A Common Data Model- Which? Overview of the OMOP Common Data Model

Peter Rijnbeek, PhD
Department of Medical Informatics
Erasmus MC, Rotterdam, The Netherlands

Observational Health Data Sciences and Informatics (OHDSI) Mission

To improve health, by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

Hripcsak G, et al. (2015) Observational Health Data Sciences and Informatics (OHDSI): Opportunities for observational researchers. Stud Health Technol Inform 216:574–578.

Objectives

1. **Innovation**: Observational research is a field which will benefit greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.

2. **Reproducibility**: Accurate, reproducible, and well-calibrated evidence is necessary for health improvement.

3. **Community**: Everyone is welcome to actively participate in OHDSI, whether you are a patient, a health professional, a researcher, or someone who simply believes in our cause.

4. **Collaboration**: We work collectively to prioritize and address the real world needs of our community's participants.

5. **Openness**: We strive to make all our community's proceeds open and publicly accessible, including the methods, tools and the evidence that we generate.

6. **Beneficence**: We seek to protect the rights of individuals and organizations within our community at all times.

Source data = source structure, source content, source conventions

Truven MarketScan Commerical Claims and Encounters (CCAE): INPATIENT_SERVICES

enrolid	admdate	pdx	dx1	dx2	dx3
157033702	5/31/2000	41071	41071	4241	V5881

Optum Extended SES: MEDICAL_CLAIMS

patid	fst_dt	diag1	diag2	diag3	diag4
259000474406532	5/30/2000	41071	27800	4019	2724

Premier: PATICD_DIAG

pat_key	period	icd_code	ic
-171971409	1/1/2000	410.71	P
-171971409	1/1/2000	414.01	S
-171971409	1/1/2000	427.31	S
-171971409	1/1/2000	496	S

JMDC: DIAGNOSIS

member_id	admission_date	icd10_level4_code
M004149337	4/11/2013	I214
M004149337	4/11/2013	A539
M004149337	4/11/2013	B182
M004149337	4/11/2013	E14-

4 real observational databases, all containing an inpatient admission for a patient with a diagnosis of 'acute subendocardial infarction'

- Not a single table name the same…
- Not a single variable name the same….
- Different table structures (rows vs. columns)
- Different ICD9 conventions (with and without decimal points)
- Different coding schemes (ICD9 vs. ICD10)

OMOP CDM = Standardized structure: same tables, same fields, same datatypes, same conventions across disparate sources

Truven MarketScan Commerical Claims and Encounters (CCAE): INPATIENT_SERVICES

enrolid	admdate	pdx	dx1	dx2	dx3
157033702	5/31/2000	41071	41071	4241	V5881

Optum Extended SES: MEDICAL_CLAIMS

patid	fst_dt	diag1	diag2	diag3	diag4
259000474406532	5/30/2000	41071	27800	4019	2724

Premier: PATICD_DIAG

pat_key	period	icd_code	icd_pri_sec
-171971409	1/1/2000	410.71	P
-171971409	1/1/2000	414.01	S
-171971409	1/1/2000	427.31	S
-171971409	1/1/2000	496	S

JMDC: DIAGNOSIS

member_id	admission_date	icd10_level4_code
M004149337	4/11/2013	I214
M004149337	4/11/2013	A539
M004149337	4/11/2013	B182
M004149337	4/11/2013	E14-

- Consistent structure optimized for large-scale analysis
- Structure preserves all source content and provenance

Truven CCAE: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID
157033702	5/31/2000	41071	Inpatient claims - primary position
157033702	5/31/2000	41071	Inpatient claims - 1st position
157033702	5/31/2000	4241	Inpatient claims - 2nd position
157033702	5/31/2000	V5881	Inpatient claims - 3rd position

Optum Extended SES: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID
259000474406532	5/30/2000	41071	Inpatient claims - 1st position
259000474406532	5/30/2000	27800	Inpatient claims - 2nd position
259000474406532	5/30/2000	4019	Inpatient claims - 3rd position
259000474406532	5/30/2000	2724	Inpatient claims - 4th position

Premier : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID
-171971409	1/1/2000	410.71	Hospital record - primary
-171971409	1/1/2000	414.01	Hospital record - secondary
-171971409	1/1/2000	427.31	Hospital record - secondary
-171971409	1/1/2000	496	Hospital record - secondary

JMDC : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID
4149337	4/11/2013	I214	Inpatient claims
4149337	4/11/2013	A539	Inpatient claims
4149337	4/11/2013	B182	Inpatient claims
4149337	4/11/2013	E14-	Inpatient claims

OMOP CDM = Standardized content: common vocabularies across disparate sources

Truven CCAE: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
157033702	5/31/2000	41071	Inpatient claims - primary position	44825429	444406

Optum Extended SES: CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
259000474406532	5/30/2000	41071	Inpatient claims - 1st position	44825429	444406

Premier : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
-171971409	1/1/2000	410.71	Hospital record - primary	44825429	444406

