



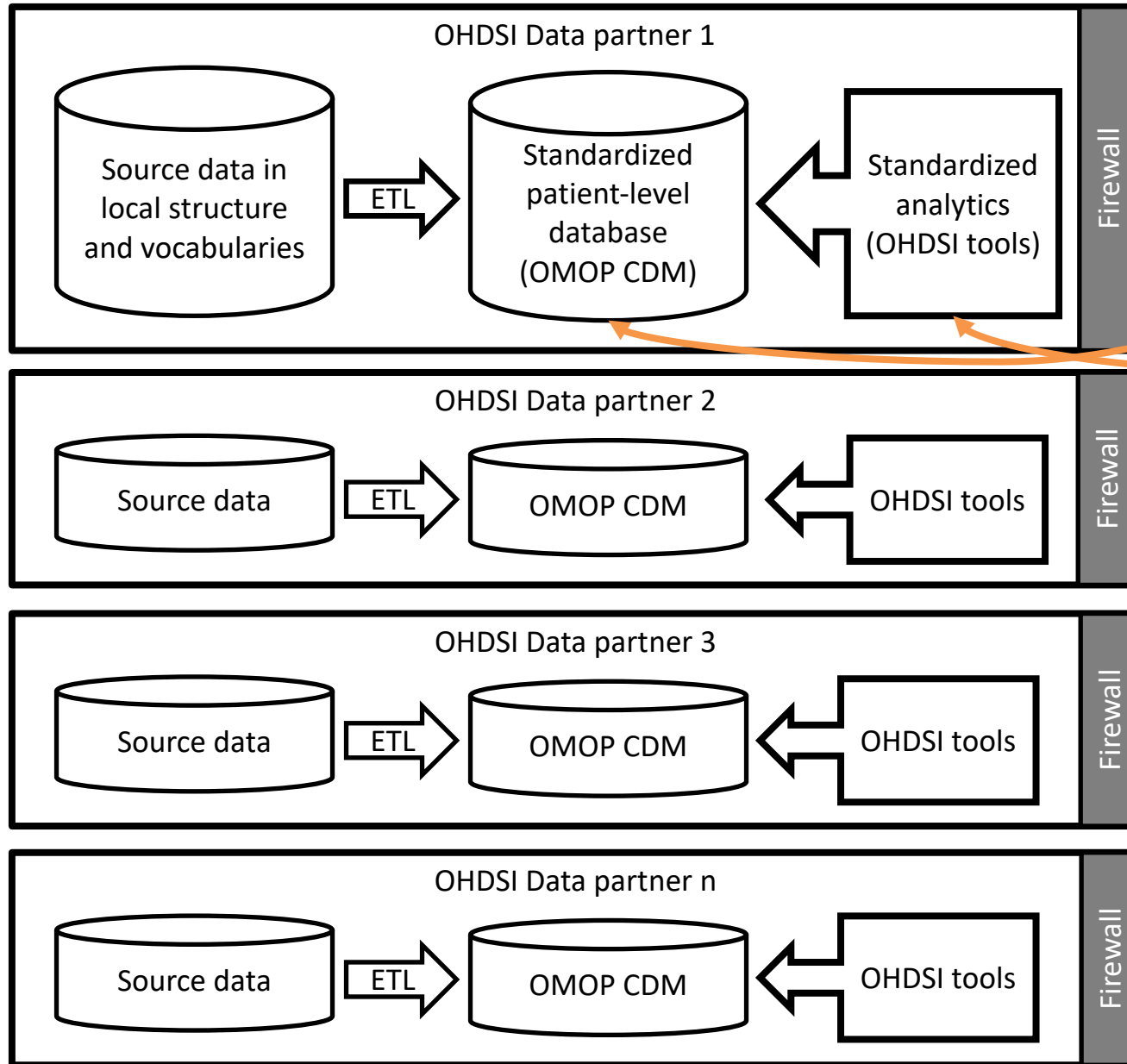
# OHDSI Evidence Network

Patrick Ryan PhD, Anna Ostropolets, PhD

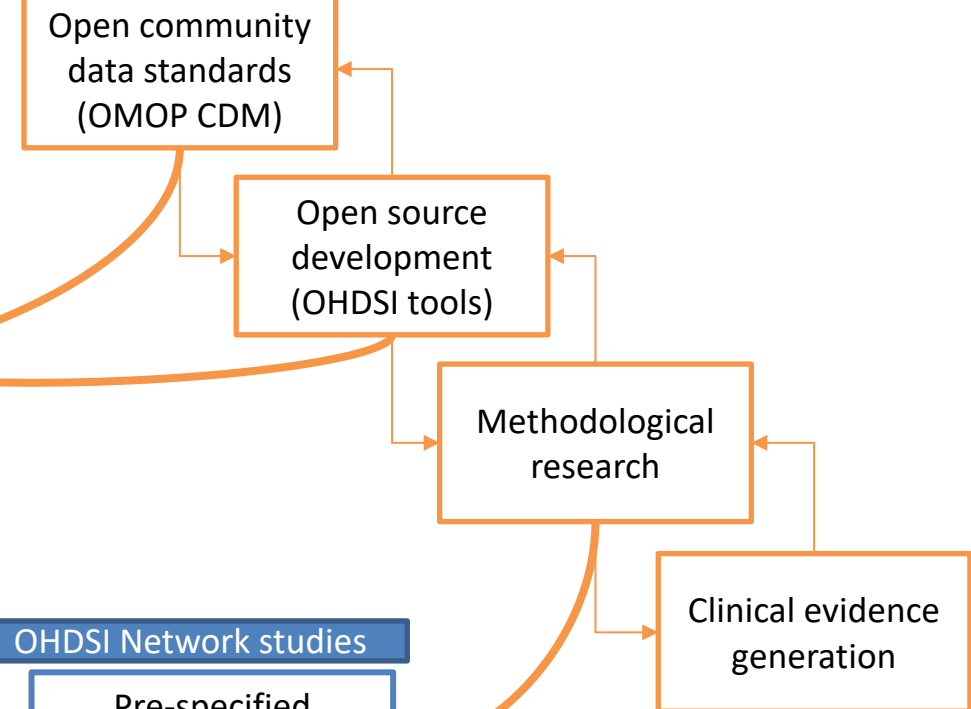
Johnson & Johnson

Columbia University Irving Medical Center

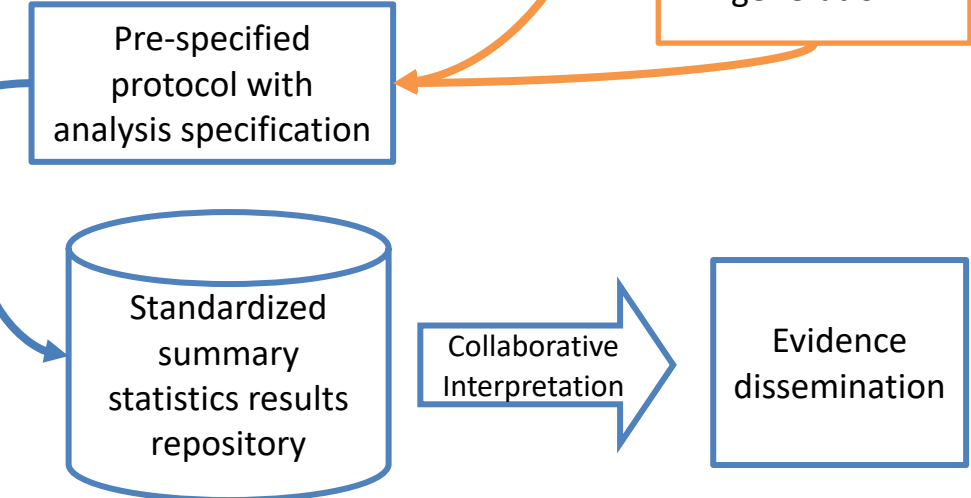
OHDSI data network



OHDSI collaborations

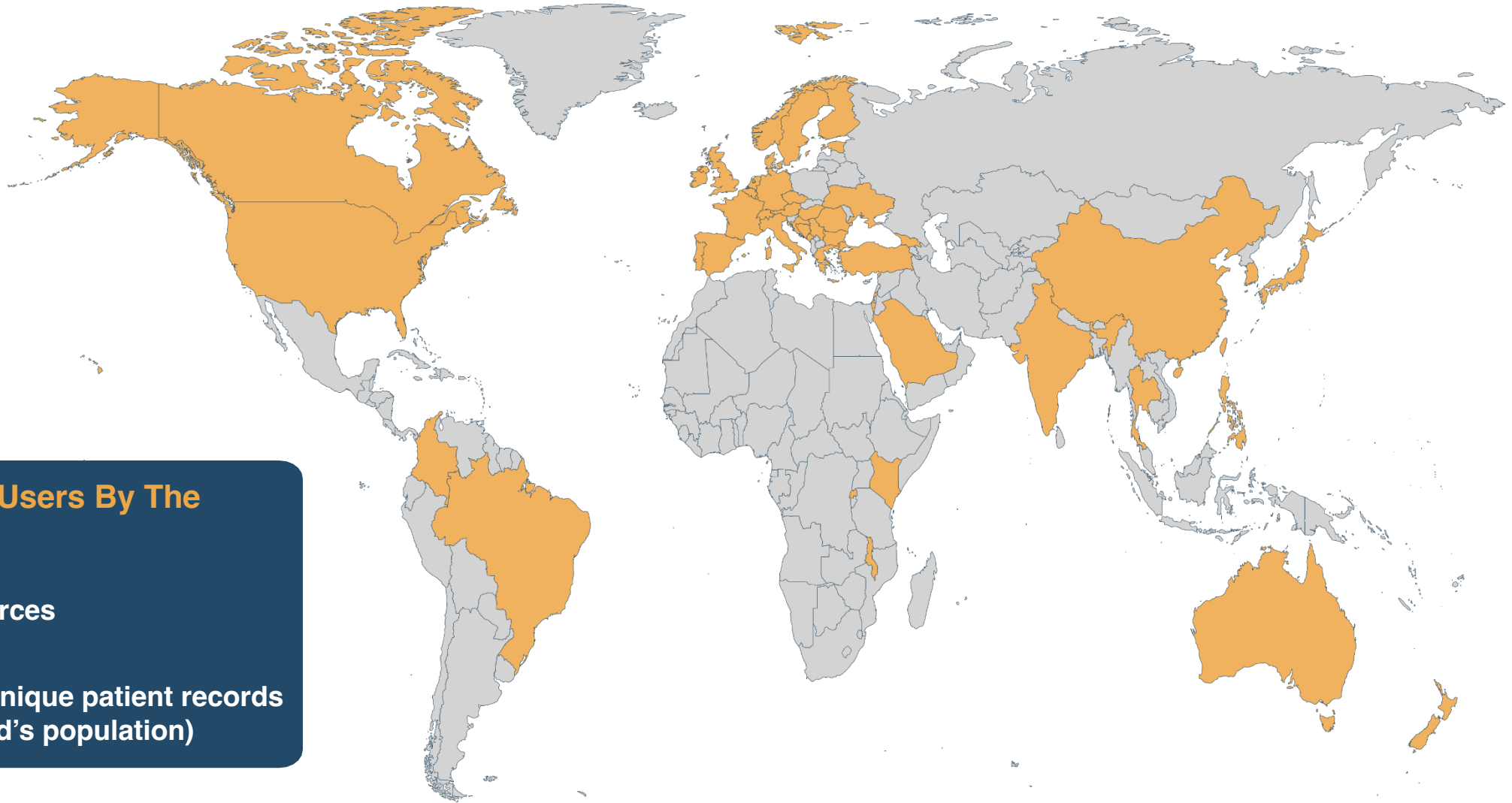


OHDSI Network studies





# OMOP Common Data Model adoption



## OMOP CDM Users By The Numbers

- 534 data sources
- 49 countries
- 956 million unique patient records (12% of world's population)

# Data Partner Organizations (DPO)



An institution



That owns or licenses data



That has been converted to the OMOP CDM v5.3+



Willingness to generate evidence and participate in network studies

