



# Managing a Research Project Using VINCI Data

**Olga Efimova and Andrew Holbrook**

VA Informatics and Computing Infrastructure (VINCI) Resource Center

December 13, 2023

# Session Outline

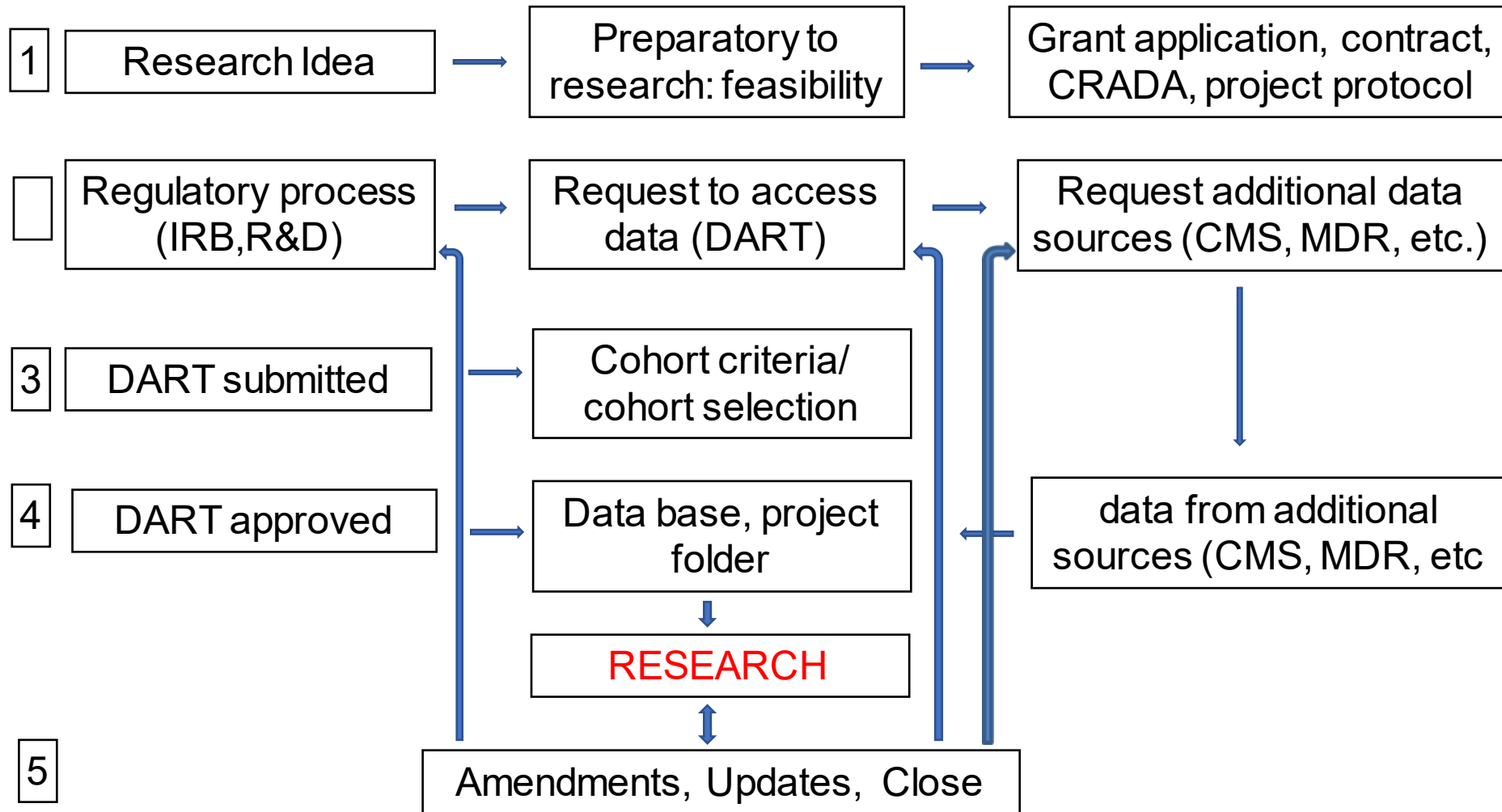
---



- Research projects life cycle and resources
  - Preparatory to Research
  - Regulatory requirements
  - The Data Access Request Tracker
- VINCI Data
  - Skills Required
  - Data sources and domains
  - Cohort
  - Time range
  - CDW structure

“There are no unrealistic goals, only unrealistic timelines.”  
-Brian Tracy

# Research Project Life Cycle



# Preparatory to Research: Data Needs Assessment



VINCI Services supports VA research by assisting customers with Data Needs questions (DNA) relating to the VA Corporate Data Warehouse (CDW) and other VA data sources.

These questions range from

- ✓ data availability
- ✓ data location
- ✓ data definitions
- ✓ assistance with completing the DART application
- ✓ assistance with completing DART forms including CDW-domain checklist,
- ✓ assistance with IRB protocol preparation,
- ✓ etc.

# Preparatory to Research: Feasibility



- Can be conducted by VINCI Services team (no DART) [Concierge Services \(va.gov\)](https://www.va.gov/concierge-services/)
- Feasibilities provide aggregate counts for researchers during the preparatory-to-research phase. The deliverables result in summarized counts to assist in grant and IRB applications.
- Feasibilities help to understand cohort size and structure
- Feasibilities help to understand impact of each criteria to cohort size and structure
- Feasibilities result cannot be used for any publication
- No regulatory documents are required

# Preparatory to Research



- Preparatory to research Can be conducted by research team (Feasibility project (cohort size and structure; result cannot be published))
- SQL skills
- DART Preparatory to Research Request Process
- <http://vaww.vhadataportal.med.va.gov/DataAccess/PreptoResearchRequestProcess.aspx#SubmitRequest>

The screenshot shows the 'DART Preparatory to Research Request Process' page. At the top is a navigation bar with links for Data Sources, Data Access, Tools & Applications, Resources, Training, Policy & Admin, and Support. Below this is a breadcrumb trail: Data Access > PreptoResearch Request Process. The main content area is divided into a left sidebar and a main panel. The sidebar contains three sections: 'What is DART?' with a 'Launch DART' link and a description; 'Did you know...' with a note about one request per protocol; and 'Quick Links' with a 'DART User Guide' link. The main panel has a title 'DART Preparatory to Research Request Process' and a sub-navigation bar with 'Overview', 'Submitting a Request', 'Documents & Forms', 'Request Metrics', and 'Resources'. The 'Overview' section explains that the DART PTR Request Process is for requesting permission to access VHA data with PHI, governed by HIPAA. It also lists data sources that can be requested: ADUSH Enrollment Files and Bereaved Family Survey (BFS) Database.

