Getting Started with
VA OMOP

@DuVallScott

Poll #1: Your role as a data user

What is your role in research and/or quality improvement?

- Research investigator

- Methodologist

- Data manager, analyst, or programmer

- Project coordinator

- Other – please describe via the Q&A function

Poll #2: Your experience with CDW data

Rate your level of experience with CDW data on a scale of 1 to 5...

- 1. Not worked with it at all
- 2. Have minimal experience with it
- 3. Have work closely with it for <6 months
- 4. Have worked closely with it for 6 months to 2 years
- 5. Very experienced with CDW

Poll #3: Your experience with OMOP data

How familiar are you with OMOP?

- 1. I currently use it
- 2. I would like to use it
- 3. I have heard of it and would like to learn more
- 4. I have not heard of it
- 5. I have my doubts

Many thanks to many hands!

VINCI Governance Board
- David Atkins, Jack Bates, James Breeling, Stephen Finn, Denise Hynes, Jonathan Nebeker, John Quinn

VINCI OMOP Team
- Michael Matheny, Steve Deppen, Kristine Lynch, Jesse Brannen, Aize Cao, Jiwon Chang, Jason Denton, Elizabeth Hanchrow, Abigail Hillard, Kevin Malohi, Daniel Park, Ben Viernes

VINCI IT Team
- Augie Turano, Hamid Saoudian

Contributing Investigators
- Makoto Jones, Valmeek Kudesia, Steve Luther, Brian Sauer, Andy Zimolzak
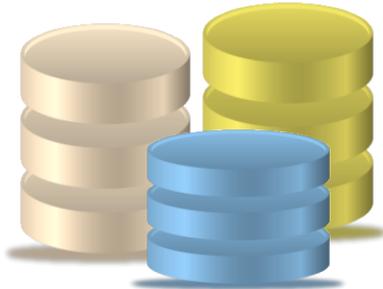
MVP Phenotyping Core
- Kelly Cho, David Gagnon, Jacqueline Honerlaw

Champions and Beta Testers
- Drs. **Whooley, Frey,** Aragam, Callaway-Lane, Cohen, Grogan, Hwang, Jeffrey, Lenert, Luther, Morano, Mudumbai, Samore, Van Houtven, Voils, Zeng

What is VA OMOP?

source data

quality assurance

community use and contribution

national terminologies

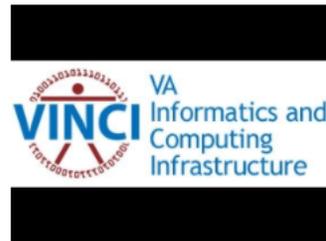extracted, transformed, derived data

best practices
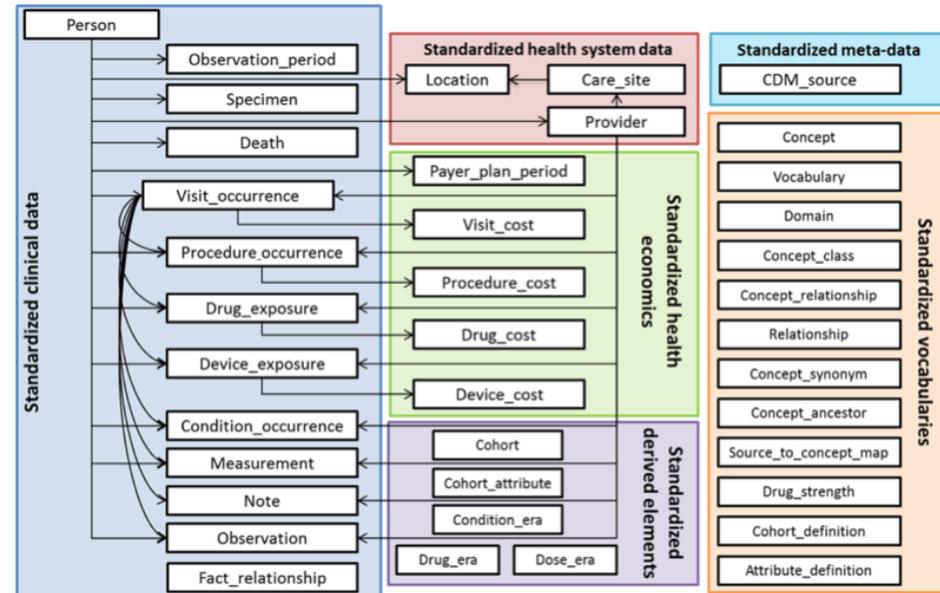
VA OMOP is available

and you can request it

How can I request VA OMOP?