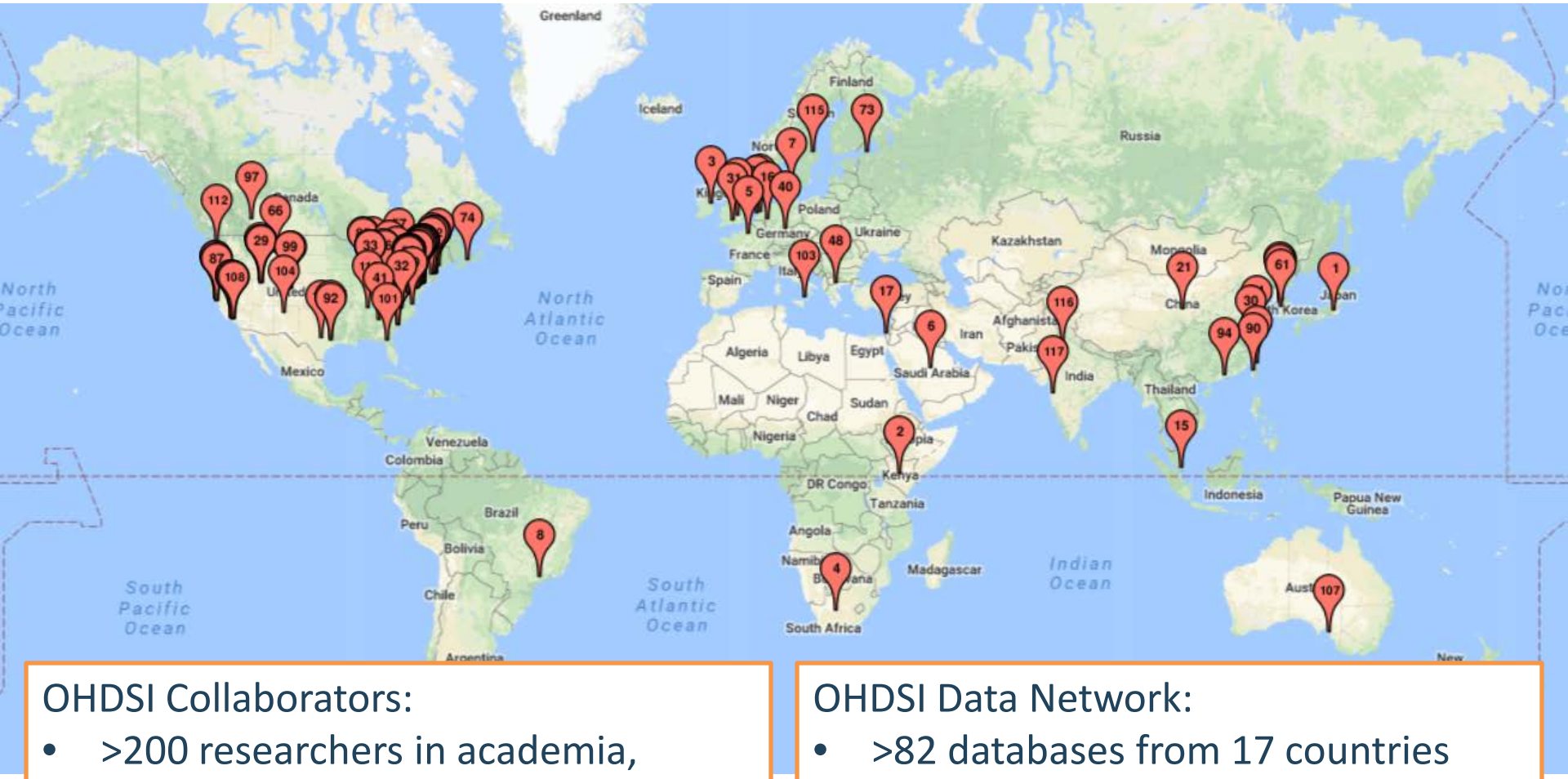
JMDC : CONDITION_OCCURRENCE

PERSON_ID	CONDITION_START_DATE	CONDITION_SOURCE_VALUE	CONDITION_TYPE_CONCEPT_ID	CONDITION_SOURCE_CONCEPT_ID	CONDITION_CONCEPT_ID
4149337	4/11/2013	I214	Inpatient claims	45572081	444406

- Standardize source codes to be uniquely defined across all vocabularies
- No more worries about formatting or code overlap

- Standardize across vocabularies to a common referent standard (ICD9/10→SNOMED)
- Source codes mapped into each domain standard so that now you can talk across different languages

OHDSI: a global community



OHDSI Collaborators:

- >200 researchers in academia, industry and government
- >17 countries

OHDSI Data Network:

- >82 databases from 17 countries
- 1.2 billion patients records (duplicates)
- ~115 million non-US patients

http://www.ohdsi.org/web/wiki/doku.php?id=resources:2017_data_network

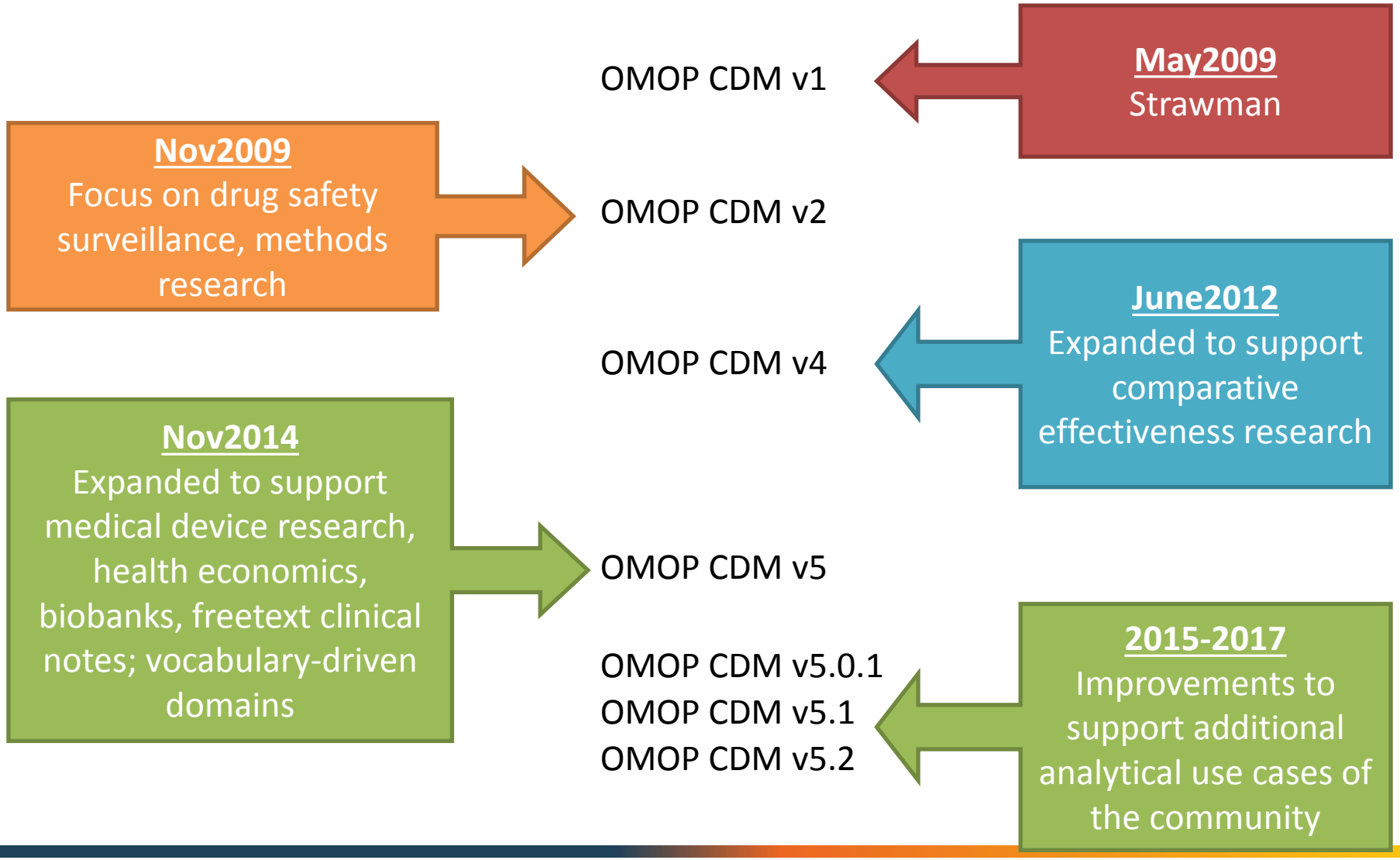
Objectives in OMOP Common Data Model development