# What you need in place

*To join the network as a DPO*

- Observational health data standardized to the OMOP CDM v5.3 or higher
- Data held in a relational database accessible by the organization
  - List of supported SQL environments here:  
<https://ohdsi.github.io/SqlRender/articles/UsingSqlRender.html#translation-to-other-sql-dialects>
- Approval from governance entity (i.e. IRB) to share metadata and concept counts with the OHDSI Coordinating Center (OCC)
  - Note: It is up to each individual DPO as owner or licensee of data to ensure all appropriate governance requirements are followed.
- The ability to run the DbDiagnostics R package against the data

# What you need to do

*To join the network as a DPO*

- Run the [DbDiagnostics package](#) `executeDbProfile` function to generate metadata and high-level concept counts about each data source submitted to the network
  - The aggregate information gathered by the package is listed here:  
<https://ohdsi.github.io/DbDiagnostics/articles/SummaryStatistics.html>
  - If the [Achilles](#) package was run previously and the results stored this step will take approximately 15-30 minutes, depending on the environment
  - If the Achilles package was not run previously or if the results were not stored this step will take approximately 1-8 hours, depending on the environment.
- Send the resulting information to the OCC via SFTP. Please contact [evidencenetwork@ohdsi.org](mailto:evidencenetwork@ohdsi.org) for the key file when you are ready