Where do I go to find information about requesting VA OMOP?

- www.vapulse.net VINCI OMOP Users Group

How can I request VA OMOP?

Where do I go to find information about requesting VA OMOP?

- www.vapulse.net VINCI OMOP Users Group

Getting_Started_with_VINCI_OMOPv5_Instructions_11152016

HOW TO GET STARTED USING VINCI OMOP v5

Welcome to VINCI OMOP. You must have **SPECIFIC OMOP IRB APPROVAL** plus either **NATIONAL OPERATIONS** or **RESEARCH** permissions granted to access VINCI OMOP v5. If you don't already have these, please initiate these requests.

HOW TO GET RESEARCH LEVEL ACCESS FOR VINCI OMOP

If you want to use OMOP v5 data in a **new** research project, please add this request to the comment section of your new DART. If you want to add OMOP v5 data views into an **existing** research project, you will first have to make the change in the research project's IRB protocol

How can I request VA OMOP?

What permission do I need to get VA OMOP data?

- VA OMOP is transformed CDW data. If you have approval for CDW data, you have approval for VA OMOP.

Do I need to have RealSSN approval to get VA OMOP?

- No. VA OMOP doesn't contain names, phone numbers, social security numbers, nor scrambled social security numbers.

How can I request VA OMOP?

Do I need to amend my IRB?

- No. VA OMOP is not a new data source. It is transformed CDW data.

Do I need to amend my DART?

- Yes. Mention "VA OMOP" in the Research Request Memo and select "OMOP Common Data Model v5 (CDW Production/Raw Source)" in the CDW Domain Checklist.

How can I request VA OMOP?

CDW DOMAIN CHECKLIST

Production

- [ ] Allergy
- [ ] Appointment
- [ ] Consult
- [ ] CPRS Orders
- [ ] Data Profiling

. . .

RAW*

- [ ] Bill Claims (CBO)
- [ ] CAPRI Audit Trail table
- [ ] CliniComp
- [ ] Echocardiogram
- [ ] Emergency Dept. Int. Software (EDIS)

. . .

Other Data

- [ ] PSSG Geocoded Files and ADUSH Enrollment Files
- [ ] OMOP Common Data Model v5 (CDW Production/Raw Source)