# Regulatory: IRB/R&D



- Research projects
- IRB and R&D
- **DART** (The Data Access Request Tracker)/ Request list of patients from VINCI Services for the prospective projects/ Request access for the data for retrospective project
- <http://vaww.vhadataportal.med.va.gov/DataAccess/DARTRequestProcess.aspx#SubmitRequest>
- Forms for the additional data sources



# Research Project: VINCI for IRB



## [VINCI Central - Home \(va.gov\)](#)

Home \ VINCI University

Workspaces Applications My VINCI Dashboard **VINCI University** Services Support Quick Links

Send by email

VINCI Central

What is VINCI?  
IRB Study How-To  
PTR Study How-To  
Ops/QI Study How-To  
Data Access Request Tracker (DART)  
VINCI Data  
VINCI SAS/Grid  
VINCI Applications  
VINCI Workspaces  
OMOP Academy  
DaVINCI Academy  
NLP Academy  
VINCI R Academy  
VINCI Training & Office Hour  
All VINCI Videos

VINCI University

top shop for all your VINCI training and educational needs.

on VINCI data, the VINCI Workspace, administrative processes, and technical  
n our VINCI Tube site to our extensive collection of user guides. Additionally,  
ing and information is provided through Cyberseminars, VINCI Training  
ednesday of the month), and VINCI Office Hour Q&A sessions (all other  
. Join at 3PM ET on MS Teams or call (872) 701-0815, conference ID 480 505  
re to add this to your Calendar.

training to large and small groups as well as one-on-one discussions using in-person and  
remote meetings. Topics include VINCI Workspace access and usage, data upload and download, data  
information, database access, data access approval processes, and DART. Email a VINCI Concierge to discuss  
you or your group's needs at [VINCI@VA.gov](mailto:VINCI@VA.gov).



## Required Information in Project Protocol

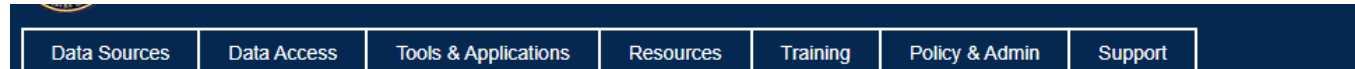
- VINCI description for IRB [IRB Study How-To \(va.gov\)](#)
  3. \*\*As part of your IRB package you are required to state where data will be stored for analysis. If you are planning to use VINCI for data storage and analysis we provide the following environment and security descriptions for use with IRB packages.
    - [VINCI Description for IRBs](#)
    - [VINCI Information Security Description](#)
- Real SSN/ Scrambled SSN should be requested (CAPRI/JLV, IVC data, USVETS, non-VINCI data sources, other)
- If project will need [CAPRI/JLV access](#), it should be described in the protocol/IRB protocol and HIPAA Waiver (“Access to patients records on national level will be conducting using CAPRI/JVL applications”)
- If project required cause of death analysis using [MDR data](#), it should be described in IRB protocol in study procedure section [Mortality Data \(va.gov\)](#)
- If project required [Medicaid /Medicare analysis using CMS data](#), it should be described in IRB protocol in study procedure section [VA/CMS Home](#)
- If targeted population when using 7332 diagnosis (drug abuse, alcohol abuse, HIV, sickle cell anemia) the following statement should be included in IRB Protocol & HIPAA: The purpose of the data is to conduct scientific research and that no personnel involved in the study may identify, directly or indirectly, any individual patient or subject in any report of such research or otherwise disclose patient or subject identities in any manner.  
<https://www.networkforphl.org/wp-content/uploads/2019/01/Snapshot-VA-Confidentiality-of-Sensitive-Records.pdf>
- Storage location
- Inclusion/ exclusion criteria
- Waiver of HIPAA Authorization



# Research Project: HIPAA Waiver



## [DART Research Request Process \(va.gov\)](#)



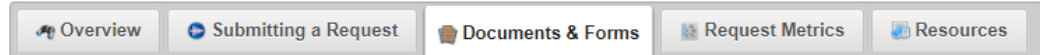
Data Access > DART Request Process

**What is DART?**

[Launch DART](#)

DART is an online application for requesting permission to access VHA data.

### DART Research Request Process



#### Documents & Forms

- IRB Approved Sample Informed Consent & HIPAA Authorization - *required if signed informed consent forms or HIPAA authorizations will be collected for any portion of the study.*
- IRB Approved [Waiver of HIPAA Authorization](#) - *required if a Waiver of HIPAA Authorization is required for any portion of the study.*

Required for...	Form
ALL DART Requests	<a href="#">Research Request Memo</a>
Real SSN Identifiers	<a href="#">Real SSN Access Request Form</a>
Bereaved Family Survey (BFS) Database	<a href="#">Program Office Data Access Approval: BFS</a>

## Inclusion/exclusion criteria

### Ethics Principles for Access to and Use of Veteran Data

- Principle 7: All parties are obligated to ensure data security, quality and integrity of Veteran data.

