



Overview

- Why download this release
- Oncology WG work and new tables in CDM v5.5
- Deep dive into roadmap activities
- Community contributions
- Deep dive into oncology



Why download the release

- **Expanded SNOMED content:**

+7k terms for diseases and procedures
and +32k new hierarchical relationships

- **More measurements in LOINC:**

+4k new labs and improved grouping

- **More drugs and drug mappings:**

+500 RxNorm codes and +20k mappings

- **Rare disease coverage:**

Mondo and Orphanet with 31k of new
rare disease terms

- **Oncology
harmonization:**

Major refreshes and improvements

- New **metadata** tables in the
bundle



Vocabularies changes in CDM v5.5



Should I move to OMOP CDM v5.5?

v5.4

Stay on v5.4

Keep doing what works today



Existing workflows and studies continue

No immediate need to rework a functioning v5.4 ETL. You can remain on v5.4 in the near term.



No access to v5.5 additions

Your database will not include the new fields and vocabulary tables.



Some future studies may require an upgrade

If a study uses a v5.5-specific field, participating databases will need to move to v5.5.



v5.5

Move to v5.5

Be ready for the next generation of OMOP



Relatively small lift

v5.5 is an incremental, additive update rather than a major redesign.



Latest vocabulary structure

Take advantage of the new vocabulary tables and metadata capabilities.



Populate new fields when relevant

Adopting the structure does not mean every new field must immediately contain data.



Our recommendation: plan your move to v5.5 when practical.

v5.4 will continue to work. v5.5 keeps you aligned with the latest vocabulary structure and prepared for future OHDSI studies.

3



PACK_CONTENT Table

Introduces the **PACK_CONTENT** table to represent the contents of drug packs and multi-ingredient products.



Table added:

- PACK_CONTENT

4



Concept Metadata Tables

Adds **CONCEPT_METADATA** and **CONCEPT_RELATIONSHIP_METADATA** tables to provide richer provenance, classification, and other metadata about vocabulary content and relationships.



Tables added:

- CONCEPT_METADATA
- CONCEPT_RELATIONSHIP_METADATA



Oncology finally arrived in the OMOP CDM

No more

“You can’t do cancer RWE in OMOP”

“I don’t know what variables they support”

“I am using ICD10 for my research”



What took us so long?

Onco WG started in 2016

The world was messy:

- Cancer needs a ton more detail
- They are spread over many vocabularies and domains
- These are badly redundant
- They are full with composite concepts
- None of the vocabs are comprehensive
- None is well suited for storing facts

-> The current solution is the result of years and years of debate, trial and error, toil, sweat, tears and blood.

30+ oncology data sources ETLed in a harmonized fashion



Cancer now has it's own mini-OMOP world

.. but it is fully integrated

It is:

- Deduplicated
- Mapped
- With a hierarchy
- With correct domains



Applause to

Andy Williams
Asieh Golozar
Dima Dimshits
Georgie Kennedy
Iryna Zherko
John Methot
Masha Khitrun
Michael Gurley
Rimma Belenkaya
Robert Miller
Anton Tarur
Thomas Falconer
Vlad Korsik



Things are deduplicated

ID	Name	Vocab	Domain	ID	Name	Vocab	Domain
36769269	Summary stage at Dx 2000 revised Cancer: Regional, regional lymph nodes only	LOINC	Observation	1346528	Regional lymph nodes	Cancer Modifier	Measurement
36769269	Colic, NOS Epicolic (adjacent to bowel wall) Mesocolic, NOS Paracolic/pericolic	NAACCR	Meas Value	35928996	Regional lymph nodes	Cancer Modifier	Measurement
36769269	Metastatic well-differentiated neuroendocrine tumor to regional lymph node	SNOMED	Condition	1076511	Regional lymph nodes	Cancer Modifier	Measurement

..4047 more!

ID	Name	Vocab	Domain	ID	Name	Vocab	Domain
36770704	Secondary nodes - inguinal/leg	SNOMED	Condition	4153892	Inguinal lymph nodes	Cancer Modifier	Measurement

Some don't even have the term "lymph" in the source.



Concepts go to the right Domain

ID	Name	Vocab	Domain	ID	Name	Vocab	Domain
37164032	IV:21 (AJCC)	SNOMED	Meas Value	1633308	AJCC/UICC Stage 4	Cancer Modifier	Measurement
4312274	Squamous cell carcinoma of lung, TNM stage 4	SNOMED	Condition				
2107454	AJCC colon cancer, Stage IV documented (ONC)	CPT4	Observation				
45532945	Follicular lymphoma grade I, lymph nodes of multiple sites	ICD10CM	Condition	1634371	Grade 1 tumor	Cancer Modifier	Measurement
1448441	G1 (UICC)	SNOMED	Meas Value				
44808015	Follicular lymphoma grade 1	SNOMED	Condition				
4183904	G1: Well differentiated (greater than 95 percent gland forming)	SNOMED	Condition				
2618047	Oncology; disease status; ovarian cancer, limited to epithelial cancer; pathologic stage ia-b (grade 1) without evidence of disease progression, recurrence, or metastases (for use in a medicare-approved demonstration project)	HCPCS	Observation				
607623	G1: Well differentiated	SNOMED	Observation				
1346617	Histologic type in Colorectal cancer specimen by WHO classification: Neuroendocrine tumor G1 (NET G1)/Carcinoid	LOINC	Measurement				