What does VA OMOP look like?
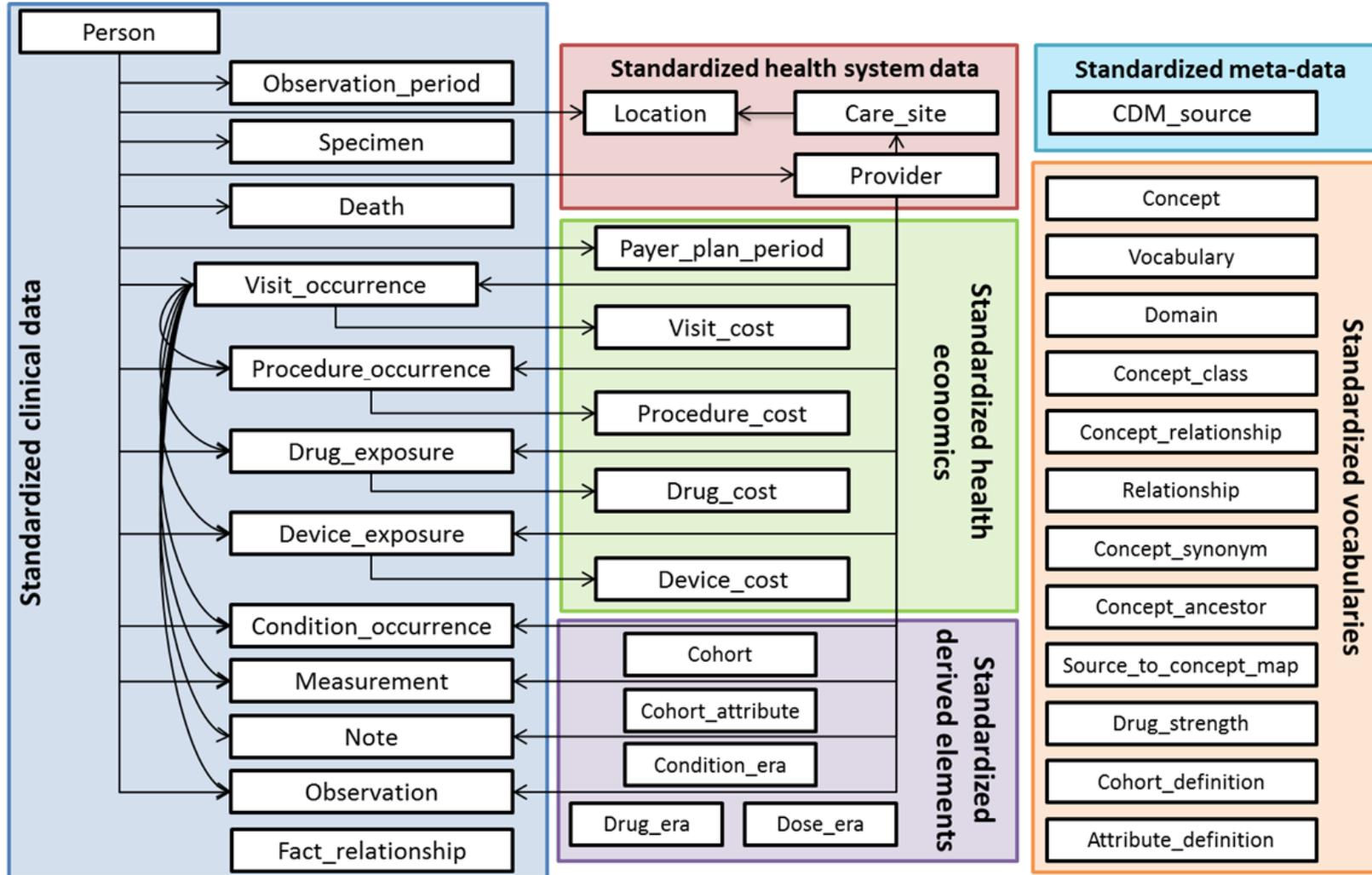
Src.OMOPV5_CARE_SITE
Src.OMOPV5_CDM_SOURCE
Src.OMOPV5_CONCEPT
Src.OMOPV5_CONCEPT_ANCESTOR
Src.OMOPV5_CONCEPT_CLASS
Src.OMOPV5_CONCEPT_RELATIONSHIP
Src.OMOPV5_CONCEPT_SYNONYM
Src.OMOPV5_CONDITION_ERA
Src.OMOPV5_CONDITION_OCCURRENCE
Src.OMOPV5_DEATH
Src.OMOPV5_DEVICE_COST
Src.OMOPV5_DEVICE_EXPOSURE
Src.OMOPV5_DOMAIN
Src.OMOPV5_DOSE_ERA
Src.OMOPV5_DRUG_COST
Src.OMOPV5_DRUG_ERA
Src.OMOPV5_DRUG_EXPOSURE
Src.OMOPV5_DRUG_STRENGTH
Src.OMOPV5_FACT_RELATIONSHIP
Src.OMOPV5_LOCATION
Src.OMOPV5_MEASUREMENT
Src.OMOPV5_NOTE
Src.OMOPV5_OBSERVATION
Src.OMOPV5_OBSERVATION_PERIOD
Src.OMOPV5_PAYER_PLAN_PERIOD
Src.OMOPV5_PERSON
Src.OMOPV5_PROCEDURE_COST
Src.OMOPV5_PROCEDURE_OCCURRENCE
Src.OMOPV5_PROVIDER
Src.OMOPV5_RELATIONSHIP
Src.OMOPV5_SPECIMEN
Src.OMOPV5_VISIT_COST
Src.OMOPV5_VISIT_OCCURRENCE
Src.OMOPV5_VOCABULARY

How can I learn how to use VA OMOP?

