



Introduction to VINCI Observational Medical Outcomes Partnership (OMOP) Project

February 4, 2016
VINCI OMOP and OHDSI Services
Stephen Deppen, PhD

Goals

- What is the OMOP common data model
- OMOP compared to CDW organization
- How to get access to VINCI OMOP tables
- Future directions for VINCI OMOP and common data models

Outline

- Background
 - OMOP-OHDSI international collaborative
- Version 4
 - Overview
 - Standard Vocabularies
 - RxNorm, SnoMed, LOINC
 - Documentation
- Version 5
- Getting Access DART and NDS process
- ODHSI – Achilles, Hermes, Atlas, etc
- Questions

Impetus Research with Common Data Model

- FDA Act of 2007-establish active surveillance systems using electronic health records data
- Medicare HER Incentive Programs
 - Clinicians must demonstrate **meaningful use** of EHR to get payment adjustment.
 - Hospitals must do the same

SEC. 905. ACTIVE POSTMARKET RISK IDENTIFICATION AND ANALYSIS.
(a) IN GENERAL.—Subsection (k) of section 505 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 355) is amended by adding at the end the following:

“(3) ACTIVE POSTMARKET RISK IDENTIFICATION.—
“(A) DEFINITION.—In this paragraph, the term ‘data’ refers to information with respect to a drug approved under this section or under section 351 of the Public Health Service Act, including claims data, patient survey data, standardized analytic files that allow for the pooling and analysis of data from disparate data environments, and any other data deemed appropriate by the Secretary.

“(B) DEVELOPMENT OF POSTMARKET RISK IDENTIFICATION AND ANALYSIS METHODS.—The Secretary shall, not later than 2 years after the date of the enactment of the Food and Drug Administration Amendments Act of 2007, in collaboration with public, academic, and private entities—

“(i) develop methods to obtain access to disparate data sources including the data sources specified in subparagraph (C);

“(ii) develop validated methods for the establishment of a postmarket risk identification and analysis system to link and analyze safety data from multiple sources, with the goals of including, in aggregate—

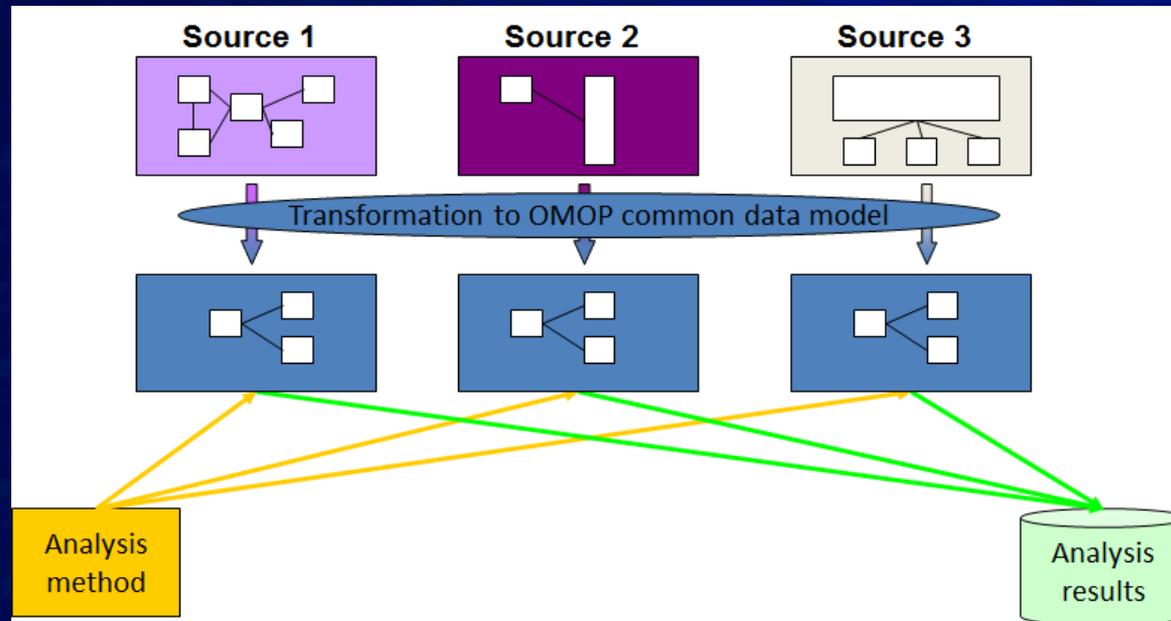
“(I) at least 25,000,000 patients by July 1, 2010; and

“(II) at least 100,000,000 patients by July 1, 2012; and

“(iii) convene a committee of experts, including individuals who are recognized in the field of protecting data privacy and security, to make recommendations to the Secretary on the development of tools and methods for the ethical and scientific uses for, and communication of, postmarketing data specified under subparagraph (C), including recommendations on the development of effective research methods for the study of drug safety questions.

“(C) ESTABLISHMENT OF THE POSTMARKET RISK IDENTIFICATION AND ANALYSIS SYSTEM.—

Common Data Model is ?



- Method for organizing data into a standard structure
- The standardized format allows for the systematic analysis of disparate observational databases

OMOP data model

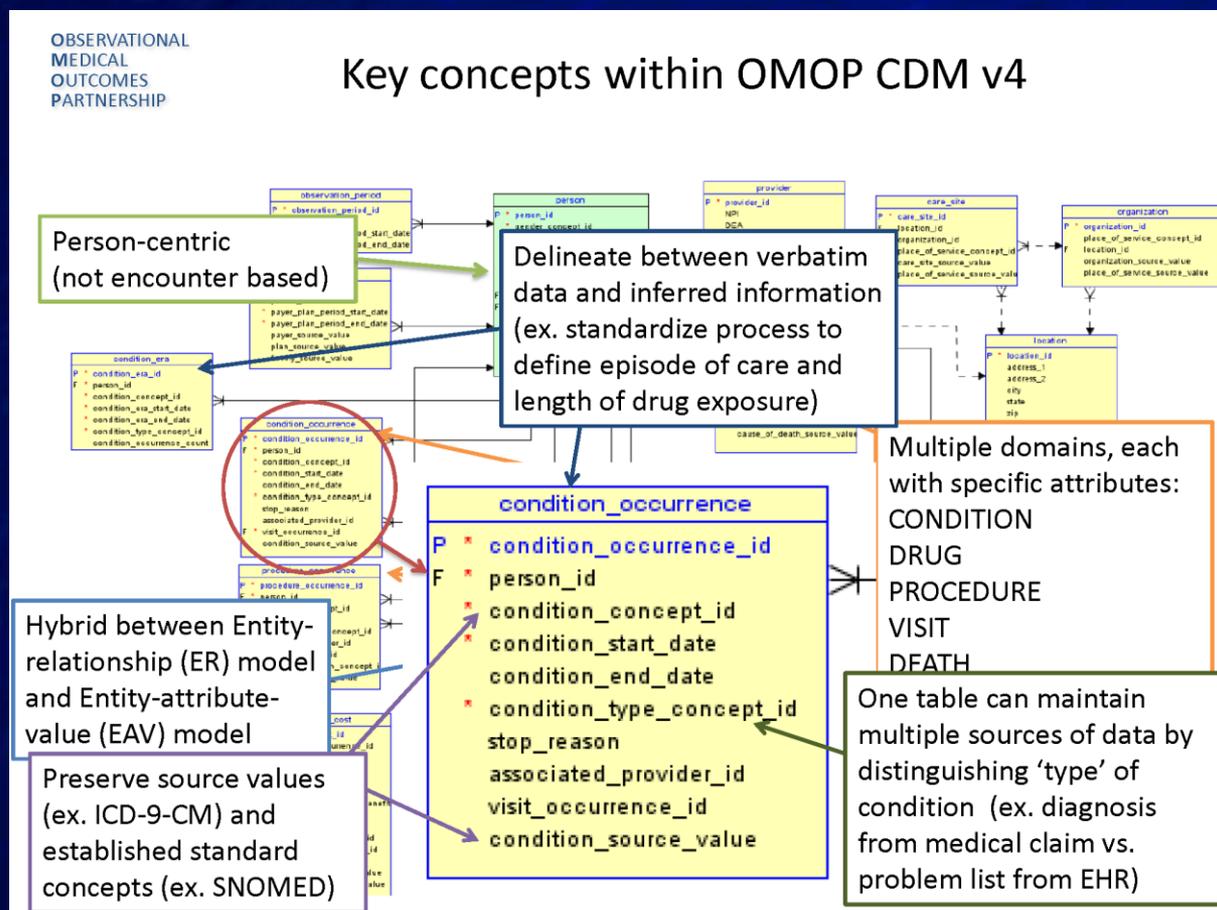
- The Observational Medical Outcomes Partnership (OMOP) CDM, now entering its fifth version.
- Open source github.com/OHDSI
- Robust development of tools for data characterization, quality, medical safety surveillance, cohort creation, comparative effectiveness and patient level predictive modeling.

Why did we choose OMOP?

