



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	icd_type
-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 Commercial Claims and Encounters (CCAE): INPATIENT_SERVICES

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

Optum Extended SES: MEDICAL_CLAIMS

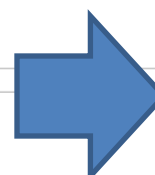
patid	fst_dt	diag1	diag2	diag3	diag4
259000474406532	5/30/2000	41071	27800	4019	2724

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_V ALUE	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_V ALUE	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_V ALUE	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_V ALUE	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



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

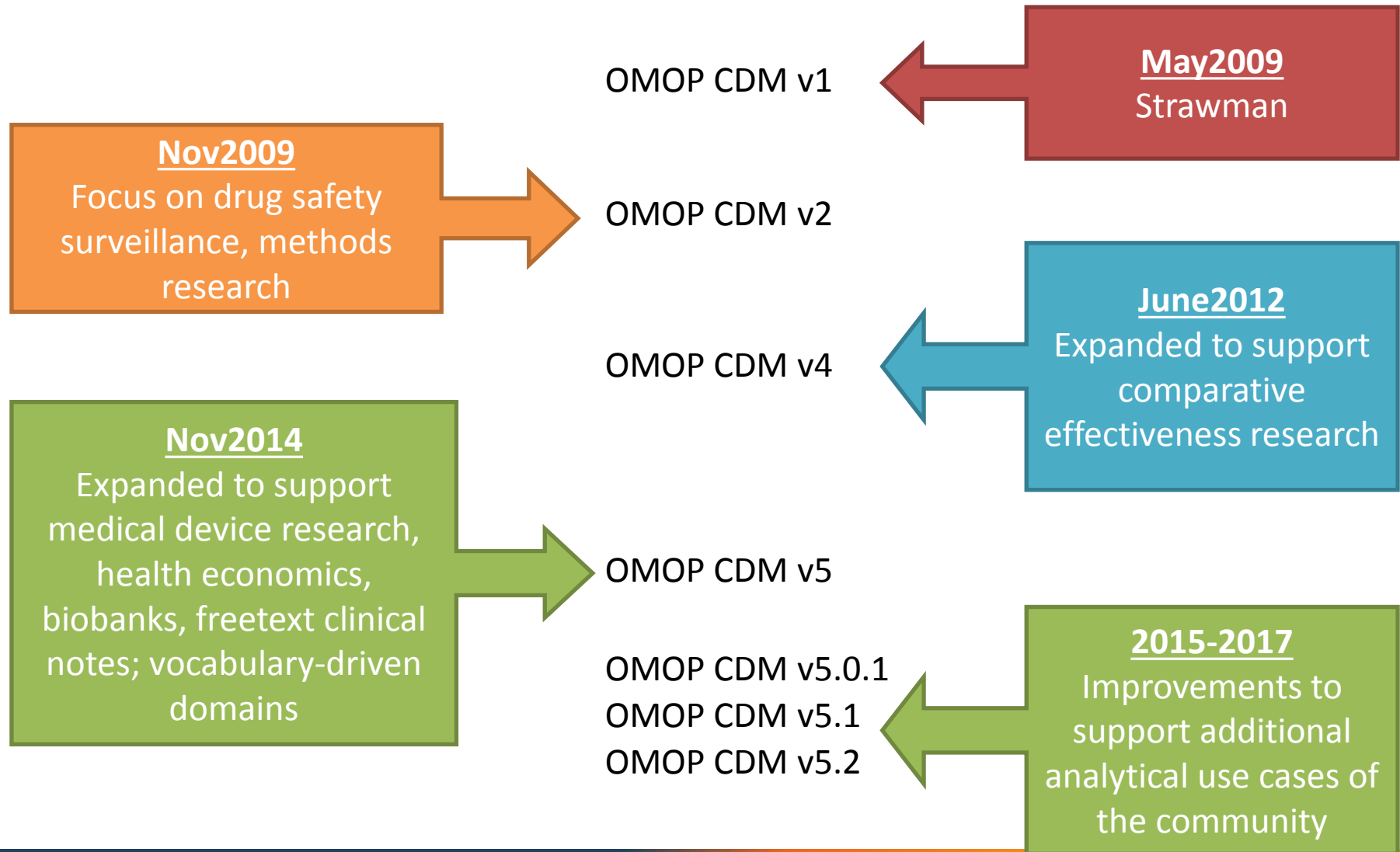


OMOP CDM Principles

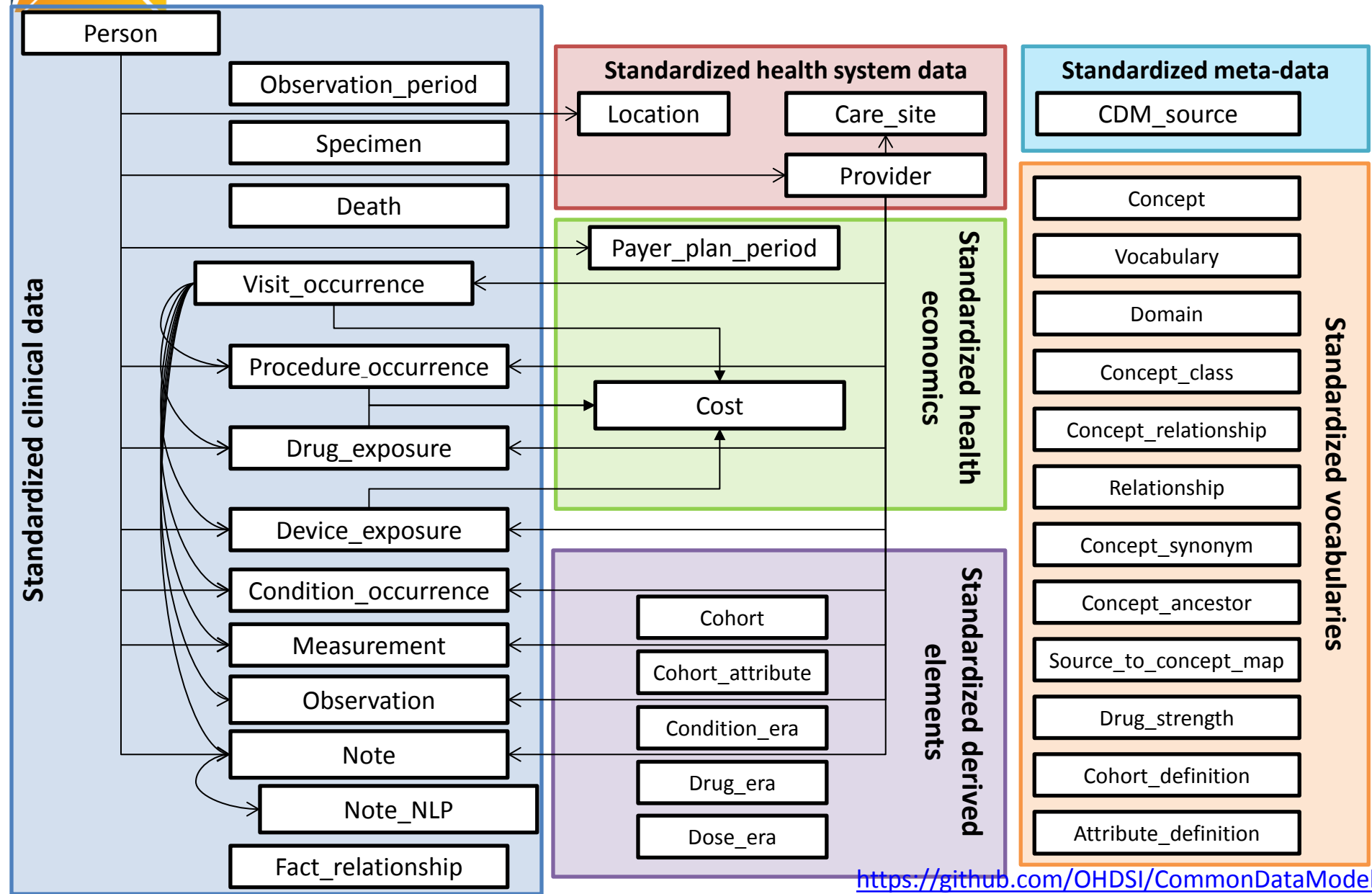
- OMOP model is an information model
 - Vocabulary (Conceptual) and Data Model are blended
 - Domain-oriented concepts
 - Patient centric
 - Accommodates data from various sources
 - Preserves data provenance
 - Extendable
 - Evolving
-




Journey of an open community data standard



OMOP Common Data Model v5.2





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 134414 patients with 37587 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