- Cyberseminars
- VINCI "Happy Hours"
- VA Pulse VINCI OMOP Users Group
- OMOP Users Call
- VINCI Central
- Concierge requests – vinci@va.gov

What does a VA OMOP table look like?

Src.OMOPV5_CONDITION_OCCURRENCE
  Columns

CONDITION_OCCURRENCE_ID (bigint, not null) ← Primary Key
PERSON_ID (bigint, null) ← Person-centric
CONDITION_CONCEPT_ID (int, not null) ← Standard Concept
CONDITION_START_DATE (date, null)
CONDITION_END_DATE (date, null)
CONDITION_TYPE_CONCEPT_ID (int, null)
STOP_REASON (varchar(20), null)
PROVIDER_ID (bigint, null) ← Provider and Visit
VISIT_OCCURRENCE_ID (bigint, null)
CONDITION_SOURCE_VALUE (varchar(50), null)
CONDITION_SOURCE_CONCEPT_ID (int, null) ← Source Concept
x_CONDITION_START_TIME (time(7), null)
x_CONDITION_END_TIME (time(7), null)
x_Source_Table (varchar(50), null) ← Multiple source tables grouped in conceptual domains
x_Source_SID (bigint, null) ← Bread crumb back to source data
x_Source_ETLBatchID_Transform (int, null) ← Custom columns
x_ETLBatchID (int, null)
x_VersionID (int, null)
x_DBUseStartDateTime (datetime2(0), not null) ← Track changes
x_DBUseEndDateTime (datetime2(0), null)

Stock columns

How does VA OMOP represent concepts?

Src.OMOPV5_CONCEPT
  Columns
    CONCEPT_ID (int, null)

	CONCEPT_ID	CONCEPT_NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT	CONCEPT_CODE	VALID_ST
1	35207365	Multiple sclerosis	Condition	ICD10CM	3-char billing code		G35	2007-01-0
2	44819583	Multiple sclerosis	Condition	ICD9CM	3-dig billing code		340	1970-01-0
3	4137855	Secondary progressive multiple sclerosis	Condition	SNOMED	Clinical Finding	S	425500002	1970-01-0
4	4145049	Relapsing remitting multiple sclerosis	Condition	SNOMED	Clinical Finding	S	426373005	1970-01-0
5	4178929	Primary progressive multiple sclerosis	Condition	SNOMED	Clinical Finding	S	428700003	2008-01-3
6	4231948	Benign multiple sclerosis	Condition	SNOMED	Clinical Finding	S	438511000	2009-01-3
7	4258676	Malignant multiple sclerosis	Condition	SNOMED	Clinical Finding	S	439567002	2009-01-3
8	40491995	Multiple mitral papillary muscles with ham...	Condition	SNOMED	Clinical Finding	S	448844001	2011-07-3
9	44784474	Dementia associated with multiple sclerosis	Condition	SNOMED	Clinical Finding	S	698626001	2014-01-3
10	45768846	Multiple sclerosing hemangioma of lung	Condition	SNOMED	Clinical Finding	S	707367000	2015-01-3
11	44782559	Dementia due to multiple sclerosis with al...	Condition	SNOMED	Clinical Finding	S	82371000119101	2014-01-3
12	40322343	Multiple sclerosis	Condition	SNOMED	Clinical Finding		155023009	1970-01-0
13	40330376	Cord multiple sclerosis	Condition	SNOMED	Clinical Finding		16092000	1970-01-0

