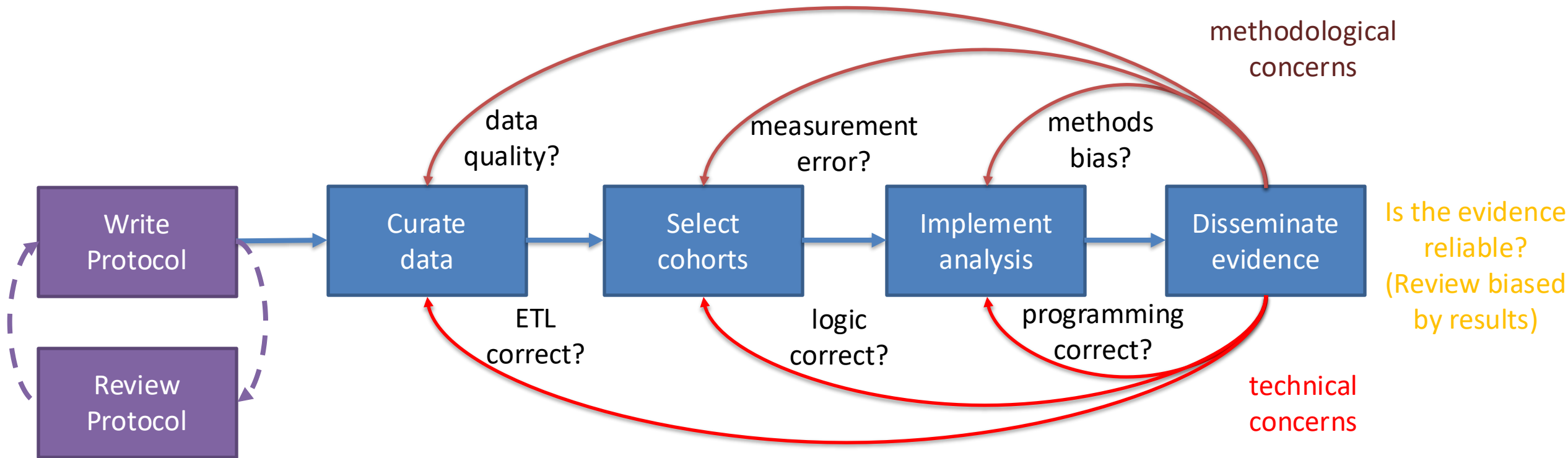# OHDSI for Real World Evidence (RWE)

Patrick Ryan, PhD

Vice President, Observational Health Data Analytics, Janssen Research and Development

Assistant Professor, Adjunct, Department of Biomedical Informatics, Columbia University Medical Center

Current status quo in observational research makes it challenging to build trust in evidence

Does the study provide an unbiased effect estimate?
Are the findings generalizable to the population of interest?

methodological concerns

data quality?

measurement error?

methods bias?

Write Protocol

Review Protocol

Curate data

Select cohorts

Implement analysis

Disseminate evidence

Is the evidence reliable? (Review biased by results)

ETL correct?

logic correct?

programming correct?

technical concerns

Can the study be fully reproduced?
Does the analysis actually do what the protocol said it would do?

Desired attributes for reliable evidence

| Desired attribute	Question	Researcher	Data	Analysis		Result
Repeatable	Identical	Identical	Identical	Identical	=	Identical
Reproducible	Identical	Different	Identical	Identical	=	Identical
Replicable	Identical	Same or different	Similar	Identical	=	Similar
Generalizable	Identical	Same or different	Different	Identical	=	Similar
Robust	Identical	Same or different	Same or different	Different	=	Similar
Calibrated	Similar (controls)	Identical	Identical	Identical	=	Statistically consistent