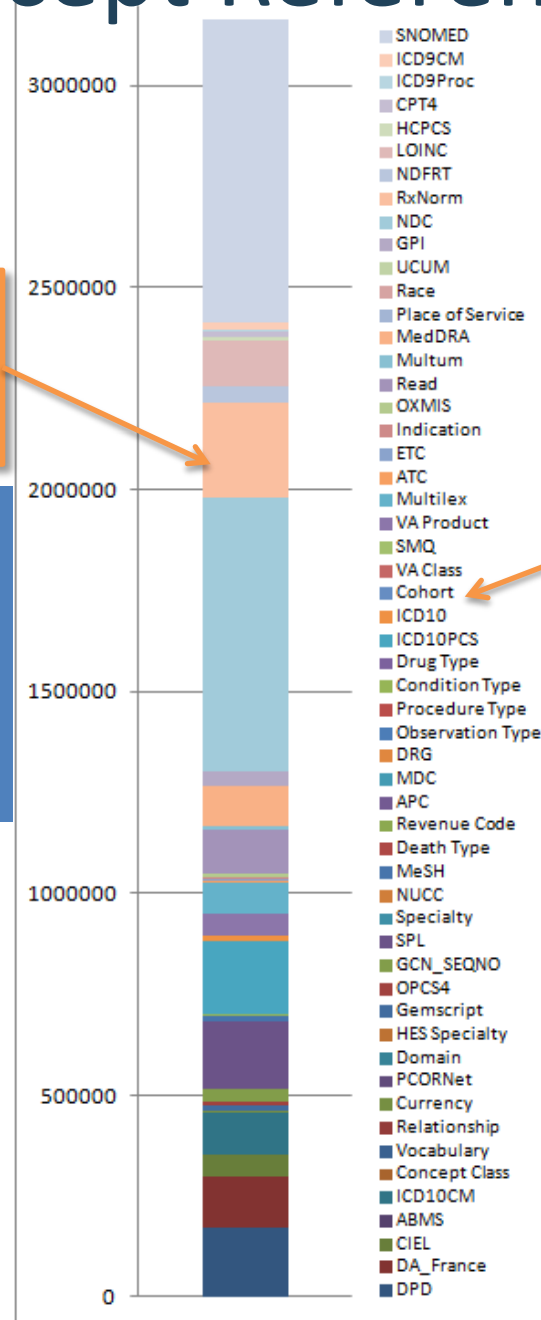
- Static dataset – the vocabulary updates regularly to keep up with the continual evolution of the sources
- Finished product – vocabulary maintenance and improvement is ongoing activity that requires community participation and support



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

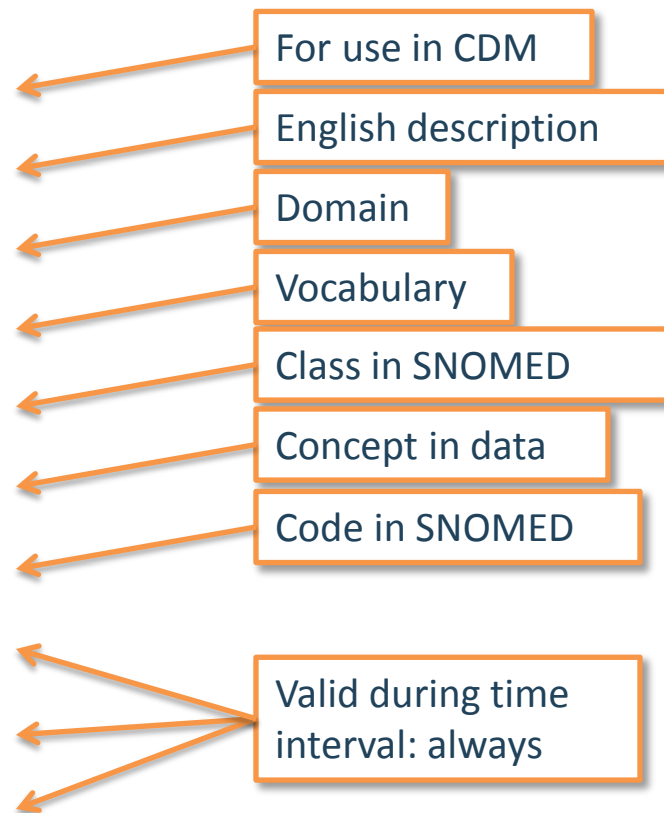


Vocabulary ID



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

CONCEPT

concept_id: 44821957
concept_name: 'Atrial fibrillation'
vocabulary_id: 'ICD9CM'
concept_code: '427.31'
primary_domain: condition
standard_concept: N