    x_DBUseStartDateTime (datetime, not null)
    x_DBUseEndDateTime (datetime, null)
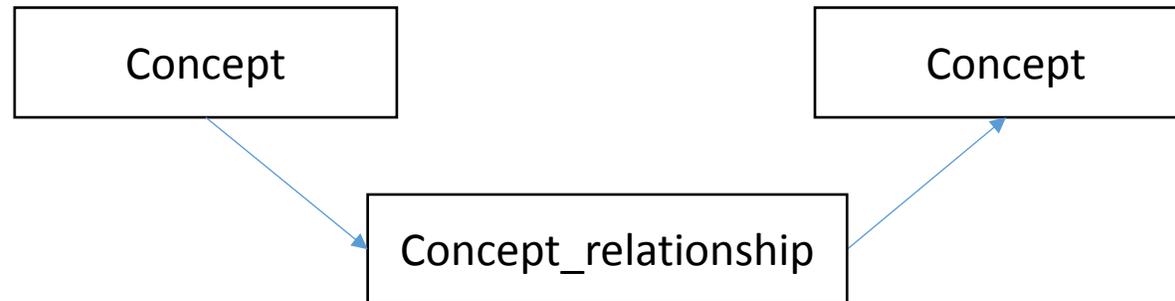
How does VA OMOP represent concepts?

Concept

Concept

Concept_relationship

```sql
SELECT
    b.[CONCEPT_ID_1] ,a.CONCEPT_CODE ,a.concept_name ,b.[RELATIONSHIP_ID] ,b.[CONCEPT_ID_2] ,c.CONCEPT_CODE ,c.concept_name
FROM
    [Src].[OMOPV5_CONCEPT] a
INNER JOIN
    [Src].[OMOPV5_CONCEPT_RELATIONSHIP] b
ON
    a.CONCEPT_ID = b.CONCEPT_ID_1
INNER JOIN
    [Src].[OMOPV5_CONCEPT] c
ON
    c.CONCEPT_ID = b.CONCEPT_ID_2
WHERE
    a.VOCABULARY_ID LIKE '%ICD%' and
    c.VOCABULARY_ID LIKE '%SNOMED%' and
    (a.concept_code = '340' or a.concept_code = 'G35')
ORDER BY
    a.CONCEPT_CODE
```

How does VA OMOP represent concepts?

```
┌─────────────┐                              ┌─────────────┐
│   Concept   │                              │   Concept   │
└─────────────┘                              └─────────────┘
        ╲                                          ╱
         ╲                                        ╱
          ╲      ┌──────────────────────┐        ╱
           ╲     │ Concept_relationship │       ╱
            ➘    └──────────────────────┘      ╱
```

	CONCEPT_ID_1	CONCEPT_CODE	concept_name	RELATIONSHIP_ID	CONCEPT_ID_2	CONCEPT_CODE	concept_name
1	44819583	340	Multiple sclerosis	Maps to	374919	24700007	Multiple sclerosis
2	45571794	G35	Multiple sclerosis	Maps to	374919	24700007	Multiple sclerosis

How does VA OMOP represent concepts?

```
]SELECT
     top 10 dim.concept_name ,dim.concept_code ,dim.vocabulary_id ,count(*) as freq
FROM
     [Src].[OMOPV5_CONDITION_OCCURRENCE] dx
INNER JOIN
     [Src].[OMOPV5_CONCEPT] dim
ON
     dx.condition_concept_id = dim.concept_id
WHERE
     condition_start_date > '2014-09-30' AND condition_s
GROUP BY
     dim.concept_name ,dim.concept_code ,dim.vocabulary_
ORDER BY
     4 DESC
```

	concept_name	concept_code	vocabulary_id	freq
1	Essential hypertension	59621000	SNOMED	7473635
2	Posttraumatic stress disorder	47505003	SNOMED	7318303
3	Type 2 diabetes mellitus	44054006	SNOMED	6808115
4	Hyperlipidemia	55822004	SNOMED	4337859
5	Depressive disorder	35489007	SNOMED	3826939
6	Low back pain	279039007	SNOMED	2483109
7	Alcohol dependence	66590003	SNOMED	2475370
8	Atrial fibrillation	49436004	SNOMED	2337367
9	Tobacco dependence syndrome	89765005	SNOMED	2144778
10	Chronic obstructive lung disease	13645005	SNOMED	2109171