OHDSI's mission

To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care
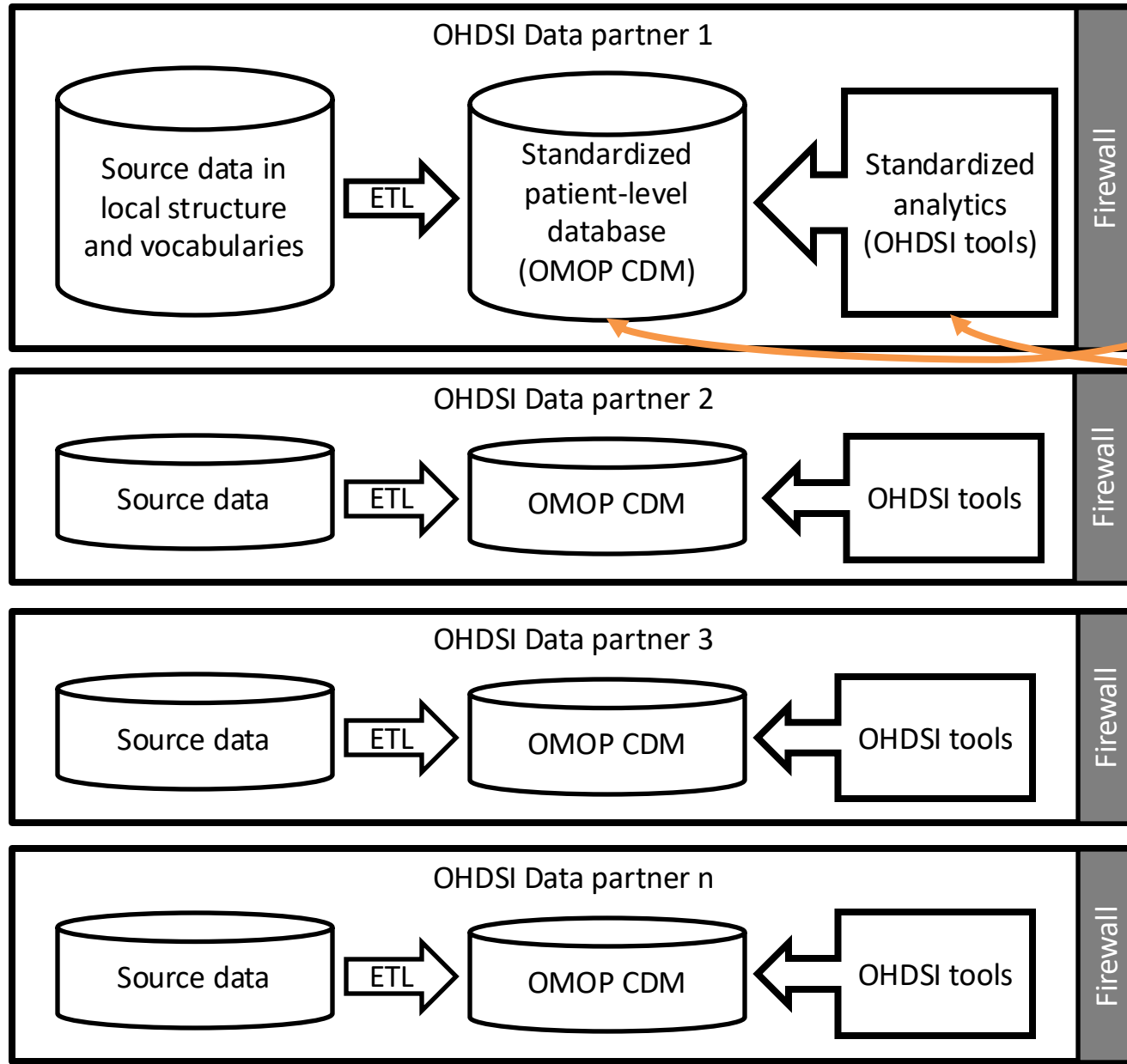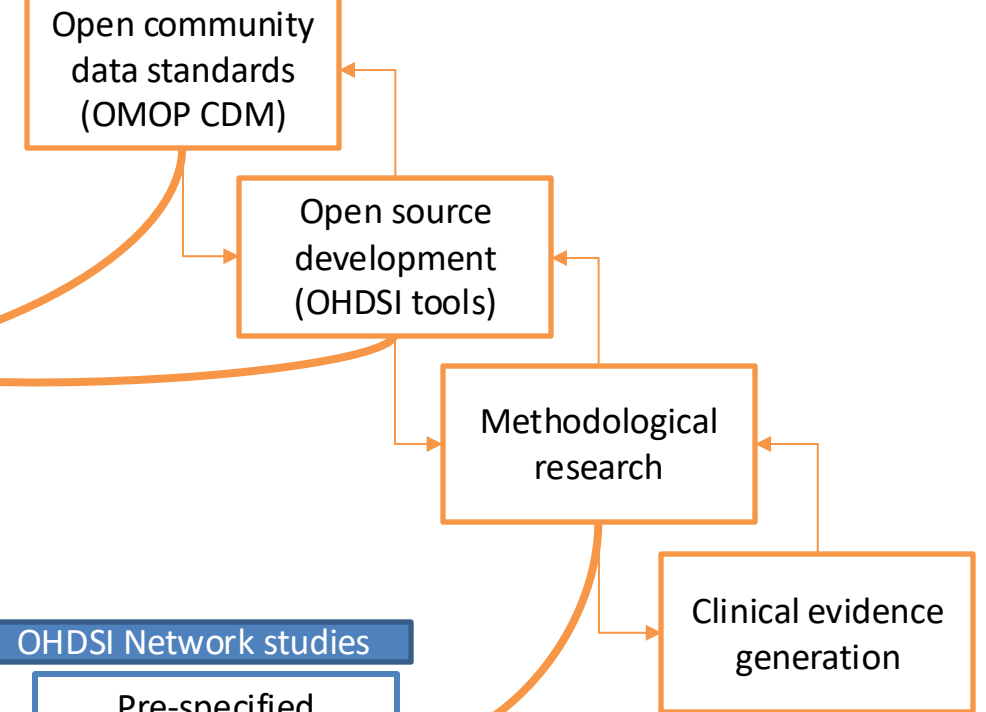
OHDSI collaborators
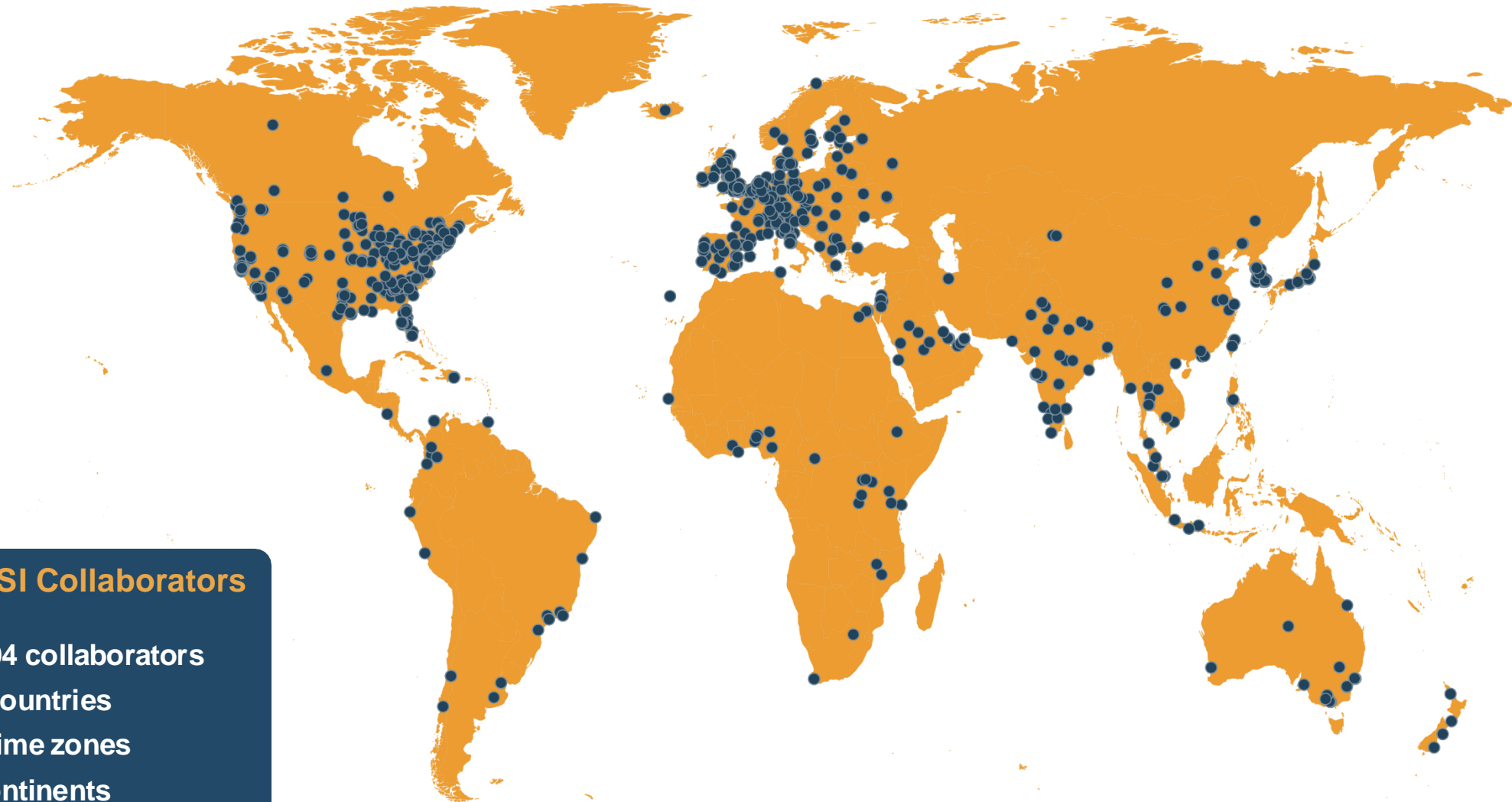
OHDSI Collaborators