# What happens after you send your DbProfile to the OCC?

1. Your organization will receive an @ohdsi.org account i.e. [VA@ohdsi.org](mailto:VA@ohdsi.org) to be used to notify you of potential network studies and other internal communications
2. Your organization will be listed as an OHDSI Evidence Network DPO on the OHDSI.org website
3. You will receive a report from the OCC putting your data source in the context of network\*

\*Once there are enough participating data partner organizations



# Pre-evidence network: concept prevalence study

Collected concepts and their use across 22 data sources

## Network study: Concept Prevalence

■ Researchers



**aostropelets** Anna Ostropelets

Apr '19

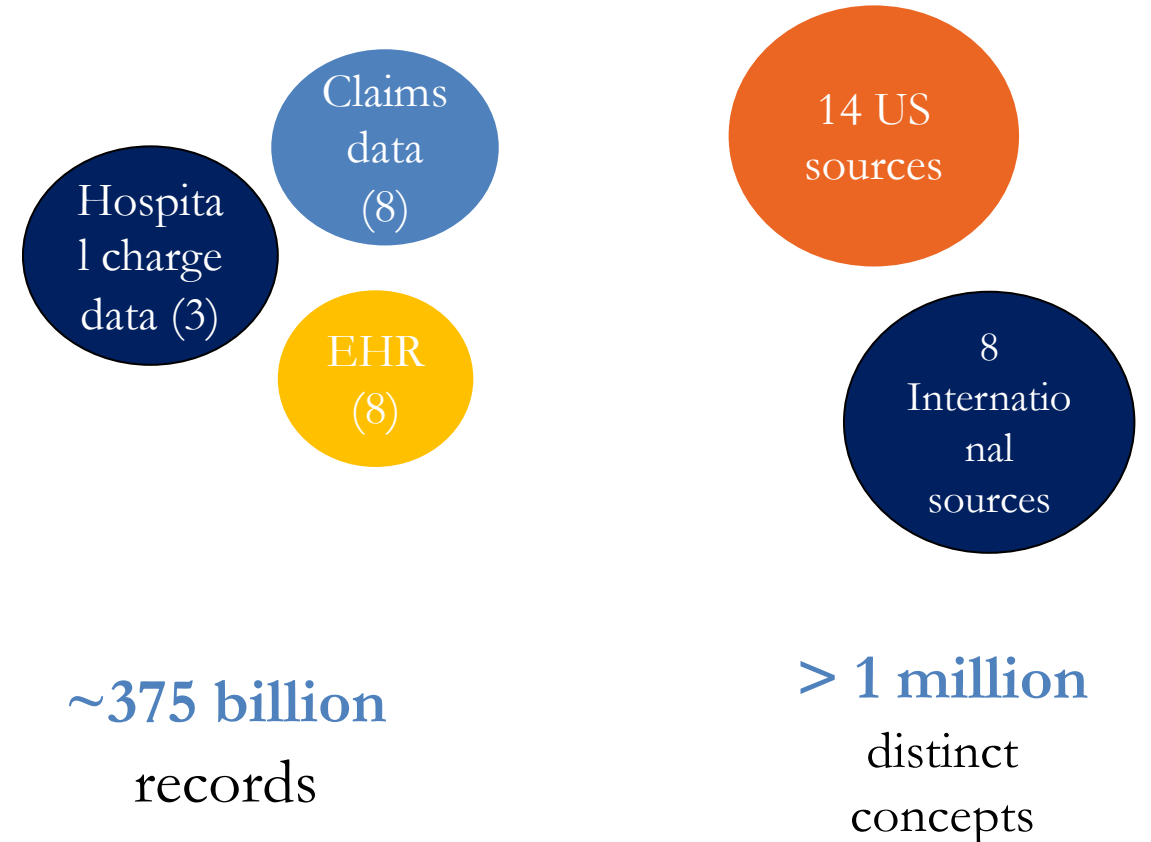
We want to announce a new network study:

<https://github.com/OHDSI/StudyProtocolSandbox/tree/master/ConceptPrevalence> 62

The full protocol can be found here:

[https://github.com/OHDSI/StudyProtocolSandbox/blob/master/ConceptPrevalence/extras/ConceptPrevalenceStudyProtocol\\_v0.1.docx](https://github.com/OHDSI/StudyProtocolSandbox/blob/master/ConceptPrevalence/extras/ConceptPrevalenceStudyProtocol_v0.1.docx) 43

