



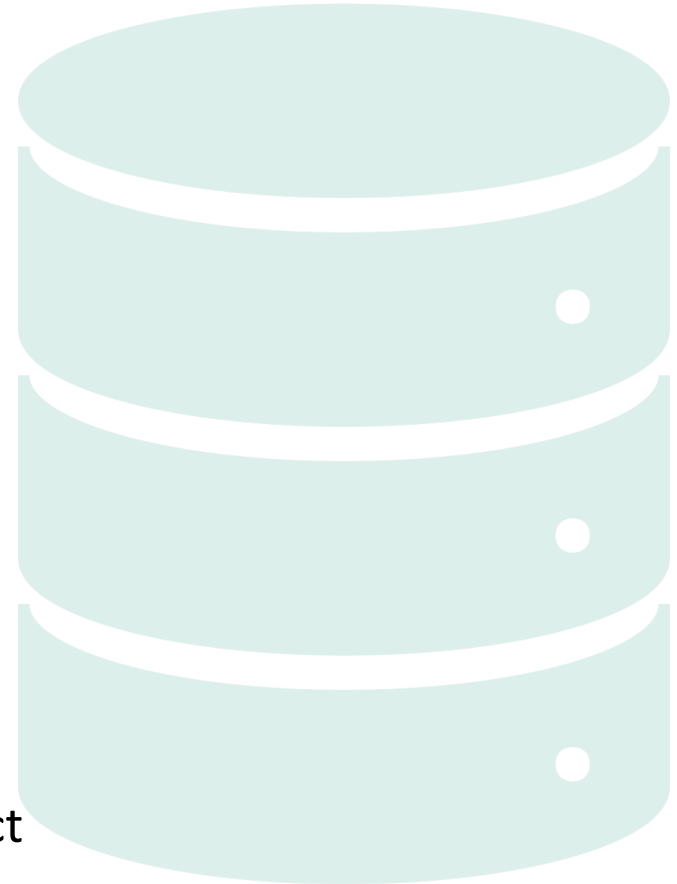
Medicare data in the OMOP Common Data Model

Kristin de Groot, MPH

Technical Director, VA/CMS Data for Research Project

VA Information Resource Center (VIReC)