- **4,294 collaborators**
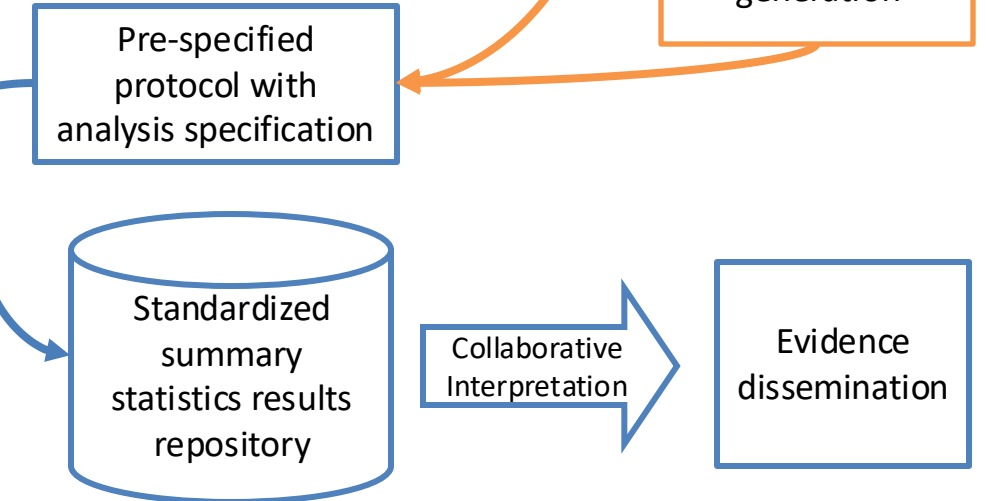- **83 countries**
- **21 time zones**
- **6 continents**

Join the Journey at https://ohdsi.org/

Workgroups led by community

Regional chapters and national nodes

Africa Chapter
- Agnes Kiragga
- Cynthia Sung

Asia-Pacific (APAC)
- Mui Van Zandt

Australia
- Nicole Pratt

China
- Hua Xu

Europe
- Peter Rijnbeek

India
- Swetha Kiranmayi Jakkuv
- Vikram Patil
- Parthiban Sulur

Japan
- Tatsuo Hiramatsu

Latin America
- Jose Posada

Republic of Korea
- Rae Woong Park

Singapore
- Seng Chan You

Singapore
- Mengling 'Mornin' Feng

Taiwan
- Jason Hsu

Node	Lead(s)
Belgium	Liesbet Peeters, Annelies Verbiest, Ilse Vermeulen
Denmark	Ismail Gögenur, Martin Høyer Rose, Andreas Weinberger Rosen
Estonia	Raivo Kolde, Sulev Reisberg
Finland	Eric Fey
Germany	Ines Reinecke, Michele Zoch
Greece	Anastasia Farmaki, Pantelis Natsiavas, Grigoris Papapostolou
Israel	Chen Yanover
Italy	Lucia Sacchi, Matteo Gabetta
Luxembourg	Claudine Backes, Andreas Kremer, Maria Quaranta
Netherlands	Renske Los, Aniek Markus
Norway	Espen Enerly, Siri Larønningen
Portugal	Patricia Couceiro, Carmen Nogueira
Spain	Miguel Angel Mayer, Talita Duarte Salles
United Kingdom	Daniel Prieto-Alhambra

OMOP Common Data Model adoption
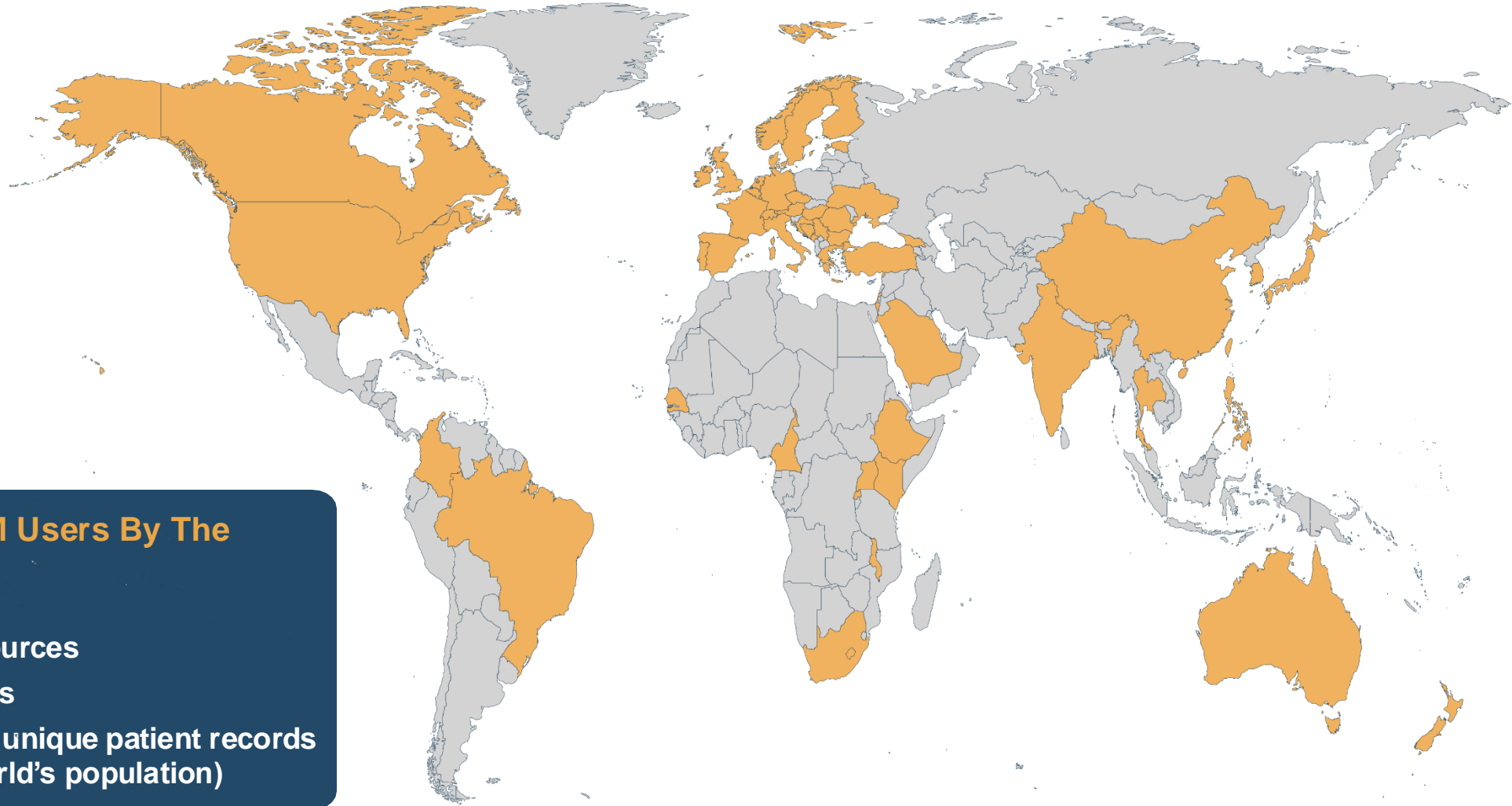
OMOP CDM Users By The Numbers

- **544 data sources**
- **54 countries**
- **974 million unique patient records (12% of world's population)**