- One model to accommodate both administrative claims and electronic health records
  - Claims from private and public payers, and captured at point-of-care
  - EHRs from both inpatient and outpatient settings
  - Also used to support registries and longitudinal surveys
- One model to support collaborative research across data sources both within and outside of US
- One model that can be manageable for data owners and useful for data users (efficient to put data IN and get data OUT)
- Enable standardization of structure, content, and analytics focused on specific use cases

OMOP CDM Principles

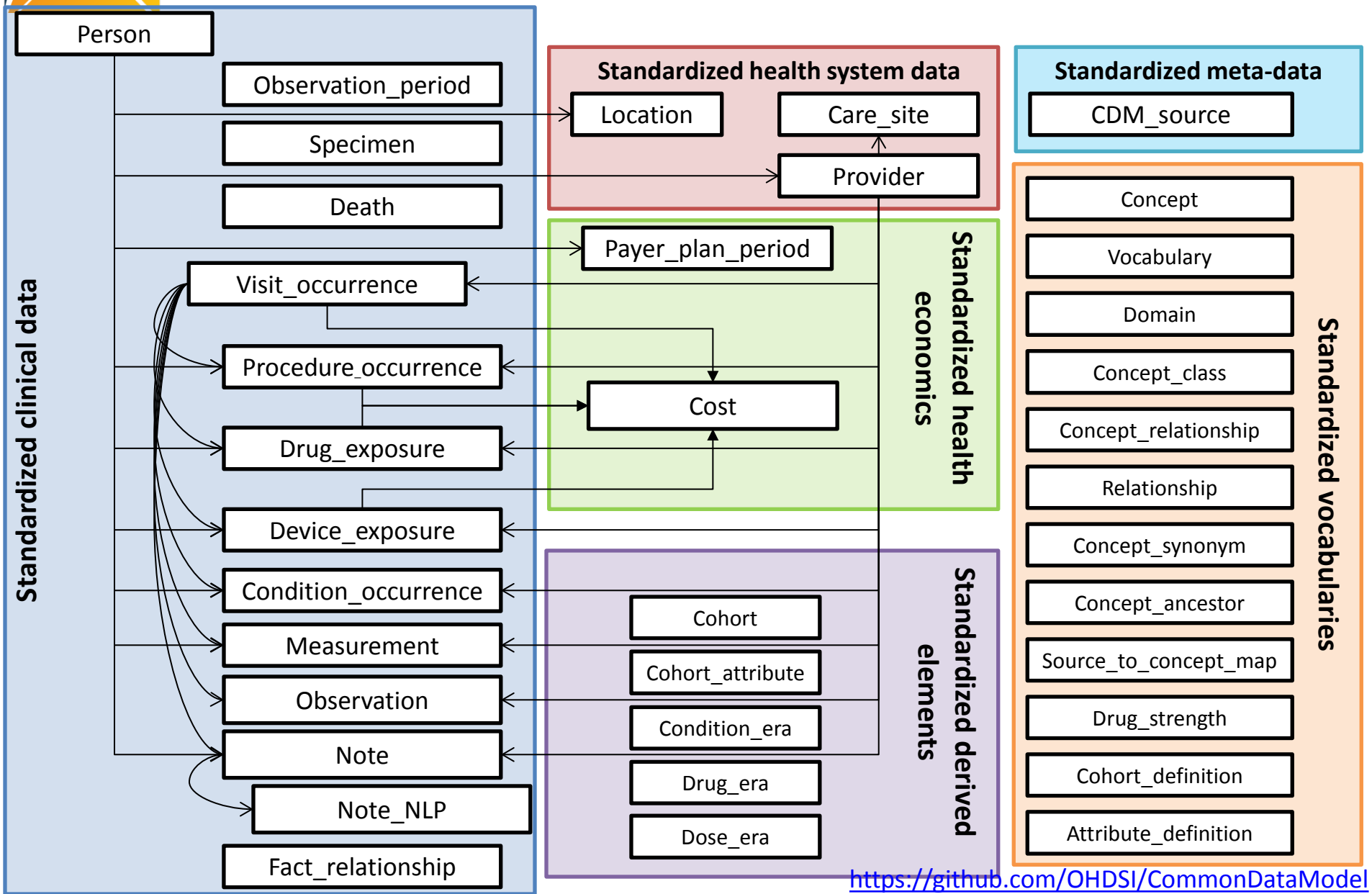
- OMOP model is an information model
  - Vocabulary (Conceptual) and Data Model are blended
  - Domain-oriented concepts
- Patient centric
- Accommodates data from various sources
- Preserves data provenance
- Extendable
- Evolving

Journey of an open community data standard



https://github.com/OHDSI/CommonDataModel

OMOP Common Data Model v5.2



https://github.com/OHDSI/CommonDataModel

Everything is a concept….everything needs to be defined in a common language

Cardiovascular, Bleeding, and Mortality Risks in Elderly Medicare Patients Treated With Dabigatran or Warfarin for Nonvalvular Atrial Fibrillation

David J. Graham, MD, MPH; Marsha E. Reichman, PhD; Michael Wernecke, BA; Rongmei Zhang, PhD; Mary Ross Southworth, PharmD; Mark Levenson, PhD; Ting-Chang Sheu, MPH; Katrina Mott, MHS; Margie R. Goulding, PhD; Monika Houstoun, PharmD, MPH; Thomas E. MaCurdy, PhD; Chris Worrall, BS; Jeffrey A. Kelman, MD, MMSc

Background—The comparative safety of dabigatran versus warfarin for treatment of nonvalvular atrial fibrillation in general practice settings has not been established.

Methods and Results—We formed new-user cohorts of propensity score–matched elderly patients enrolled in Medicare who initiated dabigatran or warfarin for treatment of nonvalvular atrial fibrillation between October 2010 and December 2012. Among 134 414 patients with 37 587 person-years of follow-up, there were 2715 primary outcome events. The hazard ratios (95% confidence intervals) comparing dabigatran with warfarin (reference) were as follows: ischemic stroke, 0.80 (0.67–0.96); intracranial hemorrhage, 0.34 (0.26–0.46); major gastrointestinal bleeding, 1.28 (1.14–1.44); acute myocardial infarction, 0.92 (0.78–1.08); and death, 0.86 (0.77–0.96). In the subgroup treated with dabigatran 75 mg twice daily, there was no difference in risk compared with warfarin for any outcome except intracranial hemorrhage, in which case dabigatran risk was reduced. Most patients treated with dabigatran 75 mg twice daily appeared not to have severe renal impairment, the intended population for this dose. In the dabigatran 150-mg twice daily subgroup, the magnitude of effect for each outcome was greater than in the combined-dose analysis.