How does VA OMOP represent concepts?

```sql
]SELECT
    top 10 dim.concept_name ,dim.concept_code ,dim.vocabulary_id ,count(*) as freq
FROM
    [Src].[OMOPV5_CONDITION_OCCURRENCE] dx
INNER JOIN
    [Src].[OMOPV5_CONCEPT] dim
ON
    dx.condition_source_concept_id = dim.concept_id
WHERE
    condition_start_date > '2014-09-30' AND conditi
GROUP BY
    dim.concept_name ,dim.concept_code ,dim.vocabul
ORDER BY
    4 DESC
```

	concept_name	concept_code	vocabulary_id	freq
1	Unspecified essential hypertension	401.9	ICD9CM	7465287
2	Posttraumatic stress disorder	309.81	ICD9CM	7318303
3	Diabetes mellitus without mention of complication, t...	250.00	ICD9CM	6282011
4	Other and unspecified hyperlipidemia	272.4	ICD9CM	4332568
5	Depressive disorder, not elsewhere classified	311	ICD9CM	3826939
6	Lumbago	724.2	ICD9CM	2481707
7	Other and unspecified alcohol dependence, unspe...	303.90	ICD9CM	2475370
8	Atrial fibrillation	427.31	ICD9CM	2337367
9	Tobacco use disorder	305.1	ICD9CM	2144778
10	Chronic airway obstruction, not elsewhere classified	496	ICD9CM	2106325

How does VA OMOP represent concepts?

	concept_name	concept_code	vocabulary_id	condition_source_value	freq
1	Essential hypertension	59621000	SNOMED	ICD9|401.9	7465287
2	Essential hypertension	59621000	SNOMED	ICD10|I10.	8348

	concept_name	concept_code	vocabula...	
1	Essential hypertension	59621000	SNOMED	7473635
2	Posttraumatic stress disorder	47505003	SNOMED	7318303
3	Type 2 diabetes mellitus	44054006	SNOMED	6808115
4	Hyperlipidemia	55822004	SNOMED	4337859
5	Depressive disorder	35489007	SNOMED	3826939
6	Low back pain	279039007	SNOMED	2483109
7	Alcohol dependence	66590003	SNOME...	
8	Atrial fibrillation	49436004	SNOME...	
9	Tobacco dependence syndrome	89765005	SNOME...	
10	Chronic obstructive lung disease	13645005	SNOME...	

	concept_name	concept_code	vocabulary_id	freq
1	Unspecified essential hypertension	401.9	ICD9CM	7465287
2	Posttraumatic stress disorder	309.81	ICD9CM	7318303
3	Diabetes mellitus without mention of complication, t...	250.00	ICD9CM	6282011
4	Other and unspecified hyperlipidemia	272.4	ICD9CM	4332568
5	Depressive disorder, not elsewhere classified	311	ICD9CM	3826939
6	Lumbago	724.2	ICD9CM	2481707
7	Other and unspecified alcohol dependence, unspe...	303.90	ICD9CM	2475370
8	Atrial fibrillation	427.31	ICD9CM	2337367
9	Tobacco use disorder	305.1	ICD9CM	2144778
10	Chronic airway obstruction, not elsewhere classified	496	ICD9CM	2106325

How does VA OMOP deal with conflicts?

OMOPV5.PERSON
Columns
PERSON_ID (bigint, not null)
GENDER_CONCEPT_ID (int, null)
YEAR_OF_BIRTH (numeric(4,0), null)
MONTH_OF_BIRTH (numeric(2,0), null)
DAY_OF_BIRTH (numeric(2,0), null)
TIME_OF_BIRTH (time(7), null)
RACE_CONCEPT_ID (int, null)
ETHNICITY_CONCEPT_ID (int, null)
LOCATION_ID (int, null)
PROVIDER_ID (int, null)
CARE_SITE_ID (int, null)
PERSON_SOURCE_VALUE (varchar(50), null)
GENDER_SOURCE_VALUE (varchar(50), null)
GENDER_SOURCE_CONCEPT_ID (int, null)
RACE_SOURCE_VALUE (varchar(50), null)
RACE_SOURCE_CONCEPT_ID (int, null)
ETHNICITY_SOURCE_VALUE (varchar(50), null)
ETHNICITY_SOURCE_CONCEPT_ID (int, null)
x_DateOfBirth (date, null)
x_ETLBatchID (int, null)
x_VersionID (int, null)
x_DBUseStartDateTime (datetime2(0), null)
x_DBUseEndDateTime (datetime2(0), null)