Composite concepts are split up

ID	Name	Vocab	Domain	ID	Name	Vocab	Domain
4138008	Philadelphia chromosome-positive acute lymphoblastic leukemia	SNOMED	Condition	134305	Acute lymphoid leukemia	SNOMED	Condition
				35977026	t(9;22)(q34;q11.2) measurement	OMOP Genomic	Measurement
				35976955	BCR::ABL1 gene fusion measurement	OMOP Genomic	Measurement
37175074	Metastatic squamous cell carcinoma to epididymis and spermatic cord	SNOMED	Condition	4300118	Squamous cell carcinoma	SNOMED	Condition
				35225740	Metastasis to spermatic cord	Cancer Modifier	Measurement
				35225540	Metastasis to epididymis	Cancer Modifier	Measurement
46284865	Malignant melanoma stage IV M1a	SNOMED	Condition	4162276	Malignant melanoma	SNOMED	Condition
				1633308	AJCC/UICC Stage 4	Cancer Modifier	Measurement
				1635100	AJCC/UICC M1a Category	Cancer Modifier	Measurement



Things have hierarchies

ID	Name	Vocab	Domain	ID	Name	Vocab	Domain
704985	Metastasis to abdomen	Cancer Modifier	Measurement	35226272	Metastasis to rectouterine pouch	Cancer Modifier	Measurement
				35225674	Metastasis to skin of abdomen	Cancer Modifier	Measurement
				35226282	Metastasis to retrocecal tissue	Cancer Modifier	Measurement
				35225547	Metastasis to mesentery	Cancer Modifier	Measurement
				35225553	Metastasis to mesocolon	Cancer Modifier	Measurement
36768587	Lymph nodes	Cancer Modifier	Measurement	36769815	Left inguinal lymph nodes	Cancer Modifier	Measurement
				36769821	Right lateral sacral lymph nodes	Cancer Modifier	Measurement
				36769832	Right superior parathyroid lymph nodes	Cancer Modifier	Measurement
				36769835	Mediastinal lymph nodes	Cancer Modifier	Measurement
36769180	Metastasis	Cancer Modifier	Measurement	36769243	Distant lymph nodes	Cancer Modifier	Measurement



Things have value_as_concept_id

ID	Name	Vocab	Domain	ID	Name	Vocab	Domain	Relationship_id
35938496	No involved regional nodes	NAACCR	Meas Value	36769269	Regional lymph nodes	Cancer Modifier	Measurement	Maps to
				9189	Negative	SNOMED	Meas Value	Maps to value
35938503	Distant metastases, NOS Stated as M1 with no other information on distant metastasis	NAACCR	Meas Value	36769180	Metastasis	Cancer Modifier	Measurement	Maps to
				9191	Positive	SNOMED	Meas Value	Maps to value
35938506	All pelvic nodes examined negative.	NAACCR	Meas Value	36769961	Pelvic lymph nodes	Cancer Modifier	Measurement	Maps to
				9189	Negative	SNOMED	Meas Value	Maps to value

You have to tell if a Cancer Attribute is positive, negative or equivocal.