January 9, 2020



VIReC-VINCI Partnership

VA/CMS Data for Research Project



VINCI's OMOP Initiative

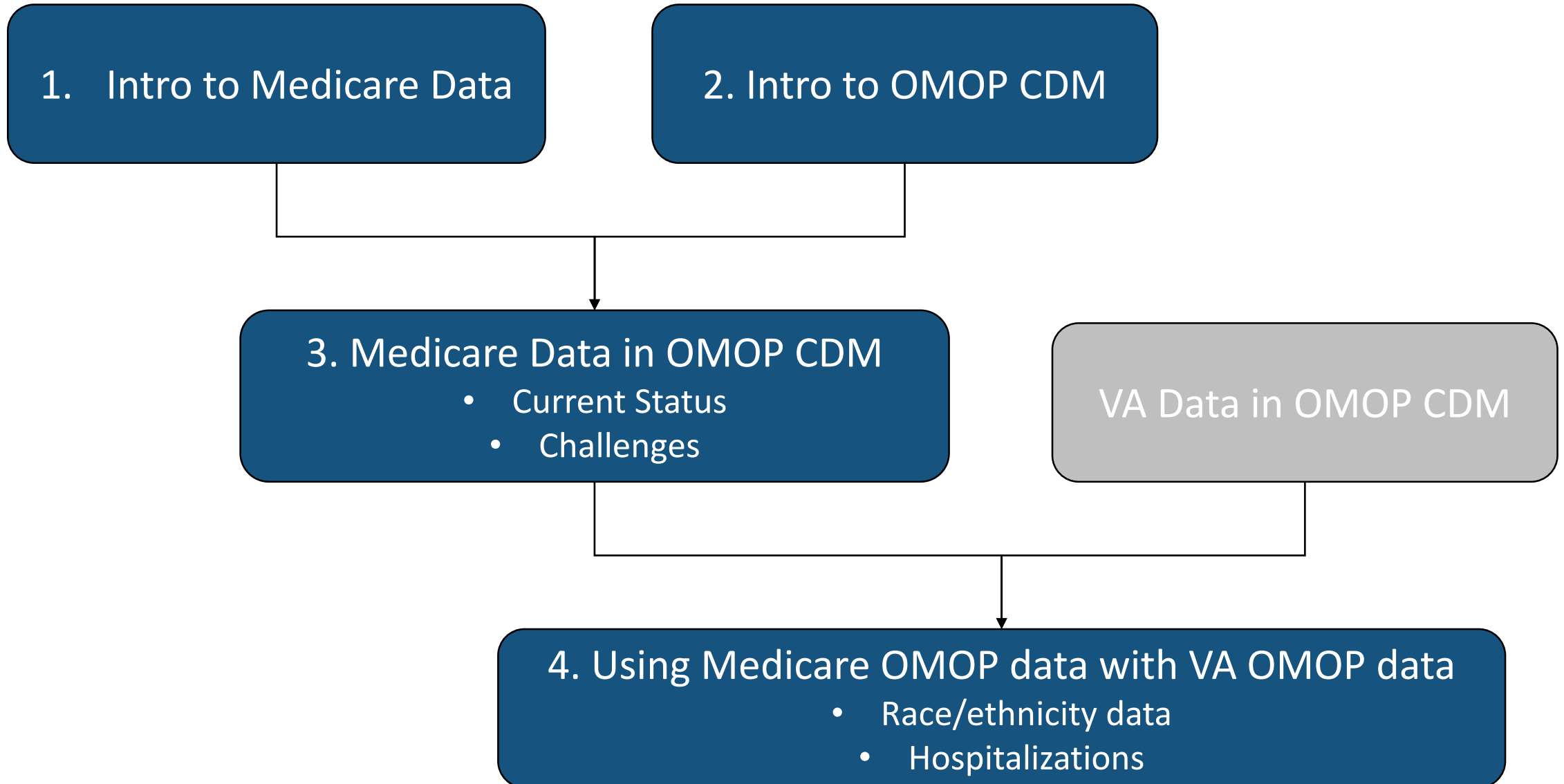


Donghui Kan
Lucy Zhang
Maria Souden
Denise Hynes

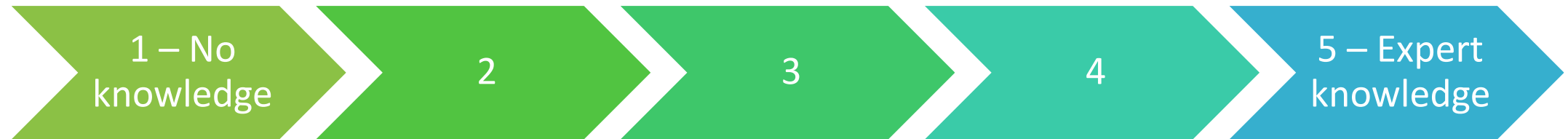


Aize Cao
Jason Denton
Fern FitzHenry
Michael Matheny
Scott DuVall

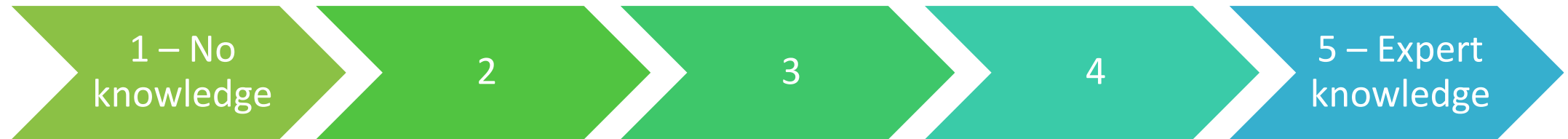
Outline for today's presentation



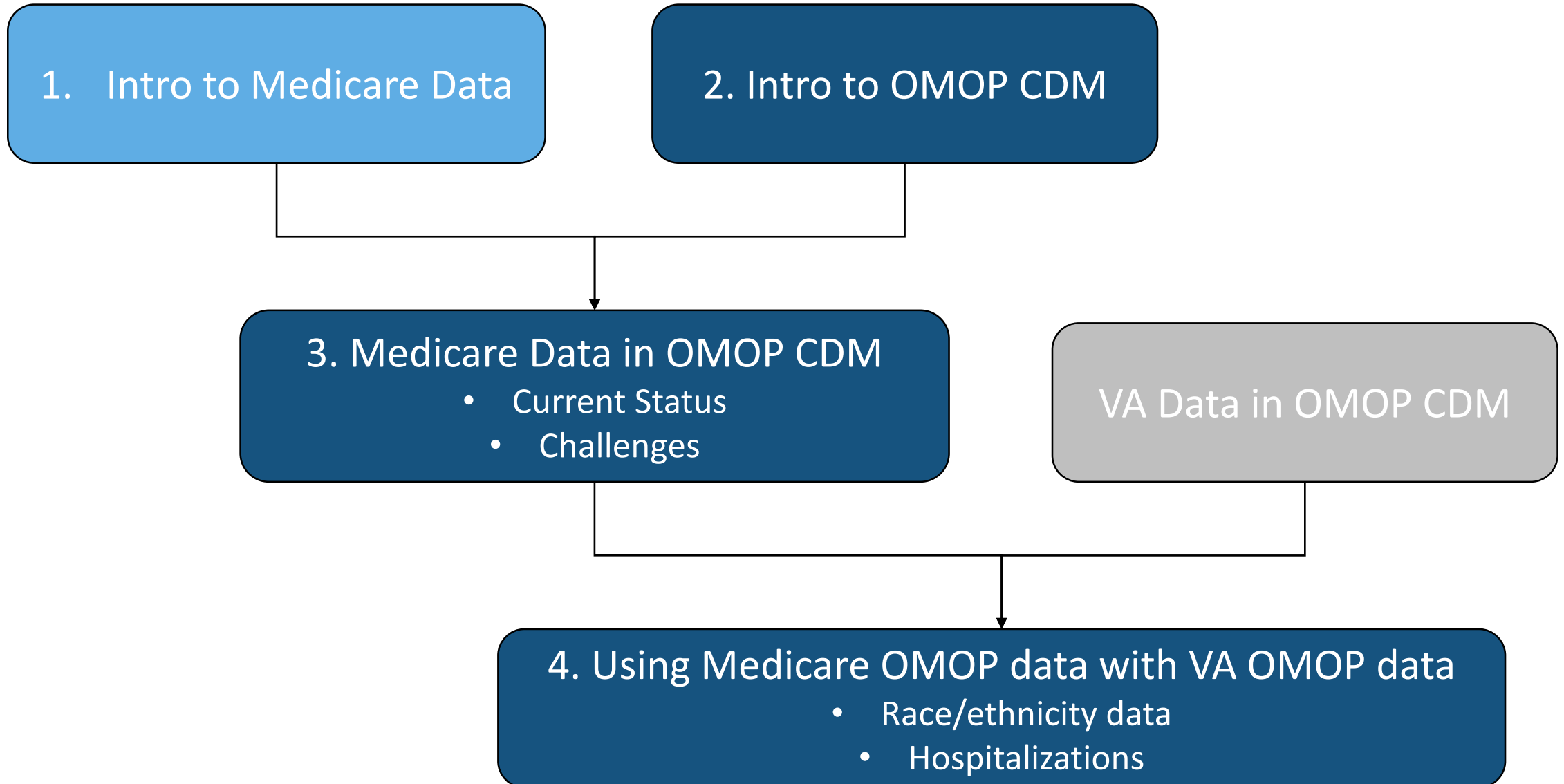
Rate your experience with Medicare data



Rate your experience with OMOP Common Data Model



Outline for today's presentation

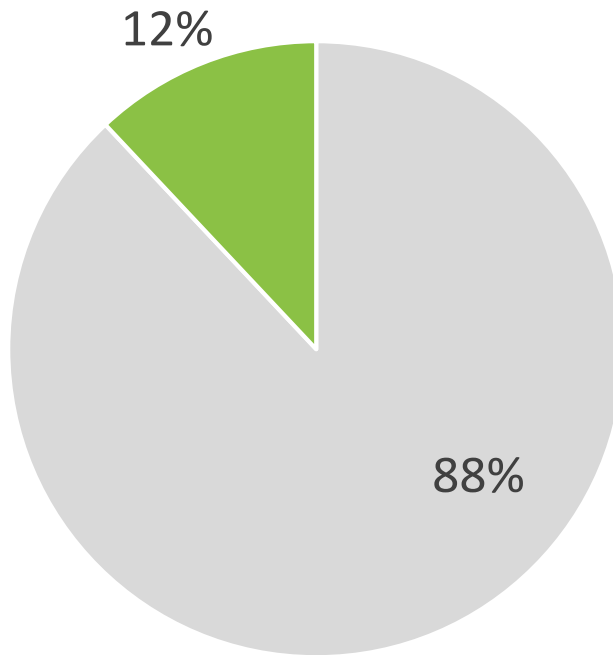


Veterans' Enrollment in Medicare

Percent of VHA Enrollees in Medicare in September 2016

Veterans age <65

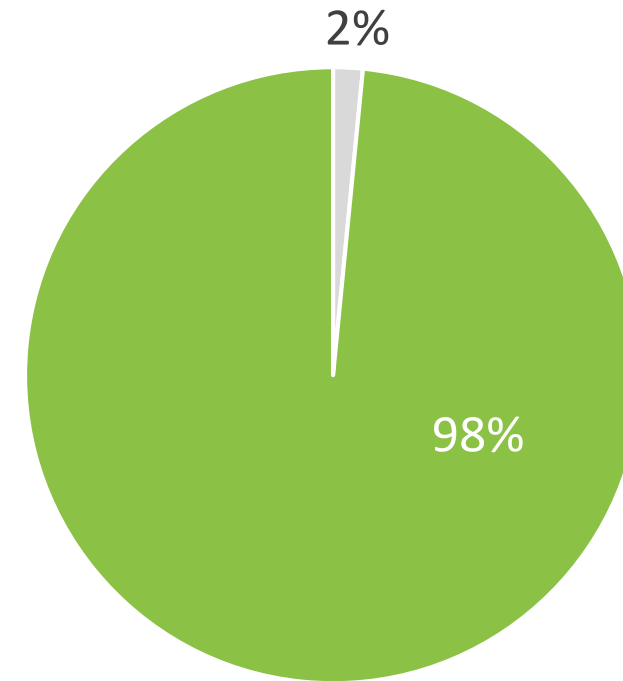
N = 4.9 million



 **VHA only**

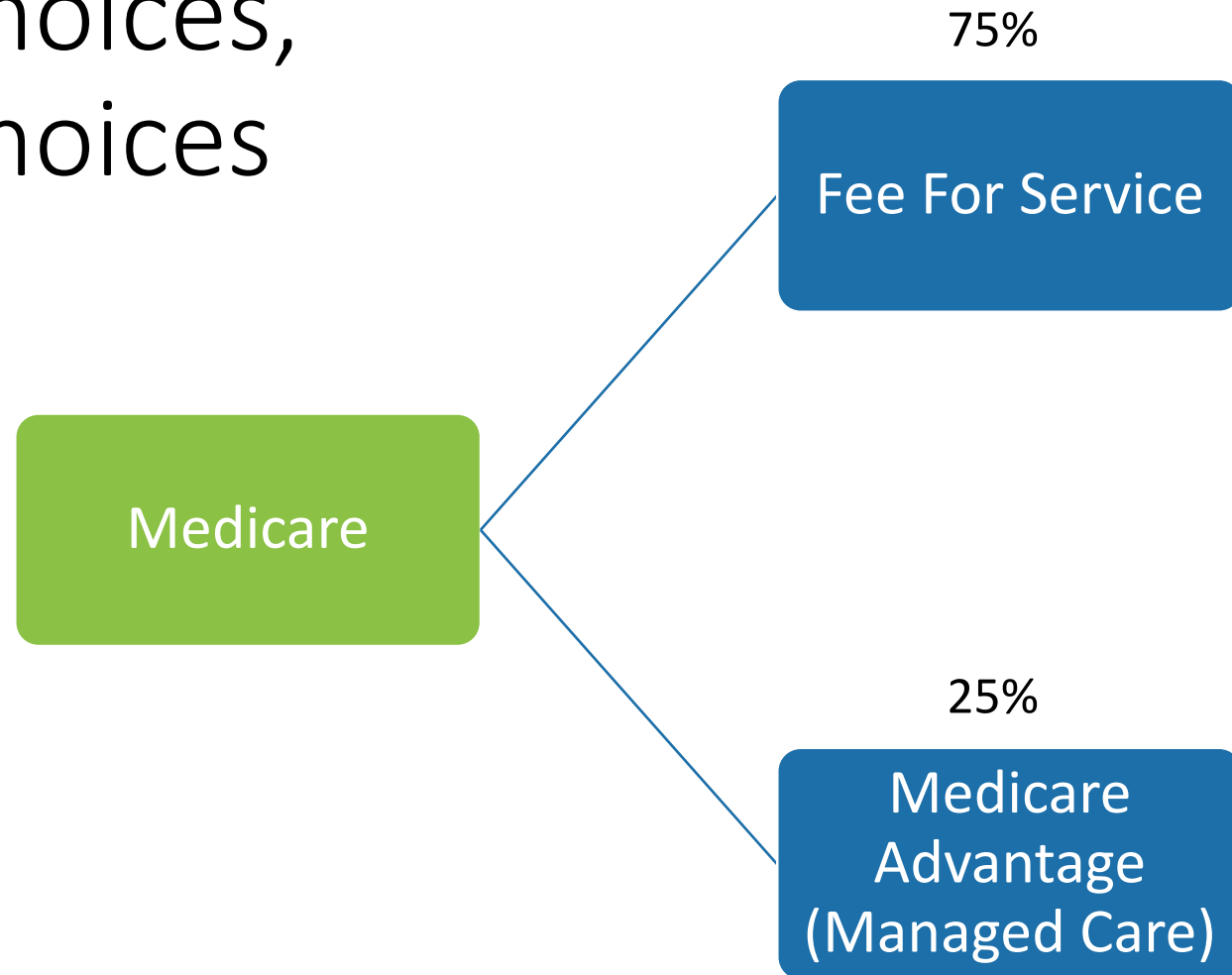
Veterans age 65+

N = 4.7 million

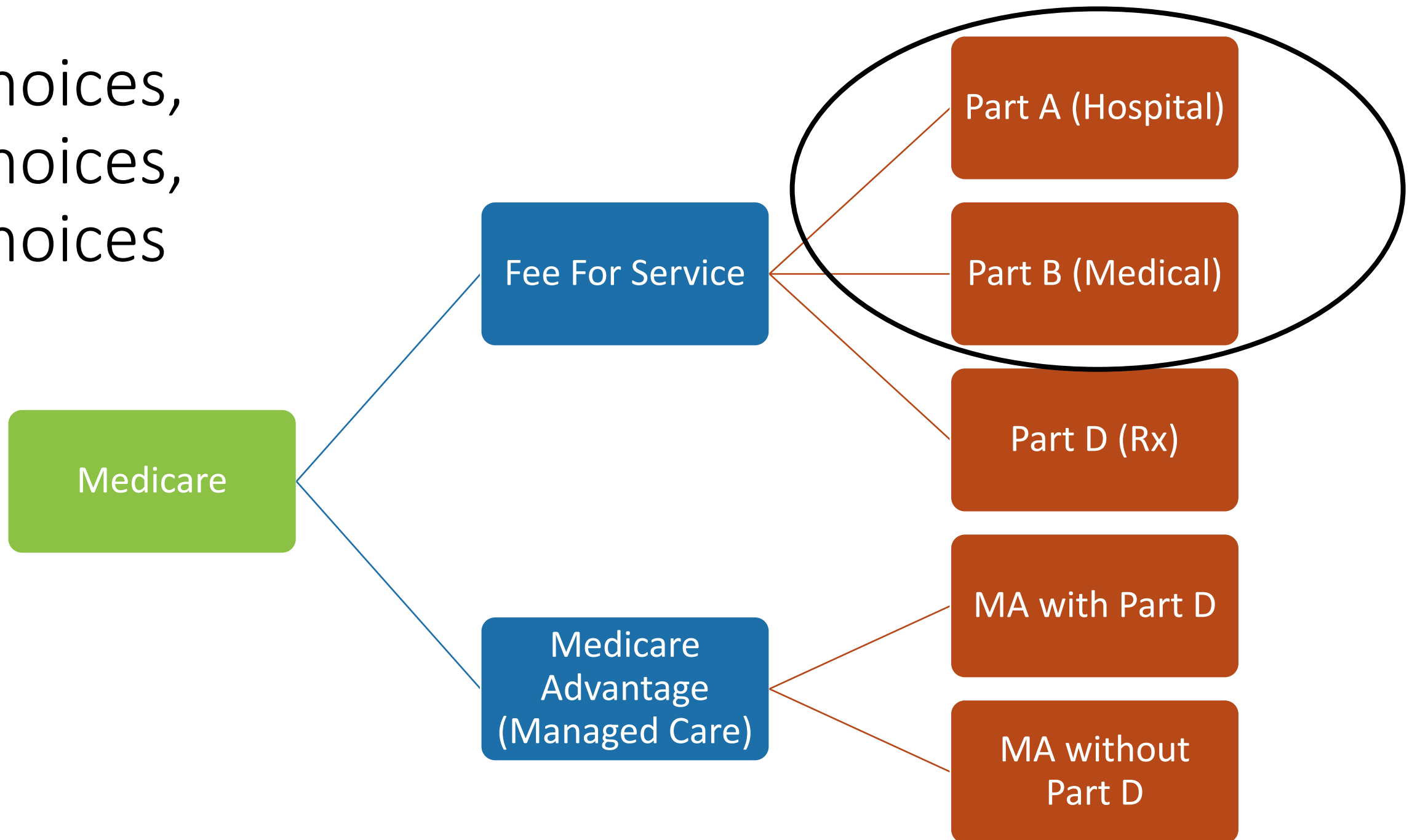


 **VHA + Medicare**

Choices,
Choices,
Choices

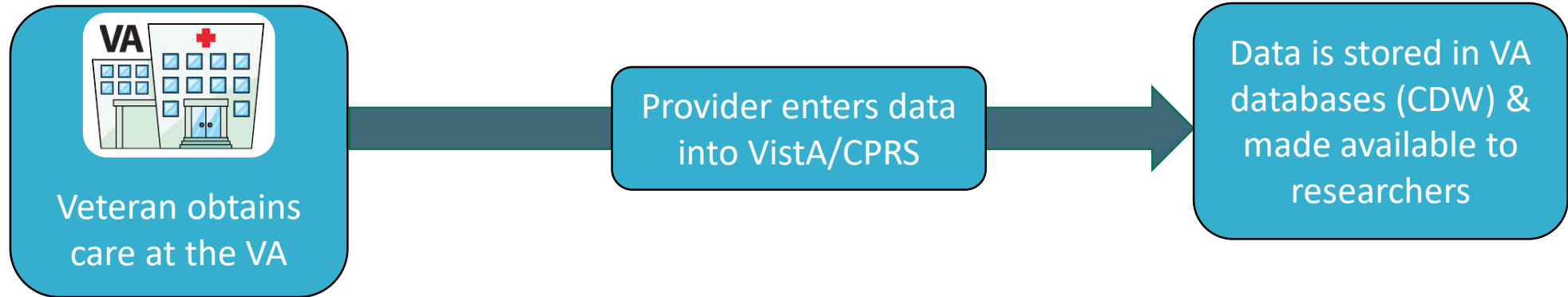


Choices,
Choices,
Choices

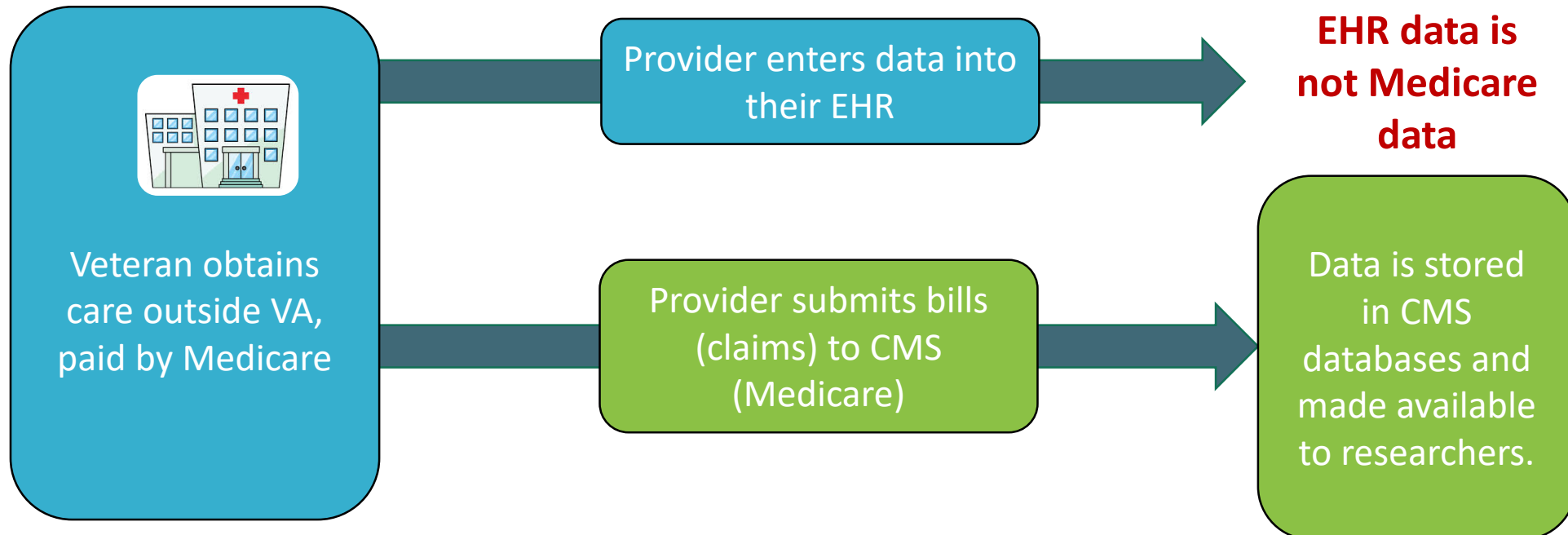


Source of VA and Medicare Utilization Data

VA Data Flow



Medicare Data Flow



Type of Provider



Type of Claim Form



Type of Claim File

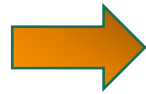
Institutional



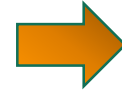
Hospital or
nursing
facility



Home health
agency or hospice



**CMS 1450/
UB-04**



5 Institutional Claim Files

Inpatient (IP)
Skilled Nursing (SN)
Home Health
Hospice
Outpatient

Non-Institutional



Clinical
laboratories



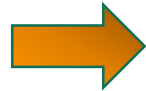
Individual
physicians,
chiropractors,
other
providers



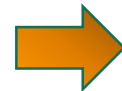
Physician
groups



Ambulances



CMS 1500



2 Non-Institutional Claims Files

Carrier
Durable Medical Equipment

What does a claim represent?

1 claim = multiple visits

- Multiple visits to same provider, for same purpose in short time period
 - Physical therapy
 - Dialysis
 - Home healthcare

1 claim = 1 visit

- Acute inpatient stay
- Office visits

Multiple claims = 1 visit

- Emergency room visits (hospital + physician)
- Long stays, especially in Skilled Nursing Facilities

Medicare Inpatient & Skilled Nursing Stays

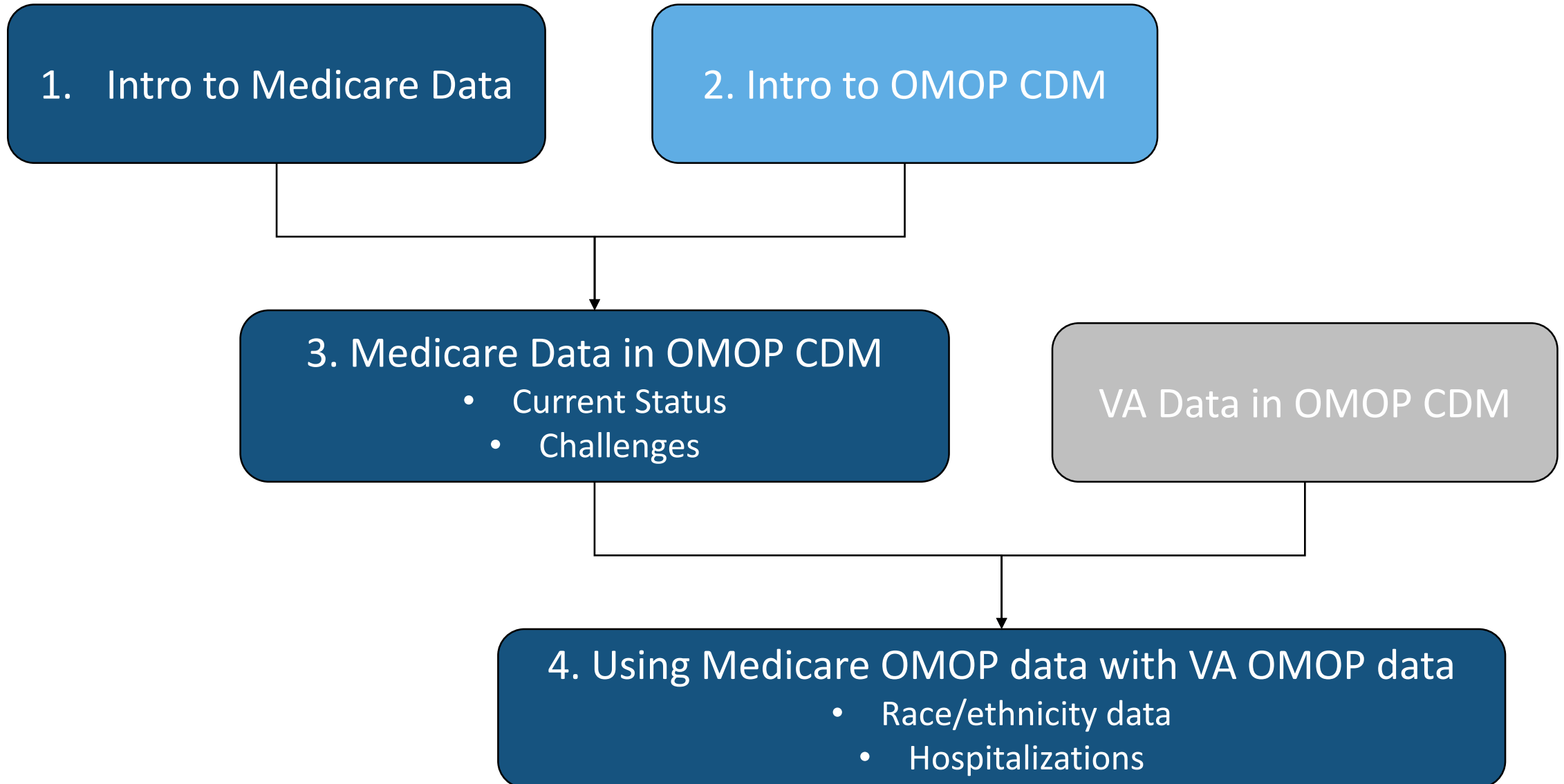
A stay may involve one or multiple claims

Stay	Claims	
Admit: July 10 Discharge: August 8	Claim 1	From July 10 Thru July 31
	Claim 2	From August 1 Thru August 8

Medicare Provider Analysis and Review (MedPAR)

- Created from Inpatient (IP) and Skilled Nursing (SN) claims
- Claims are “rolled up” to the stay level
- Doesn’t include all information from claims
- Includes some stays from Medicare Advantage enrollees

