	OMOP generated	Veterans Health Administration National Drug File	Metadata	Vocabulary	Vocabulary
115	VANDF	Veterans Health Administration National Drug File	44819120	OMOP generated	Veterans Health Administration National Drug File	Metadata	Vocabulary	Vocabulary
116	Visit	OMOP Visit	44819119	OMOP generated	OMOP Visit	Metadata	Vocabulary	Vocabulary
117	Visit Type	OMOP Visit Type	44819150	OMOP generated	OMOP Visit Type	Metadata	Vocabulary	Vocabulary
118	Vocabulary	OMOP Vocabulary	44819232	OMOP generated	OMOP Vocabulary	Metadata	Vocabulary	Vocabulary
119	CDCREC	CDC Race & Ethnicity	2002067240	CDCREC	CDC Race & Ethnicity	Metadata	Vocabulary	MSDW Vocabulary
120	MSDW Attrib Catalog	MSDW Source Attribute Catalog	2000000235	MSDW Attrib Catalog	MSDW Source Attribute Catalog	Metadata	Vocabulary	MSDW Vocabulary
121	MSDW Condition Class	MSDW Condition Class	2000000234	MSDW Condition Class	MSDW Condition Class	Metadata	Vocabulary	MSDW Vocabulary
122	MSDW Src Record Type	MSDW Source Record Type	2000000233	MSDW Src Record Type	MSDW Source Record Type	Metadata	Vocabulary	MSDW Vocabulary
123	MSDW Substance Use	MSDW Substance Use	2000000237	MSDW Substance Use	MSDW Substance Use	Metadata	Vocabulary	MSDW Vocabulary
124	MSDW Val Set Catalog	MSDW Value Set Catalog	2000000236	MSDW Val Set Catalog	MSDW Value Set Catalog	Metadata	Vocabulary	MSDW Vocabulary
125	EPIC CLM 1017	EPIC ADMISSION TYPE	2000000228	CLM 1017	EPIC ADMISSION TYPE	Metadata	Vocabulary	Epic Vocabulary
126	EPIC CLM 1018	EPIC ADMISSION SOURCE	2000000187	CLM 1018	EPIC ADMISSION SOURCE	Metadata	Vocabulary	Epic Vocabulary
127	EPIC CLM 1021	EPIC DISCHARGE STATUS	2000000178	CLM 1021	EPIC DISCHARGE STATUS	Metadata	Vocabulary	Epic Vocabulary
128	EPIC DEP 110	EPIC DEPARTMENT SPECIALTY	2000000164	DEP 110	EPIC DEPARTMENT SPECIALTY	Metadata	Vocabulary	Epic Vocabulary

OHDSI's OMOP Vocabularies and “Catalogs”

```
SELECT
  v.vocabulary_id, v.vocabulary_name, v.vocabulary_concept_id
,c.concept_code, c.concept_name, c.domain_id, c.vocabulary_id, c.concept_class_id
FROM cdm_phi.vocabulary v
JOIN cdm_phi.concept c ON v.vocabulary_concept_id = c.concept_id
WHERE v.vocabulary_name LIKE 'OMOP%'
AND v.vocabulary_id NOT LIKE '%Type' --Vocabularies replaced by 'Type Concept'
ORDER BY v.vocabulary_id
```

	vocabulary_id	vocabulary_name	vocabulary_concept_id	concept_code	concept_name	domain_id	vocabulary_id	concept_class_id
1	CDM	OMOP Common DataModel	32485	OMOP generated	OMOP Common DataModel	Metadata	Vocabulary	Vocabulary
2	Concept Class	OMOP Concept Class	44819233	OMOP generated	OMOP Concept Class	Metadata	Vocabulary	Vocabulary
3	Condition Status	OMOP Condition Status	32887	OMOP generated	OMOP Condition Status	Metadata	Vocabulary	Vocabulary
4	Cost	OMOP Cost	581457	OMOP generated	OMOP Cost	Metadata	Vocabulary	Vocabulary
5	Domain	OMOP Domain	44819147	OMOP generated	OMOP Domain	Metadata	Vocabulary	Vocabulary
6	Episode	OMOP Episode	32523	OMOP generated	OMOP Episode	Metadata	Vocabulary	Vocabulary
7	Ethnicity	OMOP Ethnicity	44819134	OMOP generated	OMOP Ethnicity	Metadata	Vocabulary	Vocabulary
8	Gender	OMOP Gender	44819108	OMOP generated	OMOP Gender	Metadata	Vocabulary	Vocabulary
9	None	OMOP Standardized Vocabularies	44819096	OMOP generated	OMOP Standardized Vocabularies	Metadata	Vocabulary	Vocabulary
10	OMOP Extension	OMOP Extension (OHDSI)	32758	OMOP generated	OMOP Extension (OHDSI)	Metadata	Vocabulary	Vocabulary
11	OMOP Genomic	OMOP Genomic vocabulary	33002	OMOP generated	OMOP Genomic vocabulary	Metadata	Vocabulary	Vocabulary
12	OMOP Invest Drug	OMOP Investigational Drugs	33051	OMOP generated	OMOP Investigational Drugs	Metadata	Vocabulary	Vocabulary
13	Provider	OMOP Provider	32573	OMOP generated	OMOP Provider	Metadata	Vocabulary	Vocabulary
14	Relationship	OMOP Relationship	44819235	OMOP generated	OMOP Relationship	Metadata	Vocabulary	Vocabulary
15	Supplier	OMOP Supplier	32574	OMOP generated	OMOP Supplier	Metadata	Vocabulary	Vocabulary
16	Type Concept	OMOP Type Concept	32808	OMOP generated	OMOP Type Concept	Metadata	Vocabulary	Vocabulary
17	Visit	OMOP Visit	44819119	OMOP generated	OMOP Visit	Metadata	Vocabulary	Vocabulary
18	Vocabulary	OMOP Vocabulary	44819232	OMOP generated	OMOP Vocabulary	Metadata	Vocabulary	Vocabulary

Different Types of Vocabularies

Type of Vocabulary	Examples	
Standard healthcare code systems & terminologies	ICD10CM SNOMED LOINC	RxNorm CPT CVX
Vocabularies created by OHDSI for OMOP content	Ethnicity Gender	OMOP Genomic Visit
“Catalog” vocabularies created by OHDSI for OMOP metadata	CDM Concept Class Domain	Relationship Vocabulary
Our Epic master files & category lists	EPIC EAP .1 EPIC EDG .1 EPIC ERX .1	EPIC EPT 145 EPIC ORD 835
Vocabularies created by us for derived data elements	MSDW Condition Class	MSDW Substance Use
Vocabularies created by us for data lineage & better OMOP metadata	MSDW Attrib Catalog MSDW Src Record Type	MSDW Val Set Catalog
Special “placeholder” vocabulary to store ATHENA version (vocabulary_version) & concept_id = 0 record	vocabulary_id = 'None'	