Conclusions—In general practice settings, dabigatran was associated with reduced risk of ischemic stroke, intracranial hemorrhage, and death and increased risk of major gastrointestinal hemorrhage compared with warfarin in elderly patients with nonvalvular atrial fibrillation. These associations were most pronounced in patients treated with dabigatran 150 mg twice daily, whereas the association of 75 mg twice daily with study outcomes was indistinguishable from warfarin except for a lower risk of intracranial hemorrhage with dabigatran. **(*Circulation*. 2015;131:157-164. DOI: 10.1161/CIRCULATIONAHA.114.012061.)**

Key Words: anticoagulant ■ pharmacoepidemiology ■ safety ■ thrombin inhibitor ■ warfarin

OMOP Common Vocabulary Model

What it is

- Standardized structure to house existing vocabularies used in the public domain
- Compiled standards from disparate public and private sources and some OMOP-grown concepts
- Built on the shoulders of National Library of Medicine's Unified Medical Language System (UMLS)

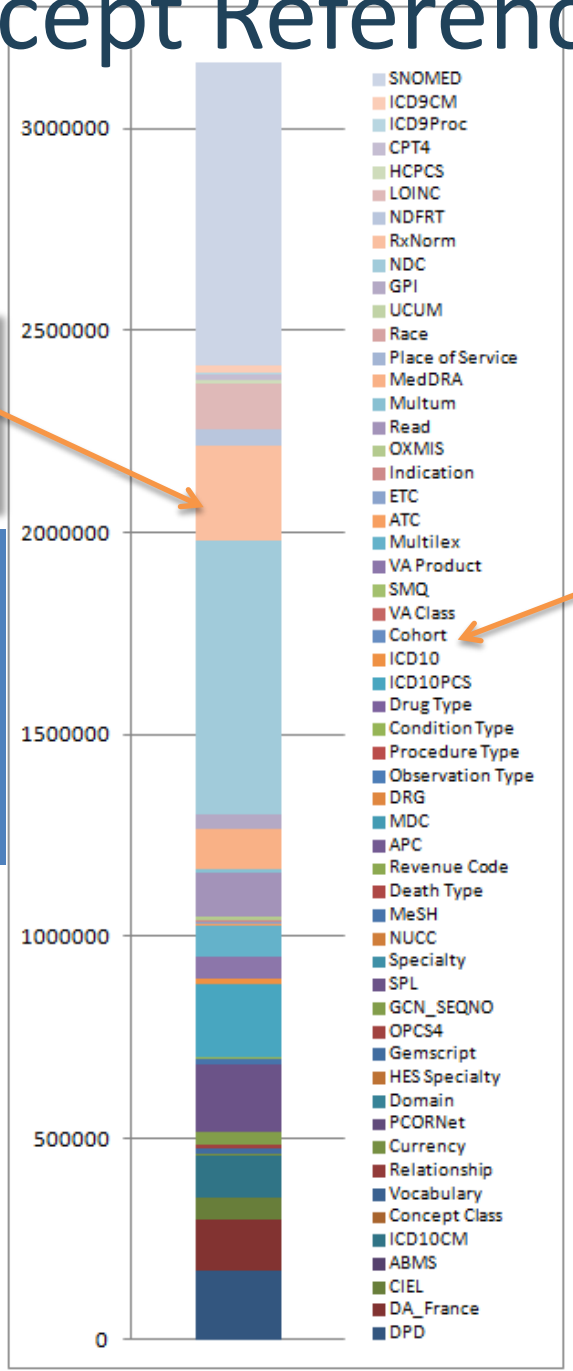
What it's not

- Static dataset – the vocabulary updates regularly to keep up with the continual evolution of the sources
- Finished product – vocabulary maintenance and improvement is ongoing activity that requires community participation and support

Single Concept Reference Table

All vocabularies stacked up in one table

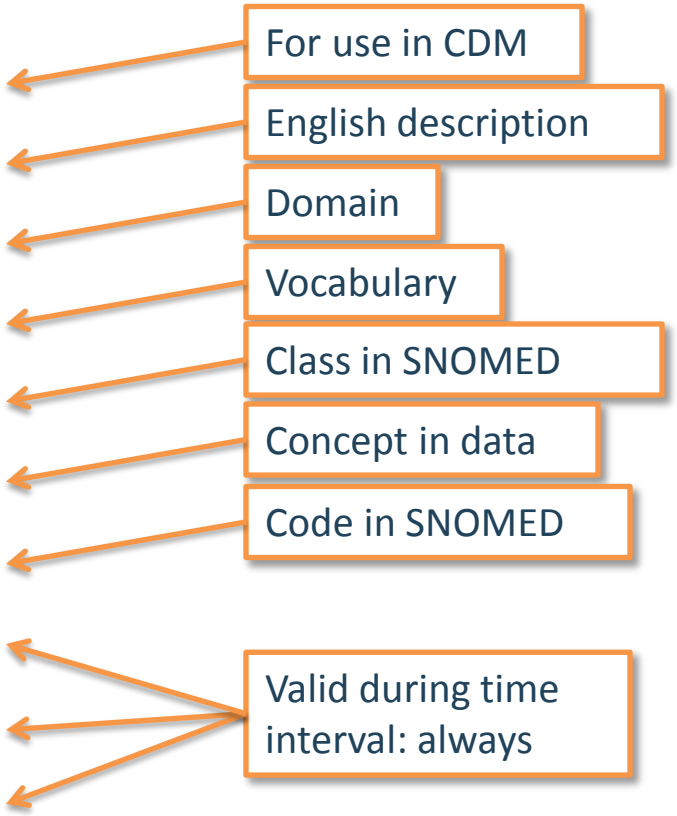
- 78 Vocabularies across 32 domains
- 5,720,848 concepts
  - 2,361,965 standard concepts
  - 3,022,623 source codes
  - 336,260 classification concepts
- 32,612,650 concept relationships



Vocabulary ID

What's in a Concept

CONCEPT_ID	313217
CONCEPT_NAME	Atrial fibrillation
DOMAIN_ID	Condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	49436004
VALID_START_DATE	01-Jan-1970
VALID_END_DATE	31-Dec-2099
INVALID_REASON	