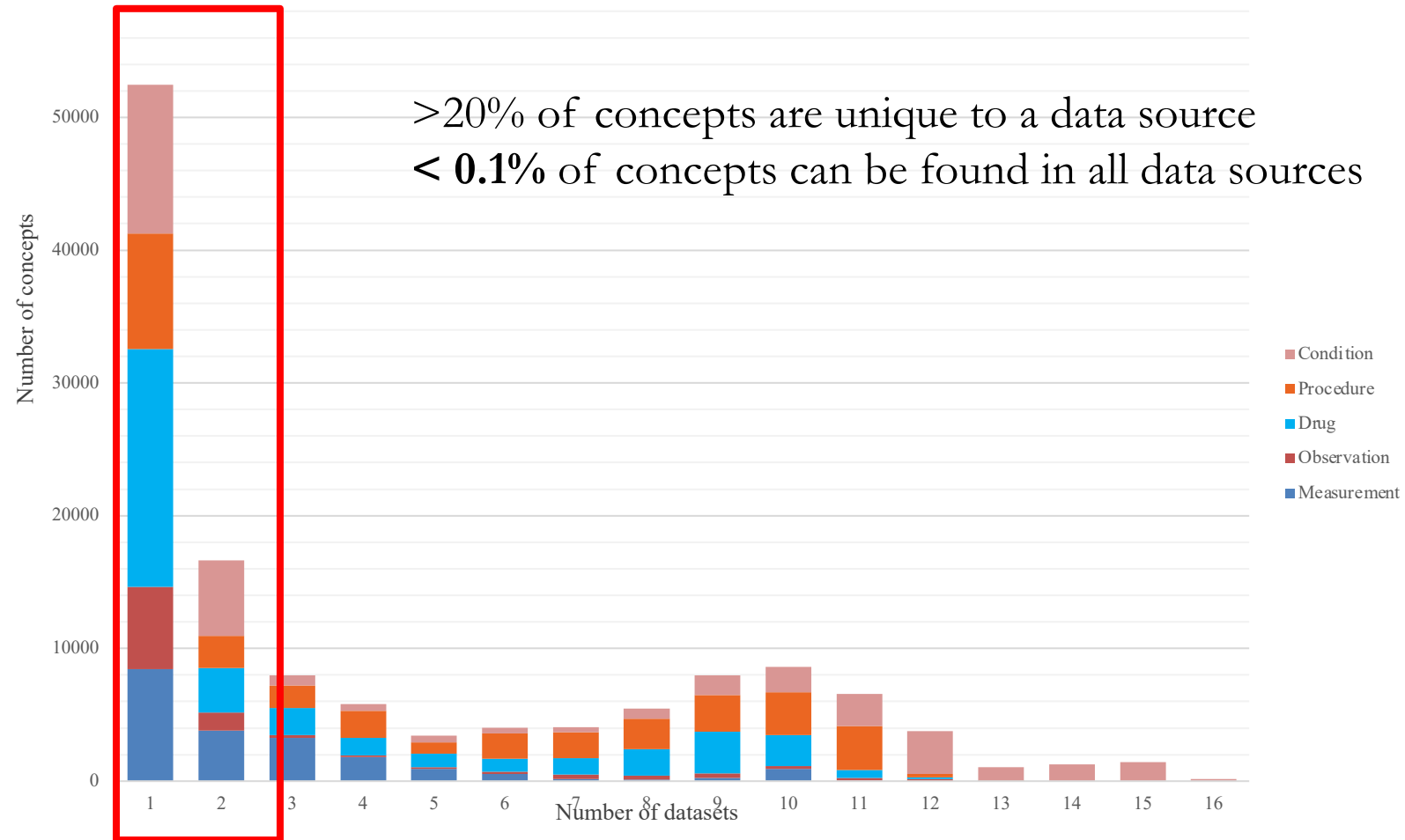
We want to study the usage patterns of Concepts across different OMOP CDM instances. This in itself could be useful information to answer many questions, but we have a concrete reason: For any one medical entity, the granularity of codes captured in a data source can vary greatly. For example, Chronic Kidney Disorder stage II can be coded as ICD9 code 585.2 Chronic kidney disease, Stage II (mild); 585.9 Chronic kidney disease, unspecified or even as 586 Renal failure, unspecified. However, this information is key for any cohort definition. Currently, researchers have no way of knowing whether a certain concept with high granularity is even available for selection, or whether they have to use a generic concept in combination with some auxiliary information to define the cohort correctly. Each data source instance is a black box and knowledge about the distribution of the concepts is limited to the very instance researchers have access to. But OHDSI Network Studies are dependent on cohort definitions that work across the network.