CONCEPT

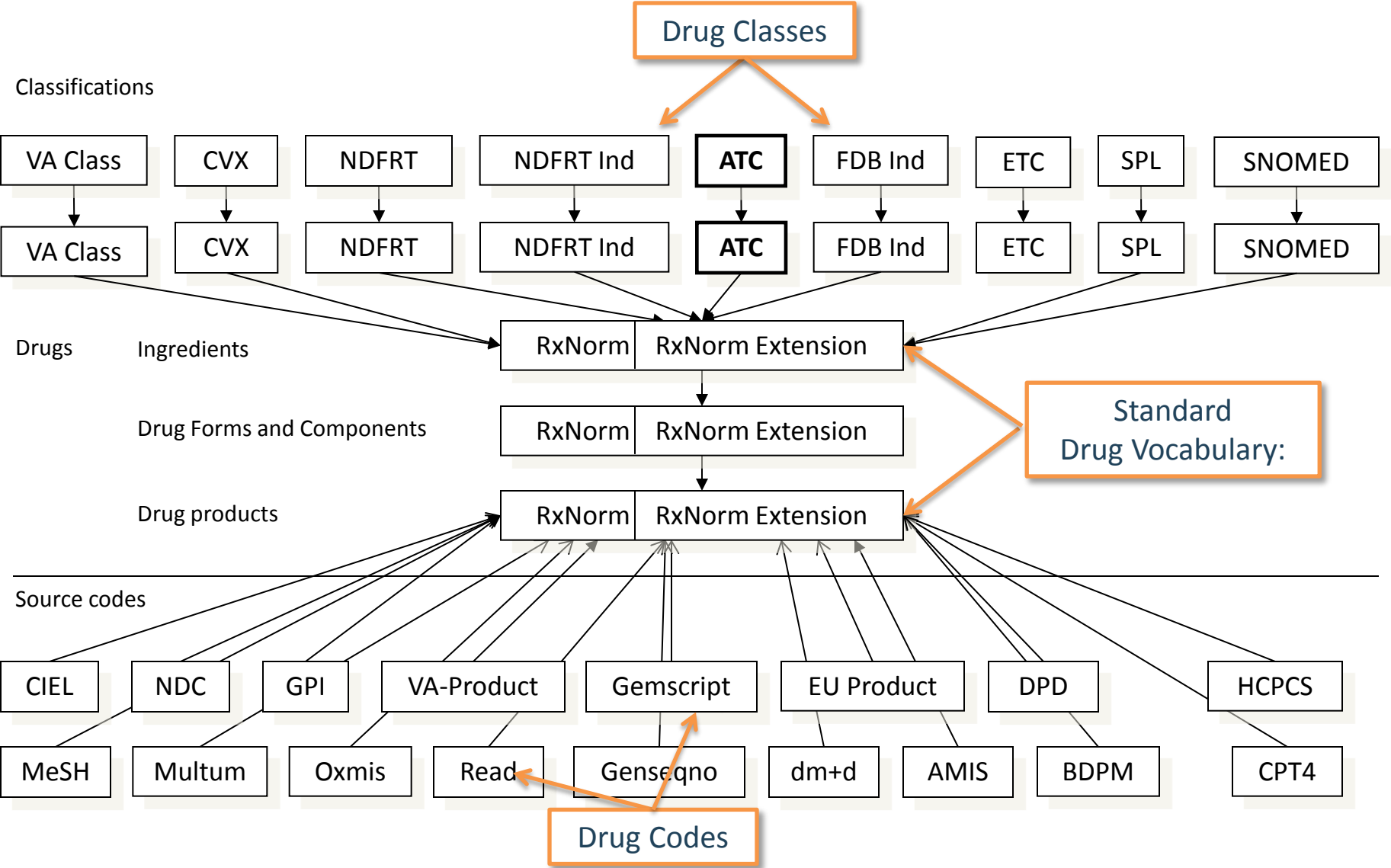
concept_id: 312327
concept_name: 'Atrial fibrillation'
vocabulary_id: 'SNOMED'
concept_code: 49436004
primary_domain: condition
standard_concept: Y