CONCEPT_CLASS

The Concept Class Catalog (vocabulary_id = 'Concept Class')

```
SELECT
  cl.concept_class_id, cl.concept_class_name, cl.concept_class_concept_id
, cn.concept_code, cn.concept_name, cn.domain_id, cn.vocabulary_id, cn.concept_class_id
FROM cdm_phi.concept_class cl
JOIN cdm_phi.concept cn ON cl.concept_class_concept_id = cn.concept_id
ORDER BY cn.concept_class_id DESC, cl.concept_class_id
```

	concept_class_id	concept_class_name	concept_class_concept_id	concept_code	concept_name	domain_id	vocabulary_id	concept_class_id
1	Code Sys Value Set	Healthcare Code System Value Set	2002067229	Code Sys Value Set	Healthcare Code System Value Set	Metadata	Concept Class	Custom Concept Class
2	Custom Concept Class	Custom Concept Class	2000000001	Custom Concept Class	Custom Concept Class	Metadata	Concept Class	Custom Concept Class
3	Custom Relationship	Custom Relationship	2002057127	Custom Relationship	Custom Relationship	Metadata	Concept Class	Custom Concept Class
4	Epic Attribute	Epic Attribute	2000000002	Epic Attribute	Epic Attribute	Metadata	Concept Class	Custom Concept Class
5	Epic Cat Value Set	Epic Category Value Set	2002067228	Epic Cat Value Set	Epic Category Value Set	Metadata	Concept Class	Custom Concept Class
6	Epic Concept	Epic Concept	2000000008	Epic Concept	Epic Concept	Metadata	Concept Class	Custom Concept Class
7	Epic Null Concept	Epic Null Concept	2000000003	Epic Null Concept	Epic Null Concept	Metadata	Concept Class	Custom Concept Class
8	Epic VCG Grouper	Epic VCG Grouper	2000000009	Epic VCG Grouper	Epic VCG Grouper	Metadata	Concept Class	Custom Concept Class
9	Epic Vocabulary	Epic Vocabulary	2000000005	Epic Vocabulary	Epic Vocabulary	Metadata	Concept Class	Custom Concept Class
10	MSDW Concept	MSDW Concept	2000000007	MSDW Concept	MSDW Concept	Metadata	Concept Class	Custom Concept Class
11	MSDW Value Set	MSDW Value Set	2000000004	MSDW Value Set	MSDW Value Set	Metadata	Concept Class	Custom Concept Class
12	MSDW Vocabulary	MSDW Vocabulary	2000000006	MSDW Vocabulary	MSDW Vocabulary	Metadata	Concept Class	Custom Concept Class
13	Unit	Unit	44819072	OMOP generated	Unit	Metadata	Concept Class	Concept Class
14	Units of Measure	Units of Measure	45905732	OMOP generated	Units of Measure	Metadata	Concept Class	Concept Class
15	US Census Division	United States Census Bureau Division	32572	OMOP generated	United States Census Bureau Division	Metadata	Concept Class	Concept Class
16	US Census Region	United States Census Bureau Region	32571	OMOP generated	United States Census Bureau Region	Metadata	Concept Class	Concept Class
17	VA Class	VA Class	44819277	OMOP generated	VA Class	Metadata	Concept Class	Concept Class
18	VA Product	VA Product	44819278	OMOP generated	VA Product	Metadata	Concept Class	Concept Class
19	Vaccine	Human drug class Vaccine	45754855	OMOP generated	Human drug class Vaccine	Metadata	Concept Class	Concept Class

Concept Classes Stratify Vocab and/or Classify Concepts

```
SELECT
cn.concept_class_id
, COUNT(DISTINCT cn.vocabulary_id) AS
vocabulary_count
, MIN(cn.vocabulary_id) AS first_vocabulary
, MAX(cn.vocabulary_id) AS last_vocabulary
, COUNT(*) AS concept_count
FROM cdm_phi.concept cn
GROUP BY
cn.concept_class_id
HAVING
COUNT(DISTINCT cn.vocabulary_id) > 1
ORDER BY
COUNT(DISTINCT cn.vocabulary_id) DESC
, cn.concept_class_id
```

- ▶ What is a concept class?
 - Stratify vocabulary
 - Classify concepts
- ▶ Most stratify a single vocabulary
- ▶ Some classify concepts across multiple vocabularies

	concept_class_id	vocabulary_count	first_vocabulary	last_vocabulary	concept_count
1	Epic Concept	109	EPIC CLM 1017	EPIC SYSTOLIC BP	2087506
2	Epic Null Concept	95	EPIC CLM 1017	EPIC SER 6000	380
3	Procedure	11	CCAM	SNOMED Veterinary	257671
4	Ingredient	10	BDPM	SPL	87621
5	Dose Form	9	BDPM	SNOMED	3018
6	Brand Name	8	BDPM	RxNorm Extension	180406
7	Epic VCG Grouper	7	EPIC VCG .1 EAP	EPIC VCG .1 ORP	8953
8	Qualifier Value	7	Cancer Modifier	SNOMED Veterinary	34856
9	Supplier	7	AMT	RxNorm Extension	14339
10	Visit	7	CMS Place of Service	Visit	629
11	Variant	6	CGI	OncoKB	23334
12	Attribute	5	Nebraska Lexicon	SNOMED Veterinary	2563
13	Clinical Drug	5	DPD	SNOMED	155228
14	Clinical Drug Form	5	DPD	SNOMED	74109
15	Clinical Finding	5	Nebraska Lexicon	SNOMED Veterinary	362310
16	Device	5	BDPM	SPL	15902
17	Drug Product	5	BDPM	VANDF	135310
18	Morph Abnormality	5	Cancer Modifier	SNOMED Veterinary	12615
19	Observable Entity	5	Nebraska Lexicon	SNOMED Veterinary	28415
20	Physician Specialty	5	ABMS	Provider	553
21	Provider	5	HES Specialty	Supplier	461
22	Social Context	5	Nebraska Lexicon	SNOMED Veterinary	16974
23	Specimen	5	CIEL	SNOMED Veterinary	4038
24	Substance	5	Nebraska Lexicon	SNOMED Veterinary	64231

Concept Class Example: Finding Genetic Variant Concepts

```
SELECT
  c.concept_class_id
,c.vocabulary_id
,v.vocabulary_name
,COUNT(*) AS concept_count
FROM cdm_phi.concept c
JOIN cdm_phi.vocabulary v
ON c.vocabulary_id = v.vocabulary_id
WHERE c.concept_class_id IN (
  'Variant'
, 'DNA Variant'
, 'Genetic Variation'
, 'Protein Variant'
, 'RNA Variant'
)
GROUP BY
  c.concept_class_id
,c.vocabulary_id
,v.vocabulary_name
ORDER BY
  c.concept_class_id
,c.vocabulary_id
```

	concept_class_id	vocabulary_id	vocabulary_name	concept_count
1	DNA Variant	OMOP Genomic	OMOP Genomic vocabulary	27678
2	Genetic Variation	OMOP Genomic	OMOP Genomic vocabulary	19297
3	Protein Variant	OMOP Genomic	OMOP Genomic vocabulary	33858
4	RNA Variant	OMOP Genomic	OMOP Genomic vocabulary	40158
5	Variant	CGI	Cancer Genome Interpreter (Pompeu Fabra University)	323
6	Variant	CIViC	Clinical Interpretation of Variants in Cancer (civcdb.org)	988
7	Variant	ClinVar	ClinVar (NCBI)	8072
8	Variant	JAX	The Clinical Knowledgebase (The Jackson Laboratory)	7855
9	Variant	NCIt	NCI Thesaurus (National Cancer Institute)	527
10	Variant	OncoKB	Oncology Knowledge Base (MSK)	5569

CONCEPT

The Concept Table

```

SELECT
  vc.concept_class_id          AS vocabulary_concept_class_id
, COUNT(DISTINCT c.domain_id) AS domain_count
, COUNT(DISTINCT c.vocabulary_id) AS vocabulary_count
, COUNT(DISTINCT c.concept_class_id) AS concept_class_count
, (CASE WHEN c.standard_concept IN ('S', 'C') THEN c.standard_concept ELSE '' END) AS standard_concept
, (CASE WHEN c.invalid_reason IN ('D', 'U') THEN c.invalid_reason ELSE '' END) AS invalid_reason
, COUNT(*)                    AS concept_count
FROM cdm_phi.concept c
JOIN cdm_phi.vocabulary v ON c.vocabulary_id = v.vocabulary_id
JOIN cdm_phi.concept vc ON v.vocabulary_concept_id = vc.concept_id
GROUP BY vc.concept_class_id
, (CASE WHEN c.standard_concept IN ('S', 'C') THEN c.standard_concept ELSE '' END)
, (CASE WHEN c.invalid_reason IN ('D', 'U') THEN c.invalid_reason ELSE '' END)
ORDER BY vc.concept_class_id
, (CASE WHEN c.standard_concept IN ('S', 'C') THEN c.standard_concept ELSE '' END)
, (CASE WHEN c.invalid_reason IN ('D', 'U') THEN c.invalid_reason ELSE '' END)