- Test patients removed
- Non-Veterans removed
- PatientICN link

VIReC documented best practice guidance
- Race
- Ethnicity
- Gender
- Date of Birth
- Date of Death
- Preferred Care Site

What if I have a different strategy?

OMOPV5.PERSON
- Columns
  - PERSON_ID (bigint, not null)
  - GENDER_CONCEPT_ID (int, null)
  - YEAR_OF_BIRTH (numeric(4,0), null)
  - MONTH_OF_BIRTH (numeric(2,0), null)
  - DAY_OF_BIRTH (numeric(2,0), null)
  - TIME_OF_BIRTH (time(7), null)
  - RACE_CONCEPT_ID (int, null)
  - ETHNICITY_CONCEPT_ID (int, null)
  - LOCATION_ID (int, null)
  - PROVIDER_ID (int, null)
  - CARE_SITE_ID (int, null)
  - PERSON_SOURCE_VALUE (varchar(50), null)
  - GENDER_SOURCE_VALUE (varchar(50), null)
  - GENDER_SOURCE_CONCEPT_ID (int, null)
  - RACE_SOURCE_VALUE (varchar(50), null)
  - RACE_SOURCE_CONCEPT_ID (int, null)
  - ETHNICITY_SOURCE_VALUE (varchar(50), null)
  - ETHNICITY_SOURCE_CONCEPT_ID (int, null)
  - x_DateOfBirth (date, null)
  - x_ETLBatchID (int, null)
  - x_VersionID (int, null)
  - x_DBUseStartDateTime (datetime2(0), null)
  - x_DBUseEndDateTime (datetime2(0), null)

Patient.Patient
- Columns
  - PatientSID (int, not null)
  - PatientIEN (varchar(50), not null)
  - Sta3n (smallint, not null)
  - RaceSID (int, null)
  - Gender (char(1), null)
  - Age (numeric(18,0), null)
  - StateSID (int, null)
  - State (varchar(30), null)
  - County (varchar(50), null)
  - CountrySID (int, null)
  - Country (varchar(100), null)
  - Deceased (char(1), null)
  - DateOfDeath (date, null)
  - PatientICN (varchar(50), null)
  - PatientICNChecksum (varchar(50), null)
  - PseudoSSNFlag (char(1), null)
  - PeriodOfServiceSID (int, null)
  - PeriodOfService (varchar(50), null)
  - CurrentMeansTestStatusSID (int, null)
  - MaritalStatusSID (int, null)
  - MaritalStatus (varchar(25), null)
  - PercentServiceConnect (int, null)
  - InsuranceCoverageFlag (char(1), null)
  - ReligionSID (int, null)
  - Religion (varchar(30), null)

How can I use VA OMOP logic now?

- OMOPV5Dim.CPT_CONCEPT
- OMOPV5Dim.ICD10_CONCEPT
- OMOPV5Dim.ICD10Procedure_CONCEPT
- OMOPV5Dim.ICD9_CONCEPT
- OMOPV5Dim.ICD9Procedure_CONCEPT
- OMOPV5Dim.Institution_CARE_SITE
- OMOPV5Dim.Institution_LOCATION
- OMOPV5Dim.LocalDrug_CONCEPT
- OMOPV5Dim.Location_CARE_SITE
- OMOPV5Dim.PatientChemLab_CONCEPT

