A Primer to VA Data
Accessing, Requesting, and Analyzing

Laura A. Graham, PhD, MPH
Health Services Researcher, VA Health Economics Resource Center (HERC)
Postdoctoral Fellow, S-SPIRE, Stanford University
Prior Steps

• Before requesting data access...
  • Write a clear protocol
  • Know your study team
  • Get VA Access (Employed, WOC’ed)
  • Get IRB Approval!
Seeking Information on VA Data?

- **VINCI Data Viewers (Dim and Meta Tables)**
  - Can be found under VINCI applications

- **VHA Data Portal**
  - Great resource for VHA data

- **VHA Data Architecture Repository**
  - [https://vaausdarmul81.aac.dva.va.gov/](https://vaausdarmul81.aac.dva.va.gov/)
  - Detailed data dictionary for CDW tables
  - Search string can be useful for finding data buried in CDW tables
Welcome to the VA Informatics and Computing Infrastructure

Applications

VINCI has a lot of applications both inside and outside the VINCI Workspace to help our customers achieve their research goals. From VINCI applications and utilities to pre-installed VINCI Standard Workspace applications, to self-installed Development Workspace applications. If you have your own government licensed software, new software is provided when possible. VINCI applications and utilities listed below are subject to change as newer software becomes available.

VINCI Applications

These are VINCI created non-VINCI Workspace applications.

Click on an application to launch it.

DART  Dim Data Viewer  Meta Data Viewer  VINCI Workspace Download  VINCI Workspace Upload  Let’s Collaborate VINCI

Showing 1 to 5 of 5 entries

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DART</td>
<td>Data Access Request Tracker for requesting research data access from National Data Systems (NDS)</td>
</tr>
<tr>
<td>Dim Data Viewer</td>
<td>For Exploring CDW Dimension Table Data.</td>
</tr>
<tr>
<td>Meta Data Viewer</td>
<td>For Exploring CDW Meta Data Table Data.</td>
</tr>
<tr>
<td>VINCI Workspace Download</td>
<td>Utility for downloading data from the VINCI Workspace</td>
</tr>
<tr>
<td>VINCI Workspace Upload</td>
<td>Utility for uploading data to the VINCI Workspace</td>
</tr>
</tbody>
</table>
# pulmonary_function_tests_700 Fields

Showing 1 to 10 of 53 entries

<table>
<thead>
<tr>
<th>Name</th>
<th>Data Type</th>
<th>PK</th>
<th>PK</th>
<th>Field Codes</th>
<th>ERDiagrams</th>
<th>VISTA File</th>
<th>VISTA Field</th>
<th>VISTA Field #</th>
</tr>
</thead>
<tbody>
<tr>
<td>smoker</td>
<td>varchar(30)</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td>Unknown</td>
<td>Unknown</td>
<td>0</td>
</tr>
<tr>
<td>Status</td>
<td>smallint(5)</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td>Unknown</td>
<td>Unknown</td>
<td>0</td>
</tr>
<tr>
<td>PulmonaryFunctionTests</td>
<td>varchar(50)</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td>Unknown</td>
<td>Unknown</td>
<td>0</td>
</tr>
<tr>
<td>Date/Time</td>
<td>datetime2(T)</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td>Unknown</td>
<td>Unknown</td>
<td>0</td>
</tr>
<tr>
<td>Summary</td>
<td>varchar(25)</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td>Unknown</td>
<td>Unknown</td>
<td>0</td>
</tr>
<tr>
<td>PatientID</td>
<td>varchar(50)</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td>Unknown</td>
<td>Unknown</td>
<td>0</td>
</tr>
<tr>
<td>PatientSID</td>
<td>int(10)</td>
<td>Y</td>
<td>N</td>
<td></td>
<td></td>
<td>Unknown</td>
<td>Unknown</td>
<td>0</td>
</tr>
<tr>
<td>Height/Cm</td>
<td>int(10)</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td>Unknown</td>
<td>Unknown</td>
<td>0</td>
</tr>
<tr>
<td>Height/inchstents</td>
<td>numeric</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td>Unknown</td>
<td>Unknown</td>
<td>0</td>
</tr>
<tr>
<td>Weight/Poundstents</td>
<td>numeric</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td>Unknown</td>
<td>Unknown</td>
<td>0</td>
</tr>
</tbody>
</table>
Seeking Information on VA Data?