```

	vocabulary_concept_class_id	domain_count	vocabulary_count	concept_class_count	standard_concept	invalid_reason	concept_count
1	Epic Vocabulary	18	116	3			2096839
2	MSDW Vocabulary	6	6	4			1122
3	Vocabulary	25	80	200			2524636
4	Vocabulary	27	53	166		D	1066419
5	Vocabulary	21	21	79		U	460227
6	Vocabulary	7	12	36	C		764500
7	Vocabulary	32	61	153	S		3320205

Different Types of Concepts

Types of Concept	Concept Record Condition
OHDSI assigns “durable” concept_id values	<code>concept_id < 2,000,000,000</code>
OHDSI defines which concepts are standard	<code>standard_concept = 'S'</code>
OHDSI defines which concepts are classifications	<code>standard_concept = 'C'</code>
OHDSI deletes concepts only logically, not physically	<code>invalid_reason = 'D'</code>
Source data have concept_id values above 2 billion	<code>concept_id > 2,000,000,000</code>
CONCEPT table contains all metadata	<code>domain_id = 'Metadata'</code>
Special “placeholder” concept for unknown/unmapped	<code>concept_id = 0</code>

Concept Example:

Type 2 Diabetes Mellitus with Ophthalmic Complications (E11.3)

```
SELECT c.concept_id, c.vocabulary_id, c.concept_code, c.concept_name
,c.standard_concept, c.concept_class_id, cr.relationship_id
FROM cdm_phi.concept c
JOIN cdm_phi.concept_relationship cr ON c.concept_id = cr.concept_id_1
WHERE cr.concept_id_2 = 1567959 --ICD-10 code E11.3
ORDER BY cr.relationship_id DESC, c.concept_code
```

	concept_id	vocabulary_id	concept_code	concept_name	standard_concept	concept_class_id	relationship_id
1	1567956	ICD10CM	E11	Type 2 diabetes mellitus	NULL	3-char nonbill code	Subsumes
2	2001147567	EPIC EDG 1	546229	Type 2 diabetes mellitus with ophthalmic complications	NULL	Epic Concept	Maps to non-standard
3	443733	SNOMED	422099009	Disorder of eye due to type 2 diabetes mellitus	S	Clinical Finding	Mapped from
4	1567960	ICD10CM	E11.31	Type 2 diabetes mellitus with unspecified diabetic retinopathy	NULL	5-char nonbill code	Is a
5	45581352	ICD10CM	E11.311	Type 2 diabetes mellitus with unspecified diabetic retinopathy with macular edema	NULL	6-char billing code	Is a
6	45581353	ICD10CM	E11.319	Type 2 diabetes mellitus with unspecified diabetic retinopathy without macular edema	NULL	6-char billing code	Is a
7	1567961	ICD10CM	E11.32	Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy	NULL	5-char nonbill code	Is a
8	45552355	ICD10CM	E11.321	Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy with macular edema	NULL	6-char nonbill code	Is a
9	37200198	ICD10CM	E11.3211	Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy with macular edema, right eye	NULL	7-char billing code	Is a
10	37200199	ICD10CM	E11.3212	Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy with macular edema, left eye	NULL	7-char billing code	Is a
11	37200200	ICD10CM	E11.3213	Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy with macular edema, bilateral	NULL	7-char billing code	Is a
12	37200201	ICD10CM	E11.3219	Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy with macular edema, unspecified eye	NULL	7-char billing code	Is a
13	45581354	ICD10CM	E11.329	Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema	NULL	6-char nonbill code	Is a
14	37200202	ICD10CM	E11.3291	Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema, right eye	NULL	7-char billing code	Is a
15	37200203	ICD10CM	E11.3292	Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema, left eye	NULL	7-char billing code	Is a
16	37200204	ICD10CM	E11.3293	Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema, bilateral	NULL	7-char billing code	Is a
17	37200205	ICD10CM	E11.3299	Type 2 diabetes mellitus with mild nonproliferative diabetic retinopathy without macular edema, unspecified eye	NULL	7-char billing code	Is a
18	1567962	ICD10CM	E11.33	Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy	NULL	5-char nonbill code	Is a
19	45537961	ICD10CM	E11.331	Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy with macular edema	NULL	6-char nonbill code	Is a
20	37200206	ICD10CM	E11.3311	Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy with macular edema, right eye	NULL	7-char billing code	Is a
21	37200207	ICD10CM	E11.3312	Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy with macular edema, left eye	NULL	7-char billing code	Is a
22	37200208	ICD10CM	E11.3313	Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy with macular edema, bilateral	NULL	7-char billing code	Is a
23	37200209	ICD10CM	E11.3319	Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy with macular edema, unspecified eye	NULL	7-char billing code	Is a
24	45591029	ICD10CM	E11.339	Type 2 diabetes mellitus with moderate nonproliferative diabetic retinopathy without macular edema	NULL	6-char nonbill code	Is a

RELATIONSHIP

The Relationship Catalog (vocabulary_id = 'Relationship')

```
SELECT
  r.relationship_id, r.relationship_name, r.reverse_relationship_id, ,r.relationship_concept_id
,c.concept_code, c.concept_name, c.domain_id, c.vocabulary_id, c.concept_class_id
FROM cdm_phi.relationship r
JOIN cdm_phi.concept c ON r.relationship_concept_id = c.concept_id
ORDER BY
  c.concept_class_id, r.relationship_id
```

	relationship_id	relationship_name	reverse_relationship_id	relationship_concept_id	concept_code	concept_name	domain_id	vocabulary_id	concept_class_id
1	Mapped from non-std	Non-standard to source concept map (MSDW)	Maps to non-standard	2002057129	Mapped from non-std	Non-standard to source concept map (MSDW)	Metadata	Relationship	Custom Relationship
2	Maps to non-standard	Source concept to non-standard map (MSDW)	Mapped from non-std	2002057128	Maps to non-standard	Source concept to non-standard map (MSDW)	Metadata	Relationship	Custom Relationship
3	AB-drug cigt of	Antibody-drug conjugate of (HemOnc)	Has AB-drug cigt	33016	OMOP generated	Antibody-drug conjugate of (HemOnc)	Metadata	Relationship	Relationship
4	Absorbability of	Absorbability of (SNOMED)	Has absorbability	33055	OMOP generated	Absorbability of (SNOMED)	Metadata	Relationship	Relationship
5	Acc device used by	Access device used by (SNOMED)	Using acc device	44818895	OMOP generated	Access device used by (SNOMED)	Metadata	Relationship	Relationship
6	Accepted use of	Accepted use of (HemOnc)	Has accepted use	32787	OMOP generated	Accepted use of (HemOnc)	Metadata	Relationship	Relationship
7	Access of	Access of (SNOMED)	Has access	44818847	OMOP generated	Access of (SNOMED)	Metadata	Relationship	Relationship
8	Active ing of	Active ingredient of (SNOMED)	Has active ing	44818859	OMOP generated	Active ingredient of (SNOMED)	Metadata	Relationship	Relationship
9	Add monitor ind of	NHS dm+d (dictionary of medicines and devices) ad...	Has add monitor ind	32748	OMOP generated	NHS dm+d (dictionary of medicines and devices) ad...	Metadata	Relationship	Relationship
10	Admin method of	Dose form administration method of (SNOMED)	Has admin method	581464	OMOP generated	Dose form administration method of (SNOMED)	Metadata	Relationship	Relationship
11	AMP of	Actual Medicinal Product of (SNOMED)	Has AMP	45905757	OMOP generated	Actual Medicinal Product of (SNOMED)	Metadata	Relationship	Relationship
12	AMP restr ind of	NHS dm+d (dictionary of medicines and devices) AM...	Has AMP restr ind	32750	OMOP generated	NHS dm+d (dictionary of medicines and devices) AM...	Metadata	Relationship	Relationship
13	Answer of	Answer of	Has Answer	45754684	OMOP generated	Answer of	Metadata	Relationship	Relationship
14	Answer of (PPI)	Answer of (PPI)	Has answer (PPI)	32443	OMOP generated	Answer of (PPI)	Metadata	Relationship	Relationship
15	Antineoplastic of	Antineoplastic of (HemOnc)	Has antineoplastic	32617	OMOP generated	Antineoplastic of (HemOnc)	Metadata	Relationship	Relationship
16	Approach of	Approach of (SNOMED)	Has approach	45754821	OMOP generated	Approach of (SNOMED)	Metadata	Relationship	Relationship
17	ARP of	ARP of (SNOMED)	Has ARP	45905797	OMOP generated	ARP of (SNOMED)	Metadata	Relationship	Relationship
18	Asso finding of	Associated finding of (SNOMED)	Has asso finding	44818868	OMOP generated	Associated finding of (SNOMED)	Metadata	Relationship	Relationship
19	Asso morph of	Associated morphology of (SNOMED)	Has asso morph	44818867	OMOP generated	Associated morphology of (SNOMED)	Metadata	Relationship	Relationship
20	Asso proc of	Associated procedure of (SNOMED)	Has asso proc	44818882	OMOP generated	Associated procedure of (SNOMED)	Metadata	Relationship	Relationship
21	Asso with finding	Associated with finding (SNOMED)	Finding asso with	44818792	OMOP generated	Associated with finding (SNOMED)	Metadata	Relationship	Relationship

