

What happens after your DART is approved

HSR&D Cyber Seminar

December 3, 2015



VA Informatics
and Computing
Infrastructure

Agenda

- Introduction (Hamid)
- Correspondence site & initial message (Ryan)
- Project types (Suga)
- Static vs. Live data + other datasets (Greg)
- Cohort selection process (Allen)
- Data delivery process (Allen)

Introduction

- You will work with the VINCI team if:
 - You selected CDW and/or other VINCI data
 - Need to use VINCI SQL Server
 - Selected to use VINCI for storing your data

- You will not work with the VINCI team if:
 - You only selected non CDW and/or VINCI data
 - And selected to use your local server for storing data

What VINCI offers users

- Secure analytical workspace for VA researchers
- High performance servers and storage
- Software tools, custom applications
- Access to data analysis software tools, SAS Grid
- RDP connection - Cloud computing
- Standard and development workspaces
- Large collection of VA datasets
- VINCI Data Managers to assist with project data needs

VINCI Data Managers

- Allen Faler
- Gregory Scavezze
- Ryan Hecker
- Sergey Barmin
- Sundaragandhi (Suga) Radhakrishnan

VINCI Data Managers

- Team of 5 dedicated to serve data to research projects (100%)
- Experienced SQL programmers
- Knowledge of VA data
- Efficient in data preparation
- Served over 1,200 VA research projects

Correspondence Site

- After DART approval by NDS and creation of an Active Directory group for the study, the VINCI Data Manager will create a study-specific correspondence site.
- The site keeps all communication in one convenient place rather than dozens of individual emails.
- Email alerts are generated whenever a message is added to the site
- Only goes to valid VA email addresses

VINCI Correspondence Site

Home

Forms & Documents

Data Description

EDIT LINKS

VINCI

EDIT LINKS

ExampleStudy_201512101D

Home

Forms & Documents

Data Description

EDIT LINKS

VINCI Data Project Correspondence

+ new item or edit this list

Message

Modified By

Created

Hecker, Ryan (LongView))

Please do not post any information containing PHI on this site

Please click on the study name link above to go to the website. You may also copy and paste the URL into your browser. Please do not send a reply to the email alert.

Welcome to VINCI

This SharePoint site has been created per your recent DART approval.

Cohort

A cohort of patients is needed for data extraction. Please post a message that indicates whether your cohort is ready by clicking on "new item" above

Welcome Letter

Please click on the study name link above to go to the website. You may also copy and paste the URL into your browser. Please do not send a reply to the email alert.

Welcome to VINCI

This SharePoint site has been created per your recent DART approval.

Cohort

A cohort of patients is needed for data extraction. Please post a message that indicates whether your cohort is ready by clicking on "new item" above.

PHI/PII

Never upload patient health information, including cohort identifiers, to this site. Doing so violates HIPAA and VA security standards in addition to affecting your study's ability to conduct research using VINCI.

Please post your NDS (DART) approved Study Date Range

Based on your date range, we will be able to recommend either a CDW live or a CDW static dataset (if applicable).

Purpose of This Site

This site has been created to facilitate and track correspondence between the VINCI Data Team and your project team.

Messages: Please post messages to the home page of this site to communicate with the VINCI data team. You may post new messages by clicking on "new item" above. All messages will be kept and managed in one convenient place on this site, which will eliminate the need for the use of email to communicate regarding data provisioning for your project. This site will send you and your VINCI data manager an email alert when a new message is posted; each email alert contains a link to this site.

Documentation: Please look under the "Data Description" tab for detailed descriptions of all of the information domains available.