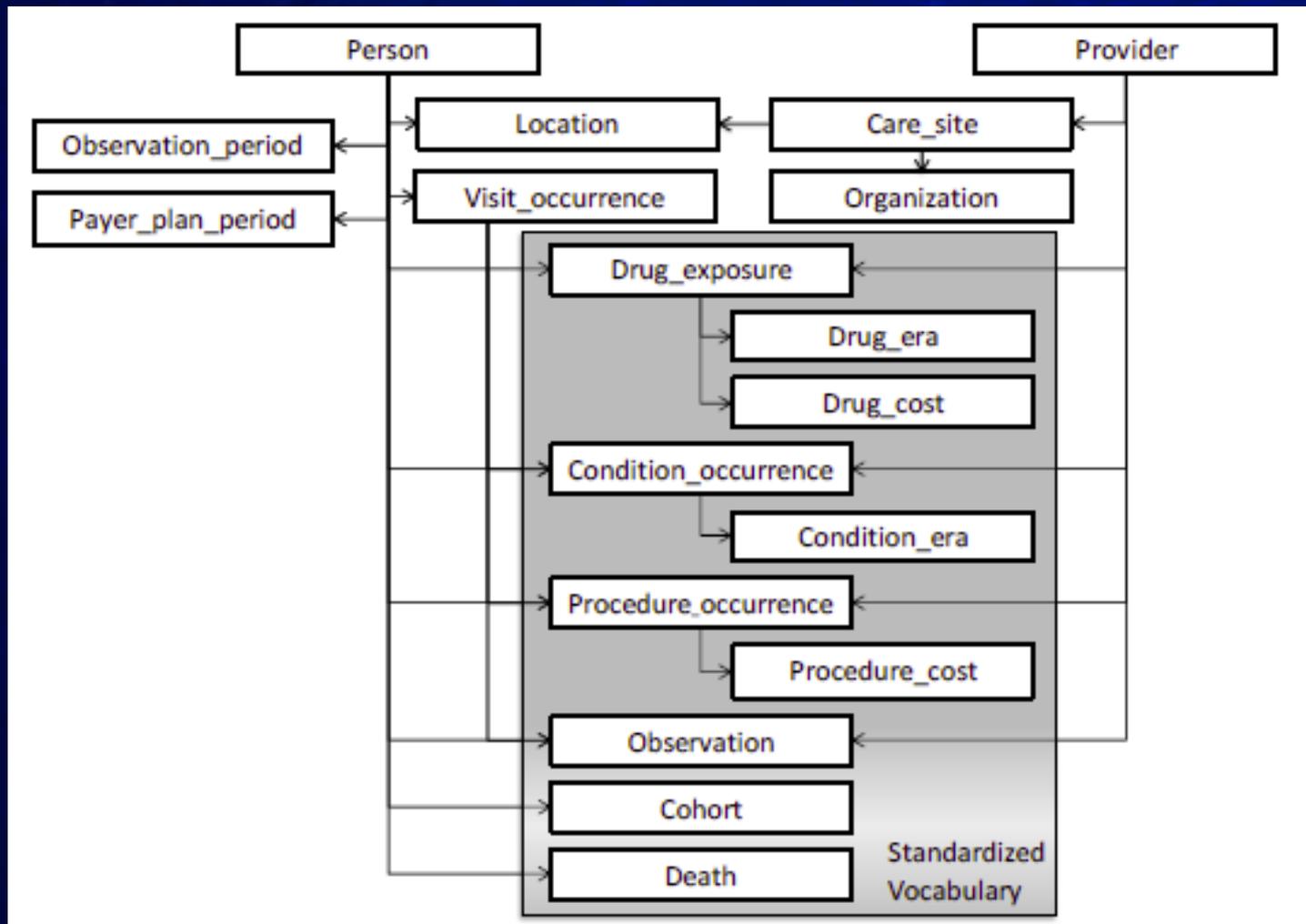
- Met the broadest needs for comparative effectiveness (1)
- Robust open source development teams for user tools to facilitate cohort creation and data visualization
- Applied across multiple databases with the least loss of data fidelity (2)

(1) Oganyemi et al, Med Care 2013; (2) Voss et al, JAMIA 2015

Gestalt of OMOP v4



Overview Version 4 (4.5) OMOP



Vocabularies in Concept Table

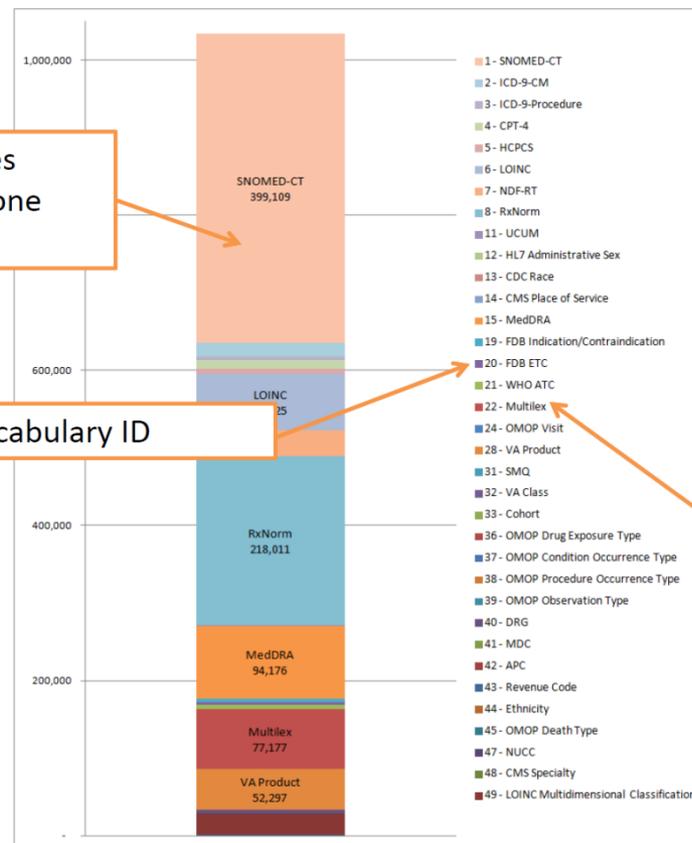
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

The concept Table

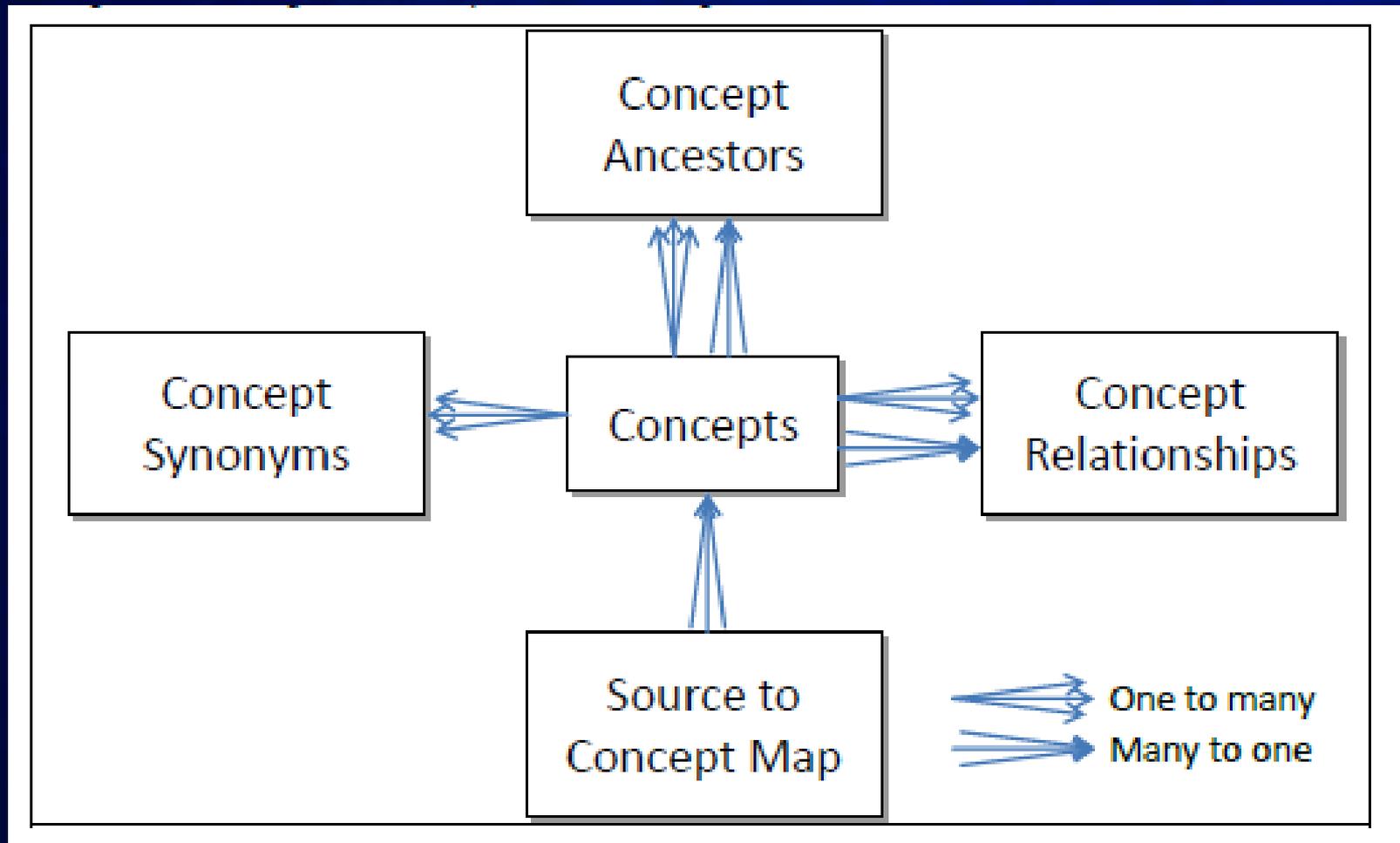
All vocabularies stacked up in one table