Bi-directional Mapping Relationships

```
SELECT
  r.relationship_id, r.relationship_name, r.reverse_relationship_id, r.relationship_concept_id
, COUNT(*) AS relationship_count
FROM cdm_phi.relationship r
LEFT JOIN cdm_phi.concept_relationship cr ON r.relationship_id = cr.relationship_id
WHERE r.relationship_id LIKE '%map%'
GROUP BY r.relationship_id, r.relationship_name, r.reverse_relationship_id, r.relationship_concept_id
ORDER BY COUNT(*) DESC, r.relationship_id
```

	relationship_id	relationship_name	reverse_relationship_id	relationship_concept_id	relationship_count
1	Maps to	Non-standard to Standard map (OMOP)	Mapped from	44818977	7354102
2	Mapped from	Standard to Non-standard map (OMOP)	Maps to	44818976	7353914
3	Mapped from non-std	Non-standard to source concept map (MSDW)	Maps to non-standard	2002057129	3197986
4	Maps to non-standard	Source concept to non-standard map (MSDW)	Mapped from non-std	2002057128	3197986
5	Maps to value	Non-standard to value_as_concept map (OMOP)	Value mapped from	44819239	10090
6	Value mapped from	Value_as_concept to non-standard map (OMOP)	Maps to value	44819242	10090
7	Maps to unit	Non-standard to unit concept map (OMOP)	Unit mapped from	44819238	3458
8	Unit mapped from	Unit concept to non-standard map (OMOP)	Maps to unit	44819241	3456
9	Maps to number	Non-standard to value_as_number map (OMOP)	Number mapped from	45754885	48
10	Maps to operator	Non-standard to operator concept map (OMOP)	Operator mapped from	45754883	48
11	Number mapped from	Value_as_number to non-standard map (OMOP)	Maps to number	45754886	48
12	Operator mapped from	Operator concept to non-standard map (OMOP)	Maps to operator	45754884	48
13	Excluded in map from	Standard to Non-standard exclude child (OMOP)	Map excludes child	44818970	1
14	Included in map from	Standard to Non-standard include child (OMOP)	Map includes child	44818973	1
15	Map excludes child	Non-standard to Standard exclude child (OMOP)	Excluded in map from	44818974	1
16	Map includes child	Non-standard to Standard include child (OMOP)	Included in map from	44818975	1
17	Maps to type	Non-standard to type concept map (OMOP)	Type mapped from	44819237	1
18	Type mapped from	Type concept to non-standard map (OMOP)	Maps to type	44819240	1

CONCEPT_RELATIONSHIP

Most Concept Relationships are Hierarchies & Value Sets

```
SELECT
  COUNT(DISTINCT c1.vocabulary_id) AS
distinct_from_vocabs
,r.relationship_id
,COUNT(DISTINCT c2.vocabulary_id) AS
distinct_to_vocabs
,COUNT(*) AS relationship_count
FROM cdm_phi.concept_relationship r
JOIN cdm_phi.concept c1
ON r.concept_id_1 = c1.concept_id
JOIN cdm_phi.concept c2
ON r.concept_id_2 = c2.concept_id
GROUP BY
r.relationship_id
ORDER BY
  COUNT(*) DESC
,r.relationship_id DESC
```

- ▶ Parent-Child relationships used for:
 - Hierarchies
 - Value sets
- ▶ Child code “Is a” parent code
- ▶ Parent code “Subsumes” child code

	distinct_from_vocabs	relationship_id	distinct_to_vocabs	relationship_count
1	63	Is a	58	38255719
2	53	Subsumes	62	38255352
3	165	Maps to	77	7354102
4	77	Mapped from	160	7353912
5	5	Maps to non-standard	4	3197986
6	4	Mapped from non-std	5	3197986
7	1	Marketed form of	2	1941193
8	2	Has marketed form	1	1941193
9	2	RxNorm has dose form	2	1762444
10	2	RxNorm dose form of	2	1762444
11	3	Has brand name	3	1451684
12	3	Brand name of	3	1451684
13	1	Module of	1	1057011
14	1	Has Module	1	1057011
15	1	Status of	1	1054577
16	1	Has status	1	1054577
17	2	Tradename of	2	1046200
18	2	Has tradename	2	1046200
19	1	SPL - RxNorm	1	638620
20	1	RxNorm - SPL	1	638620
21	1	Supplier of	1	557946
22	1	Has supplier	1	557946
23	1	Value to Schema	1	545166
24	1	Schema to Value	1	545166
25	1	Box of	2	501844
26	2	Available as box	1	501844
27	2	RxNorm is a	2	475941
28	2	RxNorm inverse is a	2	475941
29	2	Constitutes	2	432985
30	2	Consists of	2	432985

Value Sets of Epic EHR Data

```
SELECT
  c1.vocabulary_id
, v.vocabulary_name
, r.relationship_id
, c2.vocabulary_id
, COUNT(*) AS relationship_count
FROM cdm_phi.concept_relationship r
JOIN cdm_phi.concept c1
ON r.concept_id_1 = c1.concept_id
JOIN cdm_phi.vocabulary v
ON c1.vocabulary_id = v.vocabulary_id
JOIN cdm_phi.concept c2
ON r.concept_id_2 = c2.concept_id
WHERE r.relationship_id IN ('Is a', 'Subsumes')
AND (r.concept_id_1 > 2000000000
OR r.concept_id_2 > 2000000000)
GROUP BY
  c1.vocabulary_id
, v.vocabulary_name
, r.relationship_id
, c2.vocabulary_id
ORDER BY
  COUNT(*) DESC
, c1.vocabulary_id
, r.relationship_id
, c2.vocabulary_id
```