CONDITION_OCCURRENCE

person_id: 123
condition_concept_id: 312327
condition_start_date: 14Feb2013
condition_source_value: '427.31'
condition_source_concept_id: 44821957

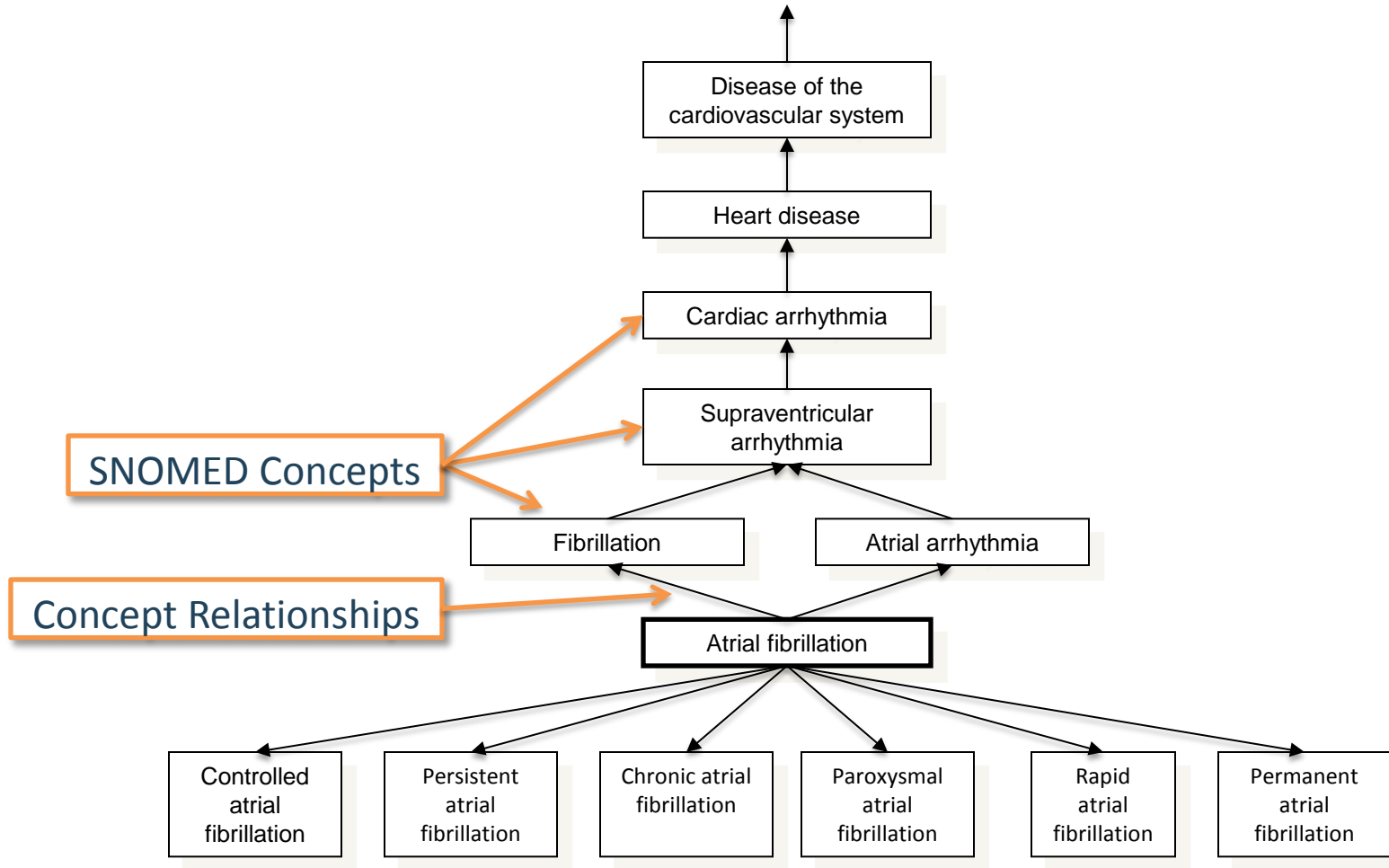


Drug Hierarchy





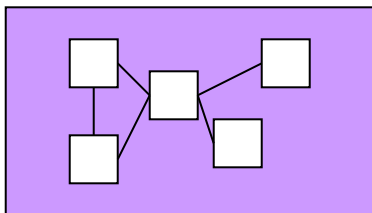
Disease Hierarchy





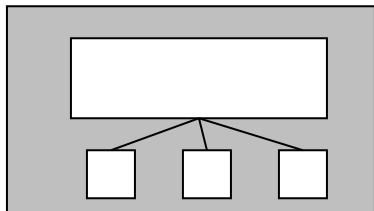
Common data model to enable standardized analytics