OMOP Common Data Model v5.4
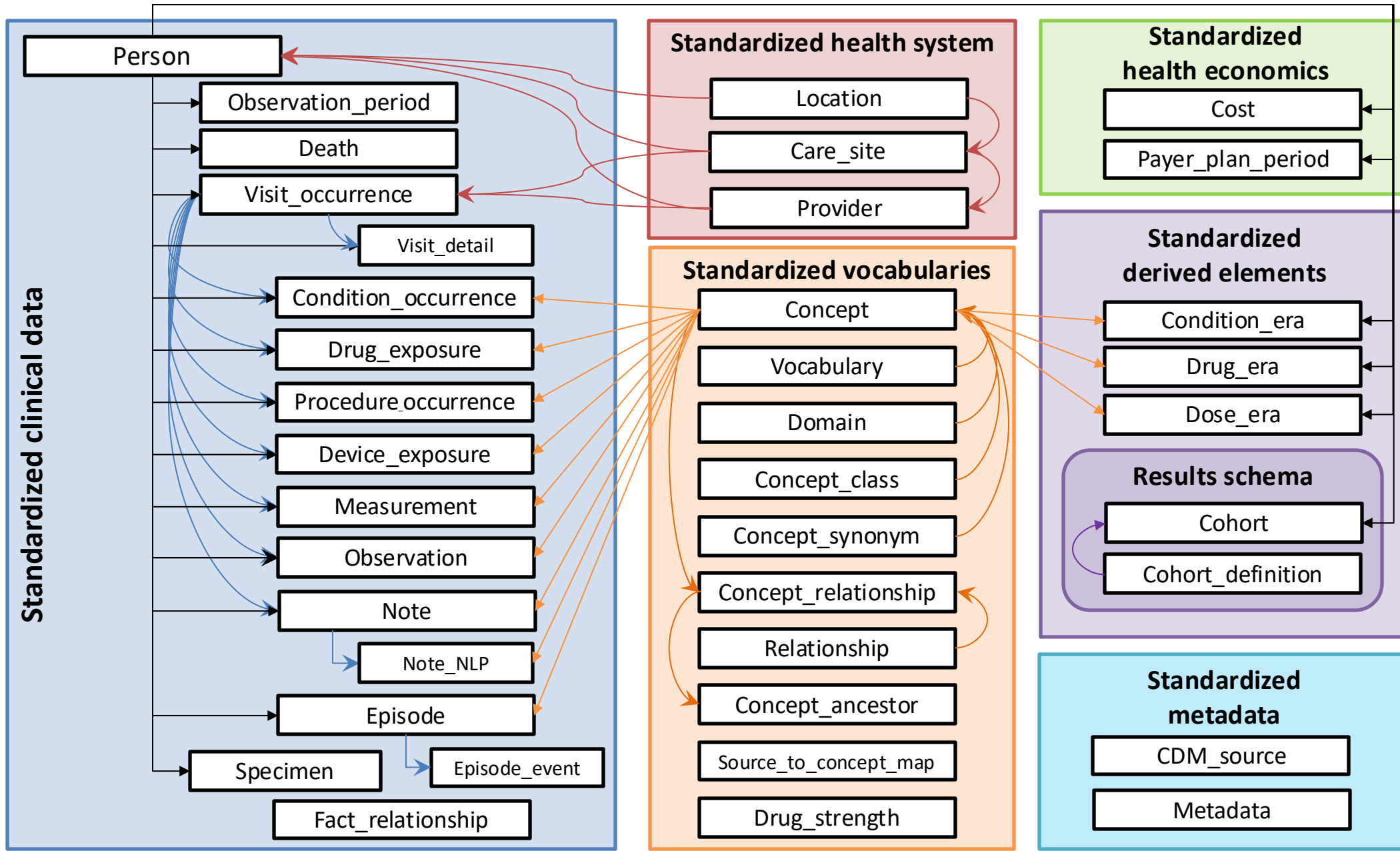
OHDSI standardized vocabularies
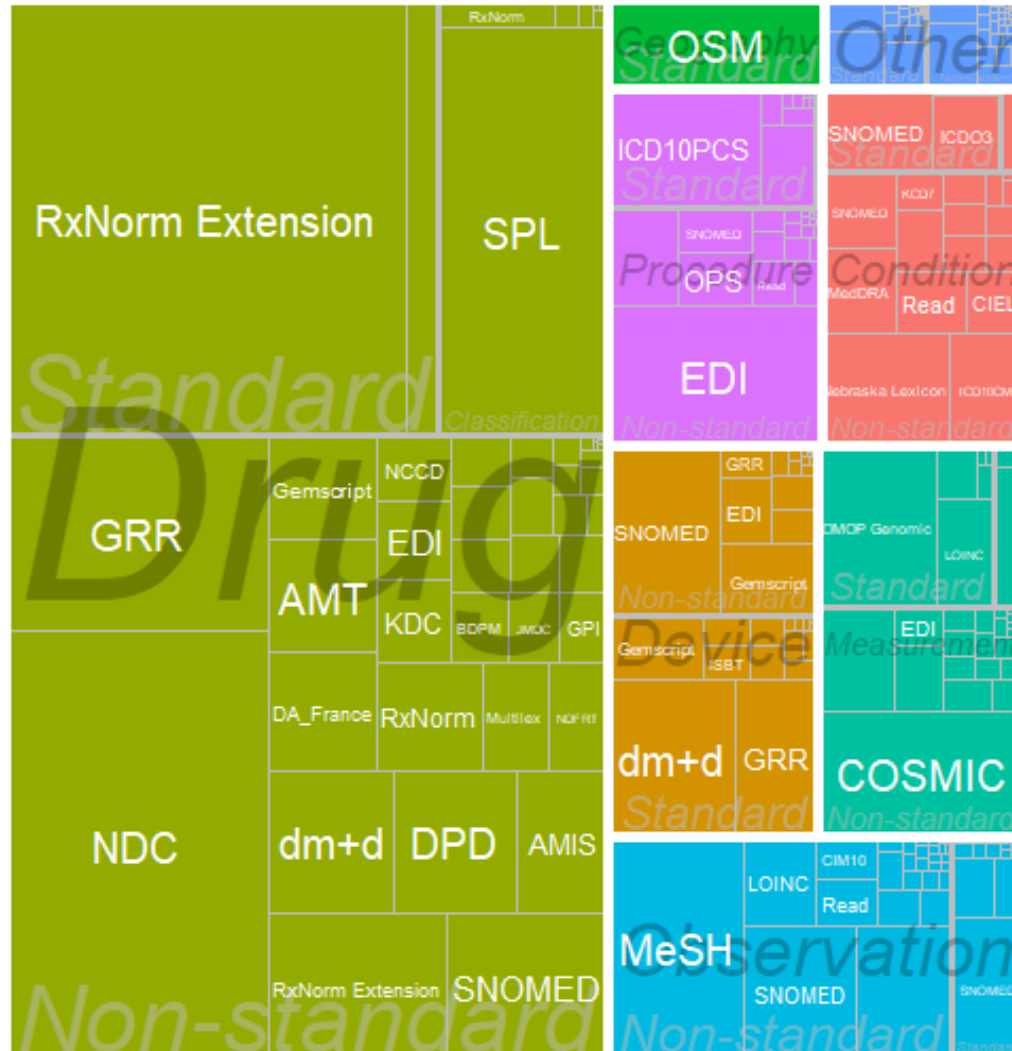
OHDSI Vocabularies By The Numbers
as of August 2024 release