	vocabulary_id	vocabulary_name	relationship_id	vocabulary_id	relationship_count
1	EPIC EDG .1	EPIC DIAGNOSIS	Is a	EPIC VCG .1 EDG	32811927
2	EPIC VCG .1 EDG	EPIC DIAGNOSIS GROUPER	Subsumes	EPIC EDG .1	32811927
3	EPIC ERX .1	EPIC MEDICATION	Is a	EPIC VCG .1 ERX	586412
4	EPIC VCG .1 ERX	EPIC MEDICATION GROUPER	Subsumes	EPIC ERX .1	586412
5	EPIC EAP .1	EPIC PROCEDURE	Is a	EPIC VCG .1 EAP	30169
6	EPIC VCG .1 EAP	EPIC PROCEDURE GROUPER	Subsumes	EPIC EAP .1	30169
7	EPIC LRR .1	EPIC LAB COMPONENT	Is a	EPIC VCG .1 LRR	4021
8	EPIC VCG .1 LRR	EPIC LAB COMPONENT GROUPER	Subsumes	EPIC LRR .1	4021
9	EPIC ORP .1	EPIC SURGICAL PROCEDURE	Is a	EPIC VCG .1 ORP	3484
10	EPIC VCG .1 ORP	EPIC SURGICAL PROCEDURE GROUPER	Subsumes	EPIC ORP .1	3484
11	EPIC LIM .1	EPIC IMMUNIZATION	Is a	EPIC VCG .1 LIM	394
12	EPIC VCG .1 LIM	EPIC IMMUNIZATION GROUPER	Subsumes	EPIC LIM .1	394
13	EPIC FLO .1	EPIC FLOWSHEET MEASURE	Is a	EPIC VCG .1 FLO	391
14	EPIC VCG .1 FLO	EPIC FLOWSHEET MEASURE GROUPER	Subsumes	EPIC FLO .1	391
15	EPIC EPT 30	EPIC CONTACT TYPE	Is a	MSDW Val Set Catalog	265
16	EPIC PRC .1	EPIC VISIT TYPE	Is a	MSDW Val Set Catalog	215
17	MSDW Val Set Catalog	MSDW Value Set Catalog	Subsumes	EPIC PRC .1	215
18	EPIC FLO .1	EPIC FLOWSHEET MEASURE	Is a	MSDW Val Set Catalog	190
19	MSDW Val Set Catalog	MSDW Value Set Catalog	Subsumes	EPIC FLO .1	184
20	MSDW Val Set Catalog	MSDW Value Set Catalog	Subsumes	EPIC EPT 30	134
21	MSDW Val Set Catalog	MSDW Value Set Catalog	Is a	EPIC EPT 30	121
22	MSDW Val Set Catalog	MSDW Value Set Catalog	Is a	EPIC FLO .1	39
23	EPIC EPT 101	EPIC PATIENT TYPE(S)	Is a	MSDW Val Set Catalog	22
24	EPIC DEP 110	EPIC DEPARTMENT SPECIALTY	Is a	MSDW Val Set Catalog	20
25	EPIC ECT 34886	EPIC CLINICAL SERVICE	Is a	MSDW Val Set Catalog	19
26	MSDW Val Set Catalog	MSDW Value Set Catalog	Subsumes	EPIC ECT 34886	19
27	EPIC INP 5010	EPIC NOTE TYPE	Is a	MSDW Val Set Catalog	18
28	MSDW Val Set Catalog	MSDW Value Set Catalog	Is a	EPIC EPT 101	11
29	MSDW Val Set Catalog	MSDW Value Set Catalog	Subsumes	EPIC EPT 101	11
30	MSDW Val Set Catalog	MSDW Value Set Catalog	Is a	EPIC DEP 110	10
31	MSDW Val Set Catalog	MSDW Value Set Catalog	Subsumes	EPIC DEP 110	10
32	MSDW Val Set Catalog	MSDW Value Set Catalog	Is a	EPIC INP 5010	9
33	MSDW Val Set Catalog	MSDW Value Set Catalog	Subsumes	EPIC INP 5010	9

Most Relationship Types are used within 1 Vocabulary

```
SELECT
  c1.vocabulary_id
, COUNT(DISTINCT r.relationship_id) AS
  distinct_relationships
, c2.vocabulary_id
, COUNT(*) AS relationship_count
FROM cdm_phi.concept_relationship r
JOIN cdm_phi.concept c1
ON r.concept_id_1 = c1.concept_id
JOIN cdm_phi.concept c2
ON r.concept_id_2 = c2.concept_id
GROUP BY
  c1.vocabulary_id
, c2.vocabulary_id
ORDER BY
  COUNT(DISTINCT r.relationship_id) DESC
, c1.vocabulary_id
, c2.vocabulary_id
, COUNT(*) DESC
```

	vocabulary_id	distinct_relationships	vocabulary_id	relationship_count
1	SNOMED	290	SNOMED	16499028
2	Nebraska Lexicon	198	Nebraska Lexicon	2425888
3	HemOnc	54	HemOnc	81808
4	SNOMED	47	SNOMED Veterinary	30405
5	SNOMED Veterinary	47	SNOMED	30405
6	SNOMED Veterinary	46	SNOMED Veterinary	107208
7	NDFRT	44	NDFRT	735902
8	LOINC	32	LOINC	2252188
9	ICD9Proc	30	SNOMED	19256
10	SNOMED	30	ICD9Proc	19256
11	RxNorm	28	RxNorm	2101218
12	RxNorm Extension	26	RxNorm Extension	14659520
13	ICD10PCS	25	SNOMED	608956
14	OPCS4	25	SNOMED	17125
15	SNOMED	25	ICD10PCS	608956
16	SNOMED	25	OPCS4	17125
17	NAACCR	16	NAACCR	1447628
18	RxNorm	16	RxNorm Extension	2898671
19	RxNorm Extension	16	RxNorm	2898671
20	HemOnc	14	RxNorm	13534
21	RxNorm	14	HemOnc	13534
22	PPI	12	PPI	42708
23	UK Biobank	12	UK Biobank	61174
24	Cancer Modifier	10	Cancer Modifier	27502
25	OMOP Extension	10	OMOP Extension	4500
26	ATC	9	RxNorm	131968
27	HemOnc	9	RxNorm Extension	196
28	RxNorm	9	ATC	131968
29	RxNorm Extension	9	HemOnc	196
30	ATC	8	RxNorm Extension	175093

Concept Relationships Example: Many in SNOMED-CT

```
SELECT
  c1.vocabulary_id
, cr.relationship_id
, r.relationship_name
, c2.vocabulary_id
, COUNT(*) AS relationship_count
FROM cdm_phi.concept_relationship cr
JOIN cdm_phi.concept c1
ON cr.concept_id_1 = c1.concept_id
JOIN cdm_phi.concept c2
ON cr.concept_id_2 = c2.concept_id
JOIN cdm_phi.relationship r
ON cr.relationship_id = r.relationship_id
WHERE c1.vocabulary_id = 'SNOMED'
AND c2.vocabulary_id = 'SNOMED'
GROUP BY
  c1.vocabulary_id
, cr.relationship_id
, r.relationship_name
, c2.vocabulary_id
ORDER BY
  COUNT(*) DESC
```

	vocabulary_id	relationship_id	relationship_name	vocabulary_id	relationship_count
1	SNOMED	Is a	Is a	SNOMED	1579779
2	SNOMED	Subsumes	Subsumes	SNOMED	1579779
3	SNOMED	Module of	Module of	SNOMED	1057011
4	SNOMED	Has Module	Has Module	SNOMED	1057011
5	SNOMED	Status of	Status of	SNOMED	1054577
6	SNOMED	Has status	Has status	SNOMED	1054577
7	SNOMED	Mapped from	Standard to Non-standard map (OMOP)	SNOMED	700831
8	SNOMED	Maps to	Non-standard to Standard map (OMOP)	SNOMED	700831
9	SNOMED	Non-avail ind of	VMP non-availability indicator of (SNOMED)	SNOMED	367188
10	SNOMED	Has non-avail ind	Has VMP non-availability indicator (SNOMED)	SNOMED	367188
11	SNOMED	VMP has prescr stat	VMP has prescribing status (SNOMED)	SNOMED	364875
12	SNOMED	VMP prescr stat of	VMP prescribing status of (SNOMED)	SNOMED	364875
13	SNOMED	CD category of	Controlled drug catheory of (DM+D)	SNOMED	359469
14	SNOMED	Has CD category	Has controlled drug category (DM+D)	SNOMED	359469
15	SNOMED	Has AMP restr ind	Has NHS dm+d (dictionary of medicines and de...	SNOMED	305409
16	SNOMED	AMP restr ind of	NHS dm+d (dictionary of medicines and device...	SNOMED	305409
17	SNOMED	Has spec active ing	Has specific active ingredient (SNOMED)	SNOMED	223388
18	SNOMED	Spec active ing of	Specific active ingredient of (SNOMED)	SNOMED	223388
19	SNOMED	Has disp dose form	Has dispensed dose form (SNOMED)	SNOMED	188527
20	SNOMED	Disp dose form of	Dispensed dose form of (SNOMED)	SNOMED	188527
21	SNOMED	Disc indicator of	Discontinued indicator of (SNOMED)	SNOMED	170034
22	SNOMED	Has disc indicator	Has discontinued indicator (SNOMED)	SNOMED	170034
23	SNOMED	Has AMP	Has Actual Medicinal Product (SNOMED)	SNOMED	163718
24	SNOMED	AMP of	Actual Medicinal Product of (SNOMED)	SNOMED	163718
25	SNOMED	Legal category of	Legal category of (SNOMED)	SNOMED	161963
26	SNOMED	Has legal category	Has legal category (SNOMED)	SNOMED	161963