• VA Informatics and Computing Infrastructure (VINCI) Data Sources
  • https://vaww.vinci.med.va.gov/vincicentral/DataSources/

• VA Health Economics Research Center (HERC)
  • http://vaww.herc.research.va.gov/

• VA Information Resource Center (VIReC)
  • http://vaww.virec.research.va.gov/

• Office of Performance Measurement (OPM)
  • https://vaww.car.rtp.med.va.gov/

• Planning System Support Group (PSSG)
  • http://vaww.pssg.med.va.gov/

• Data.gov / Corporate Databases Monograph
  • www.data.gov
Readily Available Data Sources

• VHA Support Service Center
  • [https://vssc.med.va.gov/](https://vssc.med.va.gov/)
  • Facility-level information
    • Business Operations
    • Clinical Care
    • Quality & Performance

• National Center for Veterans Analysis and Statistics
  • [https://www.va.gov/vetdata/](https://www.va.gov/vetdata/)
DART

Data Access Request Tracker
Prior Steps

• Before requesting data access...
  • Write a clear protocol
  • Know your study team
  • Get VA Access (Employed, WOC’ed)
  • Get IRB Approval!
Data Access Request Tracker (DART)

Most of your data access requests will be covered by the **Data Access Request Tracker (DART)**

- [https://dart.vha.med.va.gov/](https://dart.vha.med.va.gov/)
DART Data Sources

- ADUSH Enrollment Files
- Bereaved Family Survey
- CAPRI & Joint Legacy Viewer
- Care Assessment Need Score
- Clinical Assessment Reporting and Tracking (CART) CV
- Corporate Data Warehouse (CDW)
- COVID-19: Shared Data Resource
- DoD-VA Infrastructure for Clinical Intelligence (DAVINCI)
- Geriatrics and Extended Care (GEC) RAI/MDS
- Health Economics Resource Center (HERC) Cost Data
- Homeless Veterans Registry
- Lung Cancer Screening Demonstration Project (LCSDP) Cohort
- Managerial Cost Accounting (MCA)
- Medical SAS Files & VETSNET FILE
- Million Veteran Program (MVP)
- OMOP Common Data Model
- Patient Aligned Care Team Implementation Index (PACT Pi2)
- PSSG Geocoded Enrollee Files
- SAS Fee
- Traumatic Brain Injury Screening and Evaluation
- VA Surgical Quality Improvement Program (VASQIP)
- VINCI National Language Processing Output
- Vital Status File
- VSSC Self-Service Web-Based Products
Data Access Request Tracker (DART)
Data Access Request Tracker (DART)

- Required Documents

  - **Research Project Documents and Approvals**
    - Research study protocol
    - Research and Development Committee Approval Letter(s)
    - Institutional Review Board (IRB) Approval Letter(s)
    - IRB Approved HIPAA Authorization or [Waiver of HIPAA Authorization](#)
    - IRB Approved Sample Informed Consent or Waiver of Informed Consent
Data Access Request Tracker (DART)

- Additional documents that may be required:
  - **Data Request Forms**
    - [Research Request Memo](#) – Replaces the 9957 for **ALL** research requests submitted through DART after March 25, 2015
    - [Real SSN Access Request](#) – Required for access to real SSNs
    - [CDW Domain Check List](#) – **CDW data only**
    - [Vital Status Rules of Behavior](#) – **Vital Status File only**
    - [Special User Access Request Form for Researchers](#) – **CAPRI and VistAWeb only**
    - [Surgery Data Research Proposal Template](#) – **Surgical Quality Data Use Group (SQDUG) data only**
    - [National Surgery Office (NSO) Data Use Agreement and Security Information](#) – **SQDUG data only**
Data Access Request Tracker (DART)

• Once all required documents are uploaded and the request has been submitted, it will undergo a privacy review.

• Additional reviews are required for the following:
  • Requests for data with real SSNs will undergo review by the Office of Research & Development.
  • Requests will undergo a security review using the information provided in the Research Request Memo.
Data Access Request Tracker (DART)

• When do I need to request **Real SSNs**?
  
  • You’re planning to use chart abstraction
  • You’re linking to another existing data source by SSN
  • You’re planning to use the TIU (Text Integration Utilities)

• Everything else is linked by either **PatientIEN** or **Sta3n / PatientSID** or **SCRSSN** and you will receive a crosswalk file once your request is approved
Data Access Request Tracker (DART)

- **After data access approval** a VINCI Data Manager can help with the following:
  - Cohort selection
  - Data Extraction
  - Formatting, Indexing, and Security
VINCI
The VA Informatics and Computing Infrastructure
Why use VINCI?