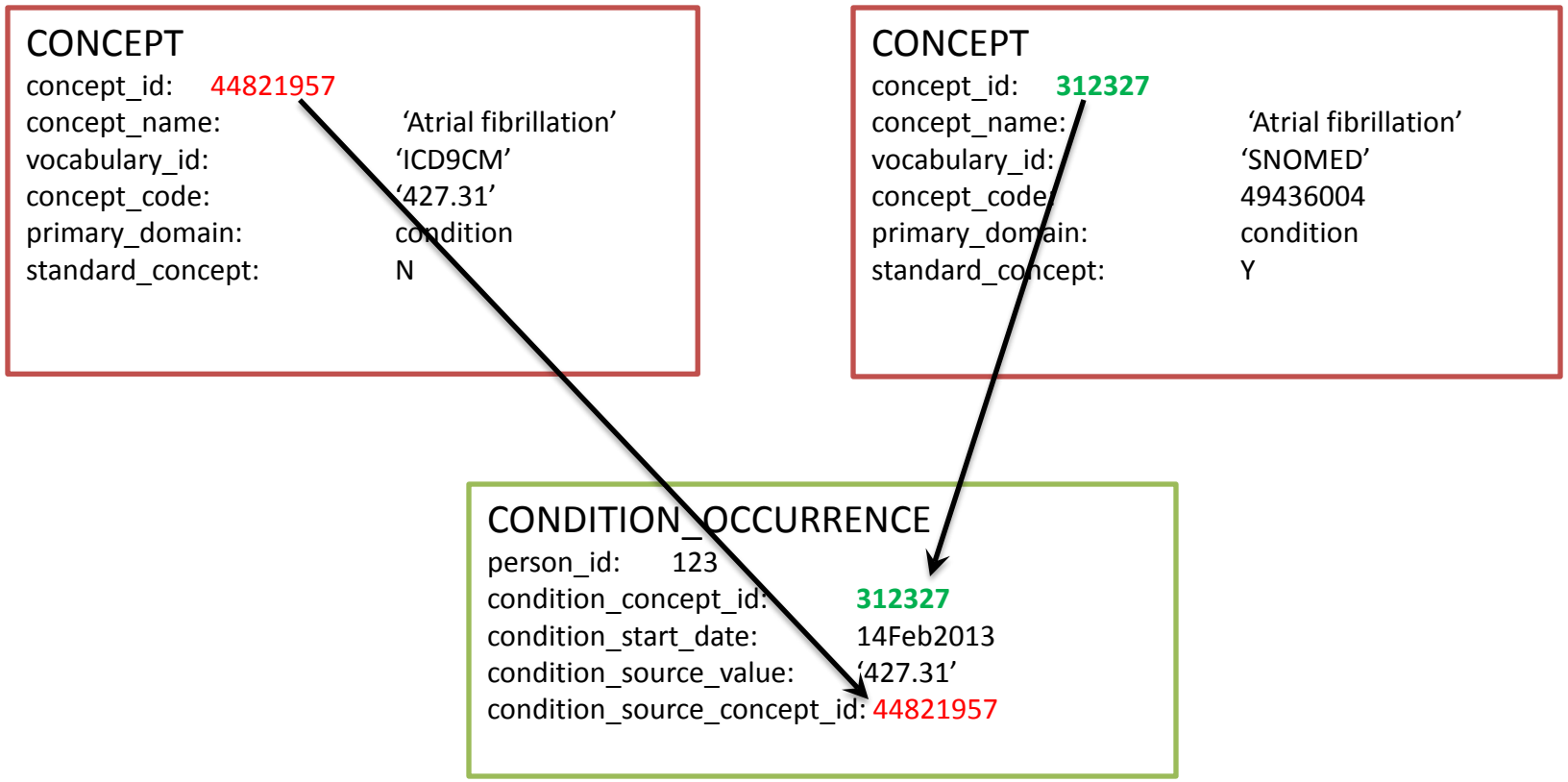


OMOP CDM Standard Domain Features

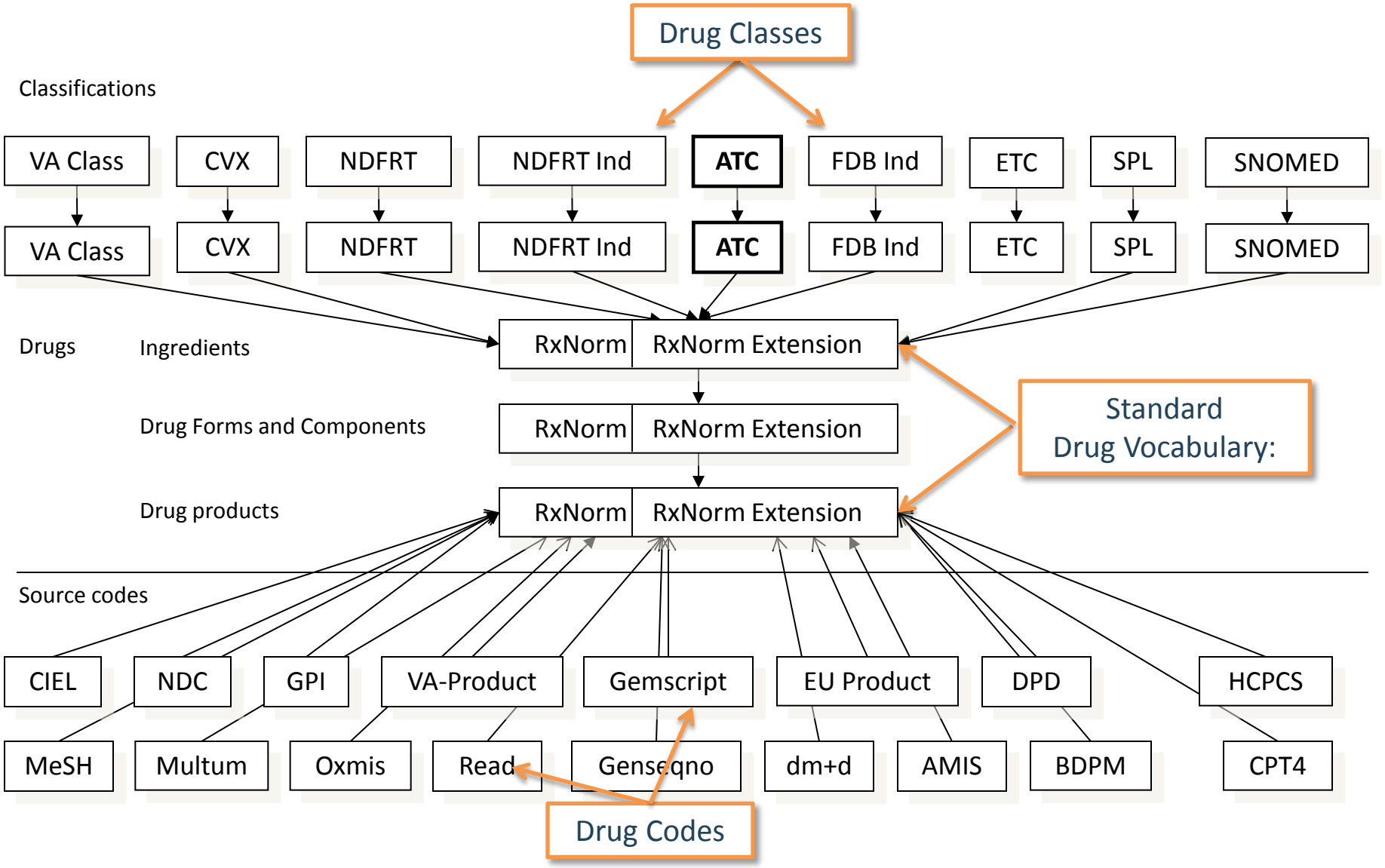
Feature	Description and purpose	Field name convention	Example
Patient centric	Every domain table has **patient identifier**. Patient data can be retrieved independently from other domains.	**person_id**	person_id 123
Unique domain identifier	Every domain table has a unique primary key to identify domain **entities**	<entity>**_id**	condition_occurrence_id 470985
Standard concept from a respective vocabulary domain	Integration with the vocabulary. Foreign key into the Standard Vocabulary for **Standard Concept**	<entity>**_concept_id**	condition_concept_id 313217 (SNOMED "Atrial Fibrillation")
Source concept from a respective vocabulary domain	Provenance. Foreign key into the Standard Vocabulary for **Source Concept**	<entity>**_source_concept_id**	condition_source_concept_id 44821957 (ICD9CM "Atrial Fibrillation")
Source value	Provenance. Verbatim information from the source data, **not to be used** by any standard analytics	<entity>**_source_value**	condition_source_value 427.31 (ICD9CM "Atrial Fibrillation")
Source type	Provenance. Foreign key into the Vocabulary for the **origin of the**	<entity>**_type_concept_id**	condition_type_concept_id 38000199 ("Inpatient header – primary")