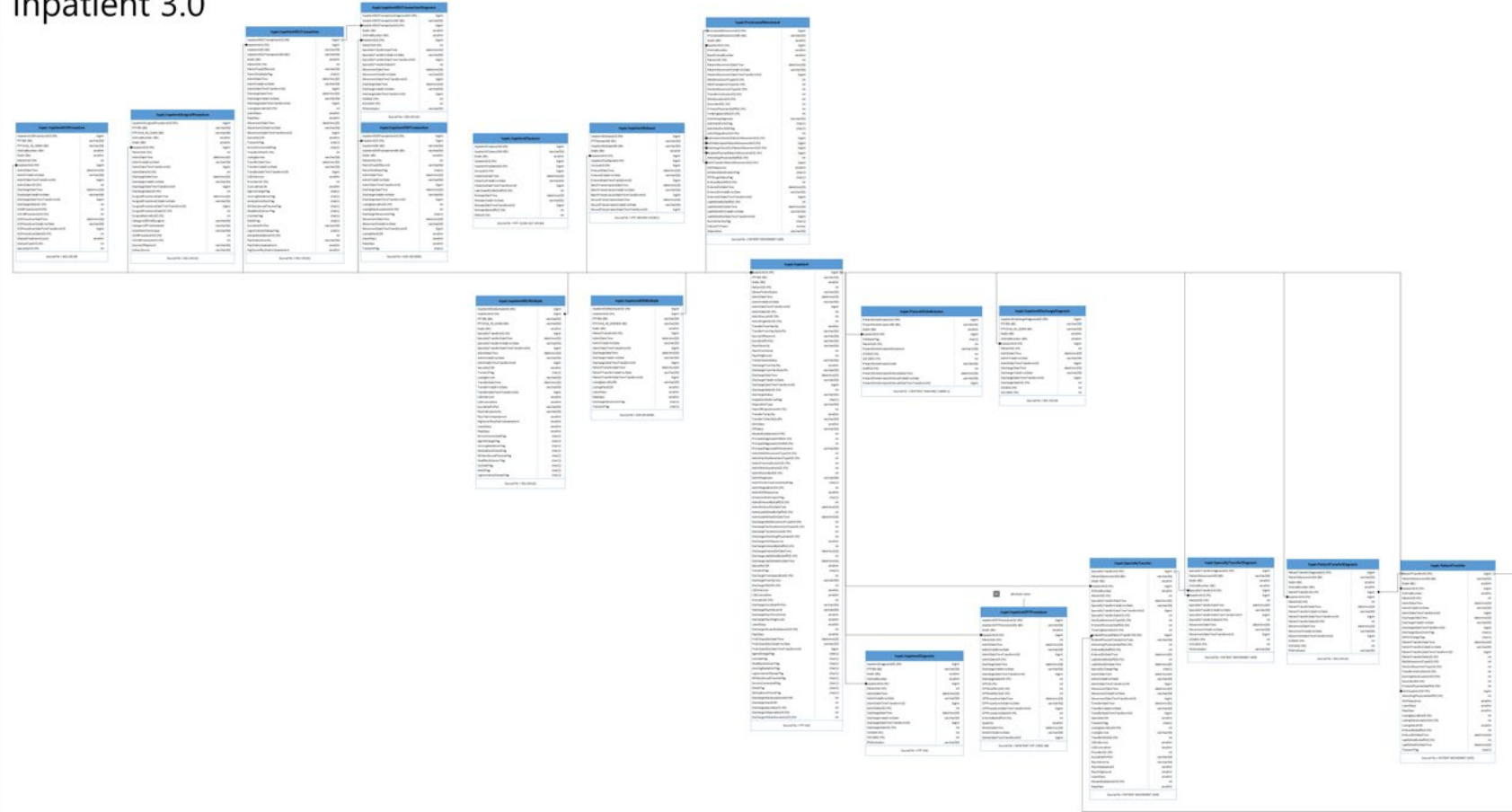
Outline for today's presentation



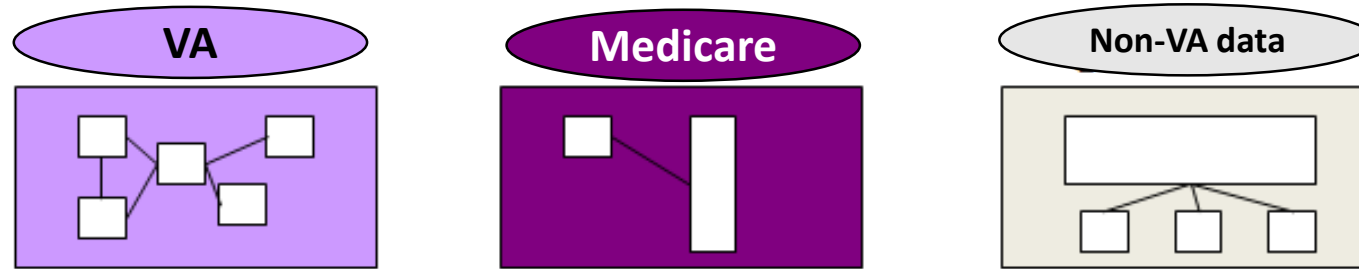
Example from VA's CDW

A data model organizes data elements and defines how they relate to one another

Inpatient 3.0

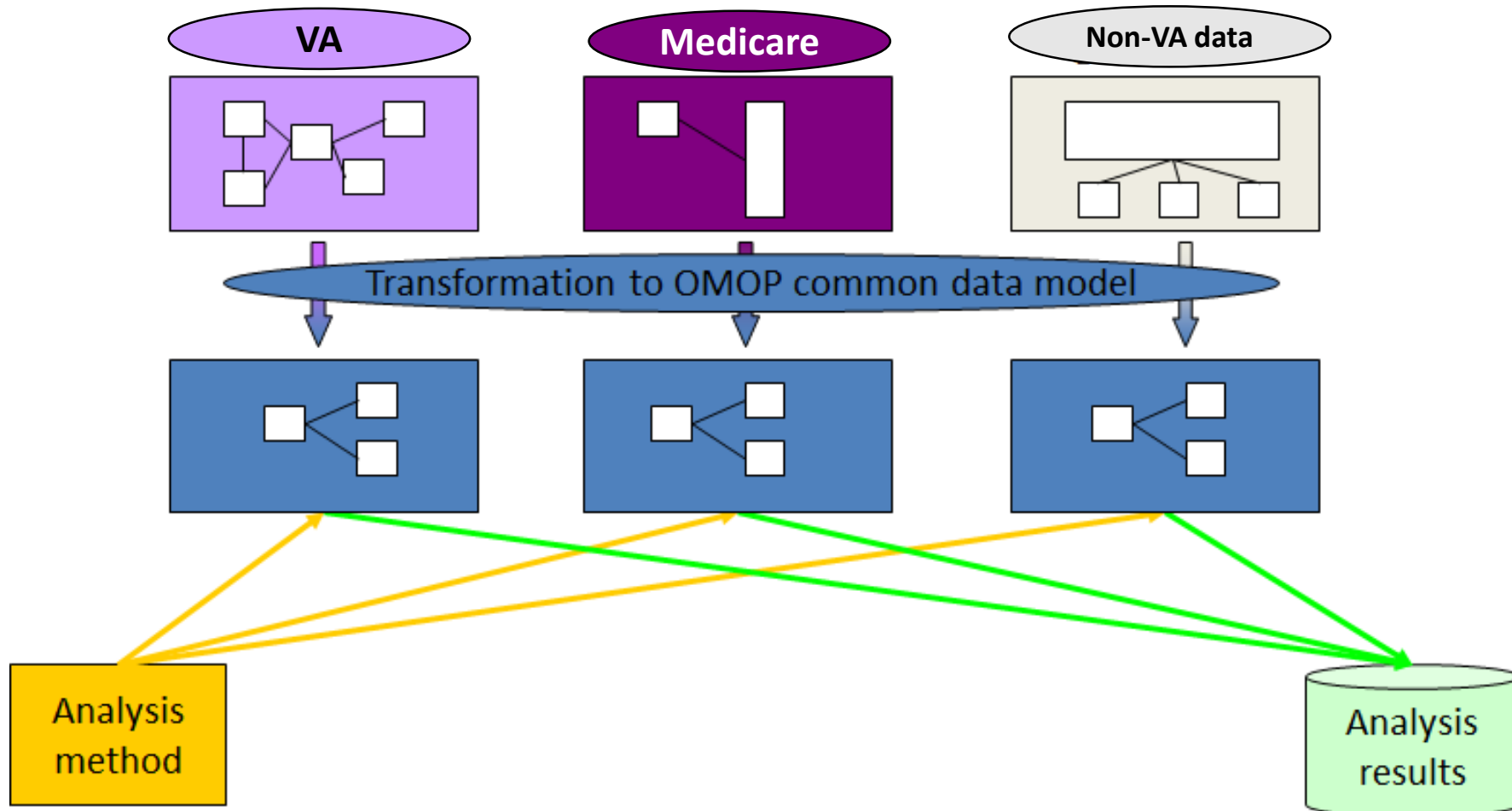


Why do we need a Common Data Model?



Source: <https://www.ohdsi.org/data-standardization/the-common-data-model/>

Benefits of Having Data in a Common Data Model



Source: <https://www.ohdsi.org/data-standardization/the-common-data-model/>

Why a Common Data Model (CDM)?

- Standardization of
 - Tables: names, contents, relationships
 - Variable/field names
 - Values
- Embed knowledge of data into the model

Why Observational Medical Outcomes Partnership (OMOP) CDM?

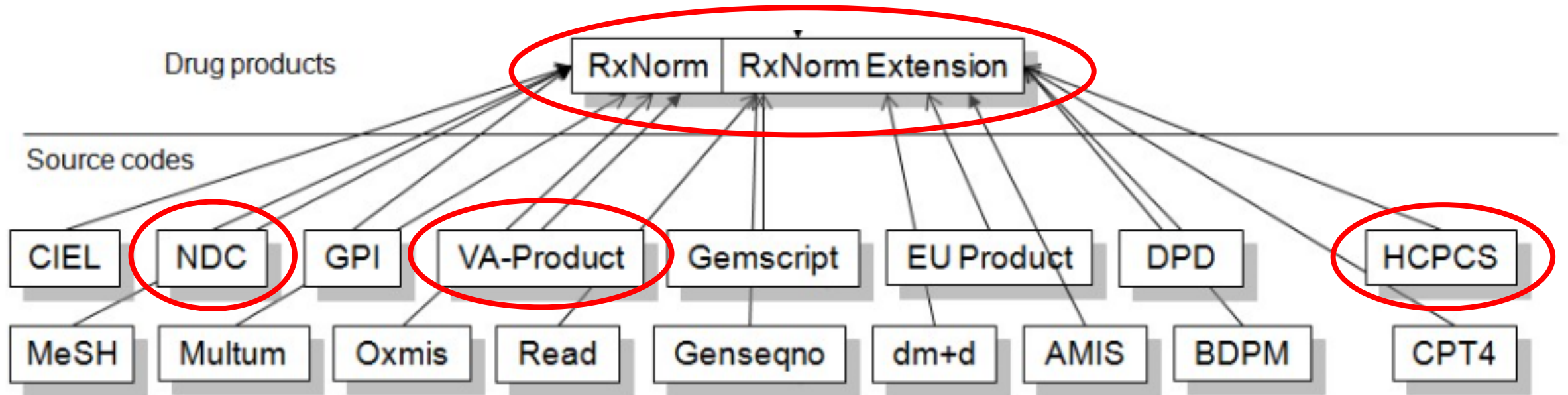
- Used by VA (CDW data) and DoD
- Benefits
 - Broad coverage
 - Open source tools
 - Standard vocabularies & concepts

OMOP Concepts

- All code sets needed to represent data
- Currently 6.3 million concepts
- All concepts are assigned a Concept ID
 - Number is not used to represent anything else
 - No decimals or special char

Concept ID	Source Code	Vocabulary	Description
44836914	250.00 (or 25000 in Medicare data)	ICD-9	Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled
2104055	25000	CPT	Incision, extensor tendon sheath, wrist

Standardizing Concepts




Search Vocabularies

Athena (web-based)

- athena.ohdsi.org

SQL tables

- Concept
- Concept Relationship
- ...and more



SEARCH BY KEYWORD

aspirin

SEARCH

DOWNLOAD

LOGIN

?

Condition x

Standard x

DOMAIN

STANDARD CONCEPT

CLASS

VOCABULARY

INVALID REASON

☐ Valid (168420)

☐ Invalid (0)

CLEAR FILTERS

DOWNLOAD RESULTS

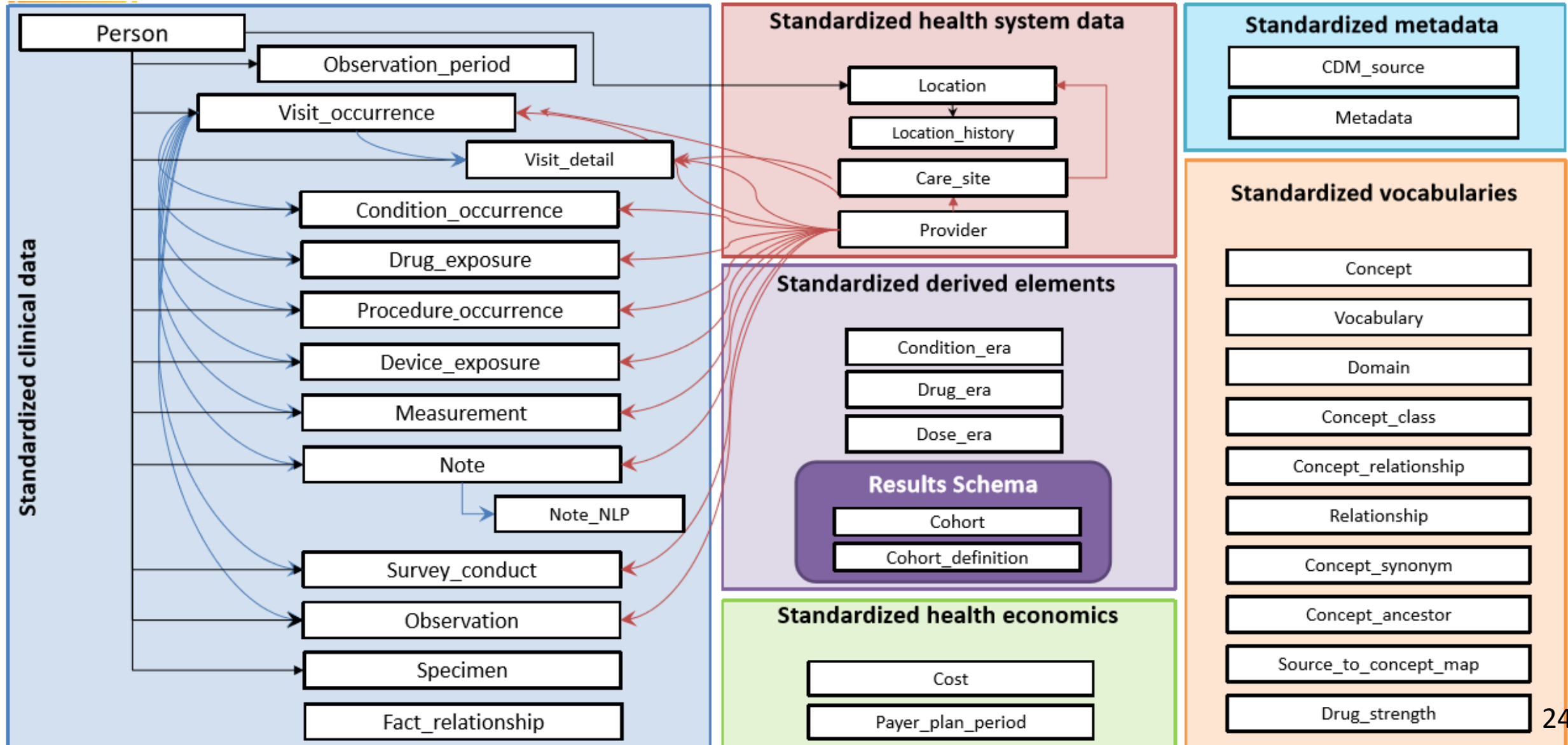
Show by 15 items

Total 168,420 items

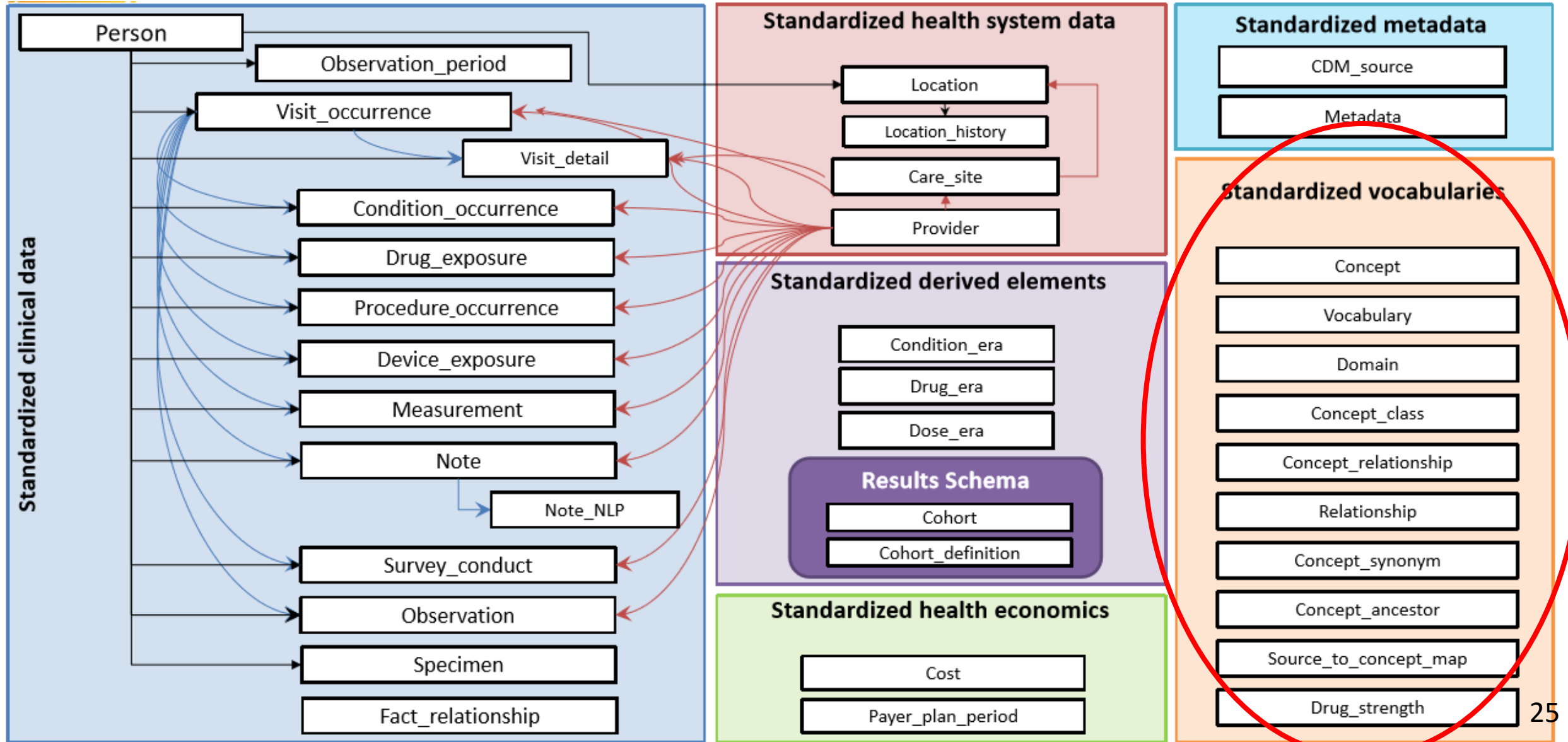
1 2 3 4 5 ... 11228 >

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
4122880	289381000	-1 level of presenting part in relation to ischial spines	Clinical Finding	Standard	Valid	Condition	SNOMED
4126418	289380004	-2 level of presenting part in relation to ischial spines	Clinical Finding	Standard	Valid	Condition	SNOMED
4128998	289379002	-3 level of presenting part in relation to ischial spines	Clinical Finding	Standard	Valid	Condition	SNOMED
4126417	289378005	-4 level of presenting part in relation to ischial spines	Clinical Finding	Standard	Valid	Condition	SNOMED
4122879	289377000	-5 level of presenting part in relation to ischial spines	Clinical Finding	Standard	Valid	Condition	SNOMED
4126419	289382007	0 level of presenting part in relation to ischial spines	Clinical Finding	Standard	Valid	Condition	SNOMED
4240068	369787008	0-5 mitoses per 10 HPF	Clinical Finding	Standard	Valid	Condition	SNOMED
4128999	289383002	1 level of presenting part in relation to ischial spines	Clinical Finding	Standard	Valid	Condition	SNOMED
4031867	129772004	1 o'clock position on mammogram	Clinical Finding	Standard	Valid	Condition	SNOMED
4288932	396447006	1 or more mitotic figure per mm2	Clinical Finding	Standard	Valid	Condition	SNOMED
4015559	170259009	1 year examination abnormal - for observation	Clinical Finding	Standard	Valid	Condition	SNOMED
4016214	170261000	1 year examination abnormal - on treatment	Clinical Finding	Standard	Valid	Condition	SNOMED
4014878	170260004	1 year examination abnormal - referred	Clinical Finding	Standard	Valid	Condition	SNOMED
4227224	420829009	1+ pitting edema	Clinical Finding	Standard	Valid	Condition	SNOMED
4306303	83170004	1,3 Indandion poisoning	Clinical Finding	Standard	Valid	Condition	SNOMED