Source 1 raw data



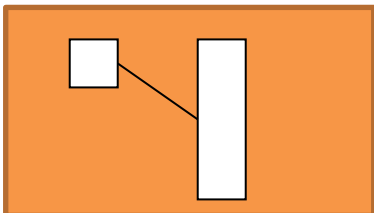
Electronic health records

Source 2 raw data

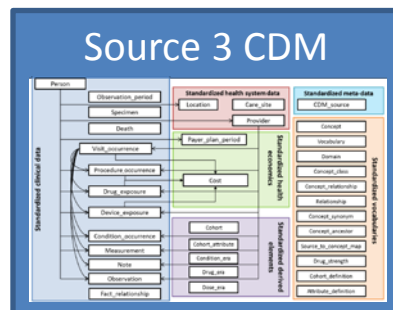
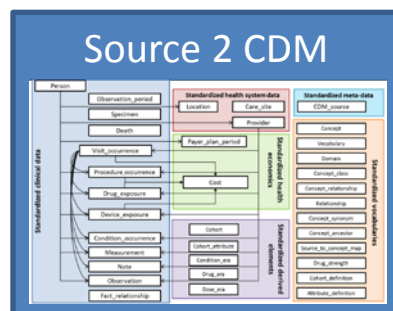
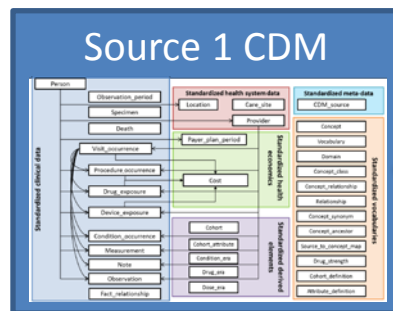
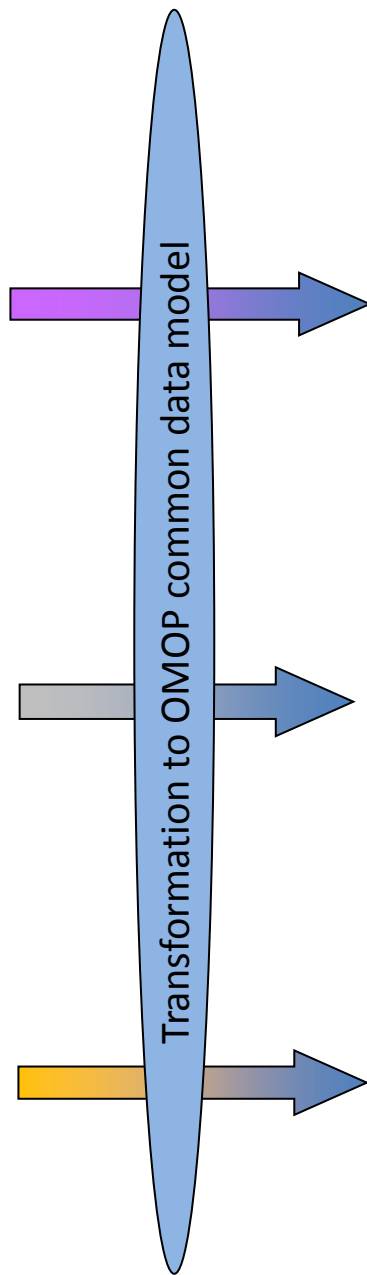


Administrative claims

Source 3 raw data



Clinical data





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.**

ATLAS	
Home	Home
Data Sources	Welcome to ATLAS.
Vocabulary	ATLAS is an open source application developed as a part of OHDSI intended to provide a unified interface to patient level data and analytics.
Concept Sets	Documentation
Cohorts	The ATLAS user guide can be found here .
Incidence Rates	Getting Started
Profiles	<div>Define a New Cohort</div> Begin performing research by defining the group of people you intend to study
Estimation	<div>Search the Vocabulary</div> Search the different ontologies used to describe patient level data around the world
Prediction	
Jobs	Release Notes
Configuration	ATLAS Version 2.2.0 Release Notes WebAPI Version 2.2.0 Release Notes
Feedback	This latest release contains 19 feature enhancements and issue resolutions:

<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

Search Results for atrial fibrillation

Vocabulary

SNOMED (44)

ICD10CM (0)

Read (0)

Indication (0)

LOINC (0)

HCPCS (0)

OXMIS (0)

Cohort (0)

Class

Clinical Finding (21)

Procedure (16)

Context-dependent (6)

Record Artifact (1)

LLT (0)

Diagnosis (0)

4-char billing code (0)

4-char nonbill code (0)

Domain

Observation (19)

Condition (17)

Procedure (6)

Measurement (1)

Type Concept (1)

Drug (0)

Meas Value (0)

Standard Concept

Standard (36)

Non-Standard (8)

Classification (0)

Showing 1 to 15 of 44 entries

Filter:

Previous 1 2 3 Next

Id	Code	Name	Class	RC	DRC	Domain	Vocabulary
313217	49436004	Atrial fibrillation	Clinical Finding	4,640,825	4,644,974	Condition	SNOMED
4108832	195080001	Atrial fibrillation and flutter	Clinical Finding	4,149	4,149	Condition	SNOMED
4181800	429211003	Maze procedure for atrial fibrillation	Procedure	0	3,085	Procedure	SNOMED
4108242	195082009	Atrial fibrillation and flutter NOS	Clinical Finding	0	0	Condition	SNOMED
44790918	248411000000105	Atrial fibrillation annual review	Procedure	0	0	Observation	SNOMED
44808026	847611000000104	Atrial fibrillation care pathway	Procedure	0	0	Observation	SNOMED
44807374	816401000000105	Atrial fibrillation excluded	Context-dependent	0	0	Observation	SNOMED
40547142	367266005	Atrial fibrillation monitoring	Procedure	0	0	Procedure	SNOMED
40546153	367037002	Atrial fibrillation monitoring	Procedure	0	0	Procedure	SNOMED
4047554	134377004	Atrial fibrillation monitoring	Procedure	0	0	Observation	SNOMED
44803386	713801000000102	Atrial fibrillation monitoring administration	Record Artifact	0	0	Type Concept	SNOMED
44803863	717221000000101	Atrial fibrillation monitoring first letter	Procedure	0	0	Observation	SNOMED
44804336	711411000000101	Atrial fibrillation monitoring invitation	Procedure	0	0	Observation	SNOMED
44803842	716981000000106	Atrial fibrillation monitoring second letter	Procedure	0	0	Observation	SNOMED
44803786	716721000000107	Atrial fibrillation monitoring telephone invitation	Procedure	0	0	Observation	SNOMED



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..



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 analytical 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>



OHDSI Home | Forums | Wiki | Github

OHDSI

all categories ▾ all tags ▾ **Categories** Latest

+ New Topic

Category	Latest	Topics
General 20 unread 2 new For general discussion about the OHDSI community and how to get involved.	🚩 Welcome to OHDSI! - Please introduce yourself 5d Weekly OHDSI Digest - 4Dec2017 • new 16h OHDSI Community Call 5Dec2017 19h	2 / day 3 / week
Implementers 5 unread For discussion about how to implement the CDM and OHDSI analytics framework in your local environment.	Achilles and Atlas integration 2d New Broadsea implementation 2d Error running Achilles: "Cannot use commit while Connection is in auto-commit mode" 5d	7 / month 60 / year
Developers 3 unread This forum is for discussion around open-source development of OHDSI applications and other tools that leverage the OMOP CDM.	🚩 Open Source Architecture Meeting Notes Jan '15 Achilles functionalities 8d ARACHNE question 15d	6 / month 70 / year
Researchers 8 unread For discussion around CDM-based research, including evidence generation, collaborative research, statistical methods, and other topics of interest to the Research Network.	Evidence Generation (LAERTES / CommonEvidenceModel) 1d Cancer study (network feasibility assessment) 2d Finding Broad Concepts 6d	6 / month 44 / year
CDM Builders 4 unread 1 new For discussion of ongoing CDM development, including requirements, vocabulary, and technical aspects.	How to populate _concept_id fields in CDM Note table? 3h Storing note data from radiology, coronary angiography, and pathology report • new 3h Themis: subgroup: measurement 20h	2 / week 8 / month
Vocabulary Users 1 unread 1 new	Athena csv files contain embedded quotes • new 3h	1 / day

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