- 11,561,982 concepts
  - 3,720,296 standard concepts
  - 883,766 classification concepts
- 143 vocabularies
- 43 domains

- 86,668,674 concept relationships
- 99,192,928 ancestral relationships
- 5,009,796 concept synonyms

1 Shared Resource to Enable Data Standards
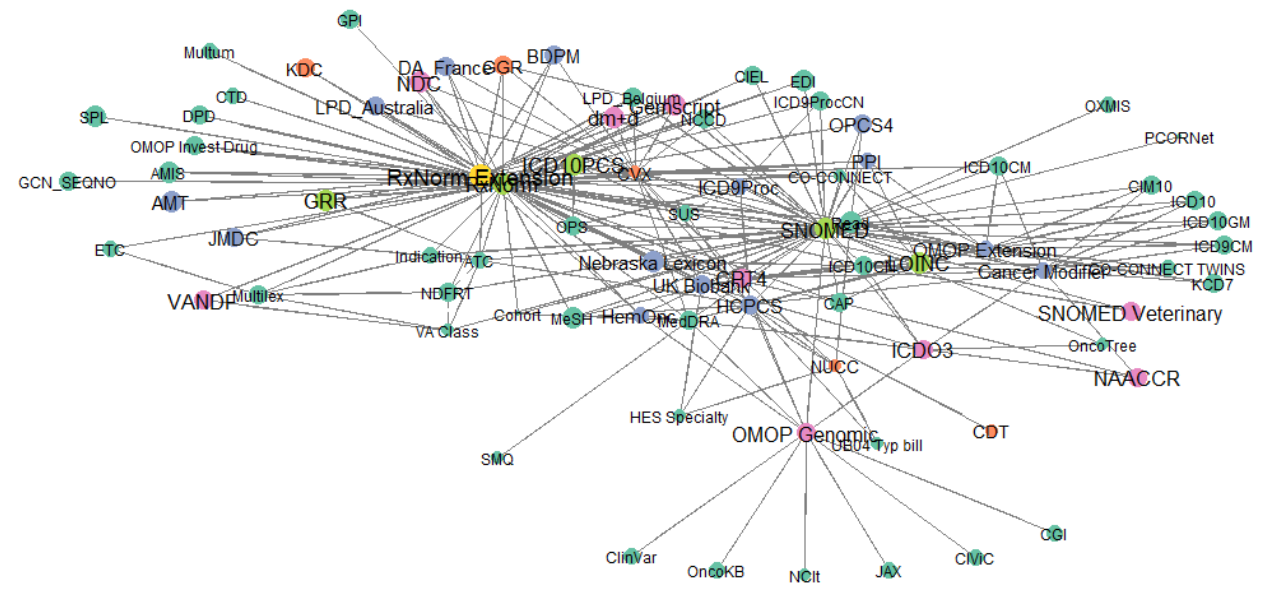
Common data model can enable standardized analytics across a distributed data network
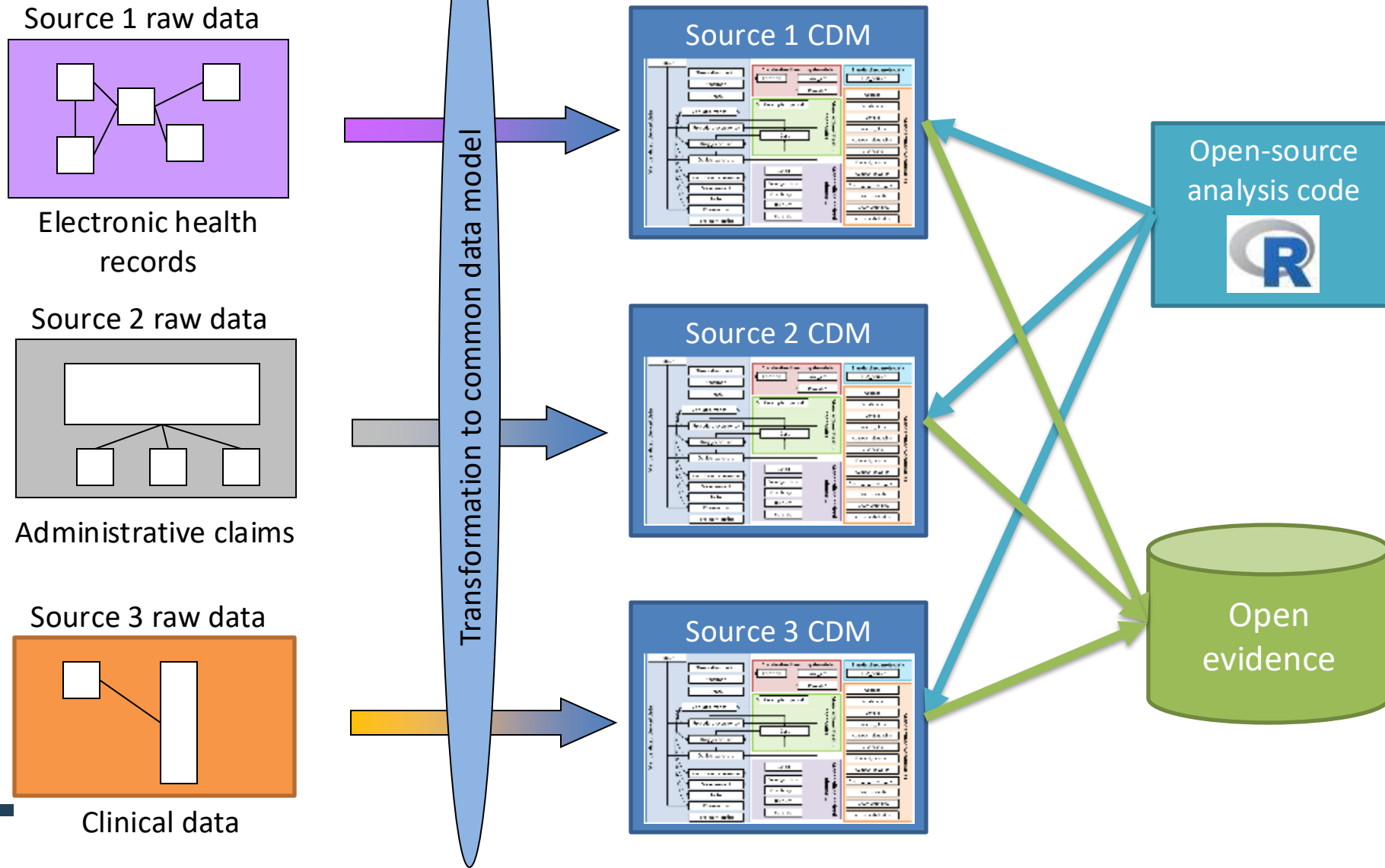
Open-source software development