• Don’t have to pay for and install software
• Simplifies data access to VA’s most common data sources
• Meets and maintains data all VA IRB data security standards
• Provides a large amount of data space for analyses and data storage
• Developmental workspaces can be used to provide more flexibility, if needed.
INCEPTION
THE ARCHITECTURE
DESIGN BY RICK SLUSHER
VINCI Access

- [https://vaww.vinci.med.va.gov/vincicentral/](https://vaww.vinci.med.va.gov/vincicentral/)

- Accessed through the **VA Intranet**

- Automatically granted with a DART request
  - Make sure to request access to ”The Grid” separately after the DART request is approved. Grid access is not automatically granted but only has to be requested once for a person.
Logging into VINCI

https://vaww.vinci.med.va.gov/vincicentral/
Logging into VINCI

Select VINCI Workspace

- Full Standard Workspace (1 Monitor)
- Full Standard Workspace (multi-monitor)
- Standard Workspace Applications
- Development Workspace (1 or 2 Monitor)

Remote Desktop Connection

Do you trust the publisher of this remote connection?

Publisher: vhacdwhfhrdrk.vha.med.va.gov
Type: Remote Desktop Connection
Remote computer: VHACDWHFHRD8KR.VHA.MED.VA.GOV
Gateway server: vhacdwhfhrd8kr.vha.med.va.gov

Enter your credentials:

User name: LAURA A GRANET
Password: [REDACTED]
Smart card credentials: [REDACTED]

Enter PIN: [REDACTED]
Logging into VINCI
VINCI Standard Workspace!
VINCI Set-up

- Windows 2012R2 Workspace
  - 4 CPUs **sharing** 128 GB RAM
  - No internet

- 3 SQL Servers
  - **RB01**: vhacdwdb01.vha.med.va.gov
  - **RB02**: vhacdwdb02.vha.med.va.gov
  - **RB03**: vhacdwdb03.vha.med.va.gov

- Your study will be assigned to one of those based on data needs and available space.
VINCI Folder Layout

- Same as on a desktop
  - Hard Disk Drives (K, L, M, N, T, S, U, V, X)
  - Network Drives
    - P:/ drive (Projects)
    - O:/ drive (Projects2)
    - H:/ drive
VINCI Folder Layout

- **P:/ (Projects) and O:/ (Projects2)**
  - All project folders
  - 100GB

- **H:/**
  - Personal folder (10GB)
  - Do not store any project data here

- All project folders have a default sub-folder named **Uploads** where uploaded information is included. You can add more folders as needed.
VINCI Folder Layout

• Data can be transferred into VINCI using the VINCI Upload Tool on the homepage

• Data can also be transferred out of VINCI using the VINCI Download Tool but make sure you are approved to do this before downloading files off of VINCI
  • Approvals are granted with your DART application. Contact VINCI if you have any questions.
More Help?

• More information on accessing and using the VINCI Workspace are found in the **VINCI Workspace User Guide**
VINCI

Applications
VINCI Standard Workspace Applications

- Accessed via **remote desktop connections** from within VINCI

- The Standard Workspace has limited software availability.
  - Log into each application separately
  - Each will be a separate desktop / server
  - No copying/pasting between servers
VINCI Standard Workspace!
VINCI Applications
VINCI Applications Access (Option 2)
VINCI Applications Access (Option 2)
VINCI Applications Access (Option 2)
VINCI

SQL Server Management Studio
VINCI Standard Workspace!
SQL Server Management Studio

• Method for accessing and viewing CDW domain tables with SQL code
• Can also be used to pull data in a .txt format if needed
• Requires some initial set up to link to servers and some SQL knowledge but also has point-and-click options that make it easy to use
SQL Server Management Studio
Set Up
SQL Server Management Studio
Set Up

• **VINCI Database User Guide**
  • Section 4.2.1 (pages 3-4) provides detailed information on SQL server set up
  • If this doesn’t work then check to see if you have Grid access.