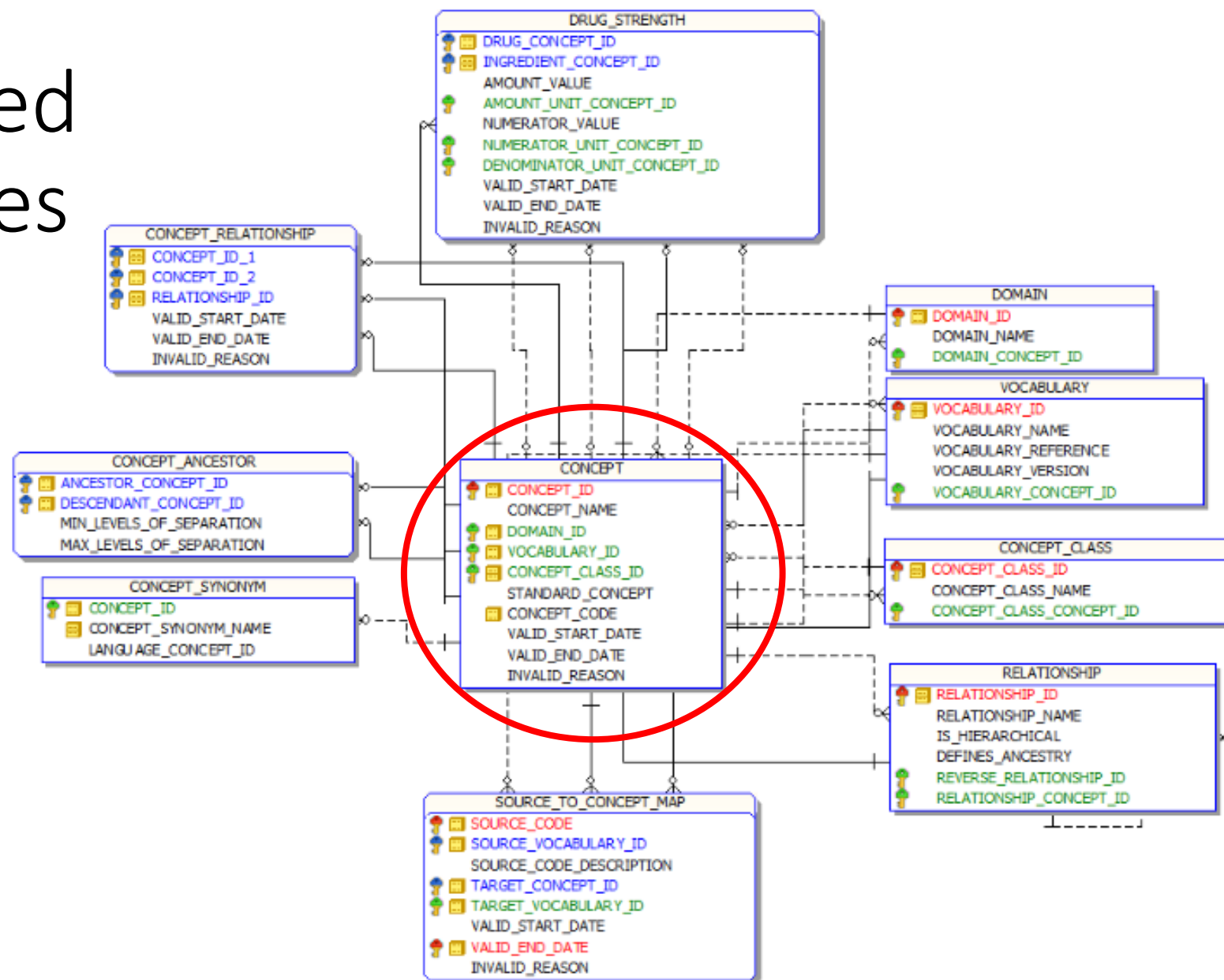
OMOP Common Data Model, Version 6



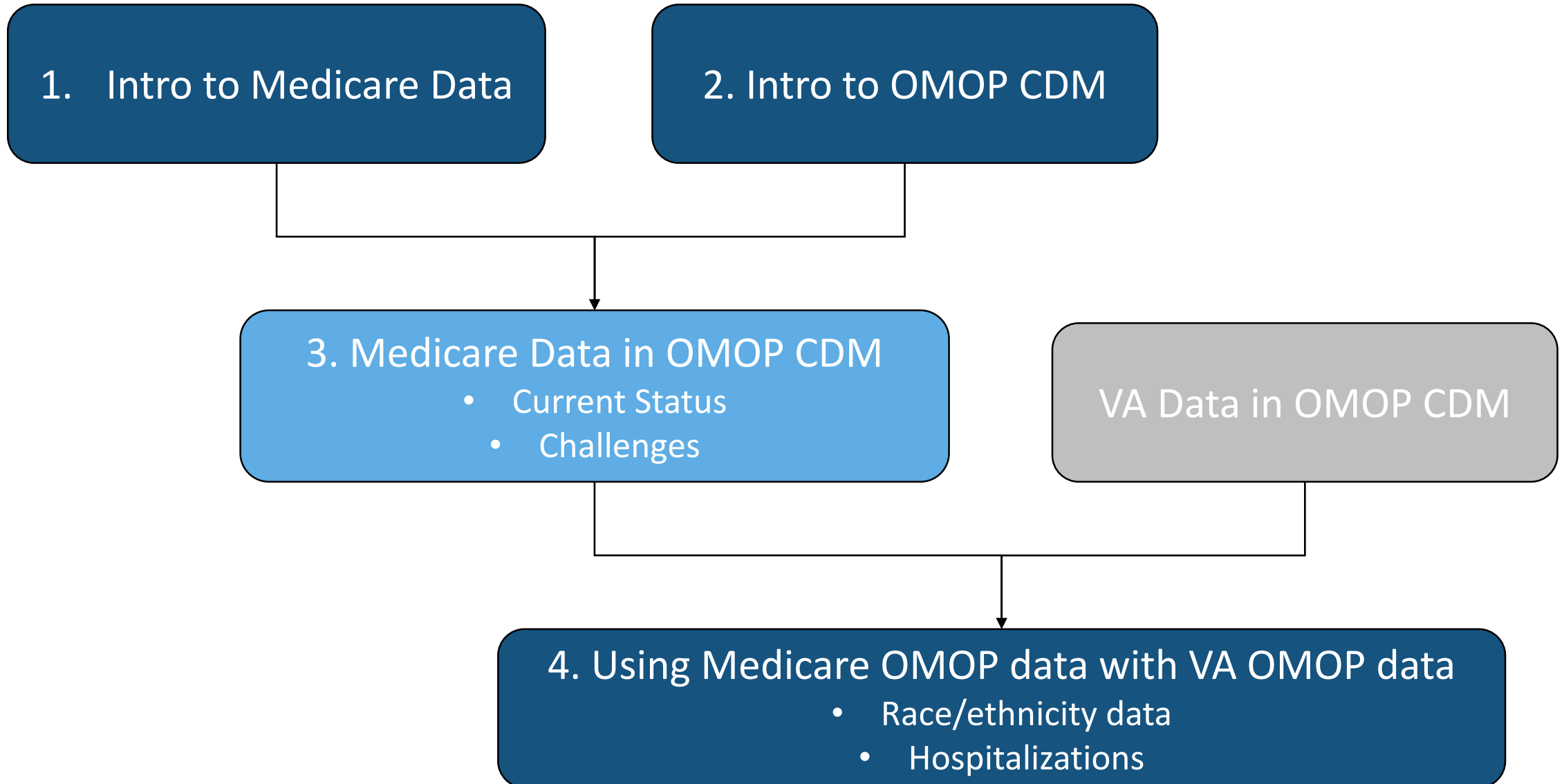
OMOP Common Data Model, Version 6



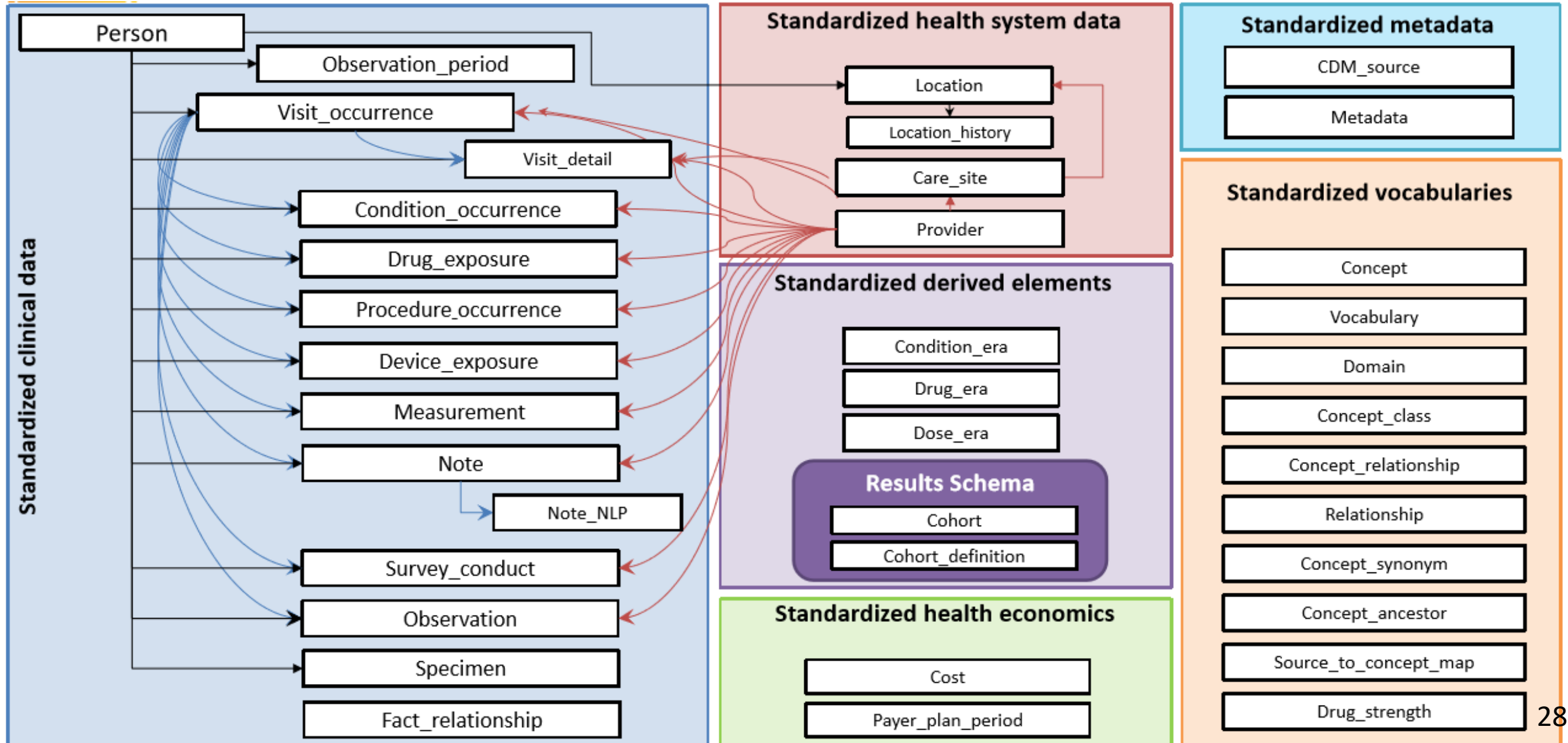
Standardized Vocabularies



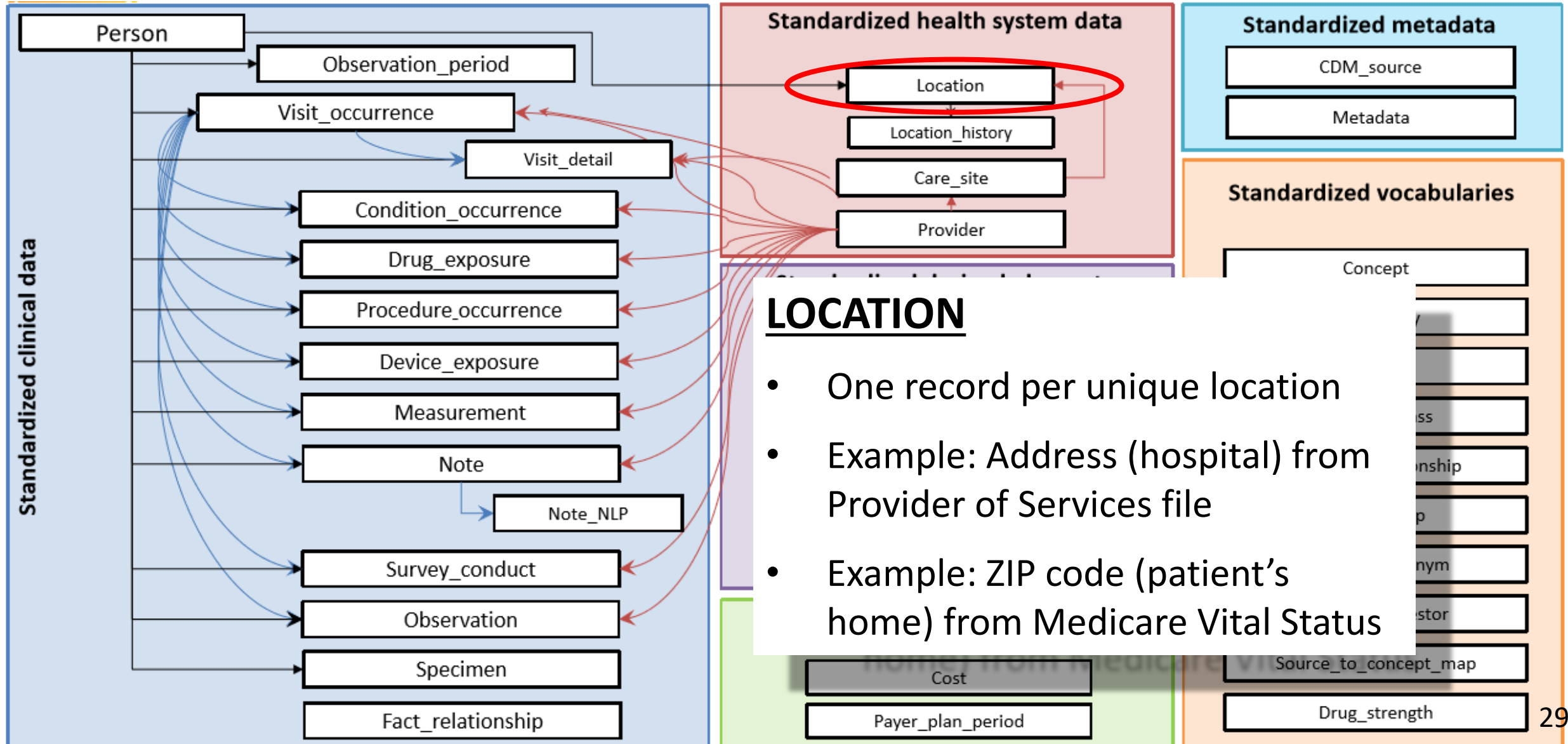
Outline for today's presentation