All parties who send, receive, or use Veteran data must ensure data security, quality and integrity. In other words, that the data remain secure; accurate; complete; and representative of the data quality, meaning, and integrity when it was received or accessed from VA. Access to data should be limited to the minimum amount needed to accomplish the stated purpose, and should be terminated when no longer required. Data that are not necessary to accomplish the purpose for which it was obtained should not be retained longer than legally required. Transparency about breaches in data security, quality or integrity is also essential to promote trust and minimize impacts to Veterans.

[Ethics Principles for Access to and Use of Veteran Data - DigitalVA](#)



## Inclusion/exclusion criteria

- Location (nationwide or some specific locations HCS/MC/VISN etc.)
- Time frame
- Age/gender/other limitation
- Cohort definition with justification

## Required Documents

- Research Study Institutional Review Board (IRB) Approval Letter
- Sample Informed Consent (for the prospective project)
- Research and Development (RD) Committee Approval Letter
- Research Protocol
- IRB Approval Waiver of HIPAA-Compliant Authorization
- DUA (if needed)
  
- *Research Request Memo\**
- *Request to Real SSN Access (if needed – specific data sources (CAPRI, IVC data, chart review))\**
- *CDW-Domain Checklist\**
- *Millennium data check list\**
- *Data source specific documents\* (if needed – CAPRI/JLV, VASQUIP, etc.)*



\* Blank documents can be found at [DART Research Request Process \(va.gov\)](#)



## The Data Access Request Tracker (DART)

<http://vaww.vhadatportal.med.va.gov/DataAccess/DARTRequestProcess.aspx#SubmitRequest>

Data Access > DART Request Process

**What is DART?**

[Launch DART](#)

DART is an online application for requesting permission to access VHA data.

### DART Research Request Process

[Overview](#) [Submitting a Request](#) [Documents & Forms](#) [Request Metrics](#) [Resources](#)

#### Documents & Forms

**Important! Tips for Viewing & Downloading Forms**

**How to download and view forms.**

- Right click on the link to a form in the table below.
- Select "Save link as..." and save form locally.
- Open PDF forms in Acrobat Reader DC or Adobe Acrobat DC

Required for...	Form
ALL DART Requests	<a href="#">Research Request Memo</a>
Real SSN Identifiers	<a href="#">Real SSN Access Request Form</a>
Repeaved Family Survey (RFS) Database	<a href="#">Program Office Data Access Approval: RFS</a>



# DART Initiated Consultation



The **Data Access Request Tracker (DART)** Research Request Process is application for requesting permission to access VHA data for the IRB - approved research project

***We strongly encourage team members to reach out to [vinci@va.gov](mailto:vinci@va.gov) and request a Data Needs Assessment (DNA)/DART consultation before submitting completed DART application.***

## **Consultation:**

- Discuss steps for the research projects using VINCI data
- VINCI data base: data availability and limitation
- Help with selection of the data sources needed for the research study

***This will help ensure that the selections made on the DART Data Sources screen and CDW Checklist will result in the creation of a comprehensive dataset for the study.***

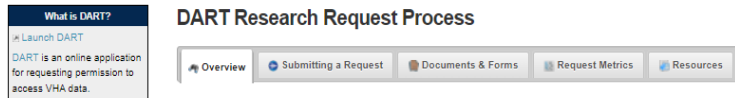
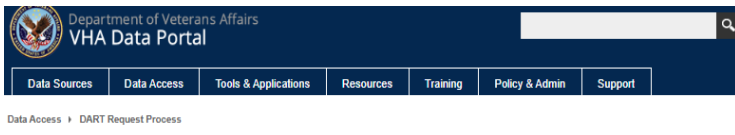
# DART: Data Sources (More on This Later)



<http://vaww.vhadatportal.med.va.gov/DataAccess/DARTRequestProcess.aspx#Overview>

[VINCI Data \(va.gov\)](http://va.gov)

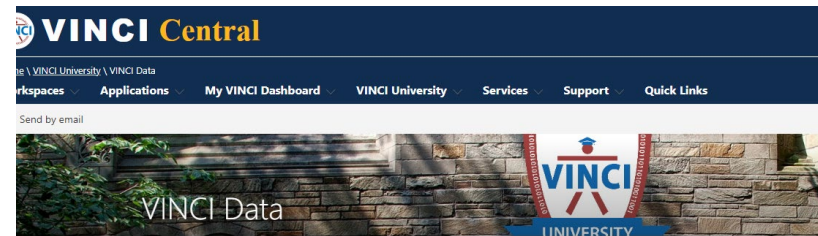
<https://vaww.virec.research.va.gov/>



**Overview**  
The Data Access Request Tracker (DART) Research Request Process is for requesting permission to access VHA data for approved research projects. Visit the [Preparatory to Research Access](#) page to learn about requesting data when preparing a research funding application or protocol.

**Which data sources can I request?**  
Many, but not all data sources and tools can be requested through DART. Here is a list of those that are available for request. Select a data source for a description of its contents and tips for requesting access.

- ADUSH Enrollment Files
- Bereaved Family Survey (BFS) Database
- CAPRI
- Care Assessment Need Score (CAN Score)
- Clinical Assessment Reporting and Tracking (CART) Program Data
- Corporate Data Warehouse (CDW)
  - Production Domains
  - Raw Domains
  - Text Integration Utility (TIU) Text Notes
  - OMOP Common Data Model
- COVID-19 Shared Data Resource
- DAVINCI (DoD-VA Health Data)
- Joint Longitudinal Viewer (JLV)
- Health Economics Resource Center (HERC) Cost Data
- Home Based Primary Care (HBPC) Masterfile
- Homeless Registry
- Lung Cancer Screening Demonstration Project (LCSOP) Cohort
- Managerial Cost Accounting (MCA)
  - National Data Extracts (NDEs)
  - Web Reports
- Medical SAS Files
- My HealtheVet Administrative Data (MHV)
- Non-VA Medical Care (formerly Fee Basis) Files



Welcome to Data on VINCI University

Here you'll learn about VINCI Data and find answers to your questions.