OMOP-CDM retains source data as verbatim and as concept code referring to source vocabulary (e.g. ICD-9CM)

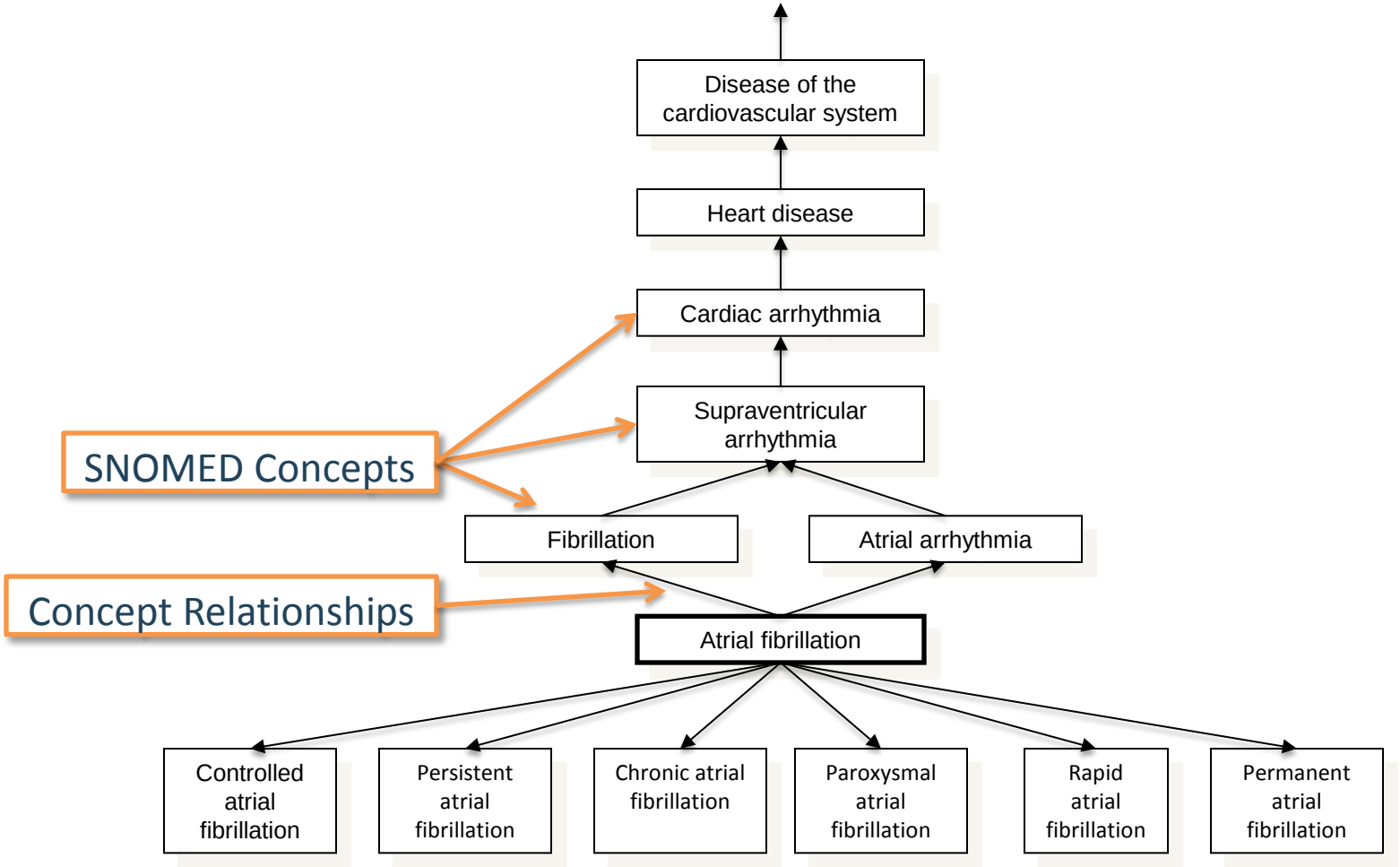
Integration of CDM and Vocabulary



Drug Hierarchy

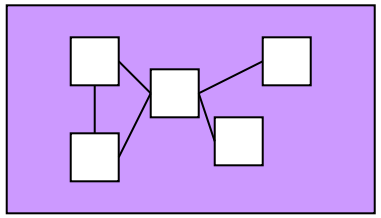


Disease Hierarchy



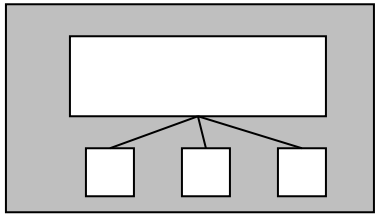
Common data model to enable standardized analytics