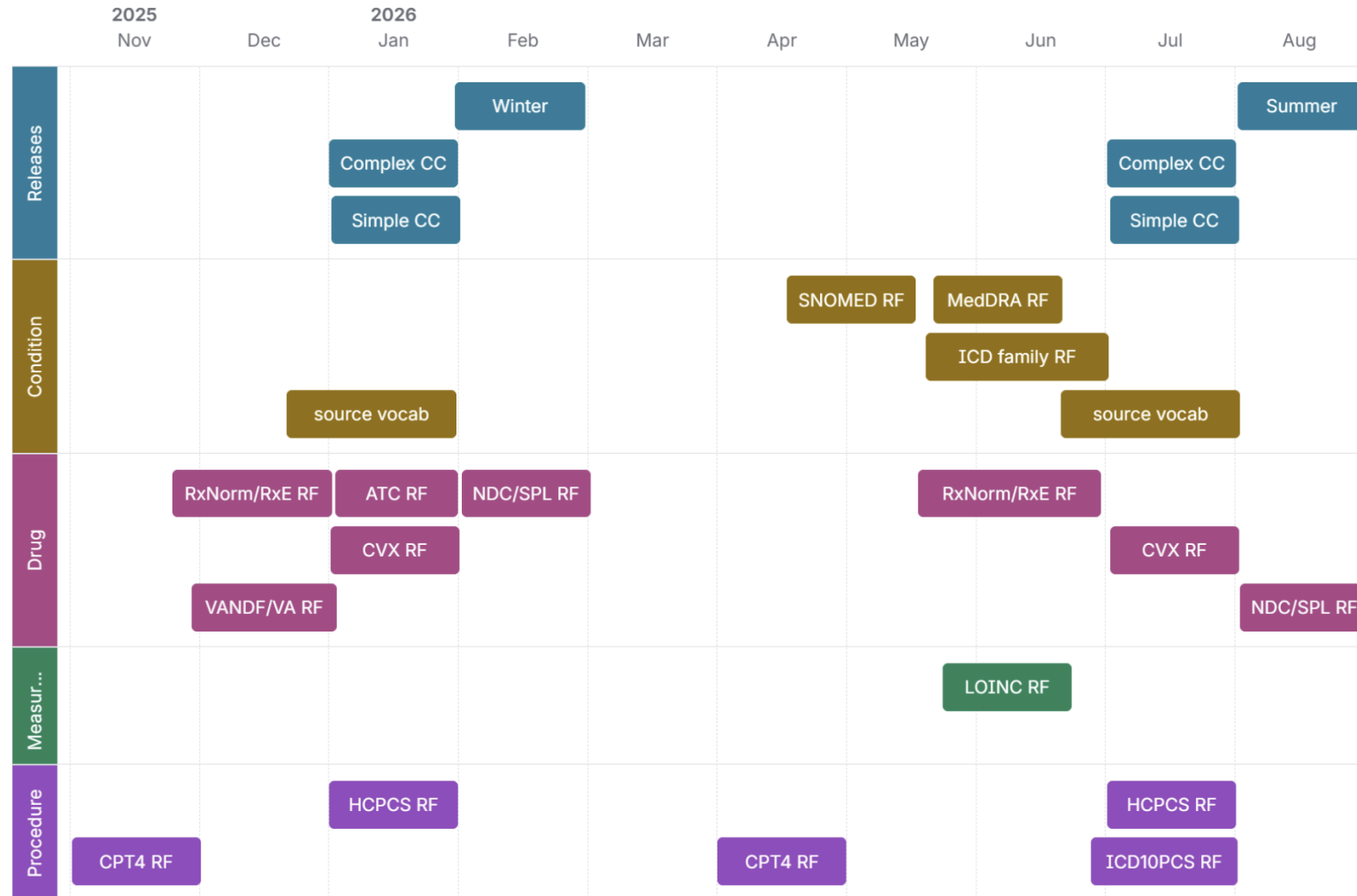
Genomic markers are extracted

ID	Name	Vocab	Domain	ID	Name	Vocab	Domain
606958	Acute myeloid leukemia with t(8;21)(q22;q22) RUNX1-RUNX1T1a	SNOMED	Condition	36017898	t(8;21)(q22;q22) measurement	OMOP Genomic	Measurement
				35977100	RUNX1::RUNX1T1 gene fusion measurement	OMOP Genomic	Measurement
607385	Primary familial polycythemia due to erythropoietin receptor mutation	SNOMED	Condition	35960541	EPOR (erythropoietin receptor) gene variant measurement	OMOP Genomic	Measurement
607405	Mixed phenotype acute leukemia with t(9;22)(q34;q11.2); BCR-ABL1	SNOMED	Condition	35976955	BCR::ABL1 gene fusion measurement	OMOP Genomic	Measurement
		SNOMED	Condition	35977026	t(9;22)(q34;q11.2) measurement	OMOP Genomic	Measurement
608277	Acute myeloid leukemia with 11q23 abnormality	SNOMED	Condition	19643443	KMT2A gene rearrangement measurement	OMOP Genomic	Measurement
609279	HER2-positive gastric cancer	SNOMED	Condition	35976979	ERBB2 protein expression measurement	OMOP Genomic	Measurement
1346611	BRAF protein Ag [Presence] in Cancer specimen by Immune stain: Negative	LOINC	Measurement	35976957	BRAF cytoplasmic protein expression measurement	OMOP Genomic	Measurement
45768522	Triple-negative breast cancer	SNOMED	Condition	35977081	PGR protein expression measurement	OMOP Genomic	Measurement
				35976980	ESR1 protein expression measurement	OMOP Genomic	Measurement
				35976979	ERBB2 protein expression measurement	OMOP Genomic	Measurement



August release

Roadmap 2025 Q4 - 2026 Q3:





August release roadmap refreshes

Vocabulary	Content	Vocabulary	Content
SNOMED Int	Refresh (February 2026 version)	RxNorm	Refresh (June 2026 version)
SNOMED US	Refresh (March 2026 version)	RxNorm Extension	Refresh (July 2026 version)
SNOMED UK	Refresh (April 2026 version)	CVX	Refresh (July 2026 version) + mapping curation
MedDRA	Refresh (March 2026 version)	NDC/SPL	Refresh (July 2026 version) + mapping propagation
ICD10CM (US)	Refresh (FY 2027 version) + mapping curation	VANDF/VA Class	Refresh (July 2026 version)
ICD10GM (Germany)	Refresh (2026 version) + mapping propagation	LOINC	Refresh (February 2025 version)
ICD9CM, ICD10, ICD10CN, CIM10, KCD7, Mesh, CIEL, OPCS4, CO-CONNECT, UK Biobank, ICD9Proc, OMOP Invest Drug, Read	Mapping curation	ICD10PCS	Refresh (FY 2027 version) + mapping curation
		CPT4	Refresh (UMLS 2026AA version)
		HCPCS	Refresh (July 2026 version)



Service vocabularies

- Two UCUM units for audiology tests: millisiemens (mS) and decapascal (daPa)
- A new Episode concept for Pregnancy
- CDM vocabulary update to the CDM 5.5 version



Improvements

SNOMED

- No major deprecations **by the source** detected
- Harmonization of domains on users' requests:

Weight findings
Height/growth findings
Mental state findings



Observation +
Condition



Condition

Spirometry
Electroretinography



Measurement



Procedure

Check your cohorts!