# Concept prevalence findings & use



Distribution of the overlapping concepts across OHDSI network



# Concept prevalence findings & use

PHenotype

University of Concept Prevalence concept

ATLAS English

- Home
- Data Sources
- Search
- Concept Sets
- Cohort Definitions

Search

Type your search

Concept Set #1884578

created by anonymous on 2024-07-31 2

type 2 diabetes\_2015-20

Concept Set Expression	Included Concepts (43)
------------------------	------------------------

Show columns

Showing 1 of 1 columns

Id	Count
Vocabulary	
SNOMED (203)	
Invalid Reason	
Valid (203)	4328
Class	
Disorder (188)	4427
Clinical Finding (12)	
Procedure (2)	
Context-dependent (1)	
Domain	
Condition (198)	



edburn Ed

25d

Thanks @Chris\_Knoll . I and @MaximMoinat can ask data partners we work with to upload the concept\_recommended.csv into the database as is.

For the longer term, what I'm a bit confused about is whether:

- phoebe is like any other vocabulary table provided by ohdsi which the data partner uploads (and then it seems like it could indeed be provided from athena with the others and maybe the code used to generate could live in <https://github.com/OHDSI/Vocabulary-v5.0> for transparency), or
- phoebe is more like achilles in that it is derived from the cdm data somebody has mapped, in which case maybe the code for it could even be added to achilles to increase uptake (as that is something people are used to running as they map a data source).

More than anything, the problem we currently have is that if a data partner does upload the concept\_recommended.csv how can we know that the results are consistent with the other vocabulary tables they have? And how would we know whether the data from that data partner was used to generate the recommendations?

Any thoughts @aostropelets @Christian\_Reich @Patrick\_Ryan @Frank ?

meimtus

Disorder due to type 2 diabetes mellitus

Disorder	264	1,171	157	712	Condition	SNOMED	3
----------	-----	-------	-----	-----	-----------	--------	---



# Phenotyping demo



# Evidence Network & OHDSI Vocabularies



# Landscape assessment of vocabularies use and needs

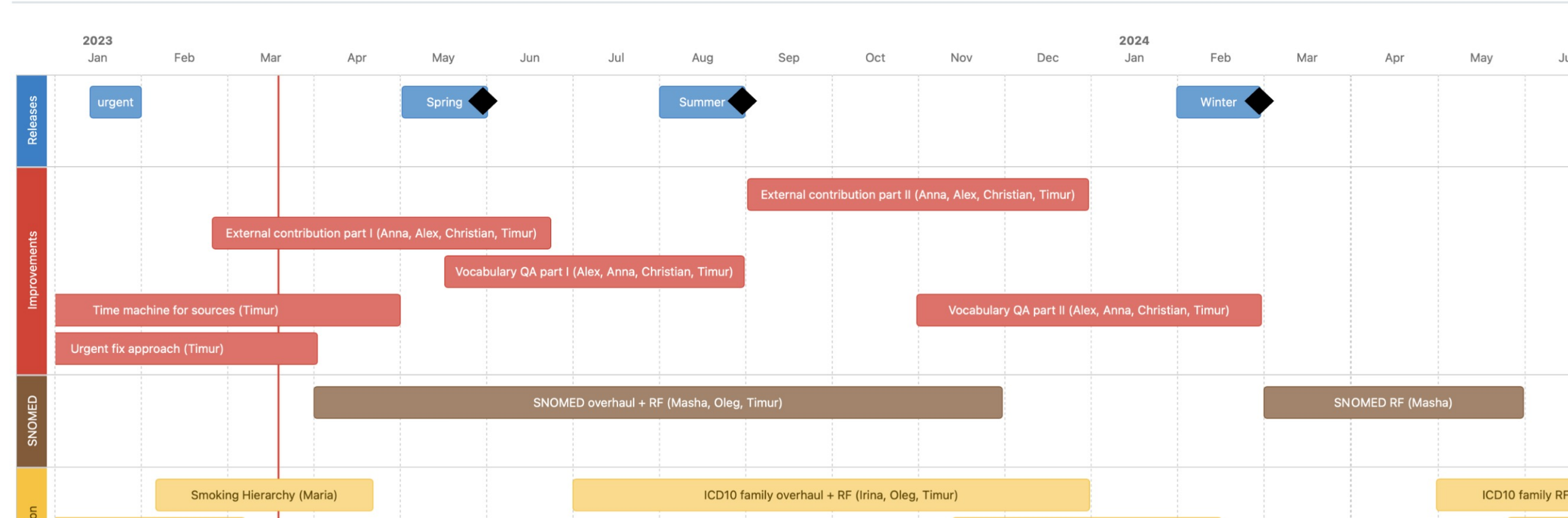
Part I: 188 responses from 144 institutions | Part II: Vocabularies use in 60 data sources