## Data Sources

- [CDW Production Data](#)
- [CDW Raw Data](#)
- [VA OMOP Data](#)
- [DaVINCI Data](#)
- [SAS Datasets](#)
- [VINCI LVEF NLP Output](#)
- [COVID-19 Shared Data Resource](#)

## VA INFORMATION RESOURCE CENTER (VIReC)

- VIReC Home
- About Us
- VA/CMS Data for Research
- VHA Data Portal
- New Users of VA Data
- FAQs
- Acronym Lookup
- HelpDesk
- Report Broken Link
- How are we doing?



- Data Issues Brief
- Upcoming Cyberseminars
- Real SSN Data Request Process
- VA REDCap Project
- Data Tip of the Month

## New from VIReC

- Attention VA Data Users! We're Hiring!**  
Check out the [Announcements page](#) on our Internet site to see descriptions of currently open positions and required skills.  
[Upcoming Cyberseminars](#)
  - May 17: [The HEART Evaluation – Partnering with National and Local Leaders to Improve Healthcare](#)
- New & Updated VIReC Products**
  - Data Review: [Patient Identifiers in Millennium](#) [released: May 2022]
  - Researcher's Notebook: [CDW Principles of Building a Data Cohort Millennium \(Part II\)](#) [updated: Apr 2022]
  - Factbook: [MVI Domain](#) [updated: Mar 2022]
  - [CDW Domain Layout & Data Details](#) [released: Mar 2022]
  - Data Review: [Procedures in Millennium Data](#) [released: Mar 2022]
  - [MCA Historical Stop Codes](#) [updated: Mar 2022]
  - Quick Peek Video: [Using the CDW Metadata Report](#)

## VIReC Resources

- [Meet VIReC, The Researcher's Guide to VA Data:](#) Learn about us, our products, and how to navigate our web environments.  
[Learn about VA data:](#)
  - Data Sources
  - Data Topics
  - Data Tools
- EHRM Resources:**
  - [Guidance on Addressing EHRM in Applications](#) [released: Nov 2021]
  - Web page: [EHRM & Implications for Data Users](#)
  - [VA Millennium EHR Data Documentation & Resources](#)
- VA/CMS Data for [Research Project](#): Data custodian for CMS and USRDS data for VA research.  
**Factbooks:** Describe tables, columns, and values in select CDW Domains.  
**Research User Guides:** Detailed information on select data sources, including variable descriptions.



# DART: CDW domain checklist



<https://vaww.virec.research.va.gov/CDW/Documentation.htm#Factbooks>

[CDW-Raw Domain Documents \(sharepoint.com\)](#)

[Metadata - Metadata Documents - All Documents \(sharepoint.com\)](#)

**VIREC INTRANET**

**VA INFORMATION RESOURCE CENTER (VIREC)**

**CDW Documentation**

**Overview**

VIREC provides documentation on select Corporate Data Warehouse (CDW) production domains. CDW documentation is designed to help new and seasoned users better understand the structure and contents of the CDW. It might be necessary to review a combination of resources to fully understand the complexities of the CDW and data use environment.

**Sign-up for VIREC product news and updates.** Email the VIREC [Help Desk](#) to receive notification of VIREC's new CDW products and product updates.

**How are we doing?** Take our [short survey](#) to share your thoughts about the CDW resources below. We value your feedback.

**Data Documentation**

Expand each type of documentation below to view these resources.

**+ Expand All**

**Getting Started with CDW**

**Factbooks**

CDW Factbooks provide descriptions of tables, columns, and values in select CDW Domains including domain-specific SQL "starter language" and sample SQL code.

Domain	Released	Factbook
Allergy	1.0 Feb 2019	
Appointment	2.2 Jul 2020	
Consult	2.1 Feb 2016	
CPRS Orders	1.1 Jan 2018	
Health Benefits Request	1.0 Feb 2019	
Health Factor	2.1 May 2019	
Immunization	2.1 Feb 2019	
Inpatient	3.0 Aug 2017	
Inpatient Fee Ledger	1.0 Feb 2019	
Lab-Chem	2.0 Aug 2017	
Mental Health	1.1 Jul 2020	
<b>NEW!</b> MVE	1.1 Feb 2021	
Non-VA Meds	1.0 Feb 2016	
Outpatient	2.1 Jul 2020	
Patient	3.1 Sep 2020	
Patient-Associated	1.0 Oct 2017	
Patient Enrollment (with EWL)	1.0 Jul 2015	

**CDW-Raw Domain Documents**

This section is read-only.

Domain	Summary
eCAMS	Summary
Echo	Contacts
eInsurance and ePay...	Change History
Enterprise Contact Ce...	2022-06-28 "Missing" ...
FBCS (deprecated)	2022-05-24 Error on Fi...
Fee Schedule	2022-05-19 Error on Fi...
FMS	2022-05-12 Added Ne...
FUM	2022-04-06 Added Ne...
HR-PAS (CDW)	2022-02-03 - Fixed Ba...
HR-PAS (NST - COVID)	2021-07-29 Added col...
ICAPA (JVPN)	2021-04-15 - Added A...
ICB	2020-11-05- 165.5 HI...
IFCAP	2019-08-21-Warnings ...
Inventory	2017-11-07- Data Bug
IV Meds	2016-04-16
MCA Pharmacy	2015-06-19
MCCF	2105-04-22
MHV	2014-09-05
<b>Oncology</b>	2012-05-22
PAID	2012-08-12
PCM	Sample Queries

**Summary**

Tuesday, December 10, 2019 5:58 AM

The V16 Cancer Registry project was selected as forward (but we can't spend for IT unless we mo

Oncology requires Privileged Read Access ("CDW

**Purpose of Data**

- Helps in the development of **evidence based per**
  - Timeliness: Veterans will receive **more tin**
- Illustrates specific **cancer costs**, enabling benchr
  - Veterans' care will consistently fall within :
- Helps **determine appropriate VERA reimburses**
  - Information about a Veteran's service can

Oncology was defined on 5/22/2012 for Wong, F  
This extract requires Privileged Read Access / SP.