**Default Settings To Use**

- Server Type: Database Engine
- Server Name: `<Database Server Name>.vha.med.va.gov`
- Authentication: Windows Authentication
SQL Server Management Studio
Database Access

• **VINCI Database User Guide**
  • Section 4.2.2 (pages 4-7) provides detailed information on accessing databases

• Overview
  • **Databases > CDWWork** contains your dimension tables
  • **Databases > ORD_XXXX > Views** contains your SQL data tables (fact tables)
  • Right-click and "Select Top 1000 Rows" for easy viewing of a data source
SQL Server Management Studio
Data Pulls

• SQL Server Management Studio is best used for viewing data structures but it can also be used to run smaller data pulls

• Once your query is set up
  • Use “Query > Results To” to change to “Results to File”
  • Then “Query > Query Options...” to set up your file but navigating through “Results > Text”
VINCI

SAS and the Grid
VINCI SAS

- VINCI SAS (SAS 9.x) can be access from the Remote Applications folder in the Standard Workspace OR directly from the Applications webpage.
VINCI SAS

• Several versions of SAS are available for use.
  • SAS 9.2
  • SAS 9.4
  • SAS Enterprise Guide 7.1
  • SAS Enterprise Guide 8.2
VINCI SAS

• **SAS 9.2** and **9.4** are not on “The Grid”
  • No access to most statistical procedures (ex. PROC TTEST)

• Can still be used for data pulls and basic data management
VINCI SAS (i.e. The Grid)

• “Grid” refers to the Linux servers where SAS Enterprise Guide runs
• The Grid is more efficient and offers better data security than other VINCI SAS options
  • ~45TB of space!
  • You will need a slightly different libname reference: “/data/dart/2014/ORD_Smith_201401001D/“
  • You will also need to use WINSCP for data transfer
• Check out “Grid – Where to Begin?” for more information
VINCI SAS Help

• Lots of **User Guides** for reference
• This course is only an intro but these will answer your more advanced questions.

<table>
<thead>
<tr>
<th>Type</th>
<th>Guide Name</th>
<th>Description</th>
<th>Last Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grid - Where to Begin?</td>
<td>This document gets you started by providing information on using the Grid for new users.</td>
<td>2016-02-28</td>
</tr>
<tr>
<td></td>
<td>Grid Computing in SAS 9.4</td>
<td>This guide is written by SAS on how to use the 9.4 SAS Grid.</td>
<td>2014-07-29</td>
</tr>
<tr>
<td></td>
<td>How to use PCFILES engine</td>
<td>Demonstrates ways for the Grid platform to work with MS Excel, MS Access and STATA data files existing on Windows network shares.</td>
<td>2017-04-25</td>
</tr>
</tbody>
</table>
WINSCP Set Up

• The first step to using SAS Enterprise Guide 7.1 is to check your Linux system registration by setting up WinSCP

• This can be done by following the instructions in Using Enterprise Guide with the Grid
VINCI Enterprise Guide 7.1

- After WINSCP is set up, you will need to set up your profile in SAS
- This can be done by following the instructions in Using Enterprise Guide with the Grid
VINCI Enterprise Guide 7.1
VINCI Enterprise Guide 8.2 (64-bit)

• New!

• Imports your profile settings from VINCI Enterprise Guide 7.1
VINCI

Data Pulls and R within SAS!
Pulling Data with SAS

• Do everything you can to avoid inefficient programs

• SQL Pass-Through queries!
VINCI SAS Data Pulls

• I initially set up the **test data pulls in SQL Management Studio** using a pull of the top 100.

```sql
/* Script for SelectTopNRows command from SSMS */
SELECT TOP 100 [Sta3n],[PatientSID],[AdmitDateTime],[DischargeFromFacility],
[DischargeDateTime],[PrincipalDiagnosisICD9SID]
FROM [ORD_Morris_201412124D].[Src].[Inpat_Inpatient]
```

• If the code works in SQL then you can use PROC SQL in SAS to connect directly to the SQL databases and pull the full data
proc sql;
    connect to olsdb as CDW(datasource="VHACDWRBR01.VHA.MED.VA.GOV" provider=SQLNCLI11
    properties="{'initial catalog':ORD_LEE_201110025D 'Integrated Security':SPSI}"
    execute(
        SELECT DISTINCT b.[Sta3n], b.[PatientSID], a.[scrssn],a.[NH_admitdate],a.[NH_disdate],
        parse(a.[NH_admitdate] AS date) as AdmitDate, parse(a.[NH_disdate] AS date) as DisDate
        INTO $POP
        FROM [ORD_Lee_201110025D].[Dfit].[CLCcohort_90day_32481] as a
        LEFT JOIN [ORD_Lee_201110025D].[Src].[SPatient SPatient] as b
        ON a.scrssn = b.scrssn
        GROUP BY b.[Sta3n], b.[PatientSID], a.[scrssn],a.[NH_admitdate],a.[NH_disdate]
    ) by cdw;
    create table POP as
    select *
        from connection to CDW(select * from $pop);
quit;
VINCI SAS Data Pulls

• You can also use SAS 9.2 or SAS 9.4 to tunnel into the CDW and pull data directly into SAS.