Thank You,

The VINCI Data Team



Project Types

- Discuss various data project types on VINCI
- What do we do on VINCI to accommodate various project types?
- How to differentiate project types on DART?
- Find out appropriate NDS forms for each type in DART

Project Types

- VINCI Extract
- VINCI Download
- CDW Extract
- Preparatory Research

VINCI Extract

- All study related data stays on VINCI servers (VHACDWRB02 or VHACDWRB01)
- All analysis takes place on VINCI environment
- PHI/PII data cannot leave VINCI environment
- Can import third party datasets to VINCI
- Studies have access to VINCI developmental workspace
- Studies can use VINCI scrub datasets for cohort creation

VINCI Download

- Source data resides on VINCI servers (VHACDWRB02 and VHACDWRB01)
- Some Analysis takes place on VINCI environment
- **Final PHI/PII data can be downloaded to local server using VINCI FTP tools**
- Can import third party datasets to VINCI
- Studies have access to VINCI developmental workspace
- Studies can use VINCI scrub datasets for cohort creation

CDW Extract

- Study data imported to local server
- Study database is on VHACDWDBS03 server
- All analysis takes place on local server
- VINCI workspace access and VINCI scrub dataset access is not available for this type

Prep to Research

- Preparatory research studies are considered as VINCI extracts. Hence...
- All study related data stays on VINCI servers (VHACDWRB02 or VHACDWRB01)
- All analysis takes place on VINCI environment
- PHI/PII data cannot leave VINCI environment
- Can import third party datasets to VINCI
- Studies have access to VINCI developmental workspace
- Studies can use VINCI scrub datasets for cohort creation

What Do We Do?

- We create source databases on VINCI servers
- VINCI download studies will be given option to download data using VINCI FTP tools
- CDW Extract studies will have a destination database on VHACDWDBS03 server which is located outside of VINCI environment
- VINCI studies will be able to create their own objects under DFLT schema
- Src is used when VINCI data team creates any database objects
- Dbo schema can be used to store data not visible to studies
- Cohorts are typically stored under Src schema tables
- All domain data extracts will be in the form of Views in Src schema

CDW Extract - DART

Information Participants **Data** Documents Submit

History Communication

2015-06-029-D Predicting Risk of 72-hour and 30-day Emergency Department Revisits

Data Sources

Data Storage Location *

VINCI

Local VA Server Location

Name of facility * Albany Stratton VA Medical Center

Address * 113 Holland Ave, Akbany NY, 12208

Building Main Building

Room # D115

Will data be transferred external to the VHA? *

No

Yes (DUA is required)

Currently CDW cannot approve a research request that would include disclosing the CDW data (identifiable or de-identified) outside of VHA without a signed research consent and HIPAA authorization from the individual because of the risk of re-identification. The data may be used (internally or within VHA) for the approved protocol. If "finder" files are to be sent out to Federal or state agencies for the purpose of the Federal or state agency disclosing information on the individual back to VA per SOR 34VA12, a DUA between the PI and the external agency must be submitted with the DART application.



CDW Extract - Research Request Memo

At least one signature is invalid. Please fill out the following form. You can save data typed into this form.

To: Director, National Data Systems

The following information is required and all signatures must be obtained before any review of this request can take place:

Is this Principal Investigator a VA employee? Yes No

Is this request for data use for a VA research study (includes pilot studies)? Yes No

Is this request for activities preparatory to research? Yes No

Check the type(s) of data needed: Real SSN Scrambled SSN PHI but No SSN

Is AITC Main Frame access being requested? Yes No

Is VSSC and/or MCA Web Reports access being requested? Yes No

Will any data be transferred outside of the VA? Yes No

Will the data be stored in the VINCI Environment? Yes No Both

Complete the fields below indicating where the data will be stored within the VA network:

Facility/Company Name

Building Number Room Number

VINCI Download - DART

https://vhacdwtst02.vha.med.va.gov:8443/vinci_dart_client/dart/

VINCI Data Project Tracking

DART

File Edit View Favorites Tools Help

VINCI Data Access Request Tracker > Dashboard > Requests

Welcome Radhakrishnan, Sundaragandhi P. (LongView)

Data Storage Location *

VINCI

Local VA Server Location

Name of facility * VA Boston

Address * 150 South Huntington Ave, Boston MA 01230

Building 1

Room # Floors 13 and 14

Will data be transferred external to the VHA? *

Currently CDW cannot approve a research request that would include disclosing the CDW data (identifiable or de-identified) outside of VHA without a signed research consent and HIPAA authorization from the individual because of the risk of re-identification. The data may be used (internally or within VHA) for the approved protocol. If "finder" files are to be sent out to Federal or state agencies for the purpose of the Federal or state agency disclosing information on the individual back to VA per SOR 34VA12, a DUA between the PI and the external agency must be submitted with the DART application.