Improvements

RxNorm/RxNorm Extension

- We continue deduplication across RxNorm Extension and RxNorm: **more than 1,000 duplicates de-standardized**

NDC

- Domain assignment improved: **diagnostic RP moved to the Device domain**

CPT4

- Mapping of temporary Category III codes to their permanent Category I equivalents: **23 concepts mapped**



Additional procedures for mapping curation in August release

- **Alathea**: assessed impact of vocabulary changes on 1,383 concept sets used in RWE generation
- **Ariadne**: tested generalizability of LLM-assisted mapping pipeline for 10k of selected condition concepts by comparing with existing mappings; 1-2% error rate; 500+ mappings incorporated into release





Simple contributions

Source concept	Previous mapping	New mapping
CPT4 69436 Tympanostomy (requiring insertion of ventilating tube), general anesthesia	4160439 Administration of anesthesia	Add mapping to 4262320 Tympanostomy
ICD10CM 45552073 Encounter for fertility testing	4049310 Ocular prosthetic service	Change mapping to 4227745 Infertility study
ICD10CM 45552415 Sanfilippo mucopolysaccharidoses	433446 Mucopolysaccharidosis	Change mapping to 4227905 Sanfilippo syndrome
ICD10CM 45561980 Morquio mucopolysaccharidoses, unspecified	433446 Mucopolysaccharidosis	Change mapping to 4292105 Morquio syndrome
ICD10CM 35209545 Metabolic disorders following an ectopic and molar pregnancy and 35209548 Unspecified complication following an ectopic and molar pregnancy	37162983 Disorder following induced termination of pregnancy	439658 Disorder of pregnancy
ICD9CM branch of 637.x Unspecified abortion	37018340 Induced termination of pregnancy	40539858 Pregnancy with abortive outcome

Many thanks to all contributors!



Complex contributions

Content:

- SNOMED Veterinary (refresh)
- Rare diseases: Mondo and Orphanet (new vocabulary); HPO (refresh)
- Oncology vocabularies: Cancer Modifier, OMOP Genomic, NAACCR (refresh); LOINC & SNOMED changes

Enablers:

- Shared development environment available to contributors (vocabulary server)
- New simple contribution validation templates
- Work with SNOMED and LOINC to add new concepts to their content
- Work on conventions for poorly harmonized spaces (surveys and instruments)





SNOMED Veterinary refresh



Wayde **pioneering new standards for contributions**: Refresh performed on shared development infrastructure (vocabulary server) with associated GitHub code

BUILD CHANGES THIS REFRESH CYCLE 3,088 new concepts • 174,900 new relationships • 8,213 concept name changes

NEW CONCEPTS BY DOMAIN

Domain	Count
Observation	3,024
Condition	24
Spec Anatomic Site	24
Drug	14
Procedure	1
Specimen	1
Total	3,088

EXAMPLES OF NEW CONCEPTS

- Hemorrhage of sole of hoof
- Equine gastric ulcer syndrome
- Live attenuated lumpy skin disease virus Ranchi strain only vaccine product
- Feline immunodeficiency virus antibody

NEW RELATIONSHIPS BY TYPE

Relationship Type	Forward	Inverse
Is a / Subsumes	3,563	3,563
Has Module / Module of	40,208	40,208
Has status / Status of	35,872	35,872
Attribute rel. (Has X / X of)	253	253
Mapping & replacement*	15,088	
Other	20	
Total	174,900	

OTHER CHANGES

Change Type	Count
Concept name changes	8,213
Synonym text changes	0
Standard-concept status changes	62
Domain reassignments	935



Rare diseases



ORPHAcode

clinical rare-disease identity

65

**Leber congenital amaurosis
[Condition]**

Role

Recognized rare disorder for clinical coding, traceability, registry use and reporting.

+ 6,191 ORPHAcode as non-standard concepts w/o mapping

Mondo

precise disease and hierarchy

MONDO_0018998
**Leber congenital
amaurosis
[Condition]**

MONDO_0100453
**GUCY2D-related
recessive retinopathy
[Condition]**

▲ Is a

▲ Is a

MONDO_0008764
**Leber congenital amaurosis 1
[Condition]**

Role

Preserves the exact subtype and places it under broader disease parents in a hierarchy.

+ 33,842 new codes as non-standard concepts w/o mappings

HPO

phenotype

HP_0030453
Abnormal visual
electrophysiology
[Observation]

▲ Is a

HP_0000512
Abnormal
electroretinogram
[Observation]

Maps to

274524001
Electroretinogram
abnormal
[Condition]

Additional HPO phenotype facts
HP_0007875 · Congenital blindness
HP_0000639 · Nystagmus
HP_0000613 · Photophobia