- HADES is an ecosystem of 37 R packages to support standardized analytics for the OMOP CDM and across OHDSI network
- OHDSI CRAN packages have been downloaded >800,000 times

| Package	Version	Maintainer(s)	Availability
Achilles	v1.7.2	Frank DeFalco	CRAN
Andromeda	v0.6.7	Martijn Schuemie	CRAN
BigKnn	v1.0.2	Martijn Schuemie	GitHub
BrokenAdaptiveRidge	v1.0.0	Marc Suchard	CRAN
Capr	v2.0.8	Martin Lavallee	GitHub
Characterization	v2.0.1	Jenna Reps	GitHub
CirceR	v1.3.3	Chris Knoll	CRAN
CohortDiagnostics	v3.3.0	Jamie Gilbert	GitHub
CohortExplorer	v0.1.0	Gowtham Rao	CRAN
CohortGenerator	v0.11.2	Anthony Sena	GitHub
CohortIncidence	v4.0.0	Chris Knoll	GitHub
CohortMethod	v5.4.0	Martijn Schuemie	GitHub
Cyclops	v3.4.1	Marc Suchard	CRAN
DatabaseConnector	v6.3.2	Martijn Schuemie	CRAN
DataQualityDashboard	v2.6.1	Katy Sadowksi	GitHub
DeepPatientLevelPrediction	v2.1.0	Egill Fridgeirsson	GitHub
EmpiricalCalibration	v3.1.3	Martijn Schuemie	CRAN
EnsemblePatientLevelPrediction	v1.0.2	Jenna Reps	GitHub
Eunomia	v2.0.0	Frank DeFalco	CRAN
EvidenceSynthesis	v0.5.0	Martijn Schuemie	CRAN
FeatureExtraction	v3.7.1	Ger Inberg	CRAN
Hydra	v0.4.0	Anthony Sena	GitHub
IterativeHardThresholding	v1.0.2	Marc Suchard	CRAN
Keeper	v0.2.0	Anna Ostropolets	GitHub
MethodEvaluation	v2.3.0	Martijn Schuemie	GitHub
OhdsiSharing	v0.2.2	Lee Evans	GitHub
OhdsiShinyModules	v3.0.2	Jenna Reps	GitHub
ParallelLogger	v3.3.1	Martijn Schuemie	CRAN
PatientLevelPrediction	v6.3.9	Egill Friogeirsson & Jenna Reps	GitHub
PhenotypeLibrary	v3.34.0	Gowtham Rao	GitHub
PheValuator	v2.2.11	Joel Swerdel	GitHub
ResultModelManager	v0.5.11	Jamie Gilbert	GitHub
ROhdsiWebApi	v1.3.3	Gowtham Rao	GitHub
SelfControlledCaseSeries	v5.3.0	Martijn Schuemie	GitHub
SelfControlledCohort	v1.6.0	Jamie Gilbert	GitHub
ShinyAppBuilder	v3.1.0	Jenna Reps	GitHub
SqlRender	v1.18.1	Martijn Schuemie	CRAN

The open-source tools that empower OHDSI research are not only available to the community, but they are DEVELOPED by the community. We thank the many developers and maintainers who empower our research initiatives around the world!