Vocabulary	Used in data	Used in research
ATC	45%	62%
CPT4	50%	40%
ICD-10(CM)	57%	<10%
ICD-9(CM)	62%	<10%
ICD-10PCS	43%	51%
ICD-9-Proc	48%	43%
LOINC	68%	25%
RxNorm	33%	79%
RxNorm Extension	<10%	53%
SNOMED	57%	86%
Cancer Modifier	<10%	25%
CVX	18%	13%
HCPCS	42%	33%



# Vocabulary refresh: Roadmap for 23/24 built based on landscape assessment

## Roadmap 2023 Q1 - 2024 Q2:





# Vocabulary refresh: Discussing the roadmap for 2025

Vocabulary	What (version)
CVX	refresh (Fall 2024 version)
HCPCS	refresh (Oct 2024 version)
ICD10PCS	refresh (2024 version)
RxNorm	refresh (Dec 2024 version)
NDC, SPL	refresh (Jan 2024 version)
RxNorm Extension	refresh (Feb 2024 version)
SNOMED UK	refresh (Oct 2024 version)
SNOMED Int	refresh (Fall 2024 version)
SNOMED US	refresh (Sept 2024 version)
ICD10(CM) and other versions	refresh (2023-24 versions)



Which local  
modifications to update?



# Vocabulary changes and enhancements (1)

## Improvement activities

Feb 24 release

Vocabulary-specific overhauls and improvements include:

### 1. SNOMED overhaul

- Stable domain and concept class id assignment.
- Alignment of the validity dates with the source.

### 3. MedDRA improvement

Feb 24 release

- Design and document the model that would allow to use MedDRA as both source and target in the Condition Domain.
- Development of system that would allow to re-use the mappings of various sources (UMLS), build our own based on the user needs, annotate them with metadata using automatically transform them using generated metadata in both horizontal and hierarchical directions.

## 2. ICD family improvement

Feb 24 release

- Mapping re-use across ICD family to identify the discrepancies and similarities improve the consistency of mappings.
- Incorporation of the mappings provided by SNOMED-CT and other sources.
- Fix of the source (CIAML) file processing to capture the ICD concepts currently missing.
- Documentation of the current procedures for mapping and quality assurance.

### 4. ATC overhaul

Aug 24 release

- Adopt the data-driven approach of attribute selection (RxNorm and RxNorm Extension) based on the data sources that have ATC codes (Z index, JMDC, others).
- Identification of discrepancies and similarities between code assignment in different data sources to ensure consistent and accurate mappings from ATC to RxNorm (Ext).
- Validation of the vocabulary using data-driven approaches (including currently existing mappings).



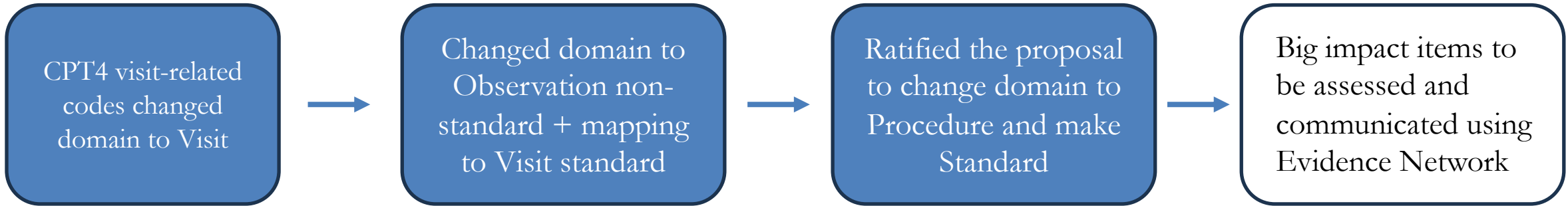


# Vocabulary changes and enhancements (2)

Jan 2023

Aug 2023

Aug 2024



Data partners use CPT4 codes to derive visits

forums.ohdsi.org/t/cpt-hierarchy-errors-lost-children-in-2023-and-changed-domains/18383

OHDSI Home | Forums | Wiki | Github

**OHDSI**

**CPT Hierarchy errors - lost children in 2023 and changed domains**

Thomas\_White Feb '23

There appears to have been a major change in the CPT hierarchy starting with the January 2023 release.

As an example, this CPT classification code subsumes the 4 Office or Outpatient visits (99202 - 99205). As expected, in the v5.0 22-JUN-22 and v5.0 31-OCT-22 Athena vocabulary releases, the ancestor\_concept\_id (45889484) in concept\_ancestor has the expected 4 values for descendant\_concept\_id.