Different Types of Concept Relationships

Types of Concept Relationship	Concept Relationship Types
Parent-Child relationships for hierarchies & value sets	<code>relationship_id = 'Is a'</code> <code>relationship_id = 'Subsumes'</code>
Mapping from source records to OMOP standard codes	<code>relationship_id = 'Maps to'</code> <code>relationship_id = 'Mapped from'</code>
Mapping from source records to OMOP units	<code>relationship_id = 'Maps to unit'</code> <code>relationship_id = 'Unit mapped from'</code>
Mapping from source records to OMOP operators	<code>relationship_id = 'Maps to operator'</code> <code>relationship_id = 'Operator mapped from'</code>
Mapping from source records to OMOP numbers	<code>relationship_id = 'Maps to number'</code> <code>relationship_id = 'Number mapped from'</code>
Mapping from Epic EHR data to “non-standard” concepts (according to OHDSI) **	<code>relationship_id = 'Maps to non-standard'</code> <code>relationship_id = 'Mapped from non-std'</code>

** Note: We added these two custom relationships specifically for MSDW.

Mapping Epic EHR Data to OMOP Concepts

```
SELECT
  c1.vocabulary_id
, v.vocabulary_name
, r.relationship_id
, c2.vocabulary_id
, COUNT(*) AS relationship_count
FROM cdm_phi.concept_relationship r
JOIN cdm_phi.concept c1
ON r.concept_id_1 = c1.concept_id
JOIN cdm_phi.vocabulary v
ON c1.vocabulary_id = v.vocabulary_id
JOIN cdm_phi.concept c2
ON r.concept_id_2 = c2.concept_id
WHERE r.relationship_id LIKE 'Maps to%'
AND c2.vocabulary_id != 'None'
AND c1.vocabulary_id LIKE 'EPIC%'
GROUP BY
  c1.vocabulary_id
, v.vocabulary_name
, r.relationship_id
, c2.vocabulary_id
ORDER BY
  COUNT(*) DESC
, c1.vocabulary_id
, r.relationship_id
, c2.vocabulary_id
```

	vocabulary_id	vocabulary_name	relationship_id	vocabulary_id	relationship_count
1	EPIC EDG .1	EPIC DIAGNOSIS	Maps to non-standard	ICD9CM	1686866
2	EPIC EDG .1	EPIC DIAGNOSIS	Maps to non-standard	ICD10CM	1365231
3	EPIC EDG .1	EPIC DIAGNOSIS	Maps to	SNOMED	1106458
4	EPIC ERX .1	EPIC MEDICATION	Maps to non-standard	NDC	145705
5	EPIC EAP .1	EPIC PROCEDURE	Maps to	CPT4	17250
6	EPIC LRR .1	EPIC LAB COMPONENT	Maps to	LOINC	12793
7	EPIC EAP .1	EPIC PROCEDURE	Maps to	HCPCS	9016
8	EPIC ERX .1	EPIC MEDICATION	Maps to	RxNorm	7213
9	EPIC ORP .1	EPIC SURGICAL PROCEDURE	Maps to	CPT4	6096
10	EPIC EPT 19360	EPIC FAMILY MEDICAL HISTORY	Maps to	SNOMED	173
11	EPIC FLO .1	EPIC FLOWSHEET MEASURE	Maps to	LOINC	159
12	EPIC FLO .1	EPIC FLOWSHEET MEASURE	Maps to unit	UCUM	156
13	EPIC EPT 155	EPIC LANGUAGE	Maps to	SNOMED	146
14	EPIC EPT 145	EPIC PATIENT RACE	Maps to non-standard	CDCREC	99
15	EPIC EPT 145	EPIC PATIENT RACE	Maps to	Race	98
16	EPIC ORD 2050	EPIC LAB COMPONENT UNIT	Maps to	UCUM	83
17	EPIC EPT 19361	EPIC FAMILY MEDICAL RELATION	Maps to	SNOMED	53
18	EPIC EPT 19928	EPIC HISTORY SDOH EDUCATION	Maps to	LOINC	48
19	EPIC EPT 19926	EPIC HISTORY SDOH PHYSICAL ACTIVITY MPS	Maps to	SNOMED	46
20	EPIC EPT 135	EPIC ETHNIC GROUP	Maps to	Ethnicity	43
21	EPIC EPT 135	EPIC ETHNIC GROUP	Maps to non-standard	CDCREC	43
22	EPIC EPT 150	EPIC RELIGION	Maps to	SNOMED	43
23	EPIC EPT 134	EPIC ETHNIC BACKGROUND	Maps to	Ethnicity	42
24	EPIC EPT 134	EPIC ETHNIC BACKGROUND	Maps to non-standard	CDCREC	42
25	EPIC ERX 330	EPIC ROUTE OF ADMINISTRATION	Maps to	SNOMED	38
26	EPIC EPT 19926	EPIC HISTORY SDOH PHYSICAL ACTIVITY MPS	Maps to number	SNOMED	32
27	EPIC EPT 19926	EPIC HISTORY SDOH PHYSICAL ACTIVITY MPS	Maps to operator	SNOMED	32

“Non-standard” Mappings to & from Epic EHR Data

```
SELECT
  c1.vocabulary_id, v.vocabulary_name, r.relationship_id, c2.vocabulary_id
, COUNT(*) AS relationship_count
FROM cdm_phi.concept_relationship r
JOIN cdm_phi.concept c1 ON r.concept_id_1 = c1.concept_id
JOIN cdm_phi.vocabulary v ON c1.vocabulary_id = v.vocabulary_id
JOIN cdm_phi.concept c2 ON r.concept_id_2 = c2.concept_id
WHERE r.relationship_id IN ('Maps to non-standard', 'Mapped from non-std')
AND (c1.vocabulary_id LIKE 'EPIC%' OR c2.vocabulary_id LIKE 'EPIC%')
GROUP BY c1.vocabulary_id, v.vocabulary_name, r.relationship_id, c2.vocabulary_id
ORDER BY COUNT(*) DESC, c1.vocabulary_id, r.relationship_id, c2.vocabulary_id
```

	vocabulary_id	vocabulary_name	relationship_id	vocabulary_id	relationship_count
1	EPIC EDG .1	EPIC DIAGNOSIS	Maps to non-standard	ICD9CM	1686866
2	ICD9CM	International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 1 and 2 (NCHS)	Mapped from non-std	EPIC EDG .1	1686866
3	EPIC EDG .1	EPIC DIAGNOSIS	Maps to non-standard	ICD10CM	1365231
4	ICD10CM	International Classification of Diseases, Tenth Revision, Clinical Modification (NCHS)	Mapped from non-std	EPIC EDG .1	1365231
5	EPIC ERX .1	EPIC MEDICATION	Maps to non-standard	NDC	145705
6	NDC	National Drug Code (FDA and manufacturers)	Mapped from non-std	EPIC ERX .1	145705
7	CDCREC	CDC Race & Ethnicity	Mapped from non-std	EPIC EPT 145	99
8	EPIC EPT 145	EPIC PATIENT RACE	Maps to non-standard	CDCREC	99
9	CDCREC	CDC Race & Ethnicity	Mapped from non-std	EPIC EPT 135	43
10	EPIC EPT 135	EPIC ETHNIC GROUP	Maps to non-standard	CDCREC	43
11	CDCREC	CDC Race & Ethnicity	Mapped from non-std	EPIC EPT 134	42
12	EPIC EPT 134	EPIC ETHNIC BACKGROUND	Maps to non-standard	CDCREC	42