No

Yes (DUA is required)

IDENTIFIERS [\(More about identifiers\)](#)

Real SSN

Scrambled SSN

Identifiable data but no real or scrambled SSNs

REQUESTED DATA SETS



VINCI Download - Research Request Memo

https://vhacdwtst02.vha.med.va.gov:8443/vinci_dart_client/dart/

VINCI Data Project Tracking DART DART vhacdwtst02.vha.med.va.g...

File Edit Go to Favorites Help

1 / 4 164%

Fill & Sign Comment

Date: 4/17/2015

From: Principal Investigator B. Price Kerfoot MD

Subj: Research Data Request Memo for: Analysis of health measure variation among patients with d

To: Director, National Data Systems

The following information is required and all signatures must be obtained before any review of this request can take place:

Is this Principal Investigator a VA employee? Yes No

Is this request for data use for a VA research study (includes pilot studies)? Yes No

Is this request for activities preparatory to research? Yes No

Check the type(s) of data needed: Real SSN Scrambled SSN PHI but No SSN

Is AITC Main Frame access being requested? Yes No

Is VSSC and/or MCA Web Reports access being requested? Yes No

Will any data be transferred outside of the VA? Yes No

Will the data be stored in the VINCI Environment? Yes No Both

Complete the fields below indicating where the data will be stored within the VA network:

Facility/Company Name MAVERIC, VA Boston Healthcare System

VINCI Download - DART

https://vhacdwttest02.vha.med.va.gov:8443/vinci_dart_client/dart/

VINCI Data Project Tracking

DART

VINCI Data Access Request Tracker > Dashboard > Requests

Welcome Radhakrishnan, Sundaragandhi P. (LongView)

Data Storage Location *

VINCI

Local VA Server Location

Name of facility * VA Boston

Address * 150 South Huntington Ave, Boston MA 01230

Building 1

Room # Floors 13 and 14

Will data be transferred external to the VHA? *

Currently CDW cannot approve a research request that would include disclosing the CDW data (identifiable or de-identified) outside of VHA without a signed research consent and HIPAA authorization from the individual because of the risk of re-identification. The data may be used (internally or within VHA) for the approved protocol. If "finder" files are to be sent out to Federal or state agencies for the purpose of the Federal or state agency disclosing information on the individual back to VA per SOR 34VA12, a DUA between the PI and the external agency must be submitted with the DART application.

No

Yes (DUA is required)

IDENTIFIERS (More about identifiers)

Real SSN

Scrambled SSN

Identifiable data but no real or scrambled SSNs

REQUESTED DATA SETS

VINCI Download - Research Request Memo

https://vhacdwtst02.vha.med.va.gov:8443/vinci_dart_client/dart/ VINCI Data Project Tracking DART DART vhacdwtst02.vha.med.va.g...

File Edit Go to Favorites Help

1 / 4 164%

Fill & Sign Comment

Date: 4/17/2015

From: Principal Investigator B. Price Kerfoot MD

Subj: Research Data Request Memo for: Analysis of health measure variation among patients with d

To: Director, National Data Systems

The following information is required and all signatures must be obtained before any review of this request can take place:

Is this Principal Investigator a VA employee? Yes No

Is this request for data use for a VA research study (includes pilot studies)? Yes No

Is this request for activities preparatory to research? Yes No

Check the type(s) of data needed: Real SSN Scrambled SSN PHI but No SSN

Is AITC Main Frame access being requested? Yes No

Is VSSC and/or MCA Web Reports access being requested? Yes No

Will any data be transferred outside of the VA? Yes No

Will the data be stored in the VINCI Environment? Yes No Both

Complete the fields below indicating where the data will be stored within the VA network:

Facility/Company Name MAVERIC, VA Boston Healthcare System

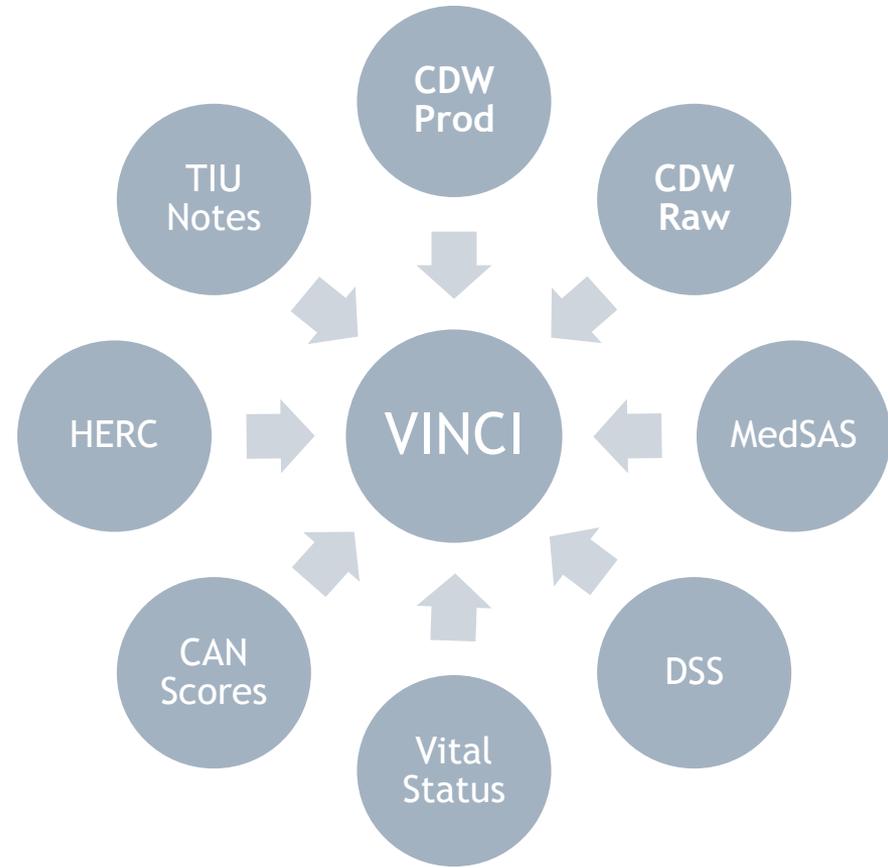
VINCI - Live Data vs. Static Data - Introduction

Projects using VINCI data have the option of choosing live or static data. This presentation will compare the two and discuss the pros and cons of each.

VINCI - Live Data vs. Static Data

VINCI offers access to data for researchers from multiple sources including:

- CDW Production
- CDW Raw
- MedSAS
- MCA (formerly DSS)
- Vital Status
- CAN Scores
- HERC
- TIU Notes



VINCI - Live Data vs. Static Data

The aforementioned data resides on two different VINCI servers:

- VINCI live server (vhacdwr02)
- VINCI “Static” server (vhacdwr01)
- VINCI Data Managers create a project database on one of these two servers based on the needs of the project.
- The “default” server a VINCI Data Manager installs a project database on is the live server, unless otherwise specified by the project.

VINCI - Live Data vs. Static Data

VINCI Live Data

- Data stored on VINCI's live server is refreshed on a regular basis. Both the data and the table structure are refreshed.
- CDW Production data is refreshed nightly.
- Other data (MedSAS, HERC, etc.) are refreshed on a quarterly or annual basis. This is partially determined by how often the stakeholder chooses to provide VINCI with updated data.

VINCI - Live Data vs. Static Data

VINCI Static Data

- Data stored on VINCI's "static" server consists of a snapshot of the data that is taken each fiscal year.
- A "snapshot" means a copy of a database (structure and data).
- A new snapshot is taken at the beginning of each fiscal year of the data and structures' state as of the end of the prior fiscal year.

VINCI - Live Data vs. Static Data

Pros of Live Data:

- Data is updated on a regular basis, in some cases, nightly.
- A project is able to take advantage of any new tables, columns, index changes, etc. introduced by CDW as they are introduced to the production environment.
- Putting a date range on the “fact” views given by a VINCI Data Manager in essence, makes the data static.
- Majority of VINCI projects are using live data with no complaints, issues.