Vocabulary ID

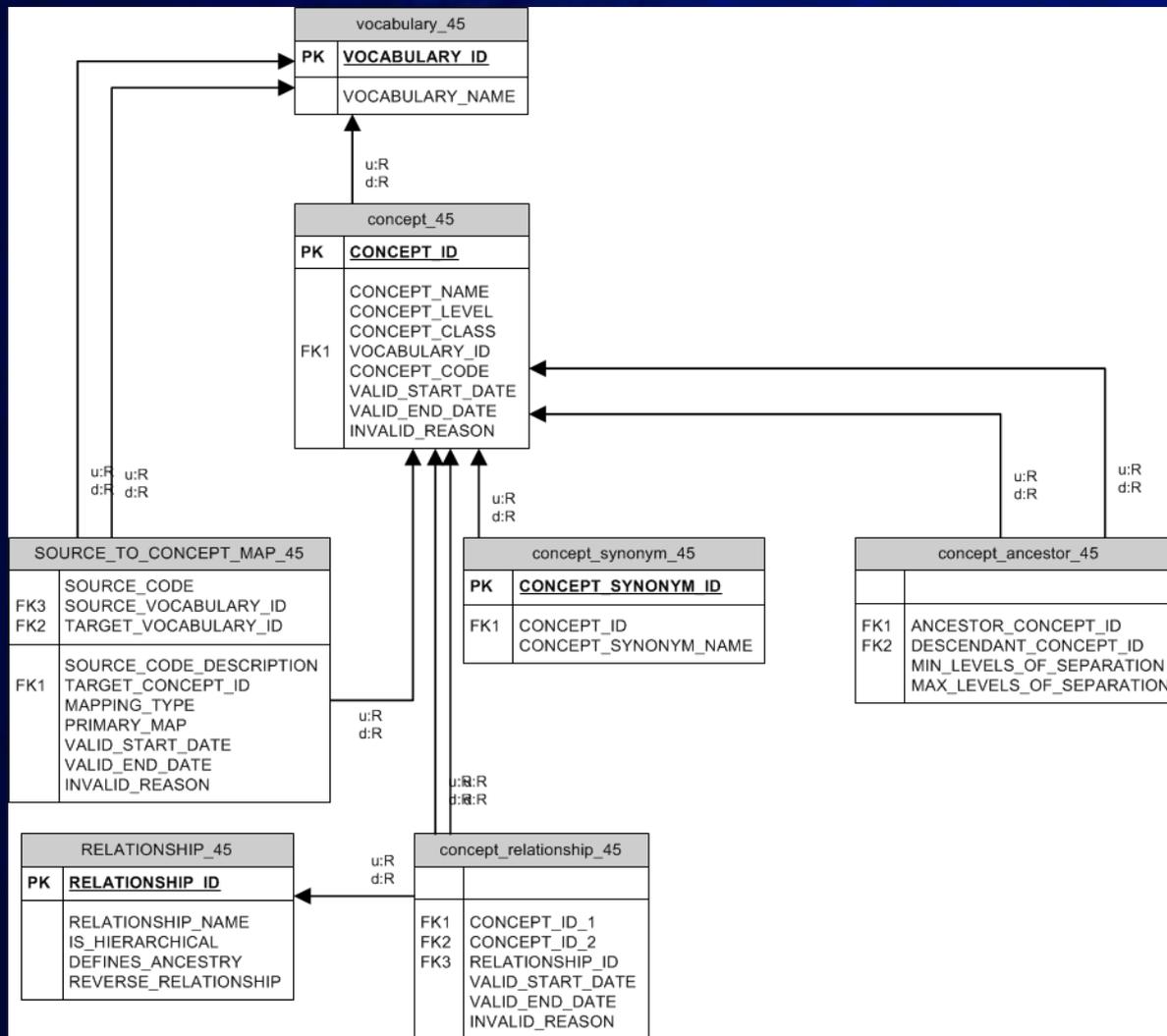
Vocabulary Name



Concepts



Vocabularies, concepts and hierarchies



Standardized vocabulary

<u>Standard Vocabulary</u>	<u>Data type or source</u>
Snomed-CT	ICD-9/10, Pathology, HCPC
LOINC	Laboratory results
RXNorm	Drugs, ingredients and
VA class, ETC and NDF-RT	Indications from RxOut, BCMA, HCPCS Drugs

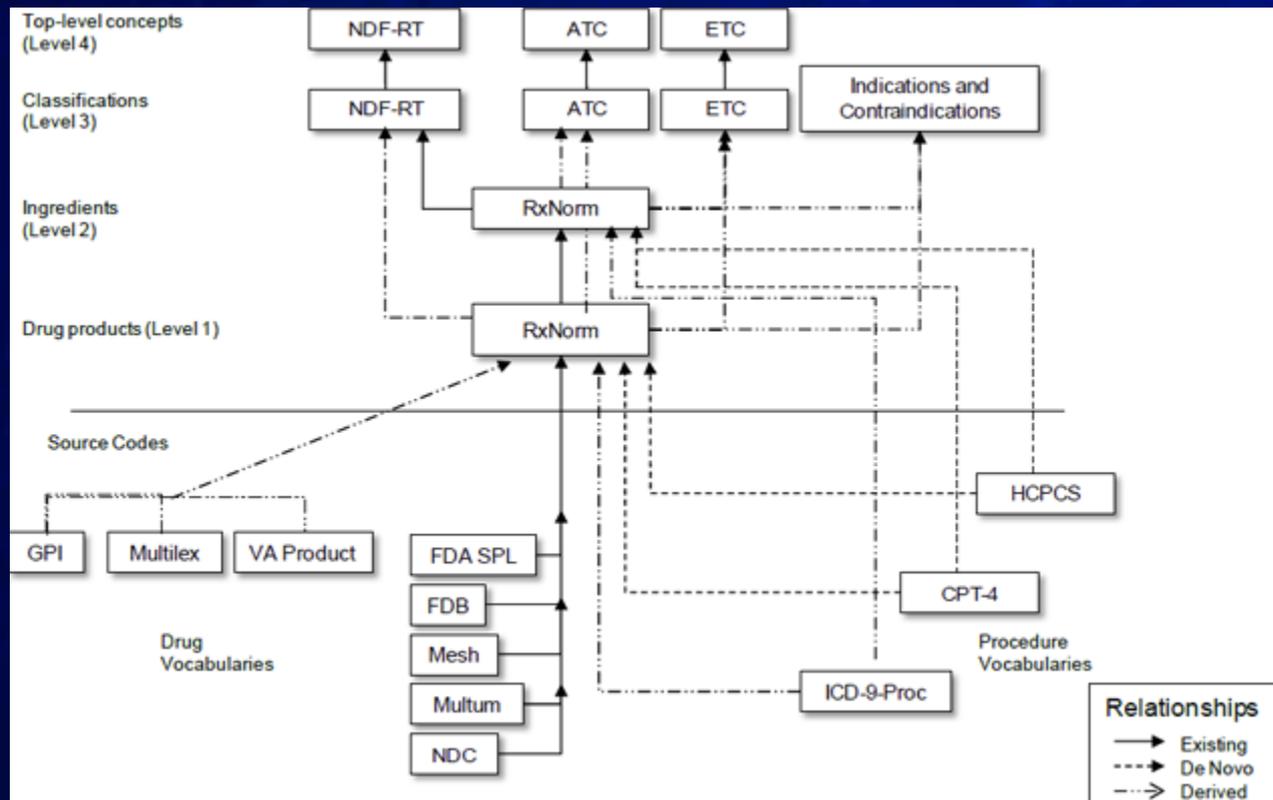
OMOP Vocabulary Tables

- Concept – Houses each individual concept from each standard vocabulary
- Concept Synonym – Alternate names for each concept
- Concept Relationship – Stores the relationship between each concept (i.e parent-child, is-a)
- Concept Ancestor – Stores ancestor-descendant relationships

OMOP Vocabulary Tables cont.

- Vocabulary – A dim table of standard vocabularies in the Concept table
- Source-to-concept map – used to map one concept to an equivalent in another vocabulary
- Relationship – A dim table of relationships used in Concept Relationship table.

Drug Domain Vocabularies



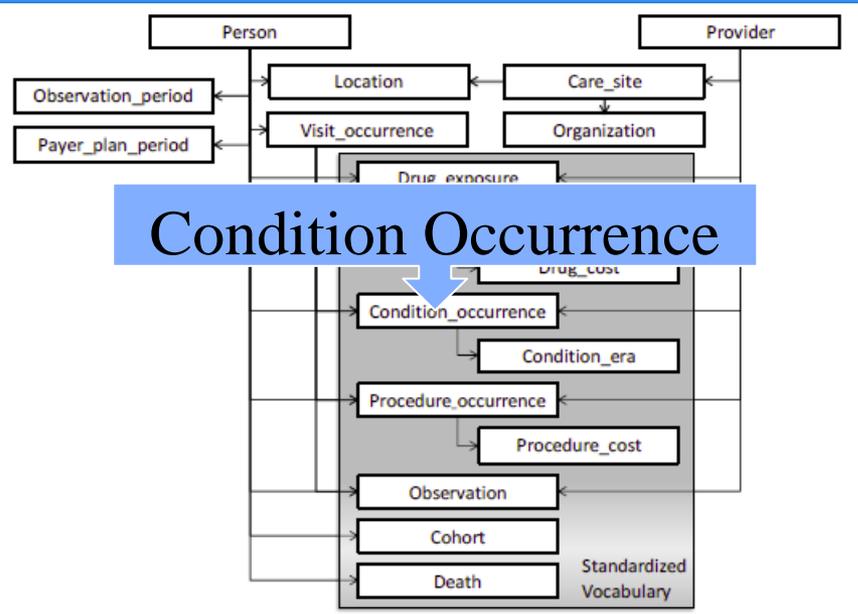
Hierarchical Ladder for Drugs, Example

Concept ID	Concept Name	Concept Level	Concept Class	Vocabulary Concept Code	
19034886	Omeprazole 20 MG Enteric Coated Capsule [Prilosec]	1	Branded Drug	8	207212
923645	Omeprazole	2	Ingredient	8	7646
4319354	Pyridinylmethylsulfinylbenzimidazoles	3	Chemical Structure	7	N0000175098
4351005	Sulfoxides	3	Chemical Structure	7	N0000008055
4350914	Heterocyclic Compounds	3	Chemical Structure	7	N0000008095
4351444	Benzimidazoles	3	Chemical Structure	7	N0000007536
4340570	Infectious Diseases	3	Indication or Contraindication	7	N0000000007
4344424	Paraneoplastic Endocrine Syndromes	3	Indication or Contraindication	7	N0000002143
4342919	Esophagitis	3	Indication or Contraindication	7	N0000001165