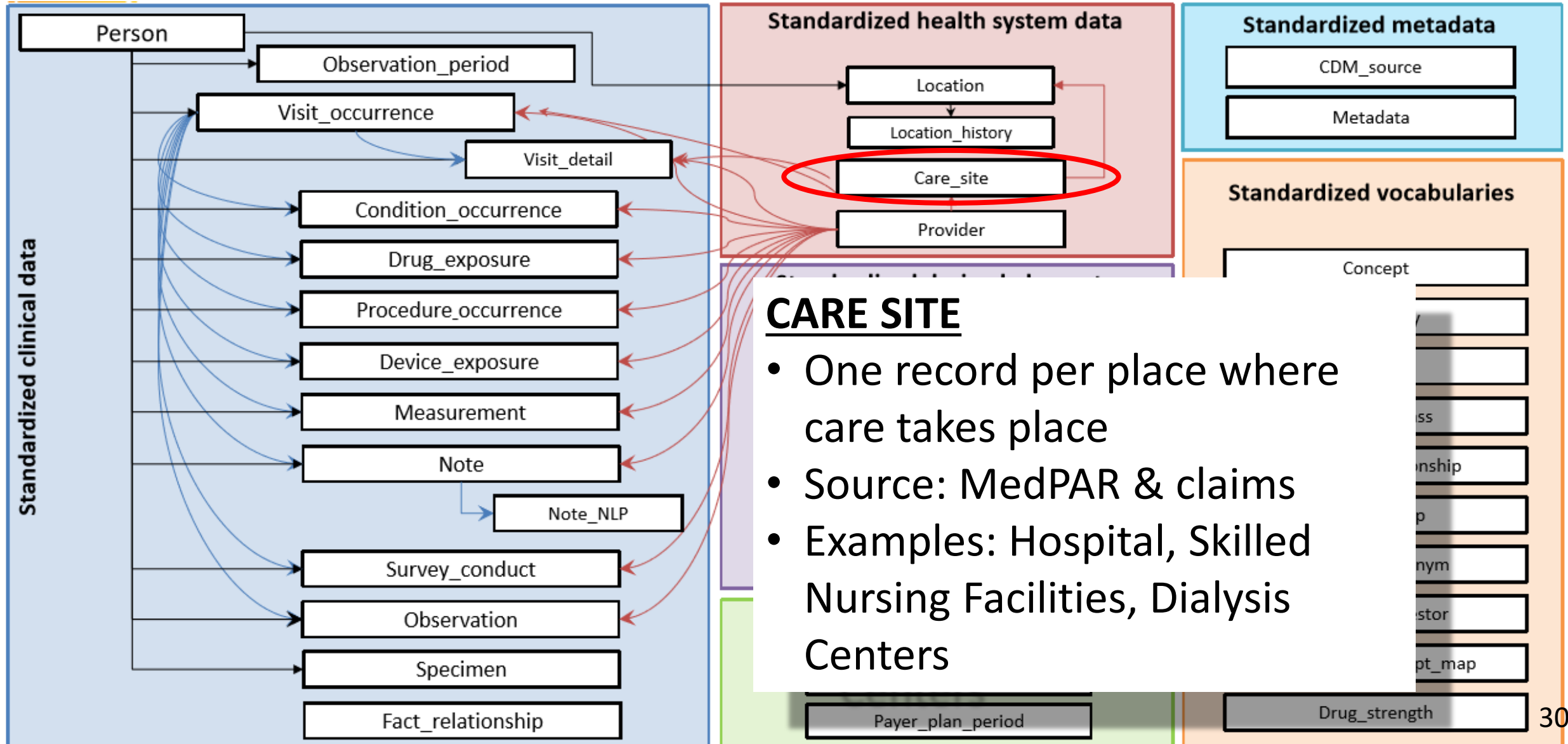
OMOP Common Data Model, Version 6



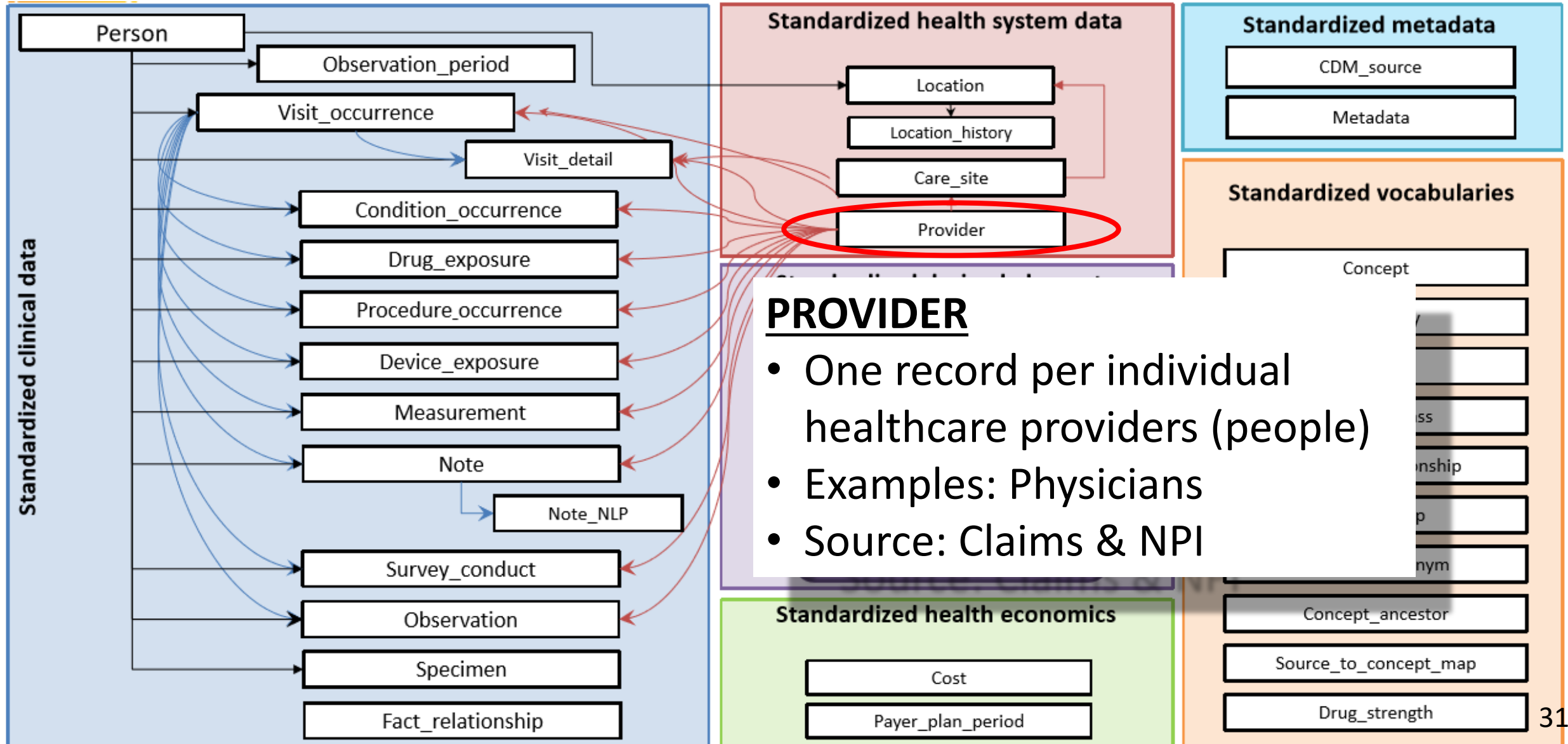
Medicare data in the OMOP Common Data Model



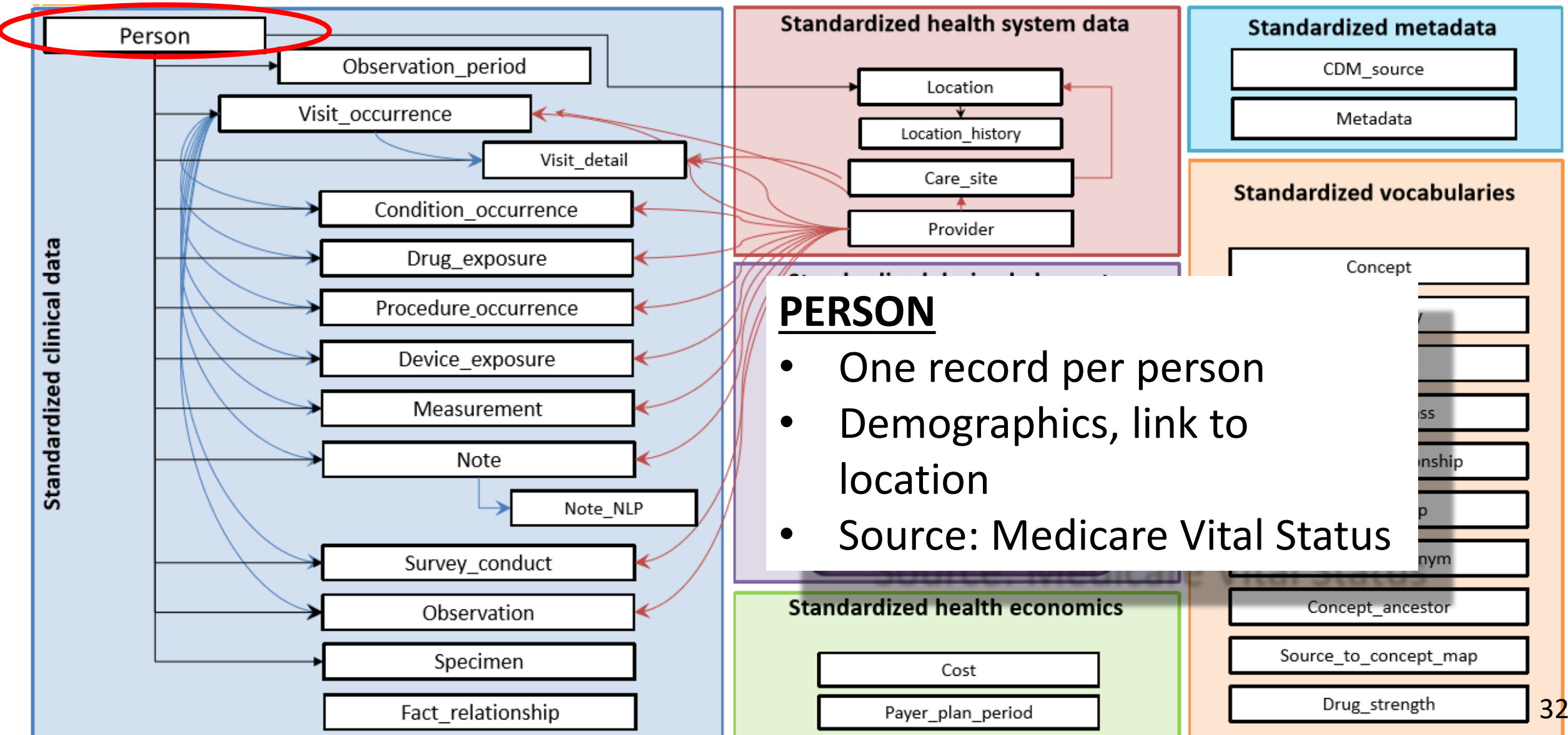
Medicare data in the OMOP Common Data Model



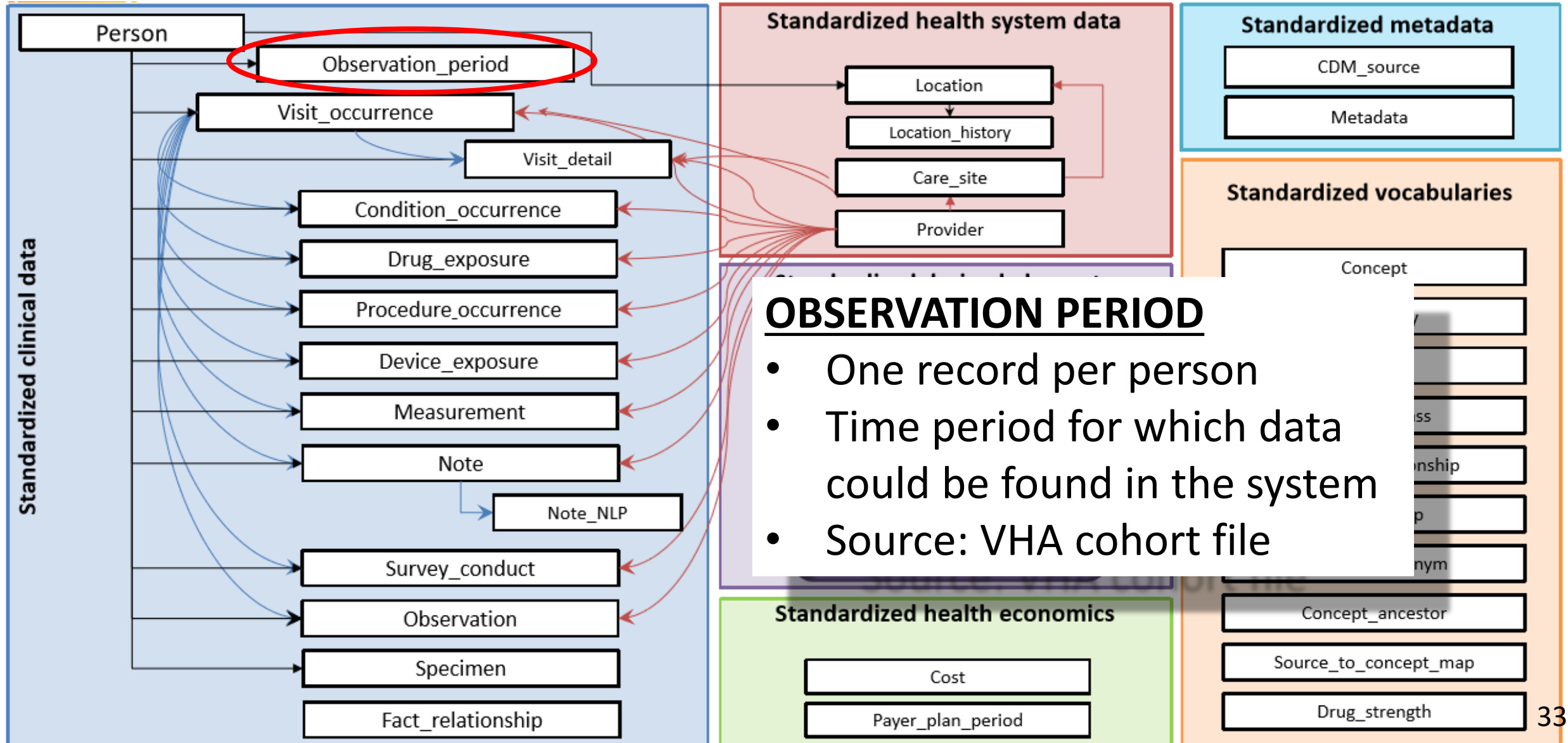
Medicare data in the OMOP Common Data Model