VINCI - Live Data vs. Static Data

Cons of Live Data:

- The data is updated nightly (CDW Production Data), which can affect counts, etc. A fixed date range on the fact views reduces this possibility.
- Changes to columns, indexes, etc. introduced by CDW can, on occasion “break” project views. VINCI Data Managers work hard to mitigate the impact of these changes and make them seamless as possible.

VINCI - Live Data vs. Static Data

Pros of Static Data:

- The data is snapshot and truly “fixed”. It will not change.
- Changes to tables introduced by CDW will not affect fixed data, so the database structure is more stable. So the risk of having broken views, etc. is reduced to almost zero.

VINCI - Live Data vs. Static Data

Cons of Static Data:

- Since the data is a snapshot of how the data existed as of the end of the last FY, the project will not be able to benefit from schema changes (new tables, etc.)
- If the project decides they want data that extends past the end of the snapshot or new tables that have been added, the project will either have to wait until the end of the current fiscal year and move to the new snapshot or will have to have their project database moved to the live server.

VINCI - Live Data vs. Static Data - Summary

Live Data

- Live data is updated on a regular basis (nightly for CDW Production data).
- Domain and table structure is also updated as CDW Makes updates and improvements
- Most projects use Live Data with no issues.

Static Data

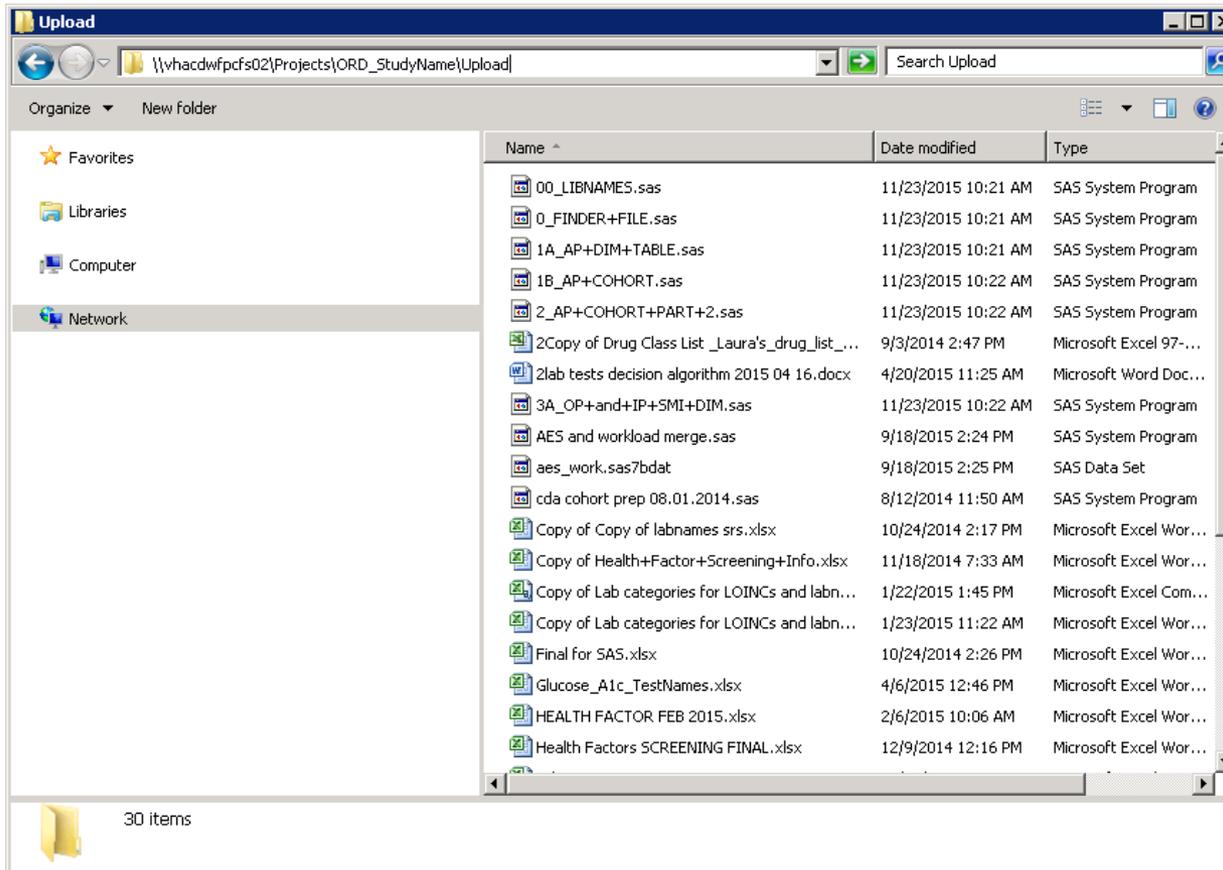
- Fixed data is a snapshot as of the end of the prior fiscal year and will not change (data and structure).
- The project will not be able to benefit from new tables added to data domains.