**CDW Metadata**

Metadata Home

CDW View Diagrams

Libraries

Special Reports

Recent

Return to classic SharePoint

**Metadata Documents**

Name	Created
Active Directory	July 22, 2016
Allergy	February 19, 2014
AntiCoag	August 14, 2020
Appointment	February 19, 2014
Beneficiary Travel	July 10, 2017
CDWWork2	August 18, 2021
Clinical Reminder Parameters (CRP)	August 26, 2022
Compensation and Pension	August 5, 2020
Consult	October 1, 2014

# DART: Submission and Cohort Creation



Your research project with DART tracking number 2022-07-003-D has been submitted.

If your study requires a cohort derived from CDW/VINCI data sources, the next step is to provide a broad study cohort criteria that will be used to filter the VA patient population. Your preliminary cohort can be defined by specific diagnoses (ICD-codes), care settings (Inpatient/Outpatient), procedures, time period, etc.

Researchers will have two options:

- The first option (**New and/or less experienced VINCI users may prefer this approach**): while your DART application is under review is to work with our consultation team ([VINCIservices@va.gov](mailto:VINCIservices@va.gov)) to review your cohort inclusion and exclusion criteria. It allows more time for cohort definition refinement if needed prior to DART approval. This consultation allows for a wide-ranging review of various data sources and broad phenotype definitions to provide a comprehensive cohort that meets your study cohort needs. Please, send inclusion/exclusion criteria to [VINCIservices@va.gov](mailto:VINCIservices@va.gov)

OR

- The second option (**More experienced VINCI users may prefer this approach as they may already be familiar with CDW/VINCI data and this process**): once your DART application is approved and correspondence site is created, is to work with the VINCI provisioning data manager directly and provide the cohort definition through the project correspondence site

Examples of **a cohort inclusion/exclusion criterion**, may include two or more of the following:

1. Location (nationwide or specific stations)
2. Time frame for each criteria
3. Diagnoses (please provide all relevant ICD9 and/or 10 codes)
4. Procedures (please provide all relevant ICD-procedure and/or CPT codes)
5. Medication (please provide all generic and brand names)
6. Lab test (topography, units, range)
7. Type of project: Prospective or Retrospective
8. Other criteria



## Inclusion/exclusion criteria for the cohort = Inclusion/exclusion criteria IRB AND DART data sources

- Location (nationwide or some specific location)
- Time frame
- Age/gender/other limitation
- Diagnosis (ICD9\*/10 codes (\*VA uses ICD10 codes since October 2015. If time frame prior to 2015, please, provide both))
- Procedures (CPT codes or/and ICD-PC)
- Lab test. (topography, units, range)
- Medications (list of all generic and brand names, strength and route of administration)

# DART Approved – Now What?

---



- “Where’s my flat file?” – What skills you need to work with VINCI Data
- Project data provisioned – What you get and how to access it
- The cohort – Who are you studying?
- The data – sources, domains and additional considerations

These slides will be more about managing the project than about managing the data

# What Skills Do You Need to Work With VINCI Data?



- Data Skills

- VINCI data is provisioned in a normalized **SQL** database – you may have millions of rows and thousands of columns across hundreds of tables
- You will need to communicate with the server in the language of SQL – this can be done directly in SSMS (SQL Server Management Studio) or as a pass-through from one of the analysis platforms below

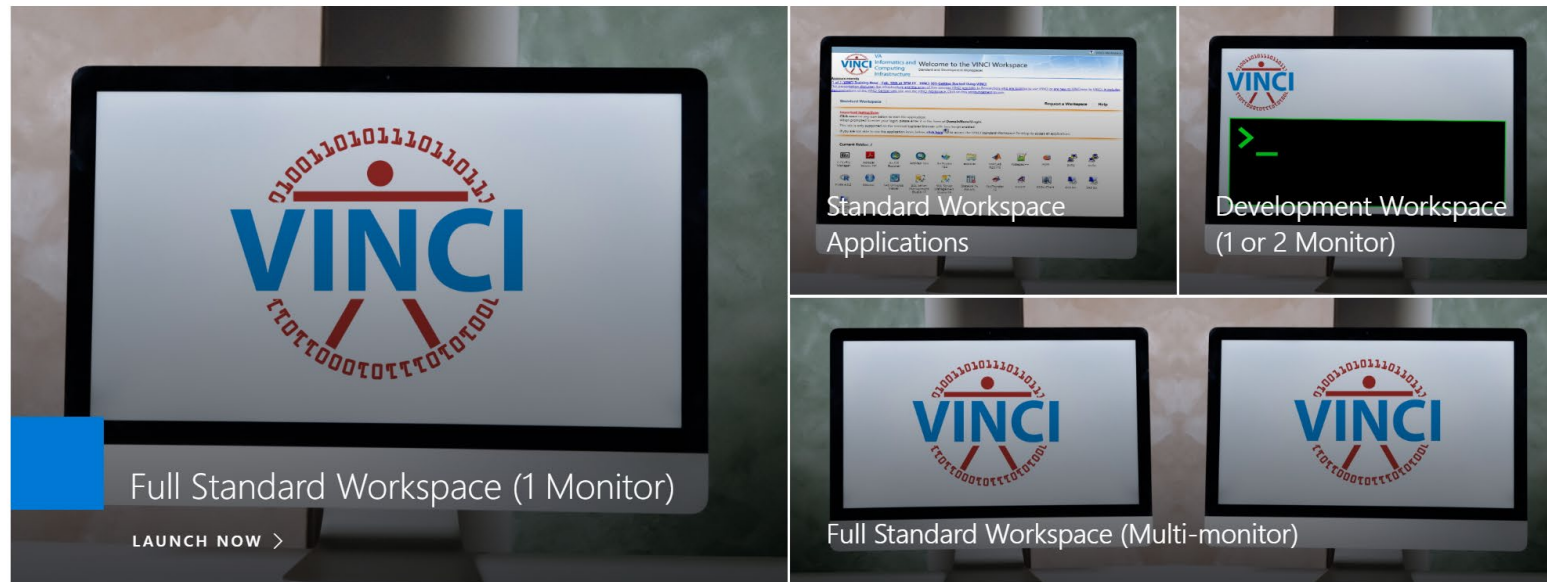
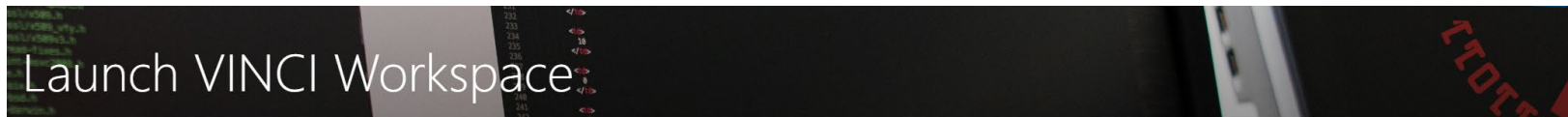
- Analysis Skills