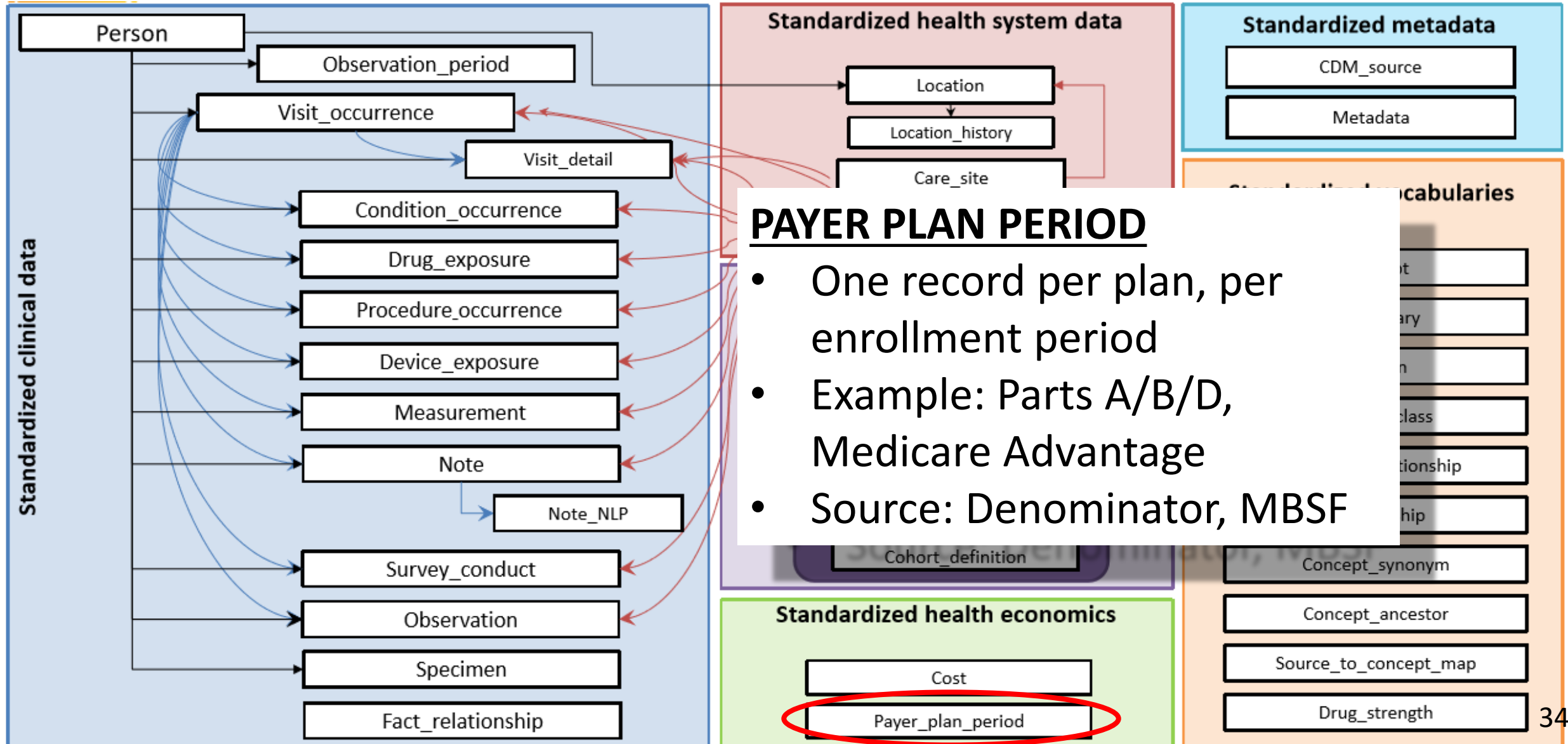
Medicare data in the OMOP Common Data Model



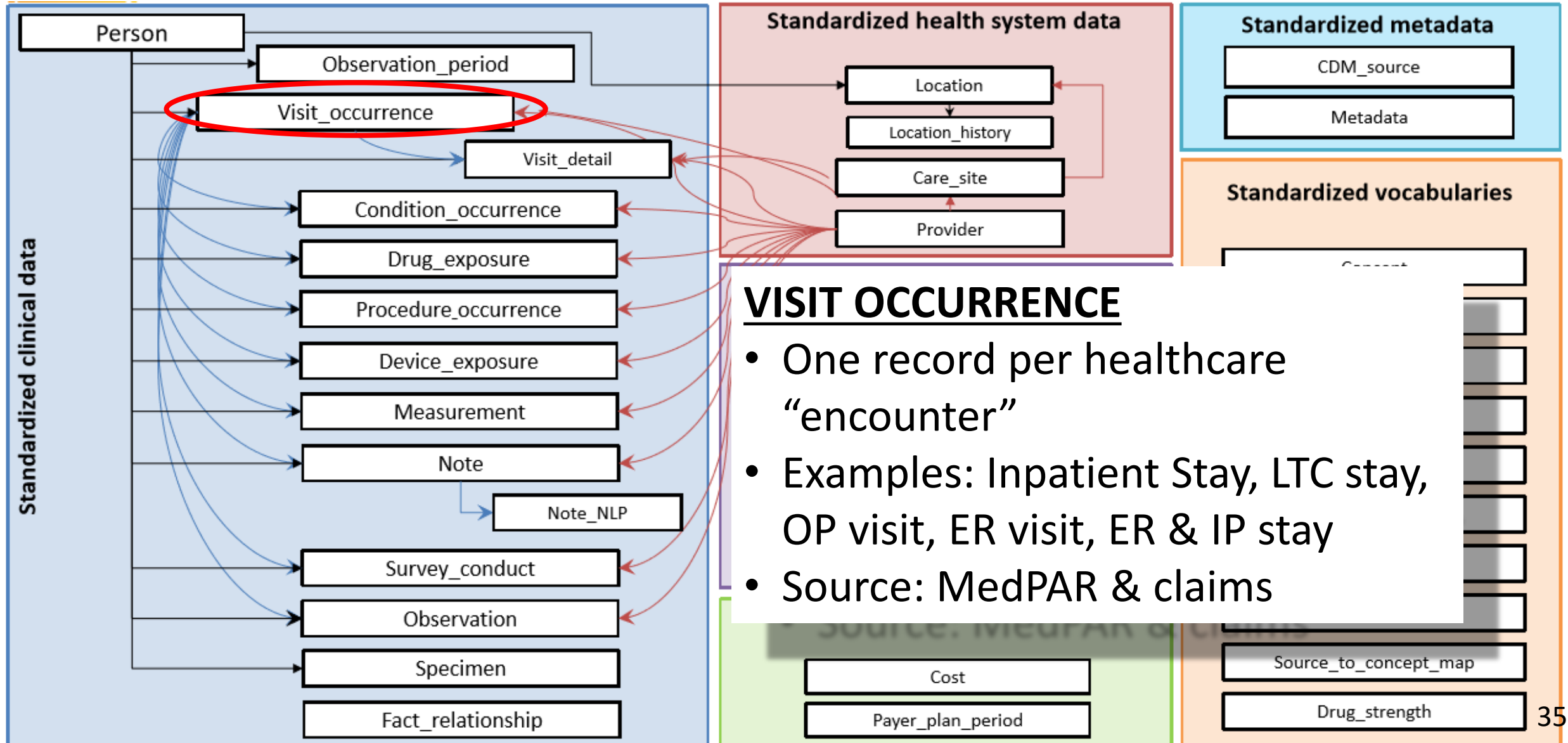
Medicare data in the OMOP Common Data Model



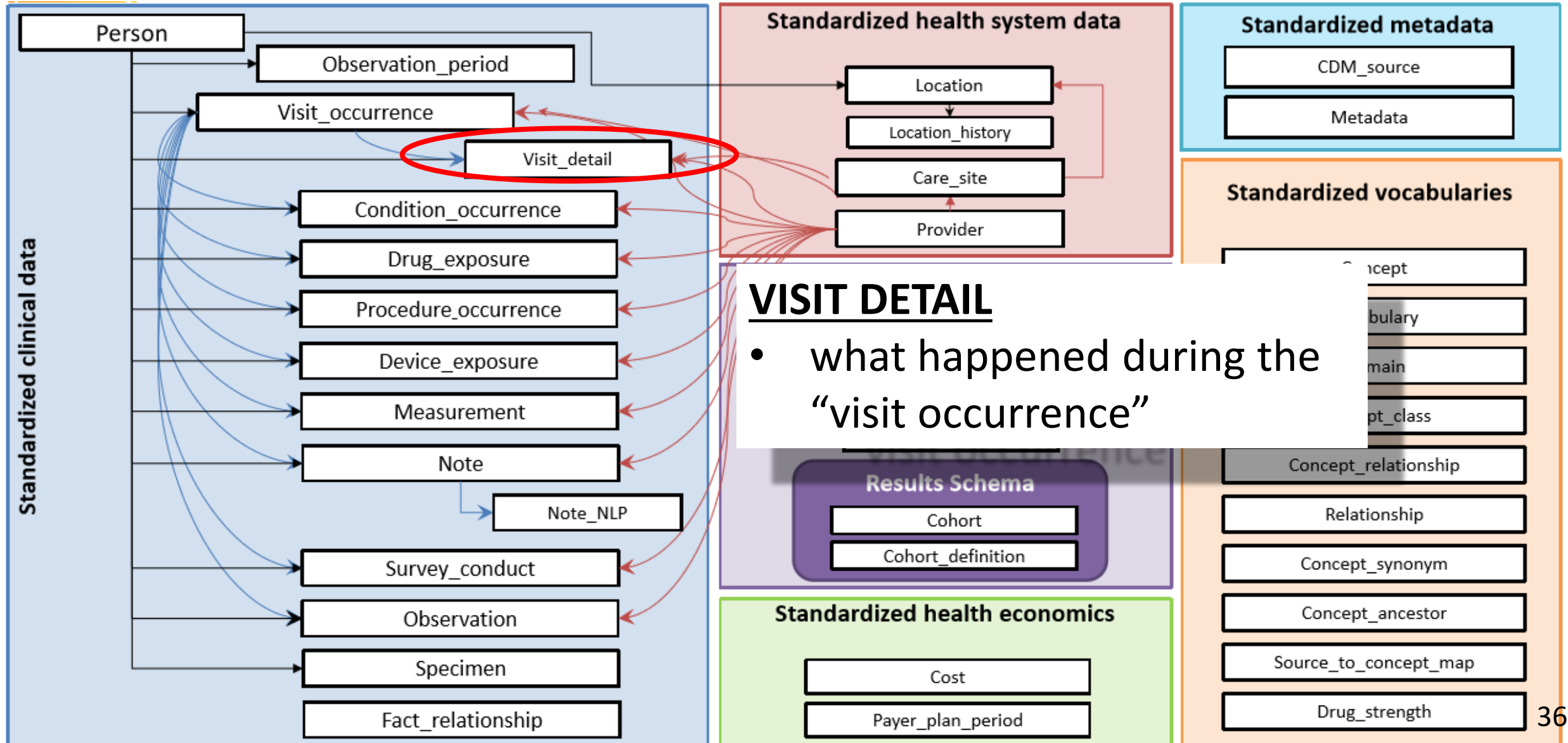
Medicare data in the OMOP Common Data Model



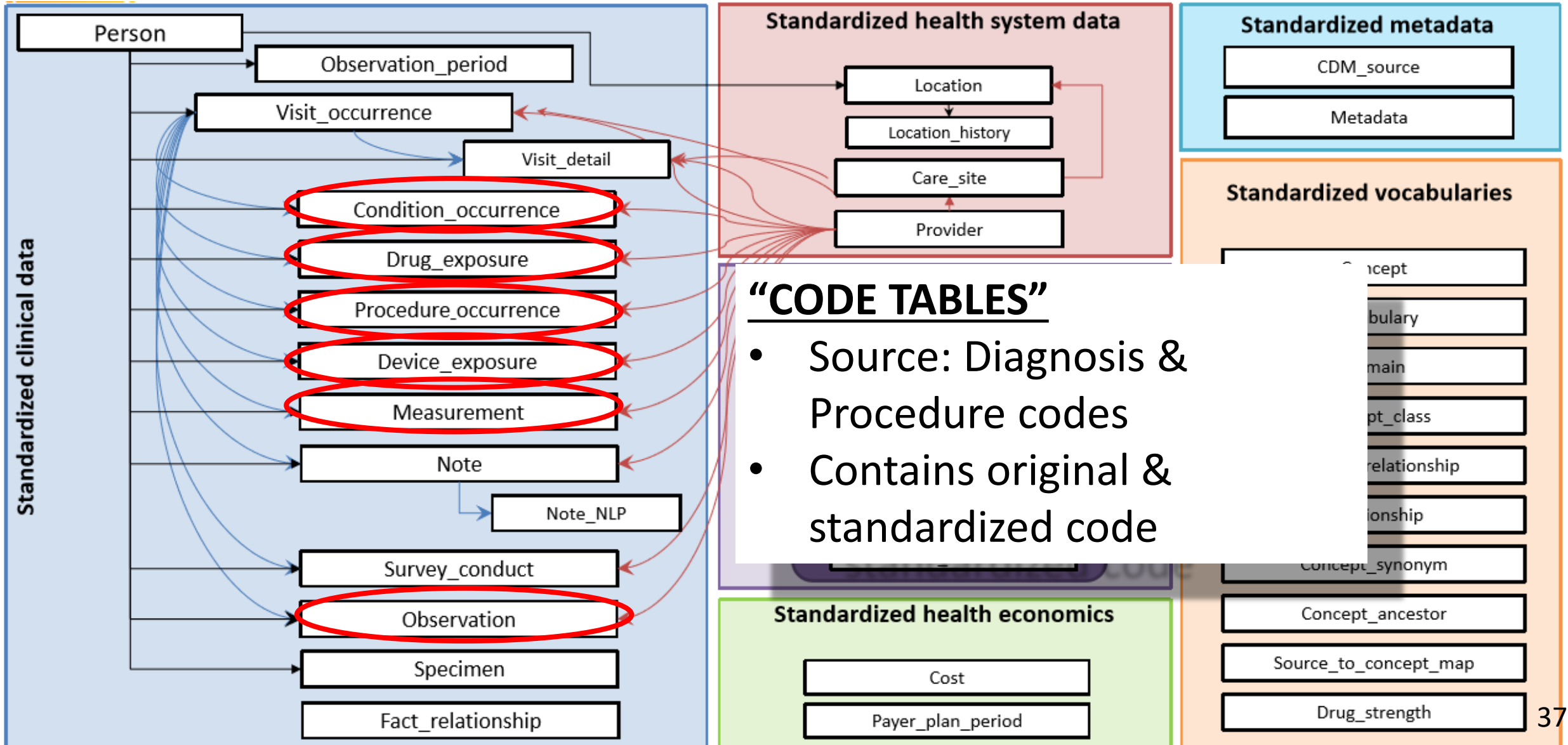
Medicare data in the OMOP Common Data Model



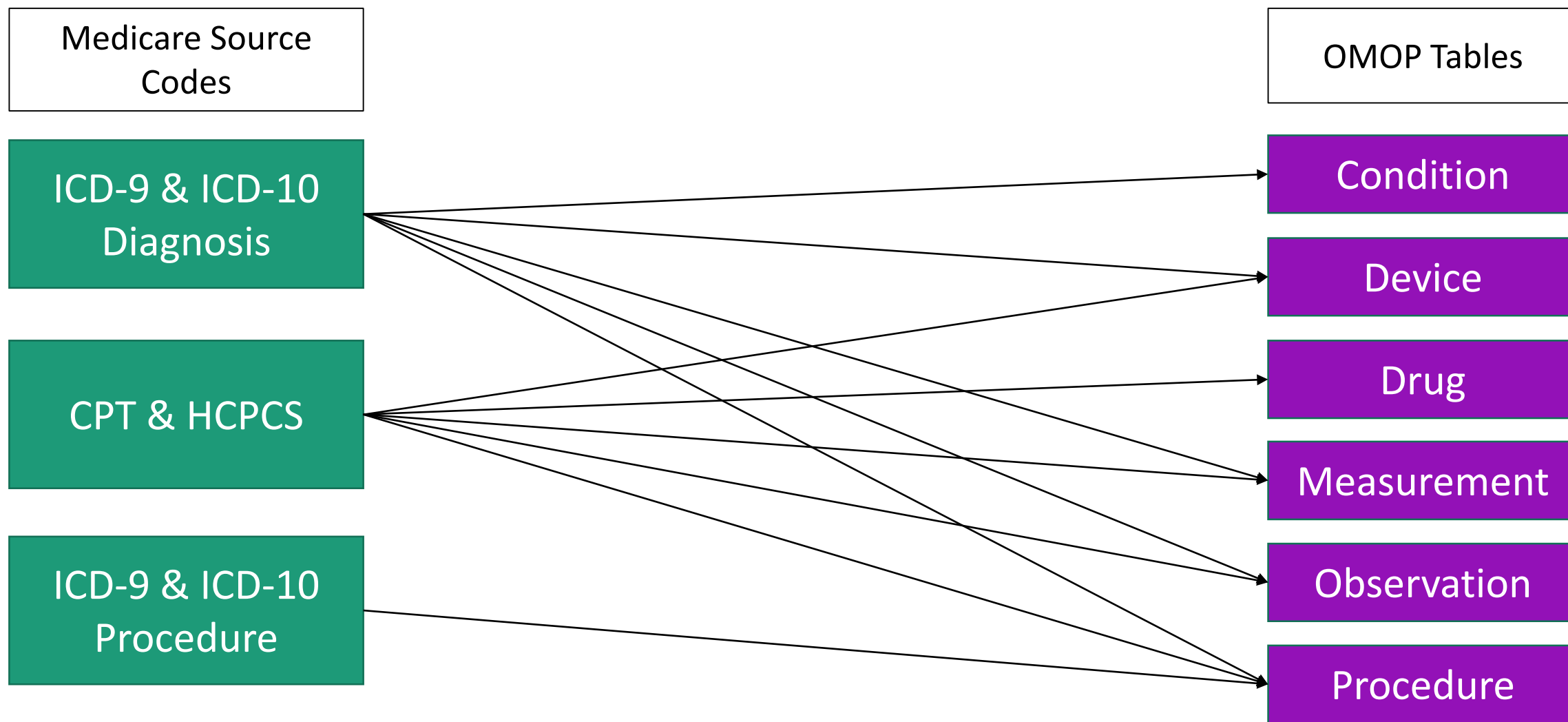
Medicare data in the OMOP Common Data Model