CONCEPT_SYNONYM

The Concept Synonym Table

```
SELECT
  c.vocabulary_id, v.vocabulary_name
, COUNT(DISTINCT c.concept_id) AS concept_count, COUNT(*) AS synonym_count
, CAST(COUNT(*) * 1.0 / COUNT(DISTINCT c.concept_id) AS NUMERIC(18,2)) AS avg_synonyms_per_concept
FROM cdm_phi.concept_synonym cs
JOIN cdm_phi.concept c ON cs.concept_id = c.concept_id
JOIN cdm_phi.vocabulary v ON c.vocabulary_id = v.vocabulary_id
WHERE cs.concept_synonym_name != c.concept_name --Names loaded as their own synonyms
GROUP BY
  c.vocabulary_id
, v.vocabulary_name
ORDER BY
  COUNT(*) DESC
```

	vocabulary_id	vocabulary_name	concept_count	synonym_count	avg_synonyms_per_concept
1	SNOMED	Systematic Nomenclature of Medicine - Clinical Terms (IHSTDO)	970231	3364878	3.47
2	LOINC	Logical Observation Identifiers Names and Codes (Regeneriel Institute)	117120	674799	5.76
3	NDC	National Drug Code (FDA and manufacturers)	223924	425527	1.90
4	OMOP Genomic	OMOP Genomic vocabulary	73076	302870	4.14
5	ICD10PCS	ICD-10 Procedure Coding System (CMS)	83265	291334	3.50
6	OMOP Invest Drug	OMOP Investigational Drugs	29287	204155	6.97
7	RxNorm	RxNorm (NLM)	56366	194827	3.46
8	MeSH	Medical Subject Headings (NLM)	11796	187353	15.92
9	CPT4	Current Procedural Terminology version 4 (AMA)	16614	166189	10.00
10	ICD10CM	International Classification of Diseases, Tenth Revision, Clinical Modification (NCHS)	67995	122360	1.80
11	OPS	Operations and Procedures Classification (OPS)	41101	104886	2.55
12	NDFRT	National Drug File - Reference Terminology (VA)	21818	91021	4.17
13	CCAM	Common Classification of Medical Acts	10205	52530	5.15
14	KCD7	Korean Classification of Diseases, 7th Revision	22506	45012	2.00
15	OSM	OpenStreetMap	18116	37151	2.05
16	VANDF	Veterans Health Administration National Drug File	14204	29858	2.10
17	SPL	Structured Product Labeling (FDA)	16189	28665	1.77
18	ICD9CM	International Classification of Diseases, Ninth Revision, Clinical Modification, Volum...	12924	25848	2.00
19	dm+d	Dictionary of Medicines and Devices (NHS)	23926	23926	1.00
20	ATC	WHO Anatomic Therapeutic Chemical Classification	5916	16390	2.77
21	HCPCS	Healthcare Common Procedure Coding System (CMS)	7329	15426	2.10
22	UK Biobank	UK Biobank	5061	10122	2.00
23	QOR	Commented Drug Directory (BCFI)	3622	9659	2.64
24	PPI	ABORUs_PPI (Columbia)	4383	9083	2.07
25	ICD9Proc	International Classification of Diseases, Ninth Revision, Clinical Modification, Volum...	3378	6796	2.00

Concept Synonym Example: Your Mileage May Vary (a lot!)

```
SELECT
  c.vocabulary_id
, c.concept_id
, c.concept_code
, c.concept_name
, cs.concept_synonym_name
FROM cdm_phi.concept_synonym cs
JOIN cdm_phi.concept c
ON cs.concept_id = c.concept_id
WHERE c.concept_name !=
cs.concept_synonym_name
AND cs.concept_synonym_name IN
('heart attack', 'heart attacks')
AND c.vocabulary_id != 'PPI'
ORDER BY
c.vocabulary_id
, c.concept_name DESC
```

- ▶ ATHENA is case sensitive
- ▶ Microsoft SQL Server is not (per our configuration)
- ▶ We excluded non-English synonyms (no Unicode)

	vocabulary_id	concept_id	concept_code	concept_name	concept_synonym_name
1	MeSH	45616764	D009203	Myocardial Infarction	Heart Attacks
2	MeSH	45616764	D009203	Myocardial Infarction	Heart Attacks
3	MeSH	45616764	D009203	Myocardial Infarction	Heart Attack
4	MeSH	45616764	D009203	Myocardial Infarction	Heart Attack
5	NDFRT	4347016	N0000002085	Myocardial Infarction	Heart Attack
6	NDFRT	4347016	N0000002085	Myocardial Infarction	Heart Attack
7	SNOMED	40323863	155304006	Myocardial infarction (& [acute]) or coronary thrombosis	Heart attack
8	SNOMED	40323863	155304006	Myocardial infarction (& [acute]) or coronary thrombosis	Heart attack
9	SNOMED	40345176	266288001	Myocardial infarction (& [acute]) or coronary thrombosis	Heart attack
10	SNOMED	40345176	266288001	Myocardial infarction (& [acute]) or coronary thrombosis	Heart attack
11	SNOMED	40436955	233824008	Myocardial infarction	heart attack
12	SNOMED	40436955	233824008	Myocardial infarction	Heart Attack
13	SNOMED	40436955	233824008	Myocardial infarction	heart attack
14	SNOMED	40436955	233824008	Myocardial infarction	heart attacks
15	SNOMED	40436955	233824008	Myocardial infarction	Heart Attacks
16	SNOMED	40436955	233824008	Myocardial infarction	heart attacks
17	SNOMED	40436955	233824008	Myocardial infarction	Heart Attacks
18	SNOMED	40436955	233824008	Myocardial infarction	heart attack
19	SNOMED	40436955	233824008	Myocardial infarction	Heart Attack
20	SNOMED	40436955	233824008	Myocardial infarction	Heart Attack
21	SNOMED	4329847	22298006	Myocardial infarction	Heart attack
22	SNOMED	4329847	22298006	Myocardial infarction	Heart attack
23	SNOMED	40623902	66514008	Coronary artery thrombosis	heart attack
24	SNOMED	40623902	66514008	Coronary artery thrombosis	Heart Attack
25	SNOMED	40623902	66514008	Coronary artery thrombosis	Heart Attack
26	SNOMED	40623902	66514008	Coronary artery thrombosis	Heart Attacks

CONCEPT_ANCESTOR

Hierarchies within each Vocabulary

```
SELECT
  an.vocabulary_id AS ancestor_vocabulary_id
,dc.vocabulary_id AS descendant_vocabulary_id
,AVG(ca.min_levels_of_separation * 1.0) AS
avg_min_levels
,AVG(ca.max_levels_of_separation * 1.0) AS
avg_max_levels
,COUNT(*) AS descendants_count
FROM cdm_phi.concept_ancestor ca
JOIN cdm_phi.concept an
ON ca.ancestor_concept_id = an.concept_id
JOIN cdm_phi.concept dc
ON ca.descendant_concept_id = dc.concept_id
WHERE --Hierarchies within each vocabulary
an.vocabulary_id = dc.vocabulary_id
AND ca.ancestor_concept_id !=
ca.descendant_concept_id
--Concepts are their own parents
GROUP BY
  an.vocabulary_id
,dc.vocabulary_id
ORDER BY
COUNT(*) DESC
```