However, in the v5.0 23-JAN-23 release, that concept has no descendant concepts. Moreover, the domain\_id has changed. In the 2022 releases, the concept is in the Procedure domain, but in the 2023 release, it is an Observation.

Here are greps of the raw downloaded files from the v5.0 22-JUN-22 release:

```

$ grep None VOCABULARY.csv
None OMOP Standardized Vocabularies OMOP generated v5.0 22-JUN-22 44819096
$ grep 2414392 CONCEPT_CPT4.csv
2414392 Procedure CPT4 CPT4 S 99203 19700101 209
$ grep 45889484 CONCEPT_ANCESTOR.csv
45889484 2414391 1 1
45889484 2414394 1 1
45889484 2414393 1 1
  
```

**Vocab WG July 30 meeting [CPT4 visits and postcoordination]**

aostroplets Anna Ostroplets 10d

Dear friends:

Please come to the July 30 noon EST [Vocab WG meeting](#) to discuss two vocabulary topics that brought multiple conversations in the community:

- CPT4 visits  
What: CPT4 codes representing visits have been de-standardized and mapped to Visit domain ([link](#), [link](#), [link](#), [link](#))
- Post-coordination for measurements in the OHDSI Vocabularies

What: some non-standard terms map to two concepts that live as a pair in different columns in OMOP CDM, [example](#), [link](#), [link](#), [link](#))



# Vocabulary changes and enhancements (3)

## 4. ATC overhaul

Aug 24 release

- Adopt the data-driven approach of attribute selection (RxNorm and RxNorm Extension) based on the data sources that have ATC codes (Z index, JMDC, others).
- Identification of discrepancies and similarities between code assignment in different data sources to ensure consistent and accurate mappings from ATC to RxNorm (Ext).
- Validation of the vocabulary using data-driven approaches (including currently existing data sources).

Review of errors based on counts from Concept Prevalence:

cnt	atc_name	rxnorm_name	source
92015665	mupirocin; inhalant, nasal	mupirocin Topical Ointment	grr
92015665	mupirocin; inhalant, nasal	mupirocin Topical Ointment	jmhc
73723924	prednisolone; ophthalmic (cortic...	prednisolone Oral Solution	dpc
55860886	leuprorelin; implant, parenteral	norethindrone Oral Tablet	z-index
31792562	betamethasone; topical	betamethasone Injectable Solution	grr
27617272	tolterodine; oral	tolterodine Extended Release Ora...	dmd



# Community contributions

Simple contributions incorporated so far:  
13 (all if submitted and processed before the deadline)

17 Open 13 Closed Author Label Projects Mi

- Autoimmune polyendocrine syndromes community contribution  
#940 by aaronabend was closed on Jul 5
- Lovastatin 60 MG Maps to Lovastatin 20 MG community contribution NDC  
#928 by mrechkem was closed on Mar 1
- ICD10CM G92.0 Immune effector cell-associated neurotoxicity syndrome (ICANS) ->Maps to SNOMED 1230414002 Immune effector cell-associated neurotoxicity syndrome community contribution ICD10CM  
#920 by dimshitc was closed on Mar 3
- OMOP Standard Concepts Deduplication Request community contribution LOINC SNOMED  
#915 by p-talapova was closed on Mar 1
- SNOMED - RxNorm/RxNorm Extension deduplication community contribution SNOMED  
#914 by makc4us was closed on Mar 1
- Sexual activity concepts in the incorrect domain of Condition.Should be Observation community contribution SNOMED  
#872 by MelaniePhilofsky was closed on Mar 1
- To fix that ICD10CM "substance use with other disorders" mapped to Mood disorder community contribution ICD10CM  
#856 by dimshitc was closed on Mar 3

Complex contributions incorporated so far:  
2 (drug vocabularies, 1 per release)

Requests for complex contributions:

	What	Content
1	Route hierarchy	Create new simple hierarchy for routes, de-standardize specific SNOMED routes.
2	SNOMED Veterinary	Refresh and de-coupling with SNOMED
3	CIEL	Refresh + modifications (
4	HemOnc	Oncology vocabulary with mixed domains
5	Vaccine ontology	Vaccines; support meaningful classification
6	ORPHANET	New non-standard vocabulary for rare diseases (can be incorporated through UMLS)
7	Spanish Drug Agency	Incorporation of Spanish drugs
8	DPD	Canadian drug vocabulary refresh
9	JMDC	Refresh of Japan drug vocab



# What we can learn from aggregated evidence network (demo)