Transformed CDW Domains

- SPatient
- BCMA
- Rxout
- Inpat
- Outpat
- Vitals
- Chem
- Staff
- Vital Status
- Locations
- Institutions
- Dental

Condition Occurrence



- Conditions suffered by Persons as extracted from source data.
- Medical claims data include ICD-9-CM diagnosis codes that are submitted as part of a claim for health services and procedures.
- Problem list

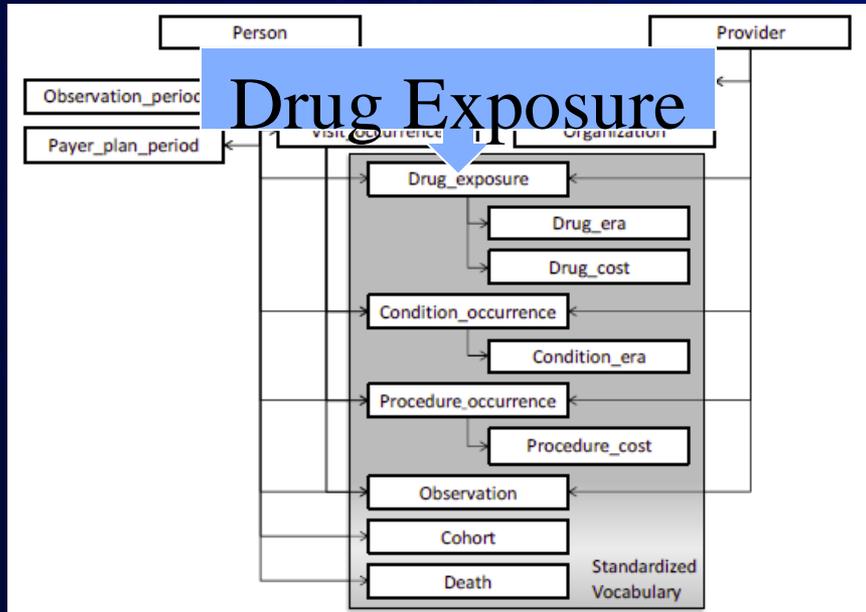


Condition Occurrence

- Derived from inpatient, fee basis, and outpatient diagnoses
- Condition concepts are mapped to ICD9 codes
- x_patientSID, x_sta3n, x_inpatientdiagnosisSID, x_inpatientSID, x_inpatientFeeDiagnosisSID, x_inpatientFeeBasisSID, x_VDiagnosisSID, x_VisitSID

CONDITION_OCCURRENCE	
12	CONDITION_OCCURRENCE_ID PERSON_ID
11	CONDITION_CONCEPT_ID
12	CONDITION_START_DATE CONDITION_END_DATE CONDITION_TYPE_CONCEPT_ID STOP_REASON ASSOCIATED_PROVIDER_ID VISIT_OCCURRENCE_ID CONDITION_SOURCE_VALUE x_PatientSID x_Sta3n x_InpatientDiagnosisSID x_InpatientSID x_InpatientFeeDiagnosisSID x_InpatientFeeBasisSID x_VDiagnosisSID x_VisitSID

Drug Exposure



- The “Prescription” section of an EHR captures prescriptions written by physicians.
- Other drugs (both non-prescription products and medications prescribed by other providers) used by a Person are recorded in the “Medications” section of the EHR.
- Administrative claim systems capture prescriptions filled at dispensing providers.
- Drug Exposure information as a by-product of certain procedure codes (i.e., procedure codes that refer to the administration of certain drugs, such as chemotherapy or vaccines).

Drug Exposure

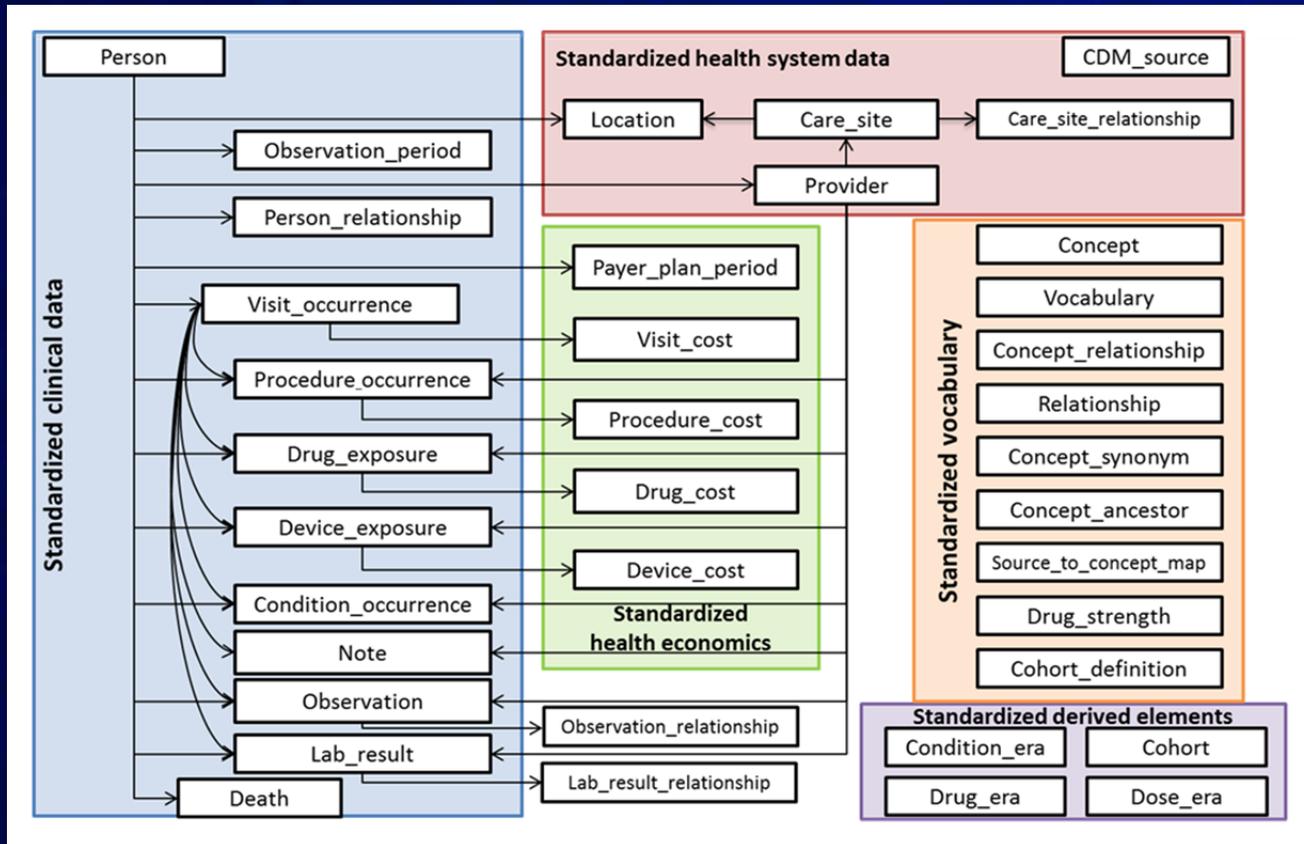
- Derived from Bar Code Med Administration (BCMA) system, and outpatient pharmacy
- Mapped to either a VA Product or RxNorm vocabulary concept
- Quantity is difficult to standardize due to a lack of units. (ex. Hydrocortisone cream has a quantity of 120, this is in milliliters, where a quantity of 25 for amoxicillin would be the number of capsules)
- x_SourceSID, x_LocalDrugSID

DRUG_EXPOSURE	
13	DRUG_EXPOSURE_ID PERSON_ID
12	DRUG_CONCEPT_ID
13	DRUG_EXPOSURE_START_DATE
	DRUG_EXPOSURE_END_DATE
	DRUG_TYPE_CONCEPT_ID
	STOP_REASON
	REFILLS
	QUANTITY
	DAYS_SUPPLY
	SIG
	PRESCRIBING_PROVIDER_ID
	VISIT_OCCURRENCE_ID
	RELEVANT_CONDITION_CONCEPT_ID
	DRUG_SOURCE_VALUE
	x_SourceSID
	x_LocalDrugSID
	x_PatientSID