- You have options! [VINCI Standard Workspace Applications](#)
  - SAS
  - R
  - Stata
  - MATLAB

# Accessing Your Data



- Log in to the [VINCI Workspace](#) to access your project folders and applications
- Everyone on the DART will automatically have access to the study folders and data



# SSMS Example



SSMS from the start menu

Connect to Server

### SQL Server

Server type: Database Engine

Server name: vhacdwr02.vha.med.va.gov

Authentication: Windows Authentication

User name: VHA19\VHASLCHolbrA

Password:

Remember password

Connect Cancel Help Options >>

Object Explorer

Connect [Icons]

- vhacdwr02.vha.med.va.gov (SQL Server)
- Databases (filtered)
  - System Databases
  - Database Snapshots
  - ORD\_Holbrook\_202202042D
    - Database Diagrams
    - Tables (filtered)
      - System Tables
      - FileTables
      - External Tables
      - Graph Tables
      - Src.CohortCrosswalk
      - Src.CohortDescription
      - Src.CohortPatientICN
      - Src.CohortPatientSID
      - Src2.CohortPersonSID
      - Dropped Ledger Tables
      - Views (filtered)
        - System Views
        - Src.Outpat\_ProblemList
        - Src.Outpat\_VDiagnosis
        - Src.Outpat\_VDiagnosis\_Recent
        - Src.Outpat\_VExam
        - Src.Outpat\_Visit



# SQL Pass-through Example (SAS)



```
PROC SQL ;
```

```
CONNECT TO OLEDB AS CDW1 ( DATASOURCE="VHACDWRB02.VHA.MED.VA.GOV" PROVIDER=SQL_Driver  
READBUFF=5000 CURSOR_TYPE=STATIC UTILCONN_TRANSIENT=YES DEFER=YES CONNECTION=GLOBAL  
AUTOCOMMIT=YES PROPERTIES=( 'INITIAL CATALOG'=ORD_Holbrook_202202042D 'INTEGRATED  
SECURITY'=SSPI ) );
```

Using this pass-through method, you can write T-SQL to be executed on the server.

Source: “Best Practices in Accessing SQL Server Data with SAS” on [VINCI SAS/Grid \(va.gov\)](https://www.va.gov/vinci-sas/grid/)



# Project Data

---



- VINCI provisions study data, capturing all records:
  - from the requested data sources and domains
  - for the cohort of interest
  - within the specified time range (optional)
  - maintaining the source data structures

# Requested Data Sources and Domains



- Most data come from CDW Production
- Some examples of non-CDW data:
  - CMS
  - Death Ascertainment File
  - ADUSH Enrollment Files
  - CAN Scores
  - DoD Data
  - PAID Data

Select data documentation links

- [Data Sources \(va.gov\)](#) (VHA Data Portal)
- [Working with VA Data](#) (VIReC)
- [Corporate Data Warehouse](#) (OIT CDW)
- [Data Sources \(va.gov\)](#) (VINCI)

For lots more data links see this recent Cyberseminar from Brian Robison:  
[Navigating CDW Data \(va.gov\)](#)



# Where the Data Come From



## DART Data Sources Screen

**Data Access Request Tracker (DART)**

HOME > DART Dashboard > Requests

Information > Participants > **Data** > Documents > Submit

2022-02-042-D-A17 VINCI training 2

AMENDMENT NARRATIVE: Adding 4 additional Participants: Ward, Motchnik, Lendval, and McCreight.

### Data Sources

Data Storage Location \*

VINCI  
 Other Server Location

Will data be transferred external to the VHA? \*

*NDS cannot approve a research request that would include disclosing data requested through DART. For example:*

*The HIPAA Privacy Rule permits the transfer of Limited Data Sets for research purposes, without obtaining a separate authorization, if "finder" files are to be sent out to Federal or state agencies for the purpose of a Federal or state agency.*

No  
 Yes (DUA is required)

Data Access Systems

SAS Grid

IDENTIFIERS (More about identifiers)

Real SSN  
 Scrambled SSN  
 Identifiable data but no real or scrambled SSNs

HEALTHCARE OPERATIONS GENERATED DATA SOURCES

ADUSH Enrollment Files  
 Bereaved Family Survey  
 CAPRI & Joint Legacy Viewer (JLV) - Individuals needing system access are selected on the Par  
 Care Assessment Need Score (CAN Score) - Scrambled SSN approval required  
 Clinical Assessment Reporting and Tracking (CART) CV (Disclaimer)  
 Corporate Data Warehouse (CDW)  
 Production Domains