Medicare data in the OMOP Common Data Model



Organization of “Codes”



Most Frequent Codes

OMOP Table	Concept ID & Description	Value Concept ID & Description
Condition Occurrence	320128 (Essential hypertension)	
Device Exposure	2614966 (Syringe, with or without needle, each)	
Drug Exposure	35605564 (Epoetin Alfa Injection)	
Measurement	2212648 (Blood count; complete (CBC), automated (Hgb, Hct, RBC, WBC and platelet count) and automated differential WBC count)	0 (Missing; 98%) 4135493 (Abnormal)
Observation	4214956 (History of clinical finding in subject)	0 (Missing; 43%) 4148030 (Aortocoronary bypass grafting)
Procedure Occurrence	2108115 (Collection of venous blood by venipuncture)	

Current Status of Medicare OMOP in the VHA

Version 1

- Status
 - Currently in Alpha testing
 - Available for all research projects users in early 2020
- Includes
 - CY 1999-2017 Medicare data
 - Person, Observation Period, Payer Plan Period, Location, Provider
 - IP/SN only: Visit Occurrence, Condition Occurrence, Device Exposure, Drug Exposure, Measurement, Observation, Procedure Occurrence, Care Site

Version 1.1

- Status
 - Currently finishing internal testing
 - Alpha testing in early 2020
 - Available for all research projects in mid 2020
- Includes all data in version 1, plus
 - Outpatient (Institutional): Visit Occurrence, Condition Occurrence, Device Exposure, Drug Exposure, Measurement, Observation, Procedure Occurrence, Care Site

Record Counts	V1 (IP/SN)	V1.1 (Adds Inst OP)
PERSON	20,809,476	No change
OBSERVATION_PERIOD	20,809,476	No change
PAYER_PLAN_PERIOD	39,072,045	No change
VISIT_OCCURRENCE	35,753,931	293,530,600
VISIT_DETAIL	n/a	1,415,687,862
CONDITION_OCCURRENCE	374,205,869	1,041,369,563
DRUG_EXPOSURE	380,819	139,157,145
PROCEDURE_OCCURRENCE	58,253,760	731,417,551
DEVICE_EXPOSURE	114,256	52,907,596
MEASUREMENT	1,790,247	475,138,850
OBSERVATION	67,788,857	210,315,799
CARE_SITE	35,360	95,147
LOCATION	12,462,857	No change
PROVIDER	3,432,661	No change

Version 1.2 and beyond

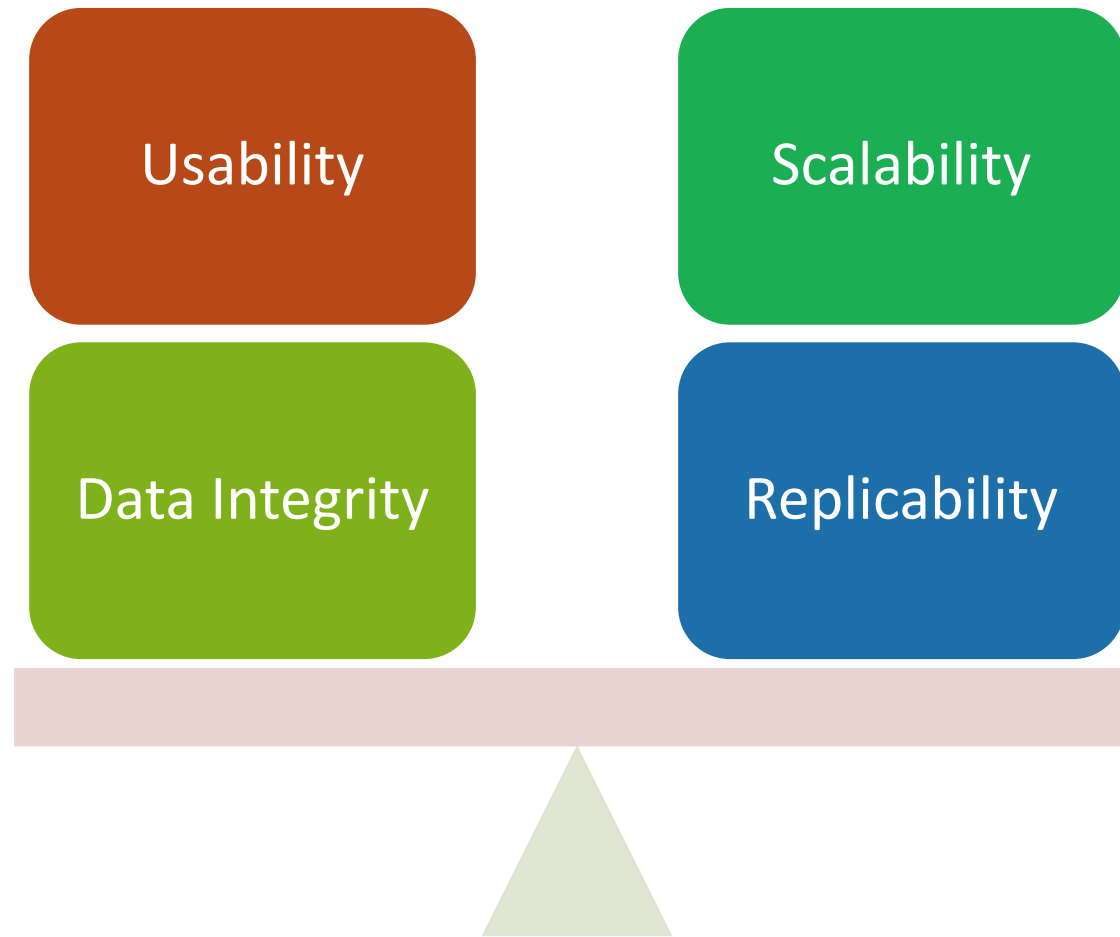
- Top priorities
 - CY2018 data
 - Carrier (physician) claims
 - Part D (prescription drug) events

Know what's in your data!

Medicare data in OMOP CDM
doesn't yet contain all visit types

Challenges when Transforming Medicare Data into OMOP CDM

Factors in decision making



Person ID

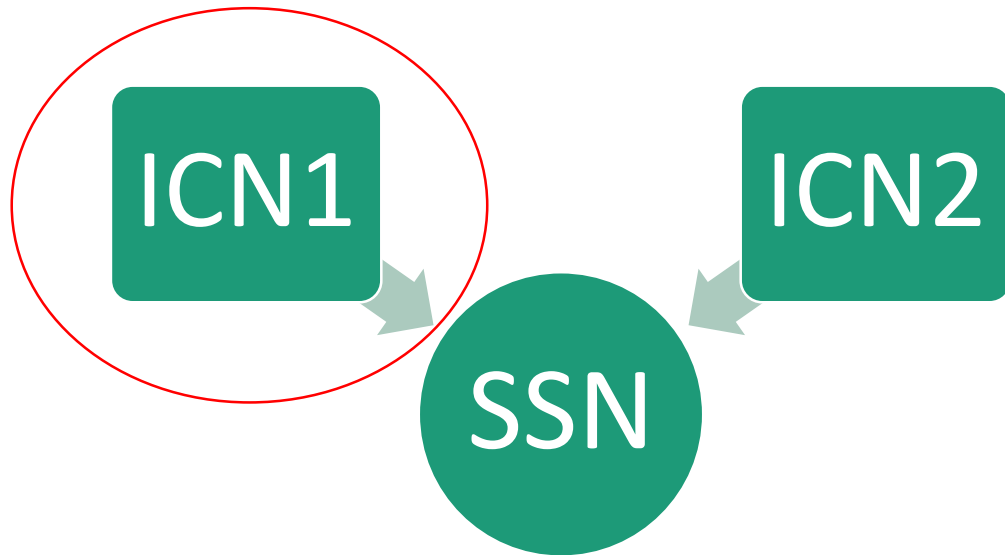
- Unique identifier
 - Medicare data uses SSN based identifier (real or scrambled)
 - CDW data uses ICN (and others)
 - OMOP recommends use of a randomly assigned Person ID
- Goal
 - Use the same Person ID in the Medicare OMOP data as is used in the VA OMOP & DoD OMOP data

Problem #1 - not all Veterans in Medicare data are found in CDW

- 1.7 million scrambled SSNs
 - Mostly VA compensation & pension recipients
- Solution: New Person IDs were created
 - Avoid overlaps with VA OMOP Person IDs
 - Start with a high number (50,000,000)
 - Reevaluate annually, to properly link data for Veterans new to VHA

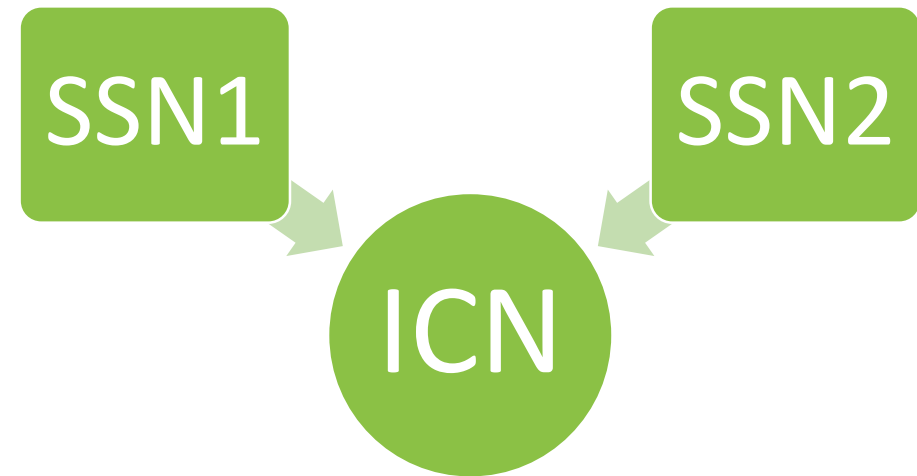
Problem #2 – scrambled SSN and ICN are not one-to-one

~70,000 scrambled SSNs were associated with 1+ ICN



Choose the smaller ICN

~300 ICNs were associated with 1+ scrambled SSN



Solution dependent on which (if any) SSN was found in the Medicare data

Should MedPAR or IP/SN claims be used to populate OMOP CDM?

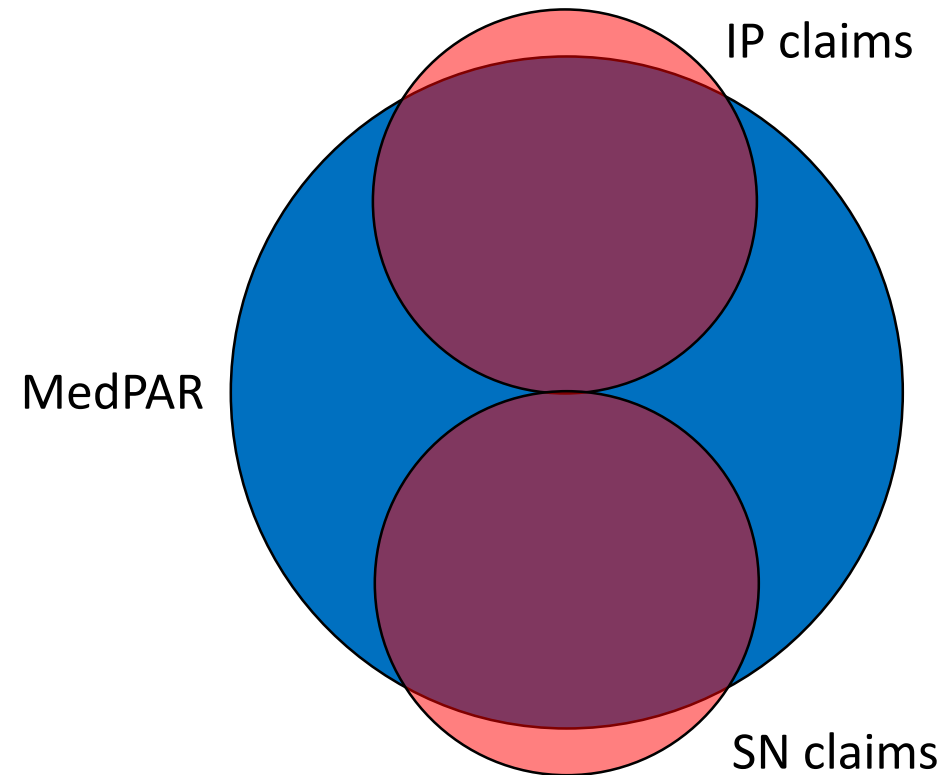
MedPAR

- Easier to use
- FFS claims & some MA encounters
- No physician info
- Diagnosis & procedure codes from last claim only

IP/SN claims

- Claims need to be converted to stays
- FFS claims only
- Includes physician's info
- All diagnosis & procedure codes

Relationship between MedPAR & IP/SN claims



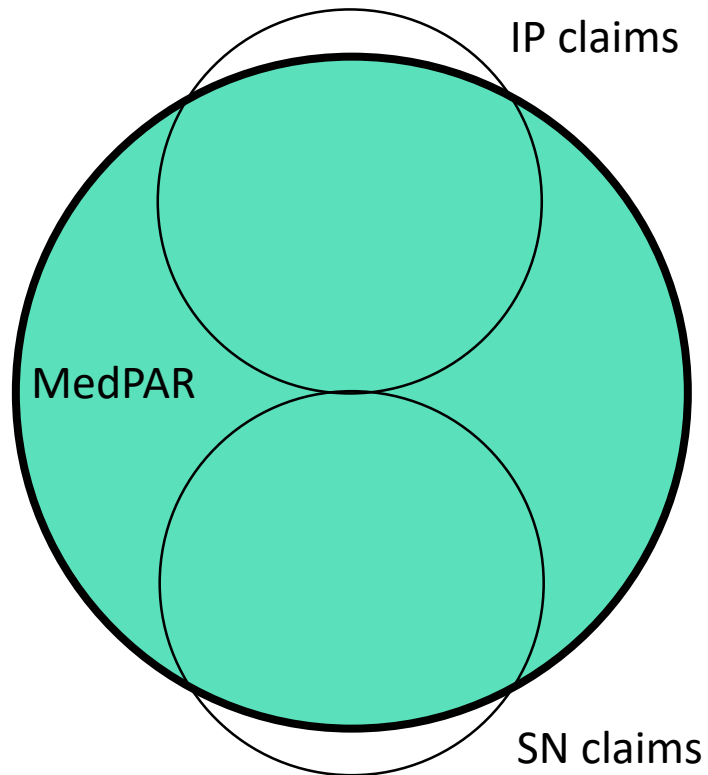
Decisions for transforming Medicare inpatient & skilled nursing data into OMOP CDM