Role

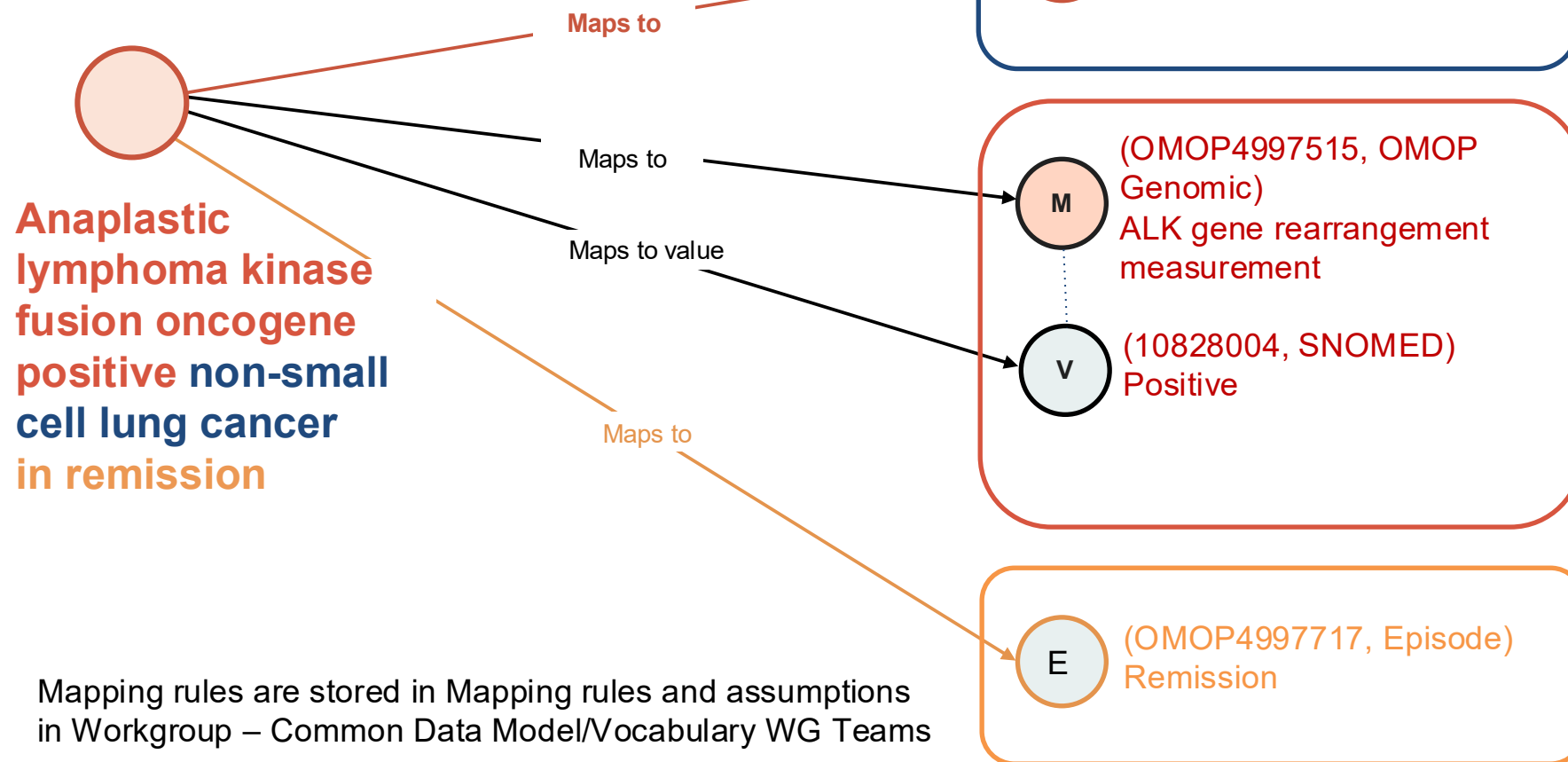
Observed abnormalities in the patient, not alternative disease identifiers.

+ 481 new concepts
+ 2k mappings and new sources (Chorus Bridge2AI , Tudor - 150)

Next steps: Rare Disease WG with Mondo and Orphanet explores data usage and designs harmonized model. Get involved!



Oncology: overall representation



Cancer diagnosis:
recorded as SNOMED or
ICDO3

CONDITION_OCCURRENCE

**Cancer
attributes/modifiers:**

- genetic variants recorded as OMOP Genomic
- nodes, grades, stages, metastasis recorded as Cancer Modifier

MEASUREMENT

Cancer dynamics:

EPISODE

To leverage the changes, follow Maps to links in CONCEPT_RELATIONSHIP as usual



Oncology Vocabulary Refresh

Standard

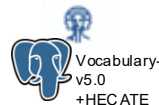
**OHDSI-built vocabulary update
(Cancer Modifier, OMOP
Genomic)**