Cumulative Downloads of OHDSI and DARWIN-EU R Packages by Year
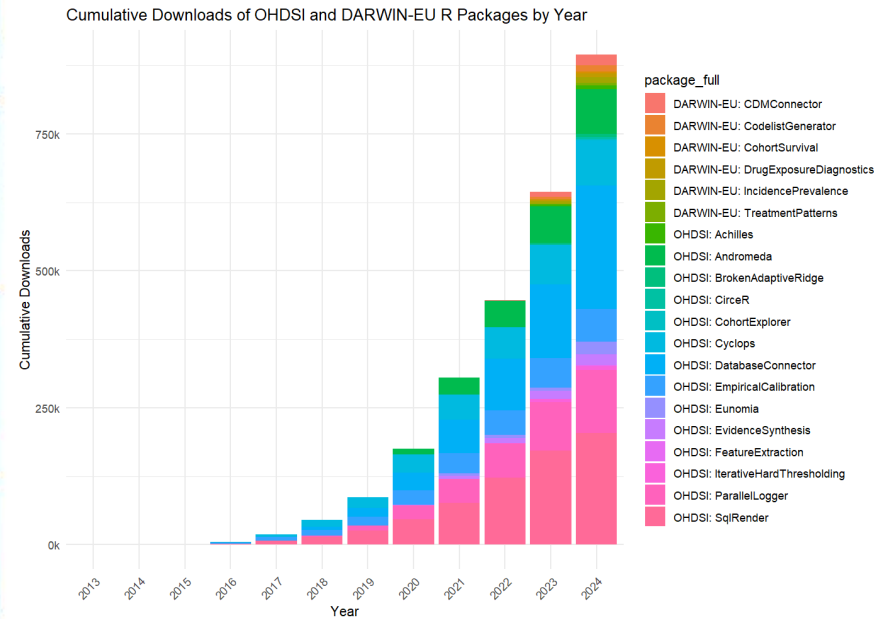
Complementary evidence to inform the patient journey
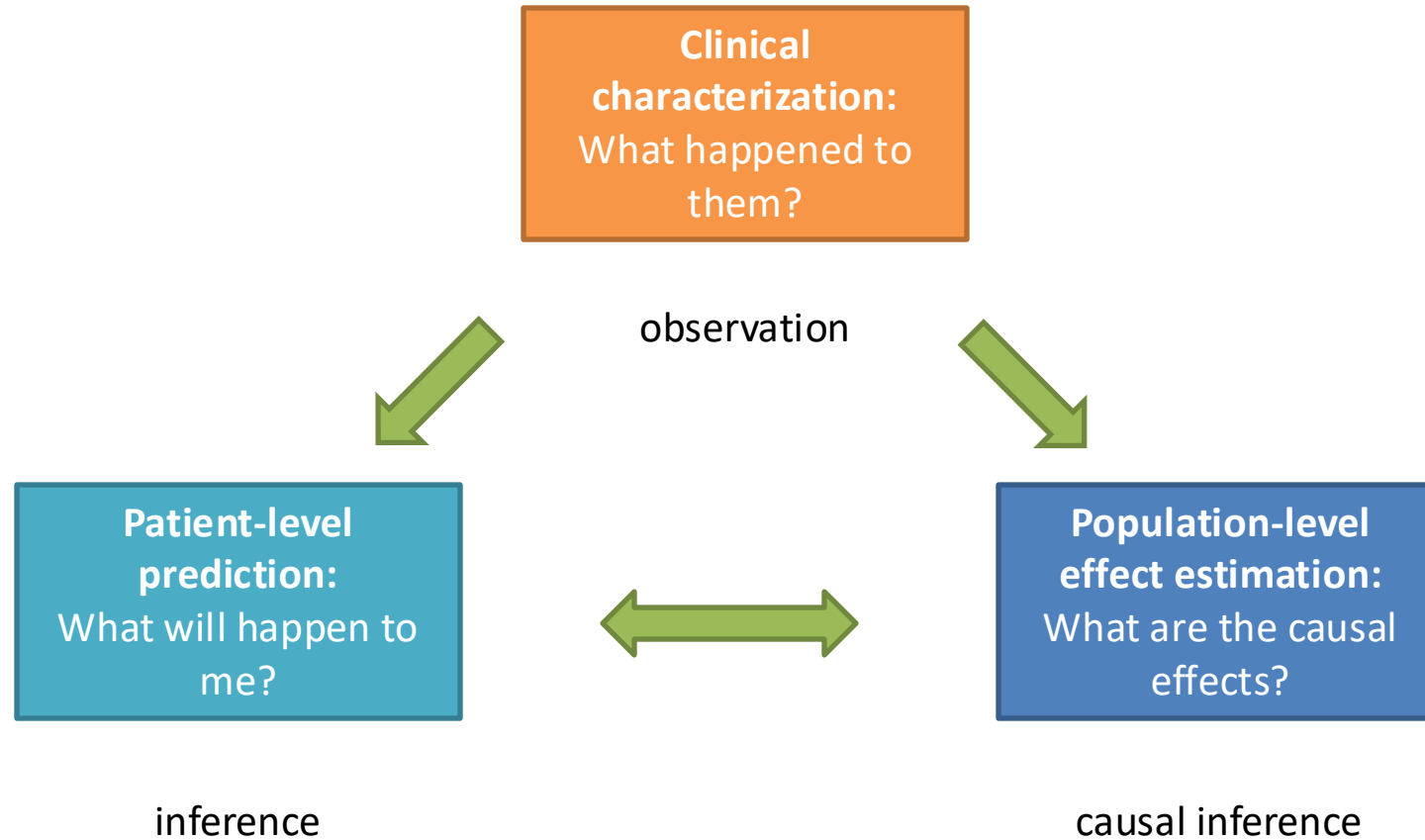
Standardizing the question makes it possible to standardize the analysis and standardize the evidence
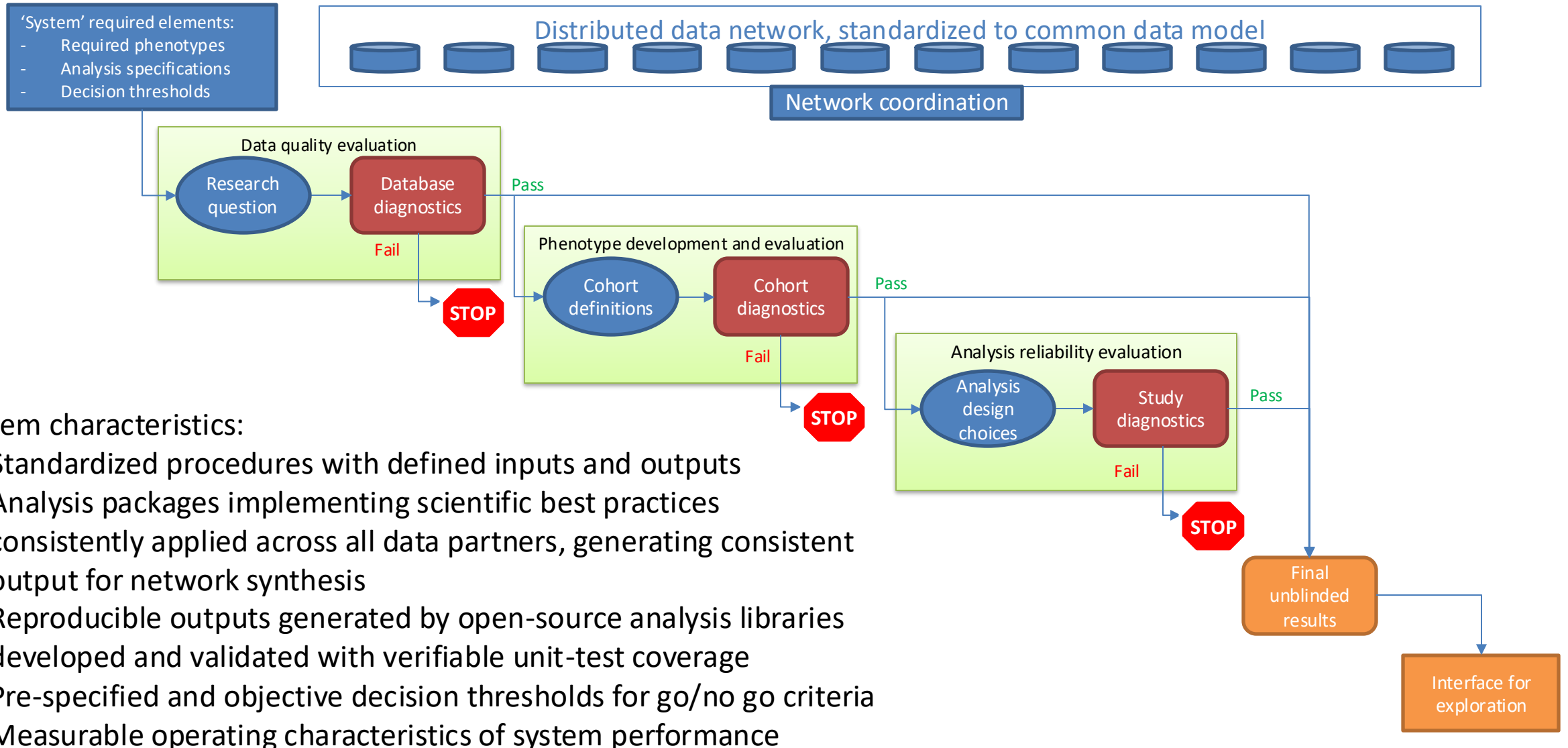
Analytic use case	Type	Structure
Clinical characterization	Disease Natural History	Amongst patients who are diagnosed with **\<insert disease of interest\>**, what are the patient's characteristics from their medical history?
	Treatment utilization	Amongst patients who have **\<insert disease of interest\>**, which treatments were patients exposed to amongst **\<list of treatments for disease\>** and in which sequence?
	Outcome incidence	Amongst patients who are new users of **\<insert drug of interest\>** among the population with **\<insert indication of interest\>**, how many patients experienced **\<insert outcome of interest\>** within **\<time horizon following exposure start\>**?
Population-level effect estimation	Safety surveillance	Does exposure to **\<insert drug of interest\>** increase the risk of experiencing **\<insert an adverse event\>** within **\<time horizon following exposure start\>**, among the population with **\<insert indication of interest\>**?
	Comparative effectiveness	Does exposure to **\<insert drug of interest\>** have a different risk of experiencing **\<insert any outcome (safety or benefit) \>** within **\<time horizon following exposure start\>**, relative to **\<insert comparator treatment\>**, among the population with **\<insert indication of interest\>**?
Patient level prediction	Disease onset and progression	For a given patient who is diagnosed with **\<insert your favorite disease\>**, what is the probability that they will go on to have **\<another disease or related complication\>** within **\<time horizon from diagnosis\>**?
	Treatment response	For a given patient who is a new user of **\<insert drug of interest\>** for **\<insert indication of interest\>**, what is the probability that they will **\<insert desired effect\>** in **\<time window\>**?
	Treatment safety	For a given patient who is a new user of **\<insert drug of interest\>** for **\<insert indication of interest\>**, what is the probability that they will experience **\<insert adverse event\>** within **\<time horizon following exposure\>**?