Future directions

- 2 versions (v4 and v5) will be available as views on VINCI
- Cohort and cohort era tables
- NPL discovered and defined conditions, occurrences and observations (example: Ejection Fraction – Scott Duvall PI)
- CMS and DOD data
- Other CDW data – as requested
 - Microbiology - M. Jones
 - Cancer Registry
- ODHSI tools – Achilles and Hermes 1st

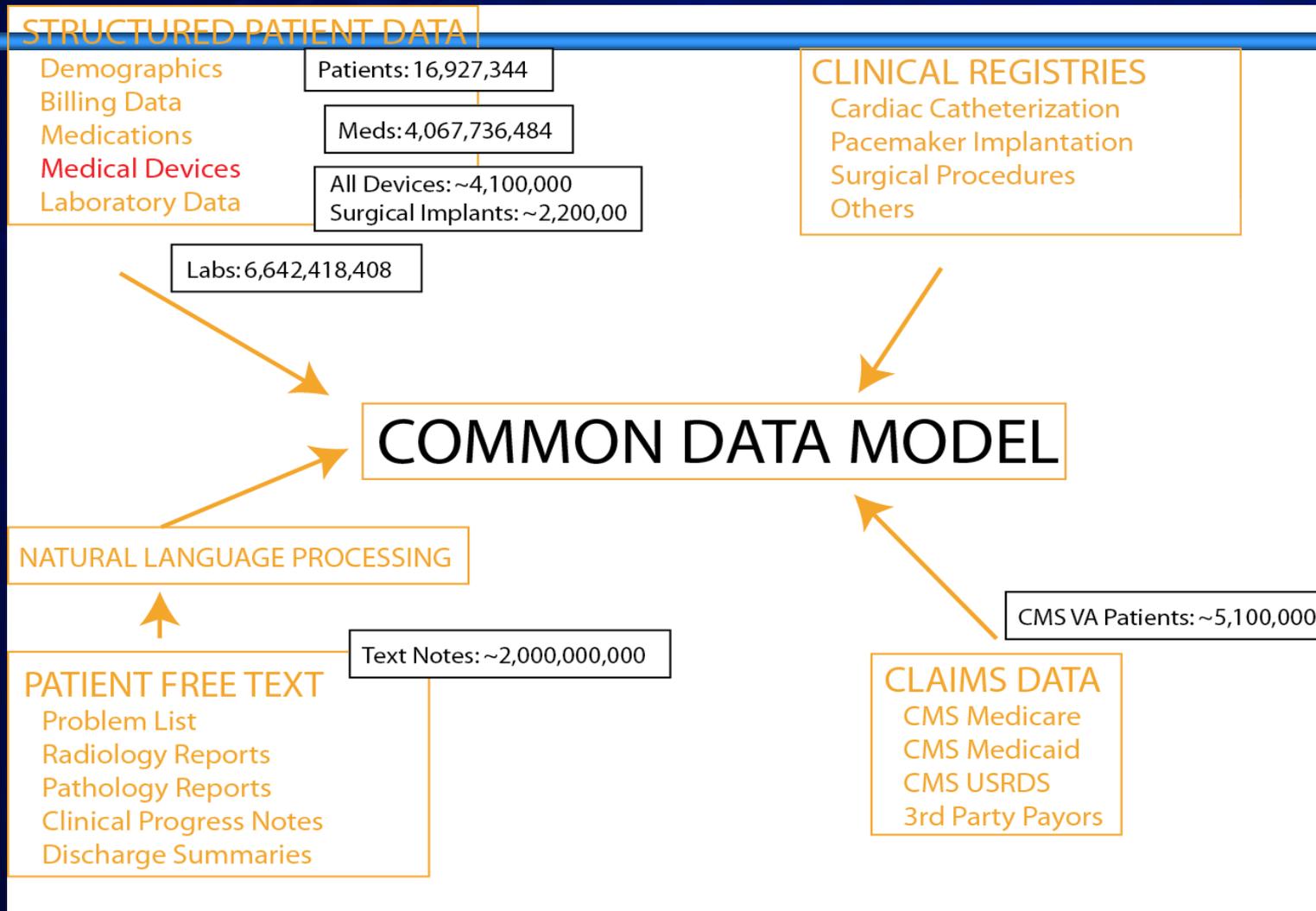
OMOP Version 5



OMOP v5

- Change of table structure to exploit the many to many mappings some concepts provide e.g some diagnoses also indicate procedures
- Necessary to use OHDSI (Odyssey) suite of tools
- Applicability to other common data models (PCORnet)
- Breaking out some of the larger tables smaller more manageable (observation into observation and measurement)

Future Plans



Tools to help with data



OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Who We Are | Who We Serve | Data Standardization | Software Tools | Resources | Join the Journey | Events

Home | Analytic Tools

Analytic Tools

Our open-source analytic tools are one of the most exciting aspects of the collaborative. These tools allow anyone with data in CDM format to take advantage of advanced visualization and analytic methodologies, explore data interactively, and discover insights into medication safety, clinical effectiveness, healthcare quality, and a wide variety of other important topics.

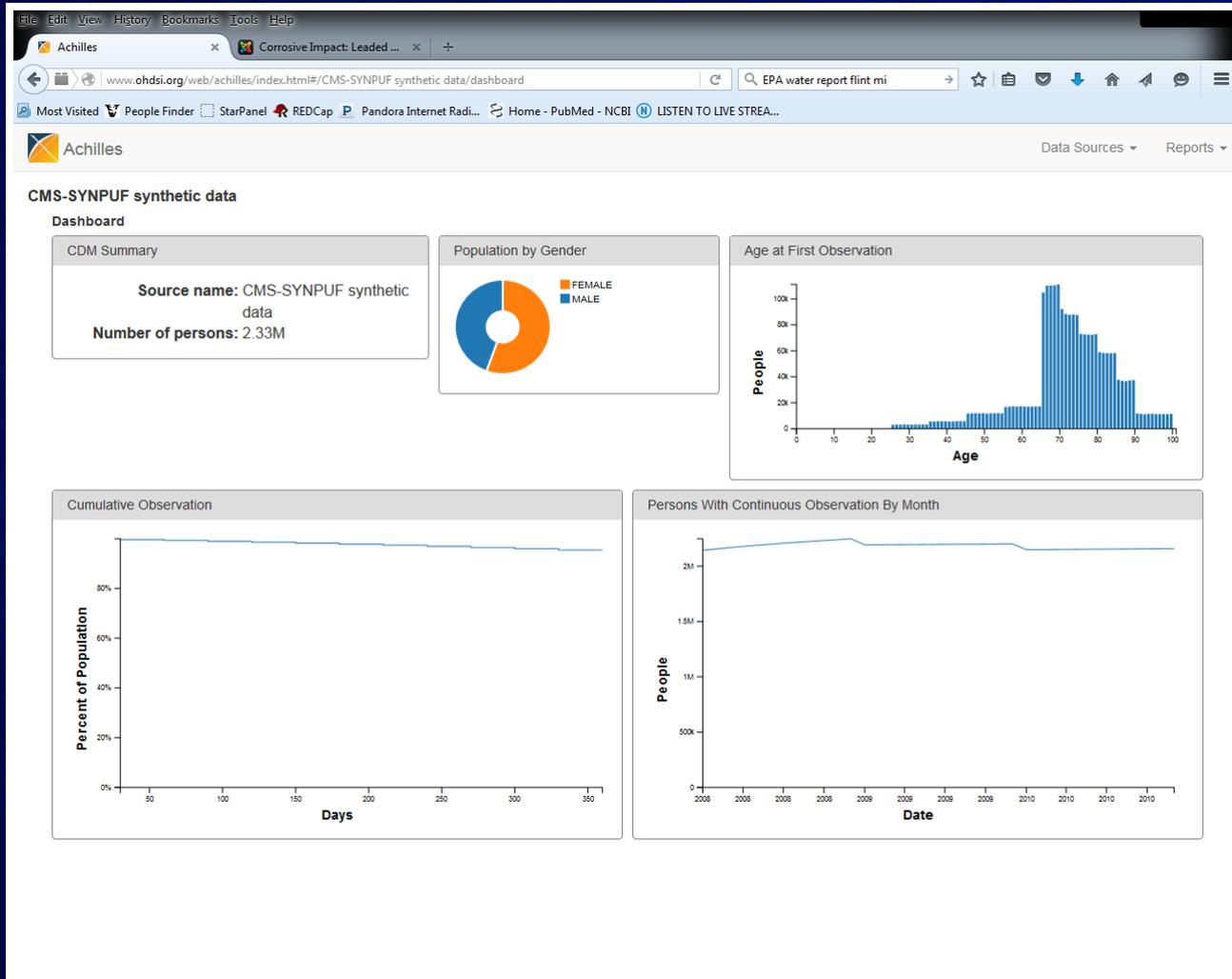
[Live Demos](#)

We encourage you not only to use these tools but to take part of their ongoing development on [GitHub](#). Explore the links below to learn more about these individual tools.

- ATHENA standardized vocabularies
- CALYPSO for study population evaluation
- OLYMPUS - A Unified Platform for OHDSI Applications
- USAGI for vocabulary mapping
- ACHILLES for data characterization
- HOMER for population-level estimation
- PLATO for patient-level prediction
- HERMES for vocabulary exploration
- HERACLES for quality of care
- WhiteRabbit for ETL design

ACHILLES
HOMER
PLATO
OHDSI
HERMES
HERCULES
WHITERABBIT

Visualizing data with Achilles



Cohort development

File Edit View History Bookmarks Tools Help

CALYPSO Corrosive Impact: Ledged ... x +

www.ohdsi.org/web/calypso/#/719 EPA water report flint mi

Most Visited People Finder StarPanel REDCap Pandora Internet Radi... Home - PubMed - NCBI LISTEN TO LIVE STREA...