• This requires the following code:

```sas
proc setinit noalias; run;

* Source Libname;
libname SRC ojdbc provider=SQLCL11 datasource="VHACDWR802.VHA.MED.VA.GOV"
    properties=('initial catalog'=ORD_Morris_201412124D 'Integrated Security'=SSPI) schema=src;

* Dimension Tables;
libname DIM ojdbc provider=SQLCL11 datasource="VHACDWR802.VHA.MED.VA.GOV"
    properties=('initial catalog'=CDUWork 'Integrated Security'=SSPI) schema=Dim;
```

• The **SRC Catalog** is changed to reference your project folder.

• **Datasource** may also need to be changed based on your project server.
R within SAS?

- You can run R from within SAS through PROC IML
- Make sure you have changed your server to the Rapp Server
R within SAS?
VINCI

Need More Help?
Policies

Protect Your Data
Policies for Electronic Data

• **VA Handbook 6500**
  • Definitions and guidance for the VHA

• **Key Points**
  • **USB drives are forbidden**
  • Do not store sensitive information on a Hard Drive (C:/ Drive). Try not to store anything on a hard drive.
  • Sensitive information can never be transported, accessed, processed or otherwise used offsite. **DON’T EMAIL DATA!**
Types of Data

1. Individually Identifiable Health Information
2. Protected Health Information (PHI)
3. Personally Identifiable Information (PII)
4. VA Sensitive Information/Data

• VHA Handbook 6500 provides guidance on definitions and management
Disposal of Data

• Governed by the VHA Record Control Schedule (RCS 10-1)

• Current guidance (Item 8300.6, Page III-8-6)
  • Temporary; cutoff at the end of the fiscal year after completion of the research project.
  • Destroy 6 years after cutoff, may retain longer if required by other Federal regulations. (DAA-0015-2015-0004, item 0032)
  • If the investigator leaves VA, all research records are retained by the VA facility where the research was conducted.
  • If the grant is ongoing and the investigator leaves one VA facility to go to another VA facility, the investigator must obtain approval for a copy of relevant materials to be provided to the new VA facility's research office.

• The investigator is not the grantee, nor does the investigator own the data.
Thank you!

Laura.Graham@va.gov
lagraham@Stanford.edu

@lagrahamepi
Corporate Data Warehouse

- National repository of clinical and administrative data

- Stored as a relational database

- Data is available beginning October 1, 1999 and refreshed on a nightly basis
Corporate Data Warehouse

• **Production Domain**
  • Modeled data domains built to make using the information easier

• **Raw Domain**
  • Not obtained through normal CDW extraction process
  • Often require more cleaning, have limited documentation, and are not extracted continuously
  • Works-in-Progress and will eventually become production domains
Corporate Data Warehouse

Production (38)
- Inpatient
- Outpatient
- Patient
- Patient Enrollment
- Surgery
- Pharmacy
- Vital Sign
- Lab Chem
- Mental Health
- Purchased Care Service
- ...

Raw (12)
- Echocardiogram
- Intravenous Meds
- Oncology
- Prosthetics
- Pulmonary Function Test
- Fee Basis Claim System
- ...
- ...
Data Layout

- Relational data have been separated out into multiple tables
  - Fact and Dimension tables
  - Joining keys are added so they can be reassembled

### Patient Table (Fact)

<table>
<thead>
<tr>
<th>PatientKey</th>
<th>Gender</th>
<th>DOB</th>
<th>Zipcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>07/01/1950</td>
<td>35698</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>09/25/1942</td>
<td>45689</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>01/01/1975</td>
<td>78236</td>
</tr>
</tbody>
</table>

### Visits Table (Fact)

<table>
<thead>
<tr>
<th>PatientKey</th>
<th>DxKey</th>
<th>VisitKey</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1/1/1999</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>6/1/2002</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>9/1/2005</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>9/15/2005</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>5</td>
<td>10/1/2007</td>
</tr>
</tbody>
</table>

### Diagnosis Code (Dimension Table)

<table>
<thead>
<tr>
<th>DxKey</th>
<th>ICD9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>250.01</td>
</tr>
<tr>
<td>2</td>
<td>250.10</td>
</tr>
<tr>
<td>3</td>
<td>250.23</td>
</tr>
</tbody>
</table>

...
Thank you!