Engineering open science systems that build trust into the real-world evidence generation and dissemination process

'System' required elements:
- Required phenotypes
- Analysis specifications
- Decision thresholds

Distributed data network, standardized to common data model

Network coordination

Data quality evaluation
- Research question → Database diagnostics — Pass
- Fail → STOP

Phenotype development and evaluation
- Cohort definitions → Cohort diagnostics — Pass
- Fail → STOP

Analysis reliability evaluation
- Analysis design choices → Study diagnostics — Pass
- Fail → STOP

Final unblinded results → Interface for exploration

System characteristics:
- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance

Academic scholarship and clinical evidence generation

- >730 publications, including in top clinical journals (JAMA, BMJ, Lancet, JAMA Internal Medicine, JACC) and leading methodological journals (JAMIA, JBI, Nature Digital Medicine)

- Clinical evidence generated to inform range of therapeutic areas, including hypertension, diabetes, COVID-19, vision care, depression, oncology
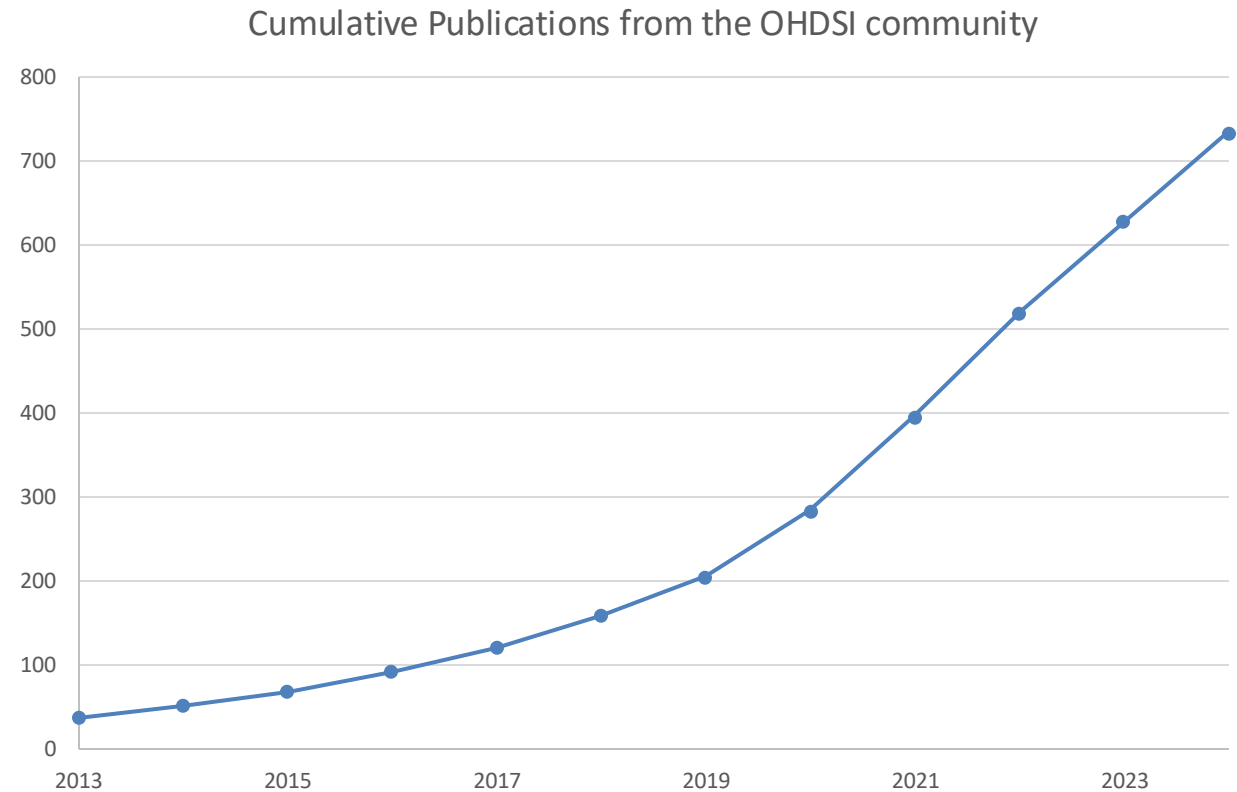
Cumulative Publications from the OHDSI community

Why OHDSI needs Singapore and APAC

THE LEGATUM PROSPERITY INDEX™ 2023
Advancing the understanding of what drives success in nations

RANK	COUNTRY												
ADJUST PILLAR WEIGHTING		x1	x1	x1	x1	x1	x1	x1	x1	x1	x1	x1	x1
17	Singapore	7	107	18	21	4	11	1	1	14	1	1	87
16	Japan	5	27	17	141	5	5	8	28	17	2	13	15
29	South Korea	37	42	30	107	25	37	21	9	26	3	3	63
20	Taiwan, China	4	26	20	44	23	7	34	12	31	4	16	73
54	China	86	162	74	31	64	38	43	24	46	5	56	139
33	Israel	124	49	22	83	15	17	33	19	12	6	26	96
3	Norway	3	1	2	3	9	14	20	5	4	7	10	9
8	Iceland	8	7	10	5	19	20	19	21	10	8	11	13
2	Sweden	10	4	7	4	6	15	7	8	3	9	8	1
5	Switzerland	2	12	4	12	12	1	11	2	5	10	4	6
6	Netherlands	9	5	5	8	3	9	3	10	1	11	6	39
7	Luxembourg	1	6	8	18	20	4	6	6	6	12	36	10
9	Germany	19	9	9	20	21	6	5	11	8	13	19	12
22	Hong Kong	16	98	28	68	1	2	2	3	27	14	9	30

Why Singapore needs OHDSI

Drug Outcome Incidence in a large US claims database

If health data for the entire Singapore population of 5.9 million were accessible, then ~5% of questions on drug-outcome pairs may have sufficient statistical power to answer alone if the incidence are comparable to US population....

...but even with a national database, >70% of drug-outcome questions are likely to have insufficient data for prediction or estimation, so an international network study would be required to generate reliable answers

Drug-outcome Incidence

- < 0.000001
- 0.000001 - 0.00001
- 0.00001 - 0.0001
- 0.0001 - 0.001
- > 0.001

Top 1000 Outcomes

Concluding thoughts

- Enabling use and establishing value of real-world evidence is a reasonable vision, which requires building trust across evidence generators and consumers
- People and processes need to be augmented with science, technology and engineering
- Community efforts today can enable a more proactive future tomorrow
  - Data network standardization and quality assessment
  - Standardized analytic tool development
  - Methodological benchmarks and objective diagnostics
  - Phenotype development and evaluation
- Open science systems that promote transparency and reproducibility can increase reliability and efficiency
- We need an international community working together in order to meet national, regional and global public health needs