CALYPSO
Criteria Assessment Logic for Your
Population Studies of Observations

Feasibility Study List Help

New Feasibility Study

Save Import Copy Generate... Delete

Show SQL

Description

Example #1

Index Cohort ID:720

Index Rule Inclusion Rules Concept Sets Print Friendly Results

Description:

People having any of the following: **Add Primary Event Filters...**

a drug exposure of Any Drug **Add Filter...** Delete Filter

with observation at least 7 days prior and 0 days after index

Limit primary events to: all events per person.

For people matching the Primary Events, include:

People having all of the following criteria: **Add New Criteria...**

with at least 1 using all occurrences of: **Add Filter...**

a specimen of Any Specimen

occurring between All days Before and All days After index **Delete Filter**

Remove Additional Filters

Limit cohort expression results to: all events per person.

Getting access to VINCI OMOP

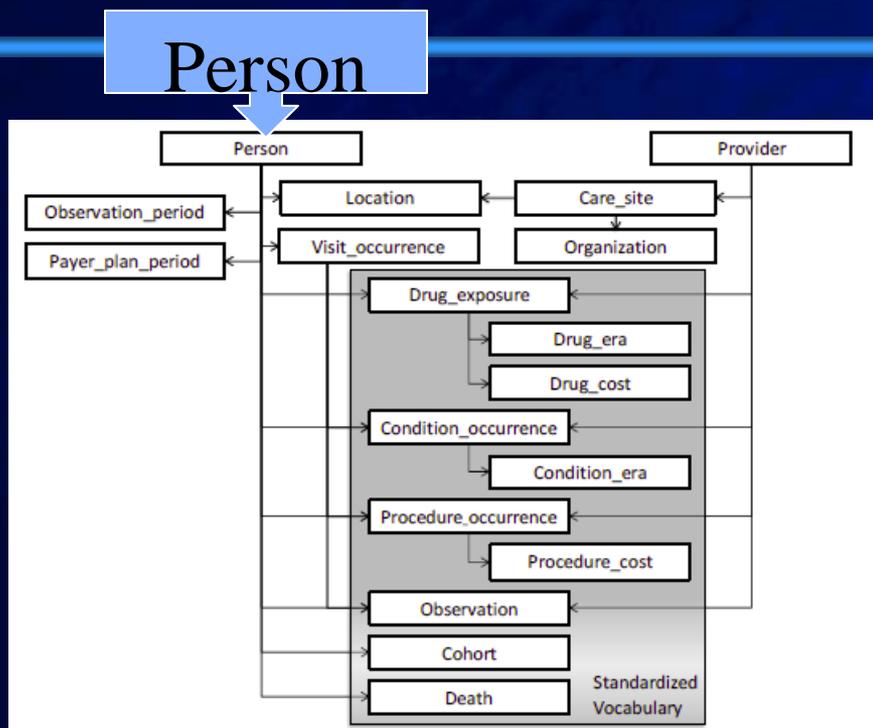
- Researcher Access - Include VINCI OMOP as part of DART. OMOP tables and OMOP metadata tables will be added like other table views. Link and crosswalk using patient id and source table IDs.
 - Will become a checkbox in the DART process
- Operational Access – As part of NDS request, include VINCI OMOP. Current CDW access grants access to OMOP tables as well.

OMOP Support

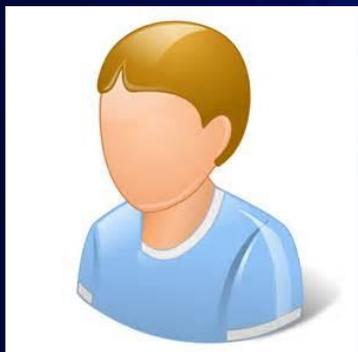
- VA Pulse VINCI OMOP group
 - Message list
 - Documentation
 - FAQ
 - Example code
 - Webinar videos
- VINCI help desk– include VINCI OMOP in your ticket descriptor.

Questions

Person



- Person table comes from person demographics data
- Data attributes are birthdate race, gender, and ethnicity
- References to other tables about the home address (location), primary care provider...



Person

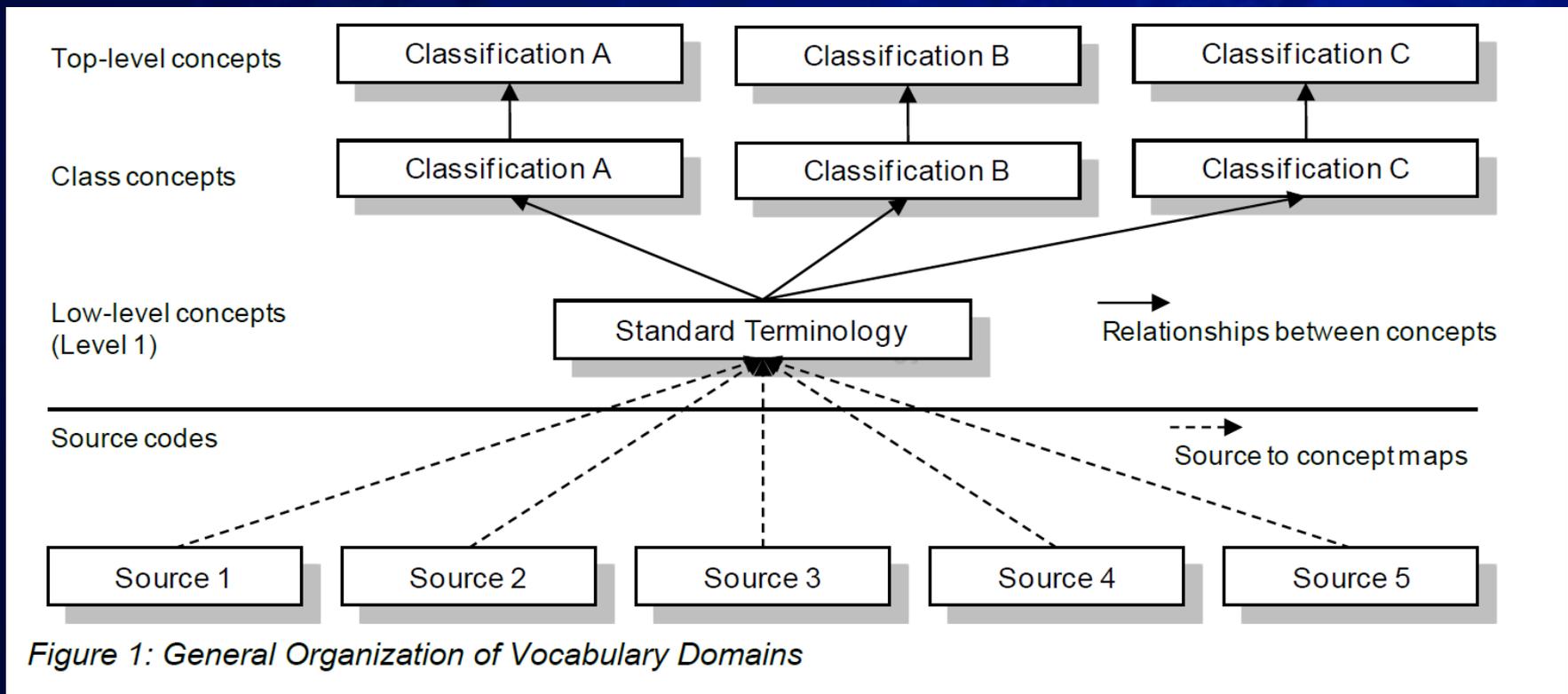
- PatientICN is the PERSON_ID
- Patients are filtered to those ICNs that are veterans, not test patients, without multiple SSNs
- DOB determined by most frequent occurring not NULL, in a tie, DOB on first visit
- Gender determined by most frequent occurring not NULL, in a tie, most recent
- Race and ethnicity determined by most frequent , in a tie, first visit

PERSON	
PK	<u>PERSON ID</u>
	GENDER_CONCEPT_ID
	YEAR_OF_BIRTH
	MONTH_OF_BIRTH
	DAY_OF_BIRTH
	RACE_CONCEPT_ID
	ETHNICITY_CONCEPT_ID
	LOCATION_ID
	PROVIDER_ID
	CARE_SITE_ID
	PERSON_SOURCE_VALUE
	GENDER_SOURCE_VALUE
	RACE_SOURCE_VALUE
	ETHNICITY_SOURCE_VALUE