Use MedPAR for
“fact of stay”,
including dates &
facility

Drop claims with
no associated
MedPAR record

Use claims to add
provider info,
diagnosis &
procedure codes

Stays included in Medicare OMOP CDM



IP claim only

112,659

IP claims with MedPAR stay

26,983,390

MedPAR stay only

3,482,382

SN claims with MedPAR stay

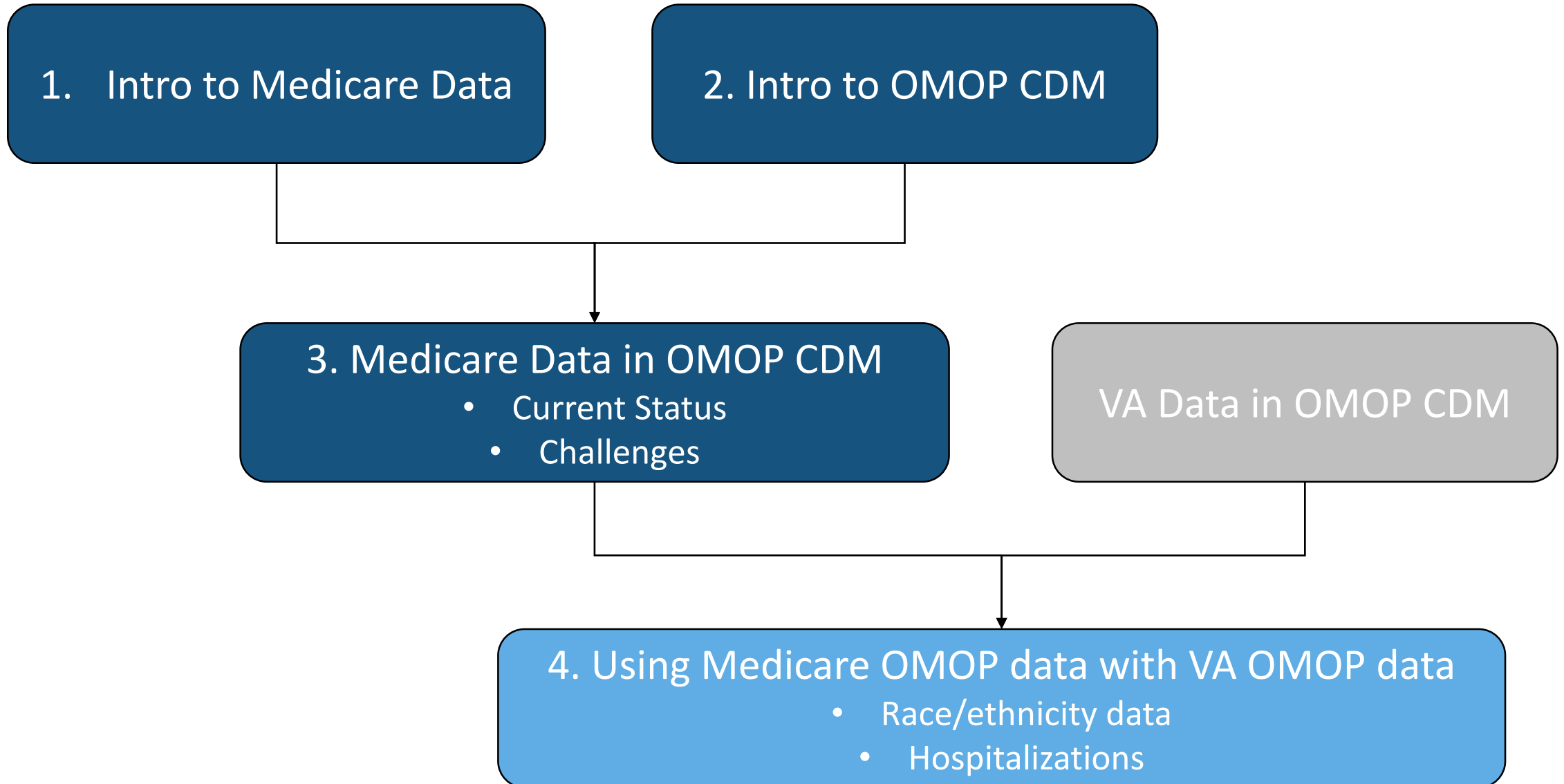
10,253,582

SN claim only

77,480

In OMOP

Outline for today's presentation



Example 1: Race/Ethnicity

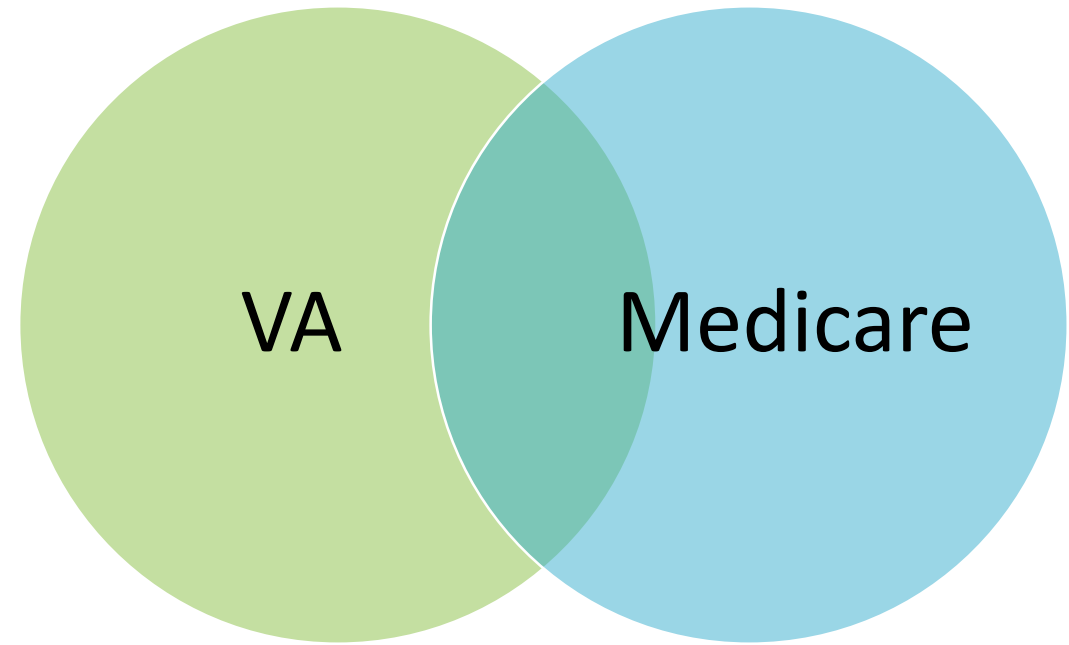
- Problem: Race/ethnicity is unknown for about half (51%) of Veterans in VA data
- How can Medicare OMOP data be used to easily “fill in” race & ethnicity data missing from VA OMOP data?

Most frequent combinations of race & ethnicity in VA OMOP

VA Race_Concept_ID	VA Ethnicity_Concept_ID	Count (%)
8552 (Race unknown)	8552 (Race Unknown)	11,564,081 (49%)
8527 (White)	38003564 (Not Hispanic)	8,304,394 (35%)
8516 (Black or AA)	38003564 (Not Hispanic)	1,707,689 (7.2%)
8552 (Race unknown)	38003564 (Not Hispanic)	543,205 (2.3%)
8527 (White)	38003563 (Hispanic)	504,737 (2.1%)
8527 (White)	8552 (Race Unknown)	462,887 (1.9%)

Methods

- Join VA OMOP Person table + Medicare OMOP Person table on Person ID
- Keep RACE_CONCEPT_ID & ETHNICITY_CONCEPT_ID from each table
- Resulting table has 1 record per person, but contains fields from both source tables



Most frequent combinations of race & ethnicity in combined VA OMOP + Medicare OMOP

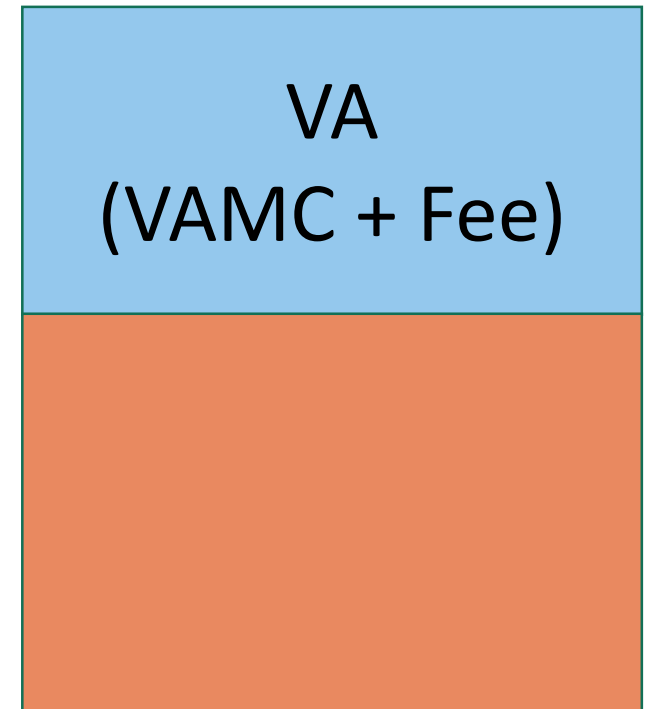
VA Race Concept ID	Medicare Race Concept ID	VA Ethnicity Concept ID	Medicare Ethnicity Concept ID	Count (%)
8552 (Race unk)		8552 (Race unk)		6,944,729 (29%)
8527 (White)	8527 (White)	38003564 (Not Hispanic)	0 (No data)	5,451,618 (23%)
8552 (Race unk)	8527 (White)	8552 (Race unk)	0 (No data)	3,864,426 (16%)
8527 (White)		38003564 (Not Hispanic)		2,683,100 (11%)
8516 (Black/AA)		38003564 (Not Hispanic)		900,847 (3.8%)
8516 (Black/AA)	8516 (Black/AA)	38003564 (Not Hispanic)	0 (No data)	782,728 (3.3%)
8552 (Race unk)	8516 (Black/AA)	8552 (Race unk)	0 (No data)	482,364 (2.0%)
8552 (Race unk)	8527 (White)	38003564 (Not Hispanic)	0 (No data)	293,632 (1.2%)
8527 (White)		38003563 (Hispanic)		266,120 (1.1%)
8527 (White)		8552 (Race unk)		253,736 (1.1%)

Example 2: Hospitalizations

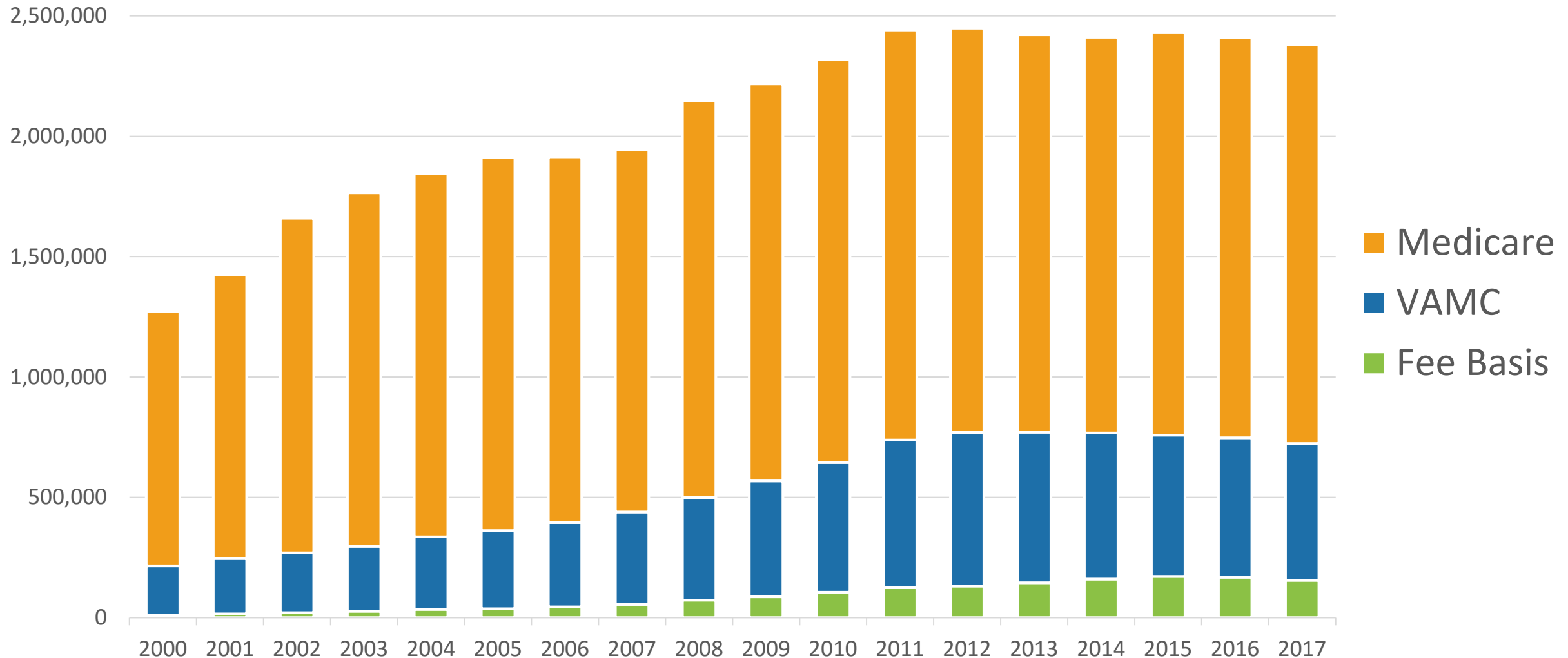
- Only using VA data doesn't provide complete picture of healthcare utilization
- Veterans may be hospitalized in
 - VA medical center (VAMC)
 - Non-VA hospital, paid for by VA (VA Fee Basis)
 - Non-VA hospital, paid for by Medicare
- Among Veterans, how many hospitalizations are in a VAMC, thru VA Fee basis, or thru Medicare?

Methods

- Union Visit Occurrence tables
 - Resulting table has same number of fields as source tables, but contains all records
- Where Visit Concept ID=
 - 9201 (Inpatient Visit)
 - 262 (Emergency Room + Inpatient Visit)
- From 1/1/2000 to 12/31/17
 - VA current range: 10/1/99 – 5/20/18
 - Medicare current range: 1/1/99 – 12/31/17
- Patients found in CDW
 - Person ID < 50,000,000



Payer/Location of Veterans' Hospitalizations, CY00-CY17



Resources


The VA/CMS Data for Research Project is based at VIREC, the data steward for CMS data used for VA research.

- Distributing data from Centers for Medicare & Medicaid Services (CMS) to VA approved projects since 2003
- Providing assistance to VA research projects using CMS data
- No cost to VA researchers

<https://vaww.virec.research.va.gov/Index-VACMS.htm>
(VA intranet only)



Requesting Medicare data (including OMOP)



INTRANET

Search All VA Web Pages

Search

Open Advanced Search

VA INFORMATION RESOURCE CENTER (VIREC)

VIREC Home

VA/CMS Home

About Us

Updates

Publications

Education

FAQs

Help

VA/CMS Data Standard Request Packet

Overview

This page outlines the documents and approvals required for the [standard request](#) process for both initial and subsequent requests for VA/CMS data.

Projects that...	Use
Are requesting VA/CMS data for the first time	Initial Request Packet
Have already submitted an initial request packet and are requesting additional data	Subsequent Request Packet

Initial Request Packet

The Initial Request Packet is required for projects requesting access to VA/CMS data for the first time. When projects are engaged in research at multiple sites, each project site that will access the VA/CMS data is required to submit the Initial Request Packet. All projects must submit the following:

- [Project Information and Authorization Form](#) signed by the Principal Investigator (PI) and approved by the PI's supervisor and the local facility Associate Chief of Staff for Research (ACOS-R).
- [Data Security Compliance Form](#) signed by the PI and the Information Security Officer (ISO) at the VHA facility where research is being conducted, and if different, the ISO at the VHA facility where the VA/CMS data will be stored. If the data will be stored on VINCI, the ISO's signature is not required.
- [VA/CMS Rules of Behavior \(ROB\) Agreement](#) signed by the PI and each project staff member who will have access to the VA/CMS data
- Data Description Form(s) pre-reviewed by VIREC then signed by the PI and approved by the local Privacy Officer (PO) or central IRB Privacy Officer. Projects may submit one or more of the following data description forms for their project:
 - [VA/CMS Data for Veterans Only](#)
 - [VA/CMS Data for Veterans and Non-Veterans](#)
 - [VA/CMS Non-Repository Data](#)

General Resources

VA/CMS Repository Data


Non-Repository Data

Provider Data


Cohorts & Identifiers

Requests

Current Data Users



VA/CMS DATA FOR RESEARCH



DATA DESCRIPTION

CMS/USRDS Data for Veterans Only

Instructions:

- This form is used to request data from the Centers for Medicare and Medicaid Services (CMS) and United States Renal Data Systems (USRDS) for use in VA research.
- This form should be used when a project is requesting VA/CMS Repository Data for the VHA cohort.
- A new form should be used for each study cohort/finder file requested.
- Researchers are encouraged to arrange a [pre-request consultation](#) for assistance in completing this form.
- This form must be reviewed by VIREC prior to obtaining other signatures. The form can be sent to VIREC at virec.vacmsdata@va.gov.
- Both wet and electronic signatures are accepted.

Section 1: General Information

VIREC Project ID (ex: Jones-01) (if known)

Principal Investigator

Project Title

Facility Name

City

State

Are you requesting data to be delivered to: (select only one)

☐ VINCI SAS Grid. Provide name of VINCI workspace:

☐ Non-VINCI server. Please provide:

Name of server:

DART ID, to facilitate data transfer:

How many subjects has the IRB approved for inclusion in this project?

(Numbers only)

Please note that:

- An IRB amendment and Privacy Officer approval will be required prior to data release if the number of subjects included in either the finder file or the data requested is larger than the number of subjects listed here.
- Projects requesting a large number of subjects will be asked to provide justification.

Contact VIREC prior to completing this form if this request is for approval to use CMS or USRDS data that has already been provided to:

- Another research project or research data repository
- A VHA program office or VHA operations project

<https://vaww.virec.research.va.gov/VACMS/Requests/Standard-Packet.htm> (VA Intranet only)

64

Learn more about CMS & Medicare data

VIReC (links are VA intranet only)

- Medicare - <https://vaww.virec.research.va.gov/VACMS/Medicare/Data.htm>
- CMS data available from VIReC - <https://vaww.virec.research.va.gov/VACMS/Intro/Data-Available.htm>

External support for CMS data users

- ResDAC - <https://www.resdac.org/>
- CCW - <https://www2.ccwdata.org/>

Learn more about OMOP CDM

Observational Health Data Sciences and Informatics (OHDSI)

- <https://www.ohdsi.org/>

CDM documentation

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

Athena – search standardized vocabularies

- <http://athena.ohdsi.org/>

Learn more about VA OMOP

VINCI OMOP Academy

- <https://vhacdwwweb02.vha.med.va.gov/prod/vincipedia/VINCIPedia/OMOP%20Academy.aspx> (VA intranet only)

VA Pulse - VINCI OMOP User Group

- <https://www.vapulse.net/groups/vinci-omop-users-group> (sign-in required)

Contact information

Kristin de Groot, MPH

Technical Director

kristin.degroot@va.gov

VA Information Resource Center

Hines VA Hospital

virec@va.gov

708-202-2413