**Semantic harmonization of standard OHDSI
vocabularies (SNOMED, LOINC)**

New vocabulary ETL pipeline

Updated content

New concept-mining pipeline



Destandardization

Source

NAACCR Update

Updated content

Semantic Schema Simplification

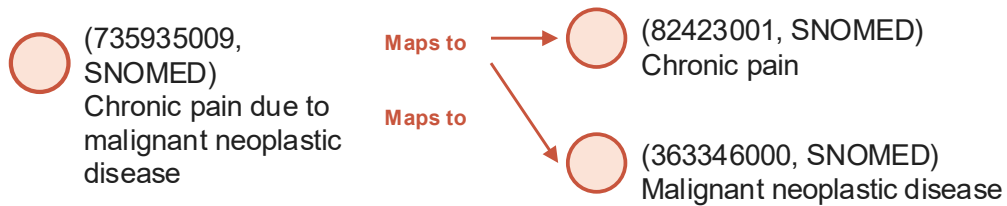
Destandardization

Nebraska Lexicon Update

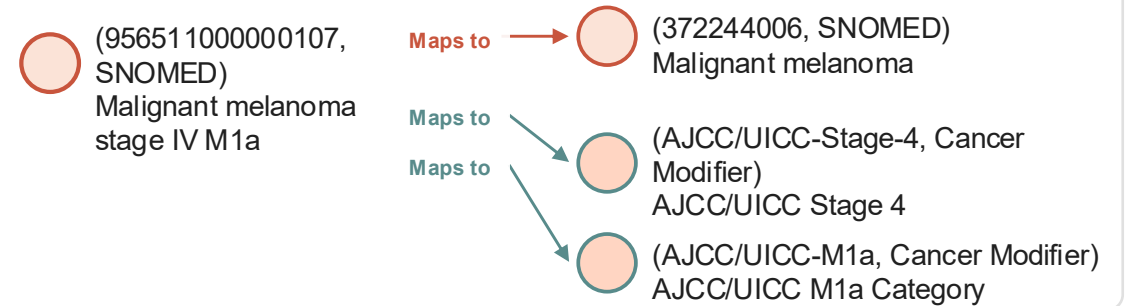
Destandardization

Compound oncology concepts are split into target components

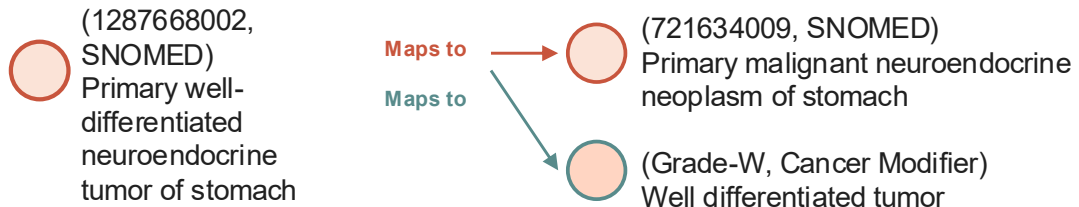
[Non-oncology disorder/symptom] due to cancer split



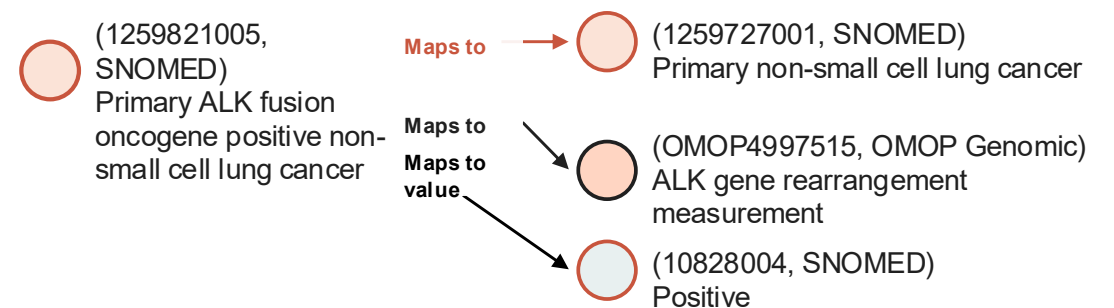
[Oncological Disease] diagnosis with stage attributes