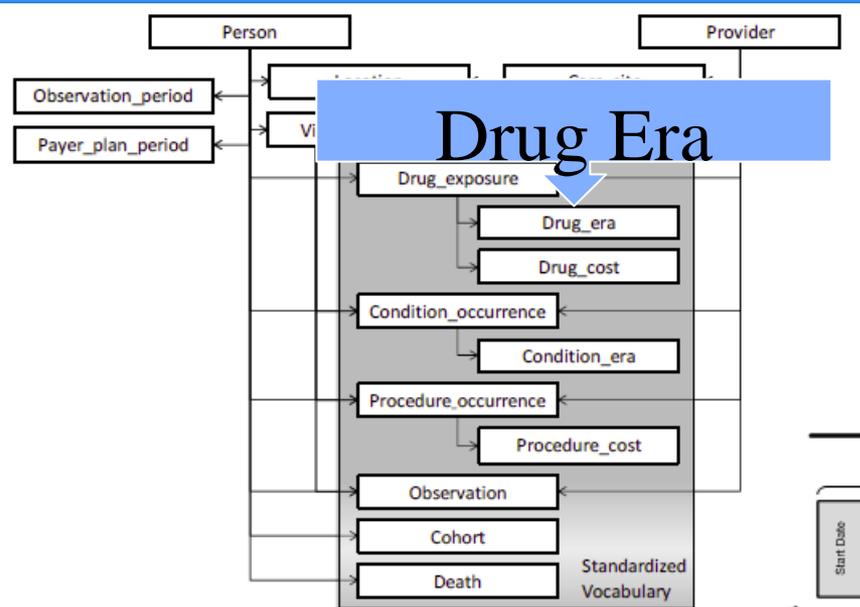
OMOP Data Tables Not Implemented

- Drug Cost
- Procedure Cost
- Cohort

Structure of Vocabulary Domains

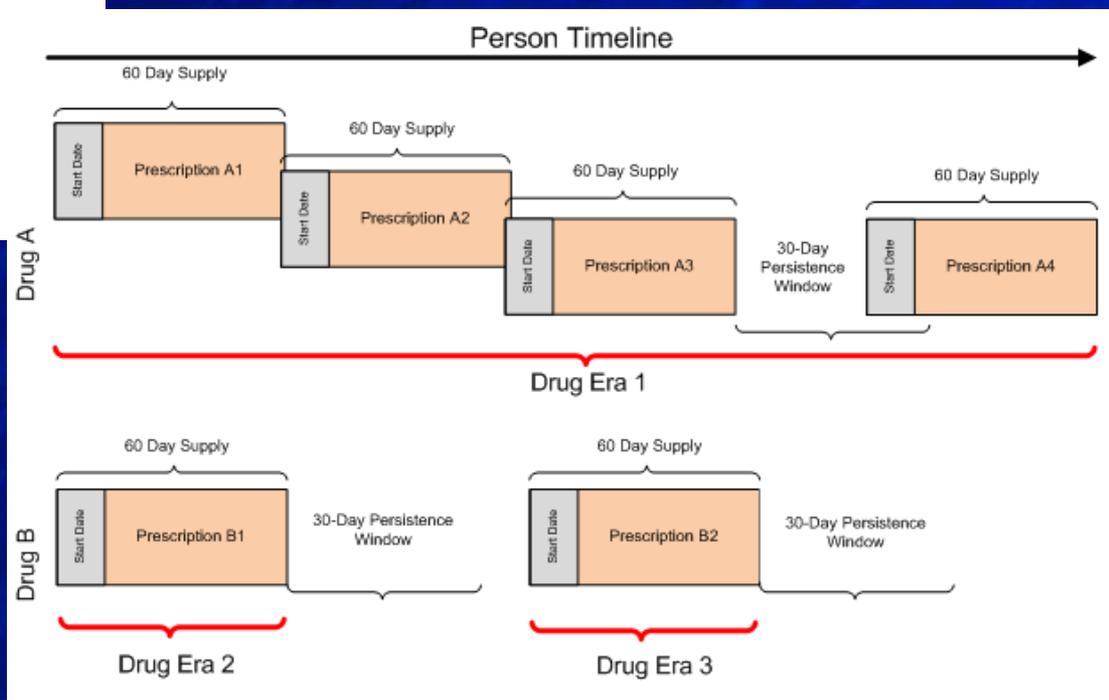


Drug Era



- Each Drug Era corresponds to one or many Drug Exposures that form a continuous interval

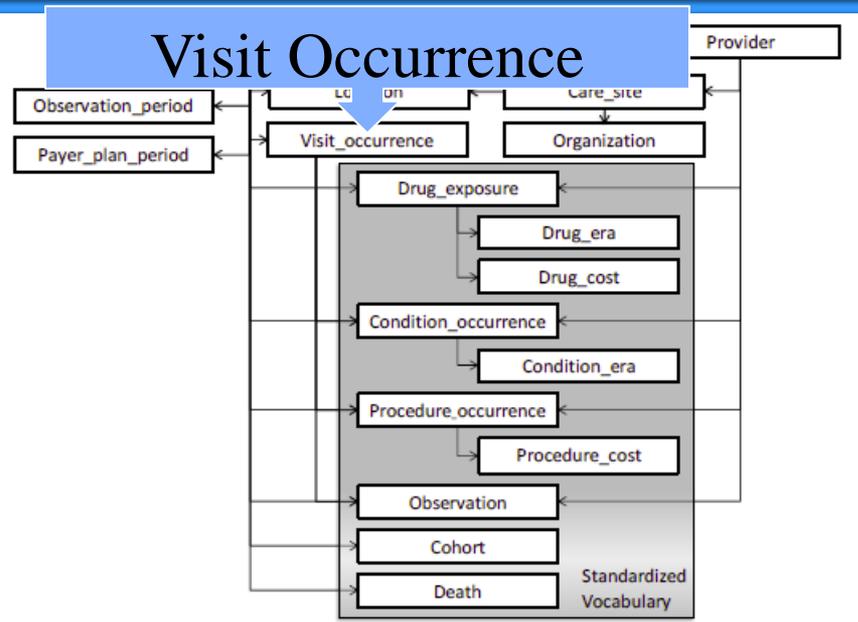
DRUG_ERA	
DRUG_ERA_ID	
PERSON_ID	
DRUG_CONCEPT_ID	
DRUG_ERA_START_DATE	
DRUG_ERA_END_DATE	
DRUG_TYPE_CONCEPT_ID	
DRUG_EXPOSURE_COUNT	
x_Start_Source_ID	
x_End_Source_ID	



Drug Era

- Because there are duplicate drug entries in the source data, a drug exposure count field can have a high number despite a short era
- Conversely, there is a 30 day window from drug end to restart, so there can be a low count for a long span

Visit Occurrence



- Visit to health care providers, including inpatient, outpatient, and ER visits
- An encounter for a patient at a point of care for a duration of time.

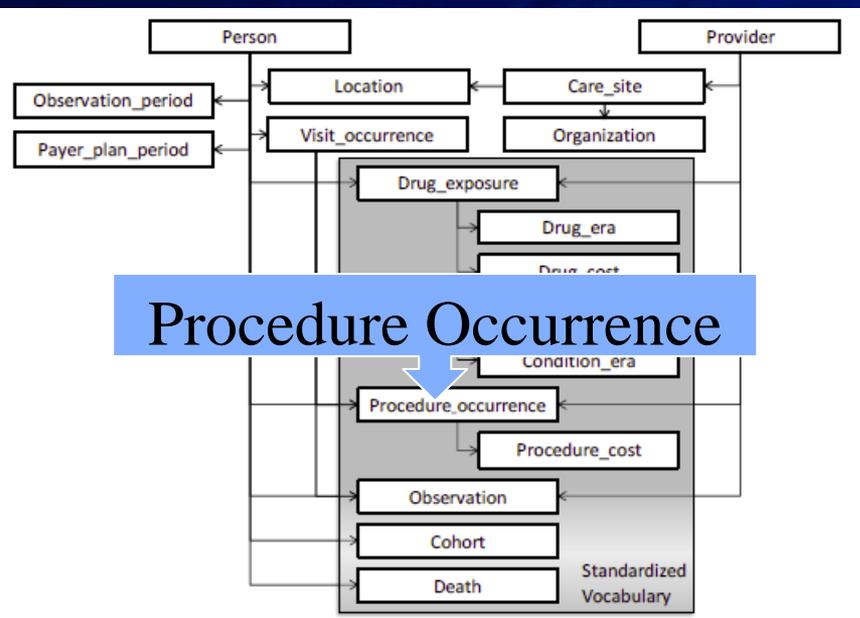


Visit Occurrence

- Includes all inpatient, inpatient fee basis and outpatient visits
- VA classifies visits to outpatient clinics during an inpatient stay as an outpatient visit
- Place of service concept ID can be used to find
 - x_patientSID, x_sta3n, x_source_table, x_source_ID, x_locationSID, x_institutionSID

VISIT_OCCURRENCE	
U2	VISIT_OCCURRENCE_ID PERSON_ID
	VISIT_START_DATE VISIT_END_DATE PLACE_OF_SERVICE_CONCEPT_ID CARE_SITE_ID PLACE_OF_SERVICE_SOURCE_VALUE x_PatientSID x_source_table x_Source_ID x_Sta3n x_LocationSID x_InstitutionSID

Procedure Occurrence



- Instances of procedures performed on Persons
- Medical Claims include CPT-4, ICD-9-CM (Procedures), and HCPCS procedure codes that are submitted as part of a claim for health services rendered, including procedures performed.
- Problem list (hx of procedure like renal transplant)