Source 1 raw data



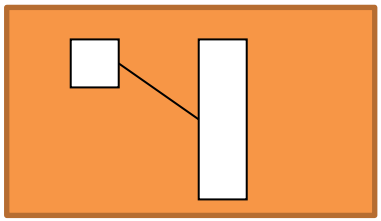
Electronic health records

Source 2 raw data



Administrative claims

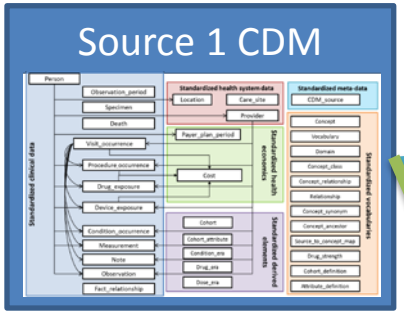
Source 3 raw data



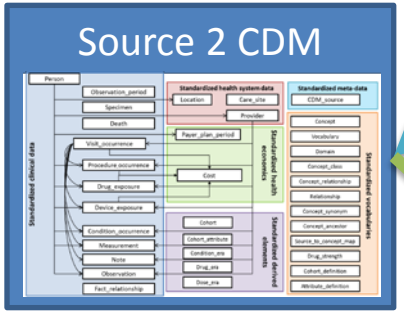
Clinical data

Transformation to OMOP common data model

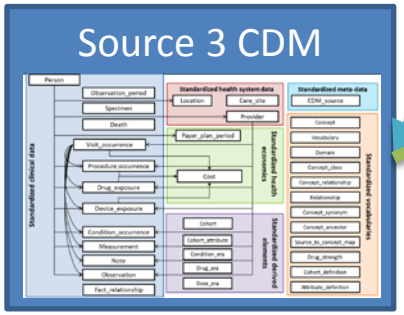
Source 1 CDM



Source 2 CDM



Source 3 CDM



Open-source analysis code

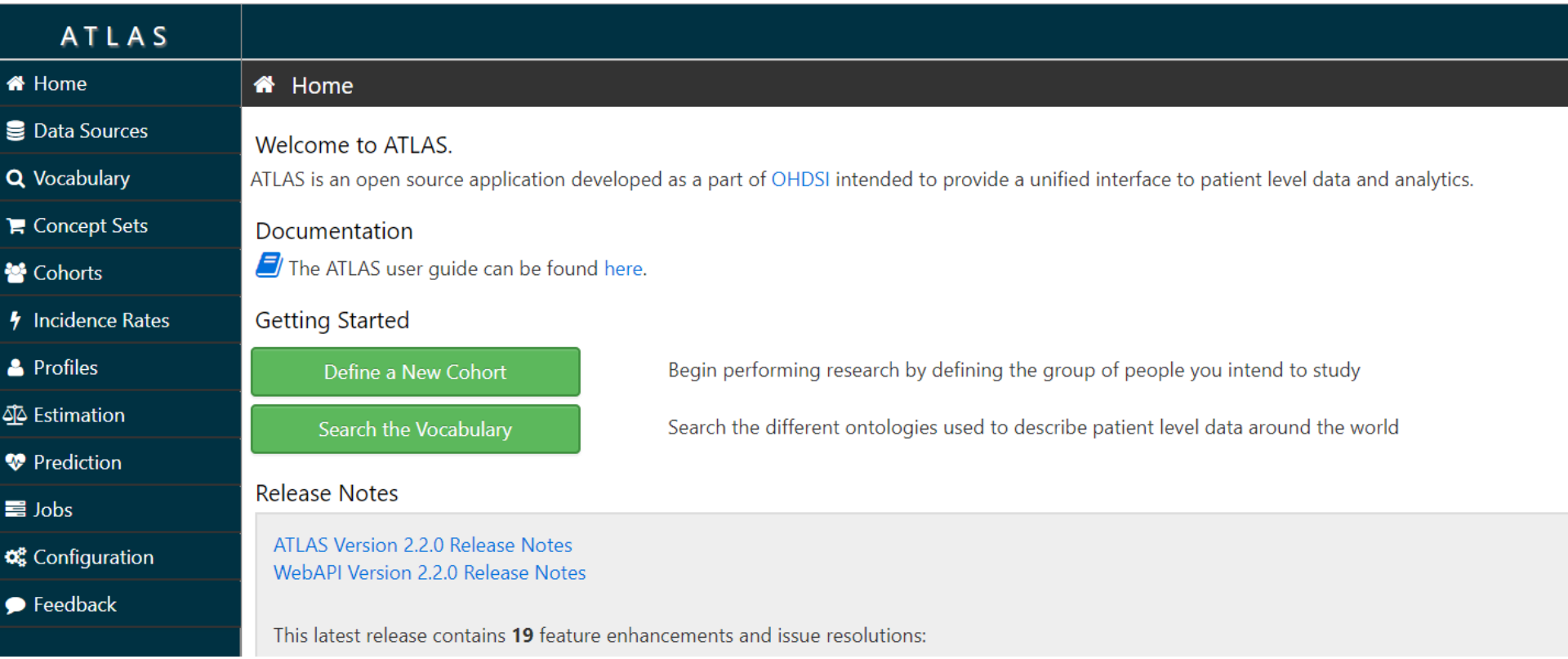
Open evidence

Evidence OHDSI seeks to generate from observational data

- **Clinical characterization**
  - Natural history: Who has diabetes, and who takes metformin?
  - Quality improvement: What proportion of patients with diabetes experience complications?
- **Population-level effect estimation**
  - Safety surveillance: Does metformin cause lactic acidosis?
  - Comparative effectiveness: Does metformin cause lactic acidosis more than glyburide?
- **Patient-level prediction**
  - Precision medicine: Given everything you know about me, now I started using metformin, what is the chance I will get lactic acidosis?
  - Disease interception: Given everything you know about me, what is the chance I will develop diabetes?

- Very active method development workgroups
- Open-source code: www.github.com/OHDSI
- Many network studies initiated