	ancestor_vocabulary_id	descendant_vocabulary_id	avg_min_levels	avg_max_levels	descendants_count
1	SNOMED	SNOMED	4.065313	5.141716	8168903
2	RxNorm Extension	RxNorm Extension	1.149614	1.297637	6764916
3	LOINC	LOINC	2.462357	2.629295	1473916
4	ICD10PCS	ICD10PCS	3.135803	3.136497	994246
5	RxNorm	RxNorm	1.166498	1.562576	993408
6	OSM	OSM	2.853151	2.853151	856099
7	MedDRA	MedDRA	2.461797	2.461797	371650
8	OMOP Genomic	OMOP Genomic	1.435241	1.436182	298524
9	CPT4	CPT4	3.009048	3.022413	71832
10	HemOnc	HemOnc	2.547220	2.879381	46336
11	SNOMED Veterinary	SNOMED Veterinary	1.798379	1.829370	33687
12	ATC	ATC	2.412834	2.412834	24683
13	Cancer Modifier	Cancer Modifier	1.874955	2.023419	16866
14	CDM	CDM	2.388351	4.007891	8997
15	UK Biobank	UK Biobank	2.698961	2.698961	6740
16	ICD9Proc	ICD9Proc	1.789208	1.971223	2780
17	Nebraska Lexicon	Nebraska Lexicon	1.304178	1.330730	2561
18	EphMRA ATC	EphMRA ATC	1.869604	1.869604	2201
19	PPI	PPI	1.634755	1.634755	1922
20	OMOP Extension	OMOP Extension	2.101748	2.117647	1258
21	NFC	NFC	1.440199	1.440199	1204
22	NUCC	NUCC	1.058823	1.058823	391
23	SOPT	SOPT	1.650000	1.650000	300
24	SMQ	SMQ	1.336322	1.336322	223
25	Type Concept	Type Concept	1.142857	1.142857	70
26	Medicare Specialty	Medicare Specialty	1.226415	1.226415	53
27	Race	Race	1.000000	1.000000	45

Hierarchies across Vocabularies (due to mappings)

```
SELECT
  an.vocabulary_id AS ancestor_vocabulary_id
,dc.vocabulary_id AS descendant_vocabulary_id
,AVG(ca.min_levels_of_separation * 1.0) AS
avg_min_levels
,AVG(ca.max_levels_of_separation * 1.0) AS
avg_max_levels
,COUNT(*) AS descendants_count
FROM cdm_phi.concept_ancestor ca
JOIN cdm_phi.concept an
ON ca.ancestor_concept_id = an.concept_id
JOIN cdm_phi.concept dc
ON ca.descendant_concept_id = dc.concept_id
WHERE --Hierarchies across vocabularies
an.vocabulary_id != dc.vocabulary_id
AND ca.ancestor_concept_id !=
ca.descendant_concept_id --Concepts are their own
parents
GROUP BY
  an.vocabulary_id
,dc.vocabulary_id
ORDER BY
COUNT(*) DESC
```

	ancestor_vocabulary_id	descendant_vocabulary_id	avg_min_levels	avg_max_levels	descendants_count
1	SPL	RxNorm Extension	1.819091	2.163466	28944394
2	ATC	RxNorm Extension	4.268952	5.083669	10114219
3	RxNorm	RxNorm Extension	2.054157	2.634150	8867188
4	SNOMED	ICDO3	4.613132	6.237203	3231618
5	MedDRA	SNOMED	4.914285	6.634080	3116612
6	MedDRA	ICDO3	5.141969	7.525547	2949741
7	HemOnc	RxNorm Extension	4.597884	5.680947	2717928
8	SNOMED	ICD10PCS	3.346389	4.280289	1984333
9	SPL	RxNorm	1.092971	1.367173	1920377
10	ATC	RxNorm	3.139798	3.798177	1150052
11	SNOMED	SNOMED Veterinary	5.606034	5.996994	342672
12	HemOnc	RxNorm	3.439881	4.281956	276720
13	MedDRA	ICD10PCS	4.111688	5.733815	272912
14	SNOMED	CPT4	3.167492	4.964391	269290
15	SNOMED	LOINC	4.210292	4.410791	143971
16	MedDRA	CPT4	3.704739	6.243450	93978
17	Cohort	RxNorm Extension	2.791786	4.479465	76431
18	SNOMED	Nebraska Lexicon	3.704362	4.691448	65357
19	US Census	OSM	3.445640	3.445640	57984
20	MedDRA	SNOMED Veterinary	4.803899	6.280267	57338
21	SNOMED	OPCS4	4.422612	5.436862	52140
22	SNOMED	ICD9Proc	3.805258	5.005448	47165
23	Cohort	ICDO3	3.242917	3.908745	33390
24	SNOMED Veterinary	SNOMED	3.954270	4.385585	27444
25	ICD9Proc	SNOMED	3.786126	4.410695	26684
26	SNOMED	HCPCS	3.788482	4.452472	20178
27	ICD9Proc	ICD10PCS	3.530789	4.271991	19633
28	MedDRA	ICD9Proc	4.252246	5.857104	18811
29	MedDRA	OPCS4	4.626108	6.112474	17933
30	ATC	HemOnc	2.585013	2.593430	15921

Are you ready for the final exam?

Combining Ancestors & Mapped Codes Example: Cancer

```
SELECT
  an.vocabulary_id AS ancestor_vocabulary_id, an.concept_code AS ancestor_concept_code
, an.concept_name AS ancestor_concept_name , dc.vocabulary_id AS descendant_vocabulary_id
, COUNT(DISTINCT dc.concept_id) AS descendant_concept_count
, ISNULL(dx.vocabulary_id, '') AS map1_vocabulary_id, COUNT(DISTINCT dx.concept_id) AS map1_concept_count
, ISNULL(edg.vocabulary_id, '') AS map2_vocabulary_id, COUNT(DISTINCT edg.concept_id) AS map2_concept_count
FROM cdm_phi.concept_ancestor ca
JOIN cdm_phi.concept an ON ca.ancestor_concept_id = an.concept_id
JOIN cdm_phi.concept dc ON ca.descendant_concept_id = dc.concept_id
LEFT JOIN cdm_phi.concept_relationship cr1
ON ca.descendant_concept_id = cr1.concept_id_1 AND cr1.relationship_id = 'Mapped from'
LEFT JOIN cdm_phi.concept dx ON cr1.concept_id_2 = dx.concept_id AND dx.vocabulary_id IN ('EPIC EDG .1', 'ICD10CM')
LEFT JOIN (cdm_phi.concept_relationship cr2
  JOIN cdm_phi.concept edg
  ON cr2.concept_id_2 = edg.concept_id AND edg.vocabulary_id = 'EPIC EDG .1')
ON dx.concept_id = cr2.concept_id_1 AND cr2.relationship_id = 'Mapped from non-std'
WHERE ca.ancestor_concept_id = 443392 --SNOMED 363346000 'Malignant neoplastic disease'
GROUP BY an.vocabulary_id, an.concept_code, an.concept_name, dc.vocabulary_id, dx.vocabulary_id, edg.vocabulary_id
ORDER BY COUNT(DISTINCT dx.concept_id) DESC, COUNT(DISTINCT edg.concept_id) DESC, COUNT(DISTINCT dc.concept_id) DESC
```

	ancestor_vocabulary_id	ancestor_concept_code	ancestor_concept_name	descendant_vocabulary_id	descendant_concept_count	map1_vocabulary_id	map1_concept_count	map2_vocabulary_id	map2_concept_count
1	SNOMED	363346000	Malignant neoplastic disease	SNOMED	676	EPIC EDG .1	28637		0
2	SNOMED	363346000	Malignant neoplastic disease	SNOMED	657	ICD10CM	1473	EPIC EDG .1	34109
3	SNOMED	363346000	Malignant neoplastic disease	SNOMED	17	ICD10CM	16		0
4	SNOMED	363346000	Malignant neoplastic disease	ICD03	50132		0		0
5	SNOMED	363346000	Malignant neoplastic disease	SNOMED	5064		0		0
6	SNOMED	363346000	Malignant neoplastic disease	Nebraska Lexicon	75		0		0
7	SNOMED	363346000	Malignant neoplastic disease	SNOMED Veterinary	34		0		0
8	SNOMED	363346000	Malignant neoplastic disease	OMOP Extension	1		0		0
9	SNOMED	363346000	Malignant neoplastic disease	PPI	1		0		0

Thank You!
For help and more information:
<https://labs.icahn.mssm.edu/msdw>