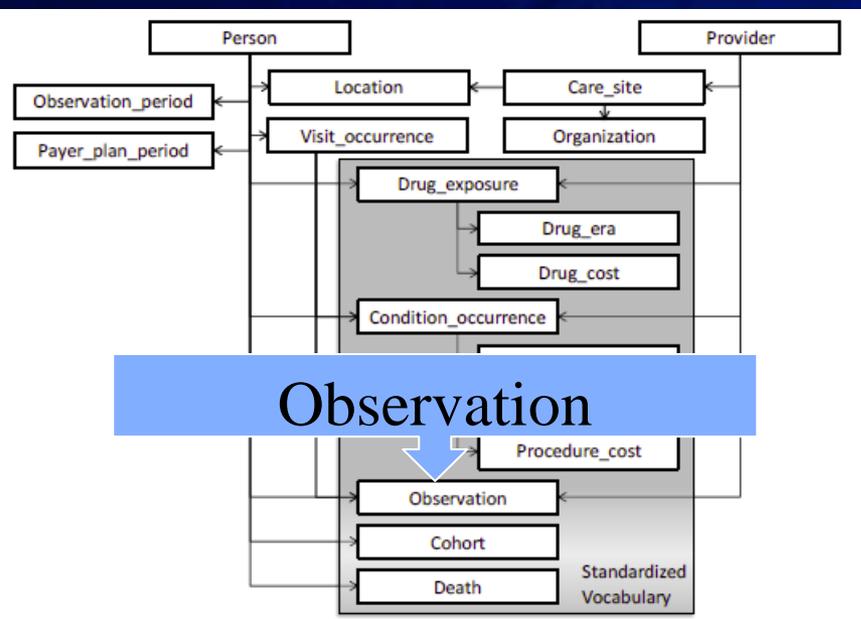


Procedure Occurrence

- Derived from Inpatient CPT, Inpatient ICD, Inpatient Surgery, and Outpatient Procedure
- The VA does not store a diagnoses with its procedure (if any), so **RELEVANT_CONDITION_ID** is NULL
- **x_PatientSID**, **x_Quantity**, **x_source_table**, **x_source_SID**

PROCEDURE_OCCURRENCE	
	PROCEDURE_OCCURRENCE_ID
	PERSON_ID
	PROCEDURE_CONCEPT_ID
	PROCEDURE_DATE
	PROCEDURE_TYPE_CONCEPT_ID
	ASSOCIATED_PROVIDER_ID
	VISIT_OCCURRENCE_ID
	RELEVANT_CONDITION_ID
	PROCEDURE_SOURCE_VALUE
	x_PatientSID
	x_Quantity
	x_Source_Table
	x_Source_SID

Observation



- Lab observations (i.e., test results) from Medical Claims
- Lab and other observations from Electronic Health Records
- Chief complaints as captured in Electronic Health Records
- General clinical findings, signs and symptoms
- Radiology and pathology reports
- General catch-all categories from various data sources that cannot be otherwise categorized within the entities provided (Drug, Condition, Procedure)



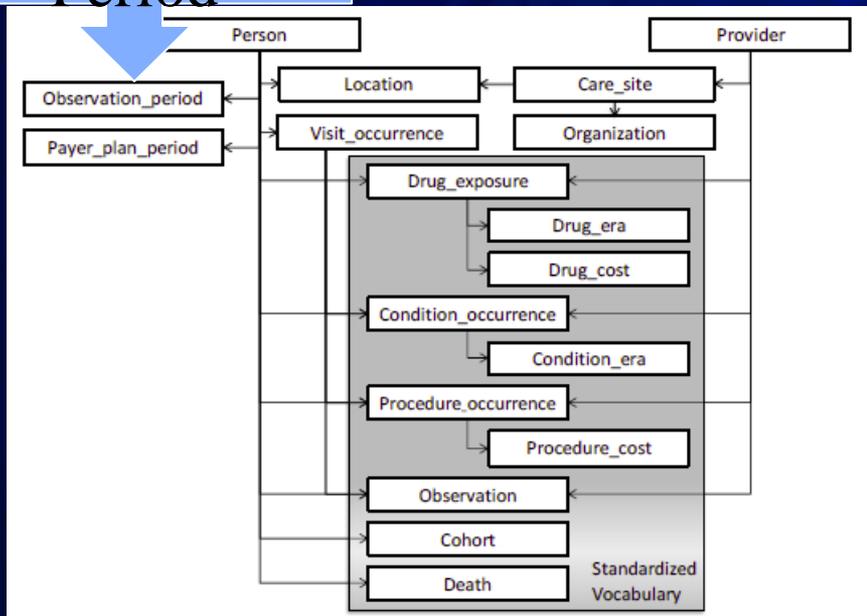
Observation

- Derived from Dental, Chem Lab, Vitals and Inpatient
- Largely functions as a catch-all table for what will not go elsewhere
- Contains tobacco history (from Dental), Chem Labs, Vital Statistics, Admitting Source, Place of discharge, type of discharge (alive or dead), Visit DRGs
- X_abnormal, x_sourceSID, x_staffSID, x_LabChemTestSID, x_source_table

OBSERVATION
OBSERVATION_ID
PERSON_ID
OBSERVATION_CONCEPT_ID
OBSERVATION_DATE
VALUE_AS_NUMBER
VALUE_AS_STRING
VALUE_AS_CONCEPT_ID
UNIT_CONCEPT_ID
RANGE_LOW
RANGE_HIGH
OBSERVATION_TYPE_CONCEPT_ID
ASSOCIATED_PROVIDER_ID
VISIT_OCCURRENCE_ID
RELEVANT_CONDITION_CONCEPT_ID
OBSERVATION_SOURCE_VALUE
UNIT_SOURCE_VALUE
x_Abnormal
x_SourceSID
x_StaffSID
x_LabChemTestSID
OBSERVATION_TIME
x_Source_Table

Observation Period

Observation
Period



- The span of time when a Person is expected to have the potential of Drug and Condition information recorded.
- For claims data, observation periods are equivalent to enrollment periods to a plan

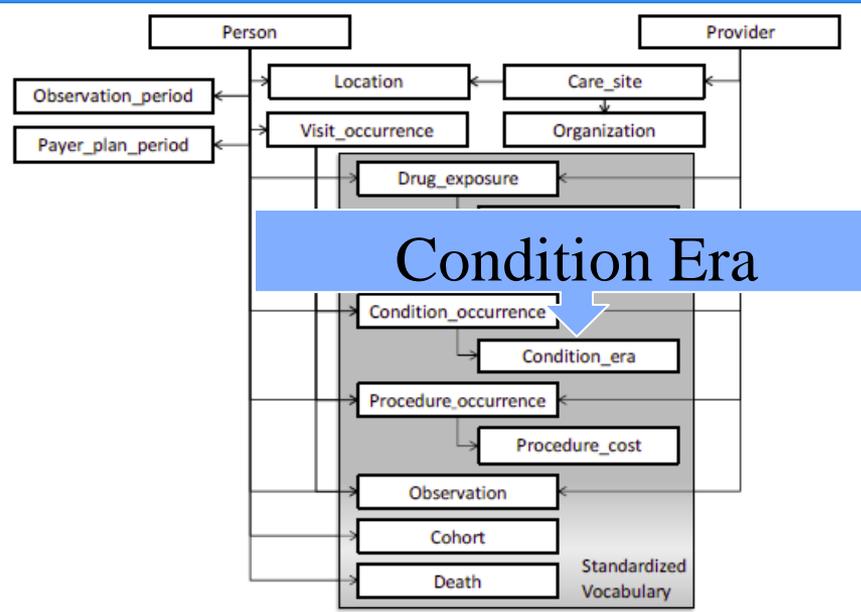


Observation Period

- Derived from OBSERVATION, DRUG_EXPOSURE, and VISIT_OCCURRENCE tables
- This is to gain a true 'period of observation' for a patient, rather than the seemingly random start and end dates that would arise from just using OBSERVATION

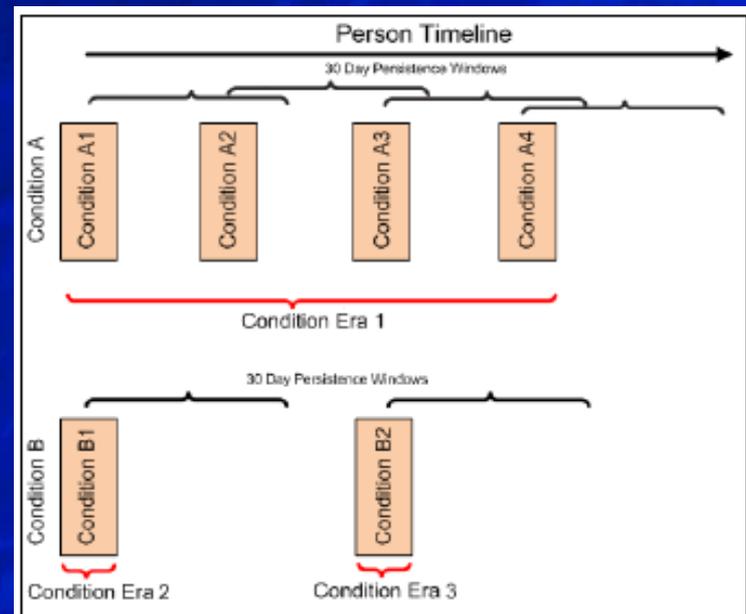
OBSERVATION_PERIOD	
	OBSERVATION_PERIOD_ID
	PERSON_ID
	OBSERVATION_PERIOD_START_DATE
	OBSERVATION_PERIOD_END_DATE

Condition Era



- Aggregation of chronic conditions that require frequent ongoing care
- Aggregation of multiple, closely timed doctor visits for the same condition to avoid double-counting the Condition Occurrences

CONDITION_ERA
CONDITION_ERA_ID
PERSON_ID
CONDITION_CONCEPT_ID
CONDITION_ERA_START_DATE
CONDITION_ERA_END_DATE
CONDITION_TYPE_CONCEPT_ID
CONDITION_OCCURRENCE_COUNT
x_Start_Source_ID
x_End_Source_ID



Condition Era

- Because not all conditions have a true end date (i.e. Diabetes, HIV, Schizophrenia), the era end date is derived from the Condition Occurrence end date, and thus the discharge date in the record.
- Because there are duplicate diagnoses in the source data, a condition occurrence count field can have a high number despite a short era