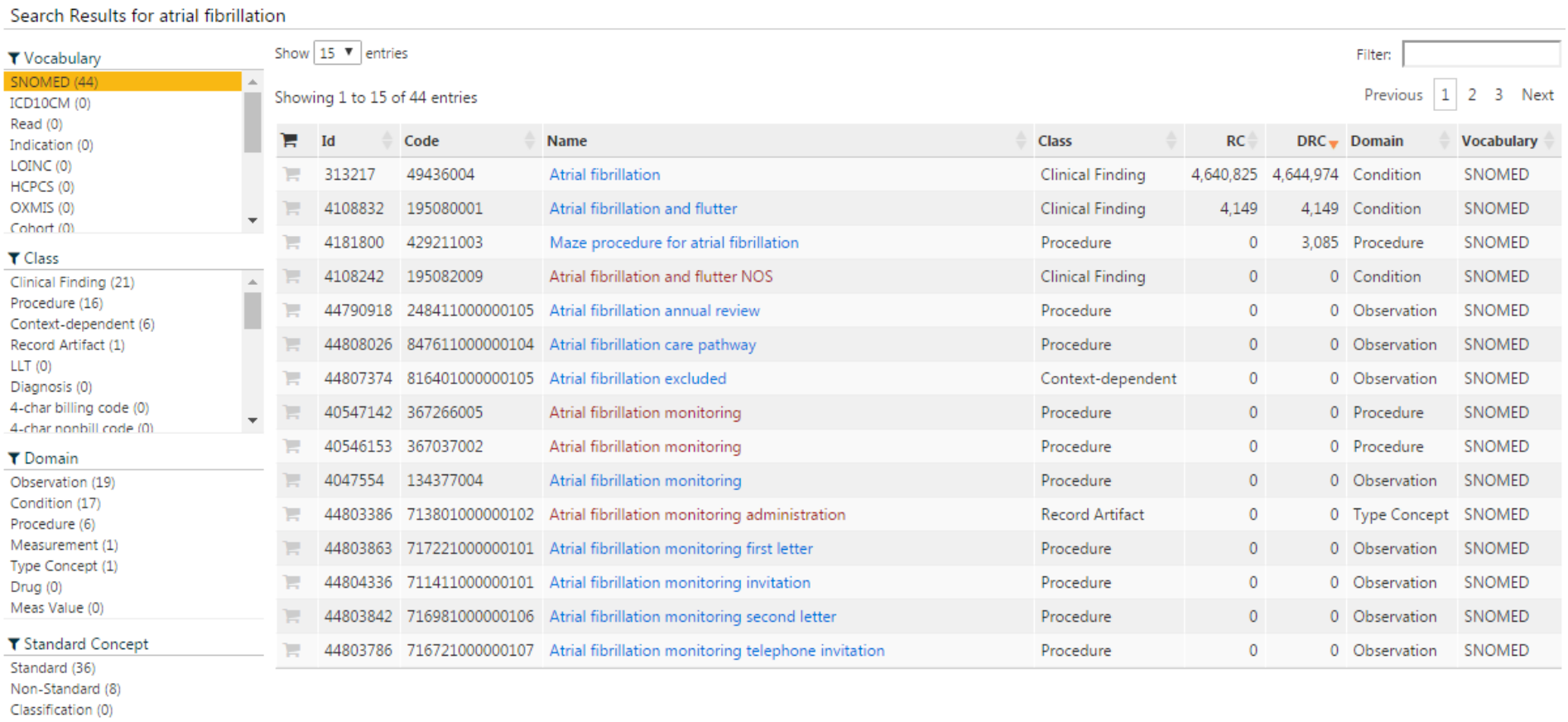
ATLAS is a free, publicly available, web based, open source software tool for researchers to conduct scientific analyses on standardized observational data.



http://ohdsi.org/web/ATLAS

ATLAS enables vocabulary browsing

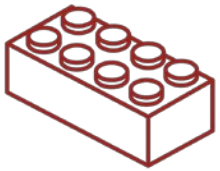
- Browsing of all vocabularies (including source vocabularies)
- Insight in concept relationships
- Transparent and reproducible Concept Set creation



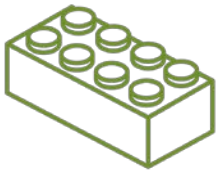
http://ohdsi.org/web/ATLAS

ATLAS enables complex phenotyping

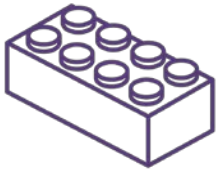
Conditions



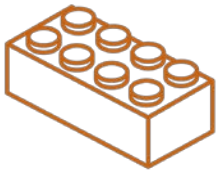
Drugs



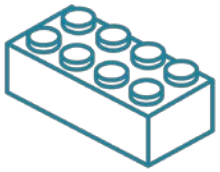
Procedures



Measurements



Observations



Visits

Diabetes Definition

A condition occurrence of **diabetes**

With drug exposure of **oral DM meds** within 90 days after index

With measurement **HbA1c** > 7.0 within 90 days before and after index

- Complex Cohort building using Standardized Vocabularies (including use of source concepts!)
- Archiving and sharing of cohort definitions in a data network
- Execution against the CDM including attrition overview
- and much more..

http://ohdsi.org/web/ATLAS

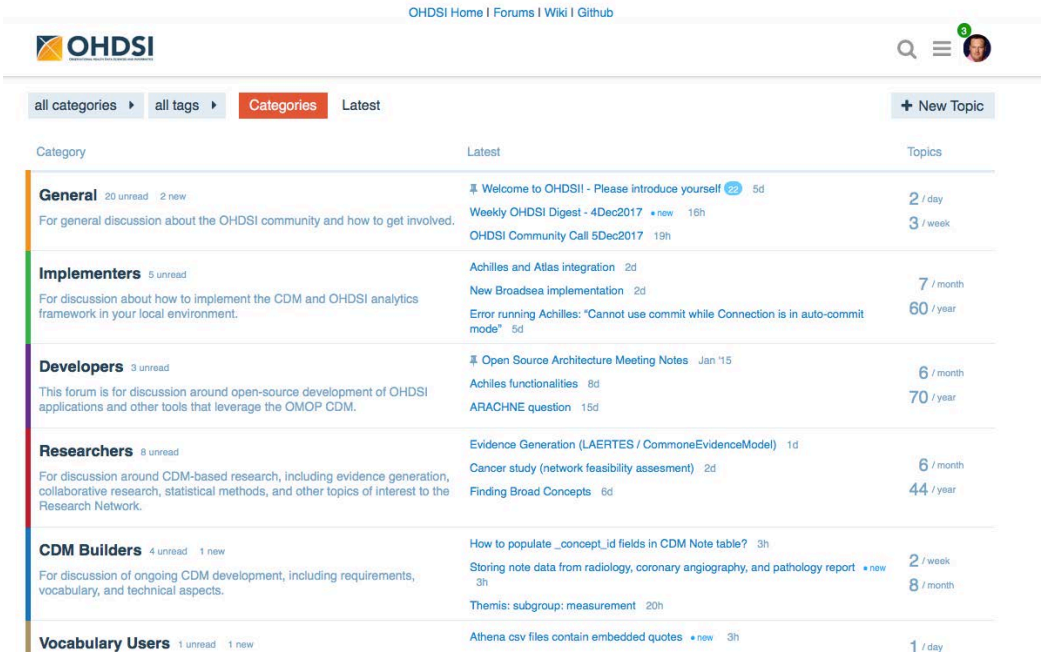
Growing European Data Network


First Annual
EUROPEAN OHDSI SYMPOSIUM
March 23th 2018
Tutorials March 24th
The European OHDSI Initiative will build a community in Europe that will improve the operability of healthcare data by standardizing to the OMOP-CDM and will support the further development of the anyalical tools to enable transparent and reproducible research.
Bridging Europe
Erasmus MC Rotterdam The Netherlands
www.ohdsi-europe.org

Questions?

OHDSI Forums:

http://forums.ohdsi.org



https://github.com/OHDSI/CommonDataModel/wiki/Frequently-Asked-Questions