## CDW DOMAIN CHECKLIST

### CDW Production

- |                                                                    |                                                                                         |
|--------------------------------------------------------------------|-----------------------------------------------------------------------------------------|
| <input type="checkbox"/> Allergy                                   | <input checked="" type="checkbox"/> Patient                                             |
| <input type="checkbox"/> AntiCoag (Anticoagulation)                | <input type="checkbox"/> Patient Associated                                             |
| <input checked="" type="checkbox"/> Appointment                    | <input type="checkbox"/> Patient Enrollment                                             |
| <input type="checkbox"/> Beneficiary Travel                        | <input type="checkbox"/> Patient Insurance                                              |
| <input checked="" type="checkbox"/> Consult                        | <input type="checkbox"/> Patient Record Flag                                            |
| <input checked="" type="checkbox"/> CPRS Orders                    | <input type="checkbox"/> PCMM (Primary Care Management Module)                          |
| <input type="checkbox"/> Dental                                    | <input checked="" type="checkbox"/> Pharmacy BCMA (Bar Code Medication Administration)  |
| <input type="checkbox"/> Emergency Dept. Int. Software (EDIS)      | <input type="checkbox"/> Pharmacy IV                                                    |
| <input checked="" type="checkbox"/> Enrollment (ADR)               | <input checked="" type="checkbox"/> Pharmacy Outpatient                                 |
| <input type="checkbox"/> Health Benefits Request                   | <input checked="" type="checkbox"/> Pharmacy Patient                                    |
| <input checked="" type="checkbox"/> Health Factors                 | <input checked="" type="checkbox"/> Pharmacy Unit Dose                                  |
| <input type="checkbox"/> Home Telehealth                           | <input type="checkbox"/> Program Integrity Tool (PIT)                                   |
| <input type="checkbox"/> Immunization                              | <input type="checkbox"/> Prosthetics                                                    |
| <input checked="" type="checkbox"/> Inpatient                      | <input checked="" type="checkbox"/> Purchased Care                                      |
| <input type="checkbox"/> Integrated Billing                        | <input checked="" type="checkbox"/> Radiology                                           |
| <input type="checkbox"/> Lab Accession                             | <input type="checkbox"/> Recall Reminders                                               |
| <input checked="" type="checkbox"/> Lab Chem                       | <input checked="" type="checkbox"/> Reengineered Primary Care Management Module (RPCMM) |
| <input type="checkbox"/> Lab Microbiology                          | <input type="checkbox"/> SPatient                                                       |
| <input type="checkbox"/> Lab Order                                 | <input checked="" type="checkbox"/> Staff                                               |
| <input checked="" type="checkbox"/> Master Veteran Index (Veteran) | <input type="checkbox"/> SStaff                                                         |
| <input type="checkbox"/> Master Veteran Index (SVeteran)           | <input type="checkbox"/> Surgery PRE, INTRA, and POST                                   |
| <input checked="" type="checkbox"/> Mental Health Assessment       | <input type="checkbox"/> VistA Compensation & Pension                                   |
| <input checked="" type="checkbox"/> Non-VA Meds                    | <input type="checkbox"/> VistA Waitlist                                                 |
| <input checked="" type="checkbox"/> Outpatient                     | <input checked="" type="checkbox"/> Vital Signs                                         |
| <input type="checkbox"/> Pathology                                 | <input checked="" type="checkbox"/> Women's Health                                      |

## SQL Server Views

Object Explorer

Connect > [Icons]

- vhacdwr02.vha.med.va.gov (SQL Server)
  - Databases (filtered)
    - System Databases
    - Database Snapshots
    - ORD\_Holbrook\_202202042D
      - Database Diagrams
      - Tables (filtered)
        - System Tables
        - FileTables
        - External Tables
        - Graph Tables
        - Src.CohortCrosswalk
        - Src.CohortDescription
        - Src.CohortPatientICN
        - Src.CohortPatientSID
        - Src2.CohortPersonSID
        - Dropped Ledger Tables
        - Views (filtered)
          - Src.Output\_ProblemList
          - Src.Output\_VDiagnosis
          - Src.Output\_VDiagnosis\_Recent
          - Src.Output\_VExam
          - Src.Output\_Visit
        - System Views



# The Names Aren't the Same!



- Did you notice? “Outpatient” vs “Outpat”
- Most of these domain/source to schema name changes are obvious, but some are not.
- Raw and external sources often have “VINCI\_”
- If you can't find one, ask about it in [VINCI's weekly office hour](#) or you can ask your data manager on the correspondence site

Some example name changes:

- PAID Data -> VINCI\_PAID
- ADUSH Enrollment Files -> VINCI\_ADUSH
- IVC Consolidated Data Sets -> VINCI\_IVC\_CDS
- Managerial Cost Accounting (MCA), National Data Extracts (NDEs) -> CDWWork\_DSS



# Cohort of Interest

---



- VINCI will provision data based on one or more study cohorts that you define
- You can have multiple cohorts
  - VINCI-provisioned cohorts (annoying for overlapping cohorts)
  - One big cohort from which you choose cohorts (best for complicated selection logic)
- Not too big
  - The larger your cohort, the more difficult and slow your queries/processing will be
  - Ethical principal 7
- Not too small
  - You can filter down to a smaller cohort yourself based on any data from any requested source/domain
  - You cannot see any data for anyone not in your study cohort – remember to consider controls

# Where does the cohort come from?



## Criteria Spreadsheet

Date of Query: 10/19/2021

Criteria Definition	Care Setting	Location	Time range	Query Definition	Domains/Data Sources Used:	Number of distinct patients remaining
<b>Inclusion criteria</b>						
Number of Veterans seen outpatient, inpatient, fee basis nationwide at period of time October 1, 2006 -present	outpatient, inpatient, fee basis	nationwide	October 1, 2006 -present	Number of Veterans, including those employed by the VA (see Additional Information for exclusion criteria) seen outpatient, inpatient, or fee basis nationwide from 10/01/2006 up to but not including the day of query.	Outpatient, Inpatient, Purchased Care, Patient	12,325,216
Number of Veteran nationwide with at least one stay in a VA CLC (nursing home)* October 1, 2006 -present	inpatient, fee basis			Number of patients from row 7 who had at least one stay inpatient in a VA CLC (see Ward Locations used in query) or one fee basis nursing home stay (see Fee Programs, yellow highlighted programs used in query), nationwide, from 10/01/2006 up to but not including the day of query.	Inpatient, Purchased Care	561,586

\*1. Specialty like %nh%

\*2. BedSection equal to 'CLC', 'NH', or 'NHCU'

\*3. MedicalService equal to 'NHCU'

### Limitation:

Cohort has been created based on DART application Data Sources screen and CDW Domain Checklist. Cohort size may be changed and/or accuracy may be increased by considering the additional data sources below.