- CDW dimension table to OMOP Concept table mappings

- Institution and Location to OMOP Care_site

- PatientChemLab to OMOP Concept

How can I use VA OMOP logic now?

```sql
SELECT
    TOP 1000 [LabChemTestSID] ,[TopographySID] ,[Sta3n] ,[LabChemTestName] ,[Topography] ,[Units]
    ,[LOINC_Original] ,[LOINC_Updated] ,[MEASUREMENT_CONCEPT_ID] ,[Freq]
FROM
    [CDWWork].[OMOPV5Dim].[PatientChemLab_CONCEPT]
WHERE
    LabChemTestName like '%creat%'
ORDER BY
    [Freq] desc
```

	LabChemTestSID	TopographySID	Sta3n	LabChemTestName	Topography	Units	LOINC_Original	LOINC_Updated	MEASUREMENT_CONCEPT_ID	Freq
1	1200066847	1200001150	573	CREATININE	PLASMA	mg/dL	2160-0	2160-0	3016723	4185158
2	1400023631	1400001955	528	CREATININE (V2)	SERUM	mg/dl	2160-0	2160-0	3016723	4091430
3	1000053873	1000000678	589	CREATININE	PLASMA	mg/dL	2160-0	2160-0	3016723	3971204
4	1000006413	1000000687	657	CREATININE	PLASMA	mg/dl	2160-0	2160-0	3016723	2935734
5	1200000852	1200001099	541	CREATININE (DCed 2.1.15	SERUM	mg/dL	2160-0	2160-0	3016723	2731575
6	1200012363	1200001119	626	CREATININE	SERUM	mg/dL	2160-0	2160-0	3016723	2540280
7	1000006070	1000000670	549	CREATININE	PLASMA	mg/dL	2160-0	2160-0	3016723	2431419
8	1200105374	1200001134	516	CREATININE	PLASMA	mg/dL	2160-0	2160-0	3016723	2362548
9	1000024305	1000000670	549	BUN/CREATININE RATIO	PLASMA	NULL	3097-3	3097-3	3018311	2331224
10	1200002718	1200001131	508	CREATININE	PLASMA	mg/dL	2160-0	2160-0	3016723	2223162
11	1400049167	1400001965	630	CREATININE	SERUM	mg/dl	2160-0	2160-0	3016723	2036484
12	1000080647	1000000598	636	CREATININE*IA	BLOOD	mg/dL	2160-0	38483-4	3051825	2015116
13	800024929	800000762	605	CREATININE	SERUM	MG/DL	2160-0	2160-0	3016723	1981040

How can I use VA OMOP logic now?

```sql
SELECT
    [LOINC_Updated] ,[MEASUREMENT_CONCEPT_ID] ,b.concept_name ,sum([Freq]) as Freq
FROM
    [CDWWork].[OMOPV5Dim].[PatientChemLab_CONCEPT]
LEFT OUTER JOIN
    [Src].[OMOPV5_CONCEPT] b
ON
    a.MEASUREMENT_CONCEPT_ID = b.CONCEPT_ID
WHERE
    LabChemTestName like '%creat%'
GROUP BY
    [LOINC_Updated] ,[MEASUREMENT_CONCEPT_ID] ,b.concept_name
```

	LOINC_Updated	MEASUREMENT_CONCEPT_ID	concept_name	Freq	spb
1	2160-0	3016723	Creatinine serum/plasma	153122677	1
2	38483-4	3051825	Creatinine [Mass/volume] in Blood	9631723	1
3	14682-9	3020564	Creatinine [Moles/volume] in Serum or Plasma	79	1
4	2161-8	3017250	Creatinine [Mass/volume] in Urine	16536454	0
5	3097-3	3018311	Urea nitrogen/Creatinine serum/plasma	11748818	0
6	14959-1	3001802	Microalbumin/Creatinine [Mass Ratio] in Urine	6870019	0

What can I contribute?

- Data mappings

- Variable definitions

- Code snippets / Logic
  - Validation, erroneous data, transforms

- Cohort lists

vinci@va.gov

Scott DuVall

scott.duvall@va.gov

scott.duvall@utah.edu

@DuVallScott