[Oncological Disease] diagnosis with grade/differentiation split

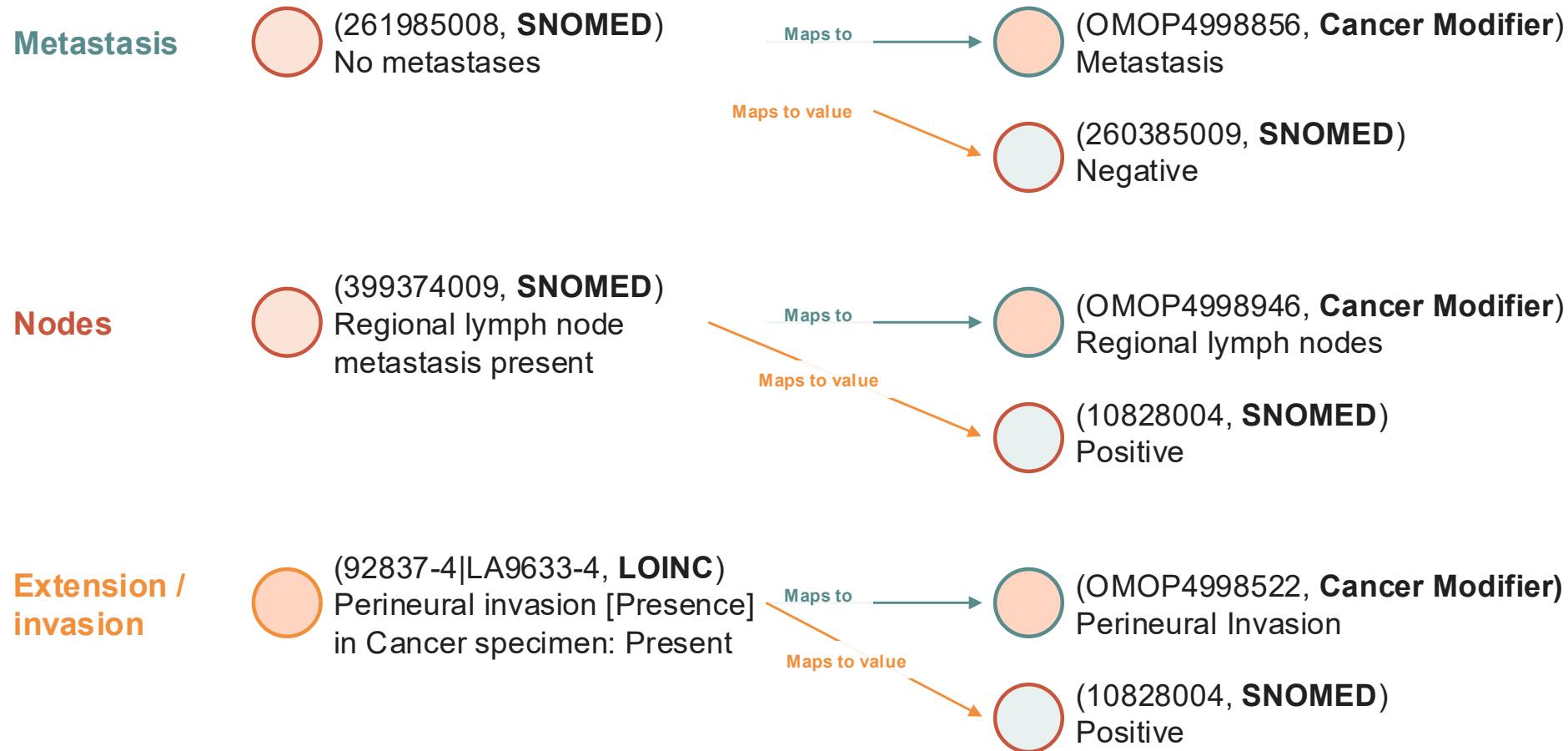


Biomarker-positive/negative [Oncological Disease] diagnosis



NB! Decompose compound oncology concepts when the source explicitly combines independently useful clinical facts and no single standard target preserves the full meaning.

Metastasis, nodes and invasion/extension mapping approach



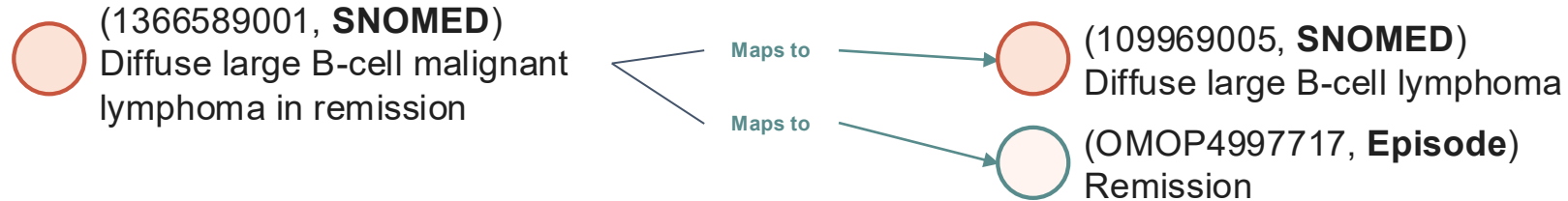
NB!

- Target modifier identifies what was observed.
- Value target states whether the finding is present or absent.
- LOINC examples use the assembled source pair when the answer is needed for the clinical meaning.

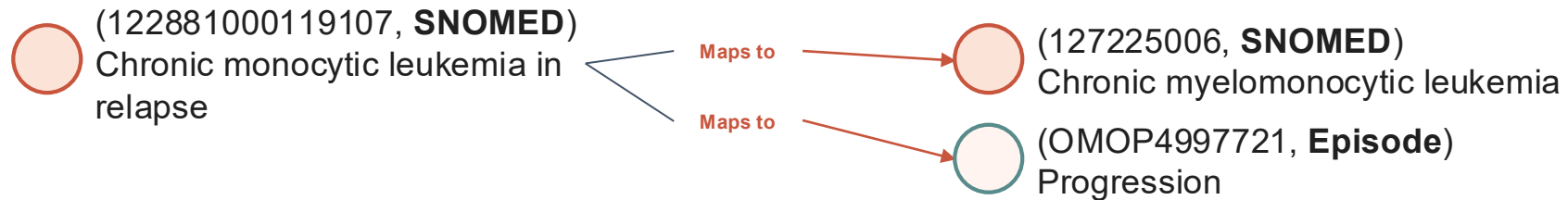
NB! Represent metastasis, nodes, stage, grade, and extension or invasion using dedicated Cancer Modifier concepts. Use Positive, Negative, or Equivocal values when that result is explicitly present in the source meaning.

Disease dynamics approach

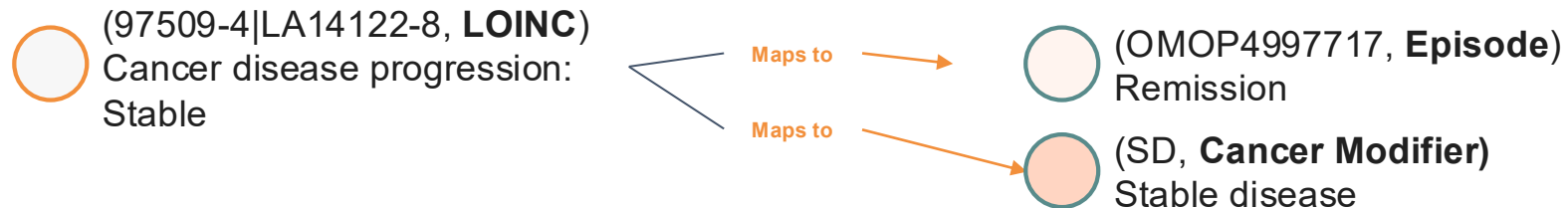
Remission



Relapse / progression



Stable disease



NB! Represent remission, relapse, progression, and stable disease as disease-dynamics concepts when the source meaning is unambiguous.