Enrollment ADR (additional veteran flag)

From communication with researcher

Data sources we couldn't use

Usually more information on other tabs (CPT or ICD codes used, drug names, LOINC codes, etc.)

Data sources we used

Final cohort number should approximately match patient count in provisioned cohort and feasibility

# Time Range



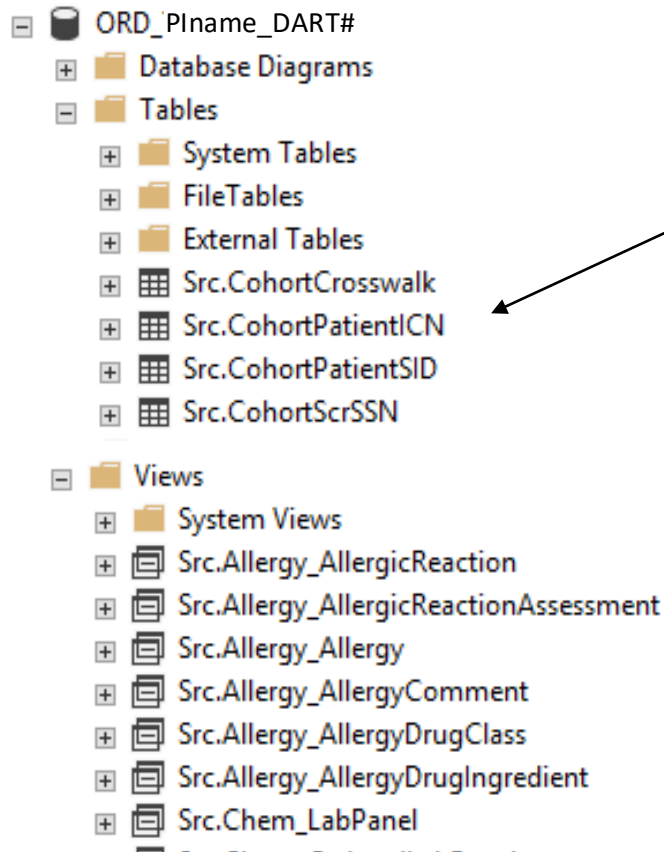
- The time range used to define the cohort is separate from the time range used for provisioning
  - For the cohort: How do we decide who gets included? E.g. Patients with a knee procedure in CY2022
  - For the provisioning: How do we decide which records to put in your database? E.g. Records from 2012 - present
- Same not-too-big, not-too-small consideration as for the cohort



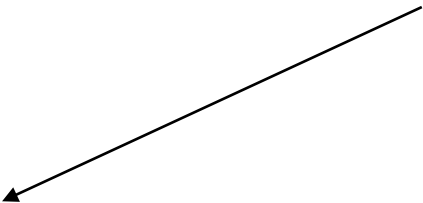


- All fact table views provisioned from CDW Production have had a name change from CDWork (the “true source” containing all VHA data) to your study database, which is a subset of the “true source”:
  - CDWork.Schema.Table -> ProjectDB.Src.Schema\_Table
  - Example: CDWork.Outpat.Visit -> ORD\_Holbrook202202042D.Src.Outpat\_Visit
- Besides that, all CDW metadata and best practices apply!
  - [VINCI:CDW Meta Data Search SOP, VINCI - VA Phenomics Library](#)
  - [Data Services & Field Support \(sharepoint.com\)](#)
    - Six Simple Steps
    - CDW Best Practices

# Cohort Creation vs Data Provisioning



- The cohort
  - Created from methodology shown previously.
  - Static, unless you request otherwise.



- The provisioned views
  - Created from your cohort and time range.
  - Contain all records for those patients and time range.
  - Dynamic...sort of.



# Tip: Request Cohort Updates If You Need Them



- If you have requested data up to present, then new records will appear over time...
- But only for your cohort!
  - Because remember that the cohort itself was static.
- And only for stations at which your cohort was seen at the time of cohort creation

VINCI checks provisioned data sources for new records **for your cohort's PatientSIDs\*** daily.

\*PatientSIDs are a station-level patient identifier. One patient will have different PatientSIDs at each station they've visited

# Examples



- Patient has been seen in two stations at time of ORD creation. Study window is “up to present.”
  - Both PatientSIDs are captured.
  - All outpatient diagnoses are captured from those two stations, with new ones appearing daily, as the patient completes encounters.
  - If the patient goes to a VA where they have never been seen before, a new PatientSID will be created in the CDW.
    - It will not appear in your cohort, because the cohort table is static.
- A different patient never got the diagnosis of interest, and so was not captured in the cohort.
  - If the patient gets a visit today with the diagnosis of interest, they will not appear in the cohort.

# You Can Request Cohorts Be Updated

---



- Use the correspondence site, just like for other communications.
- This will “fix” both of the issues from the previous slide.
- Alternatively, you can request static data on the rb01 server

# Other Tips

---



- Remember that your project protocol/IRB approval should match what you request in the DART
- If you do need to update your DART for a new data source, remember to post to the correspondence site after approval

# Final Tips

---



- We can help!
  - Feasibilities/DNAs are valuable.
  - Join office hours: [VINCI Training & Office Hour \(va.gov\)](#)
  - Ask questions or request DNAs/Feasibilities at [VINCI@VA.gov](mailto:VINCI@VA.gov)

# Acknowledgements

---



- VINCI Data Services and VINCI IT
- SAS/Grid Admins
- VIREC
- CDW Data Services and Field Support



**QUESTIONS?**



**THANK YOU!**