Statistics on changes in SNOMED and LOINC

Compound diagnosis patterns

■ Due-to-cancer disorder/symptom	43 code	87 maps
■ Stage attributes split	64 codes	136 maps
■ Grade / differentiation split	115 codes	243 maps
■ Biomarker-defined diagnosis	58 codes	187 maps

Modifier / disease-state content

■ Metastasis / nodes / invasion	278 codes	475 maps
■ Disease dynamics	441 codes	889 maps
■ Performance status / scales	19 codes	19 maps
■ Genomic alteration mappings	88 codes	100 maps
■ Disease extent	23 codes	23 maps
■ Dimensions / histopattern / resistance	21 codes	21 maps

Stages and Grades remapping

■ Direct Cancer Modifier remap	1,598 codes	1,953 maps
--------------------------------	-------------	------------

Please [Refer to CDE at OHDSI Vocabularies Drive to lookup changes](#)

NB! Propagation and/or script-based (i.e. LoadStage) changes are not displayed at document



Overview of semantic debulking and Cancer Modifier refresh



Cancer Modifier

new standard + classification concepts
internal hierarchy and replacements
relationships maintained

Updated Cancer Modifier

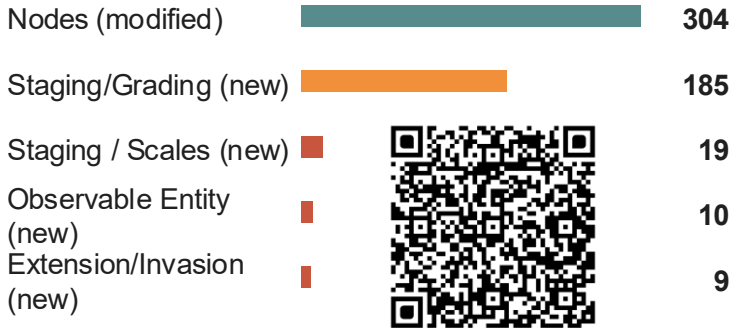
OMOP Genomic

gene-variant measurements added
concept_class_id corrections and Is a relationships
biomarker value model aligned

SNOMED

compound oncology terms split
and oncology codes remapping with
destandardization

Cancer Modifier by class



Examples are here

476 standard, 11 classification; hierarchical relationships: 551 Cancer Modifier + 30 OMOP Genomic.

LOINC

question + answer assembled as a
source pair to permit clinical fact
mappings

NAACCR refresh overview

+29,016

active non-standard concepts
added across Schema, Variable, Value,
Procedure and Range

36,034

deprecated concepts
mainly Value and Variable

1,031

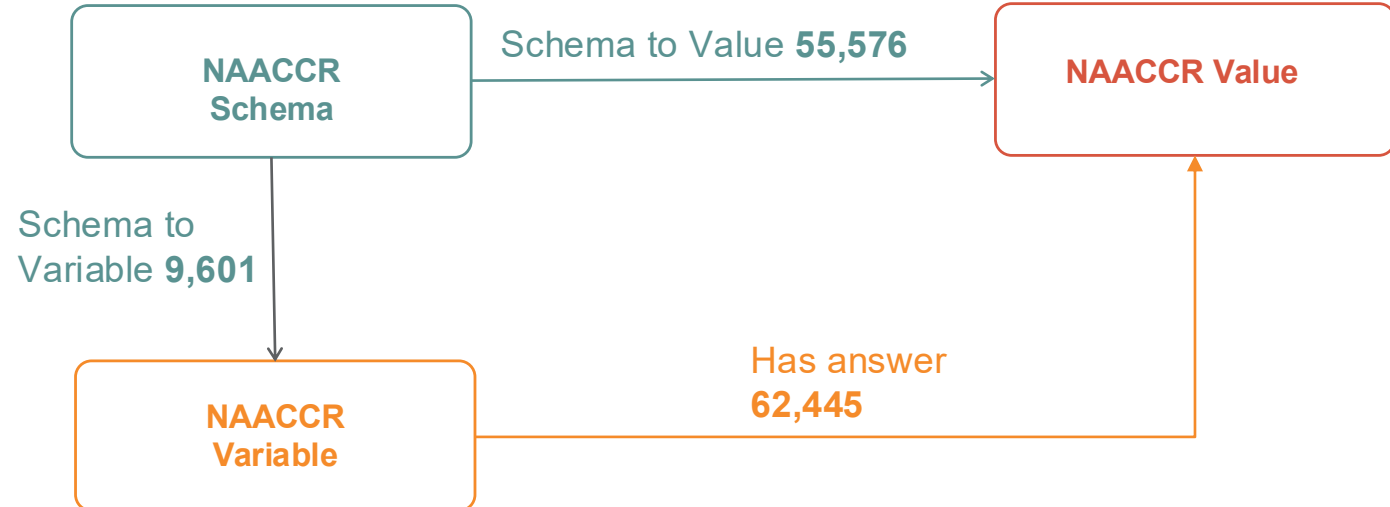
replacement pairs
Concept replaces / replaced by

262,916

relationship rows

Active concept deltas

- NAACCR Value: +22,680
- NAACCR Variable: +5,386
- NAACCR Procedure: +727
- Schema + Range: +223





Why download the release

- **Expanded SNOMED content:**

+7k terms for diseases and procedures
and +32k new hierarchical relationships

- **More measurements in LOINC:**

+4k new labs and improved grouping

- **More drugs and drug mappings:**

+500 RxNorm codes and +20k mappings

- **Rare disease coverage:**

Mondo and Orphanet with 31k of new
rare disease terms

- **Oncology**

- harmonization:**

Major refreshes and improvements

- New **metadata** tables in the bundle