Cohort/Data Provisioning

- Study Cohort Options
- Cohort Tables
- Domain Views
- Database Connection
- Recommendations
- Third-party Data Sources

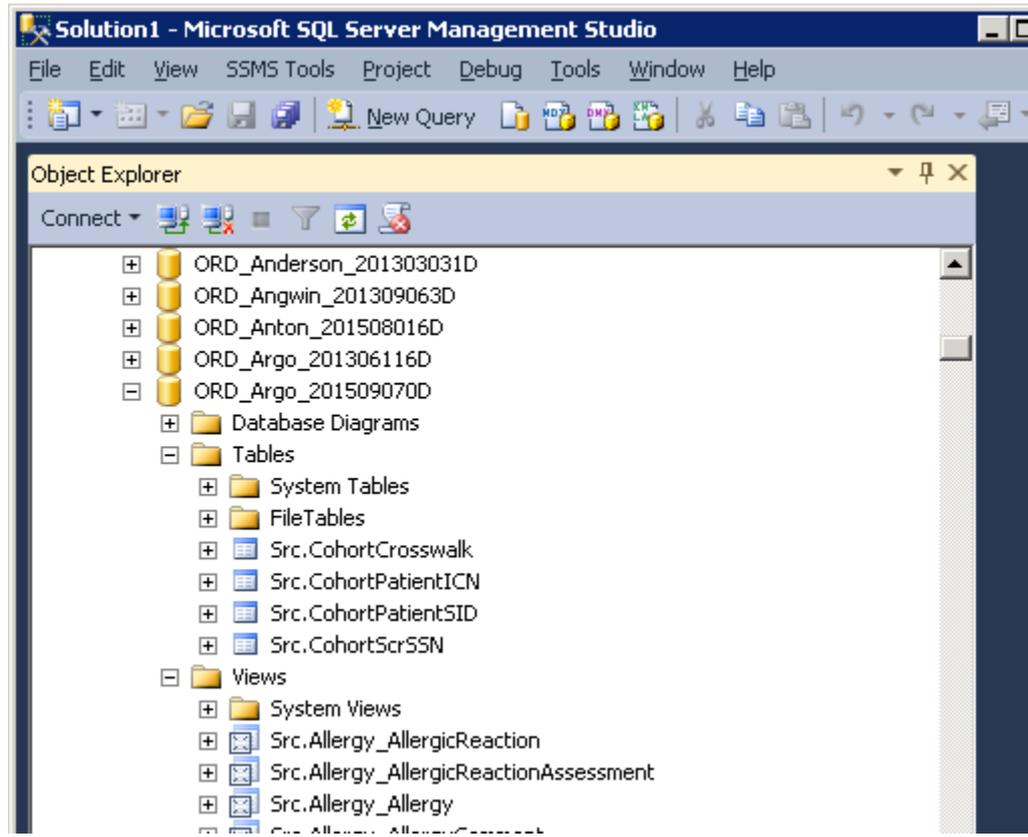
Folder vs. Database

■ Project Folder:



Folder vs. Database

■ Project Database:



Cohort Options

- Upload existing cohort
- Cohort creation based on criteria
- Create your own cohort

Upload existing cohort

■ <http://vaww.vinci.med.va.gov/vincicentral/default.aspx>

VINCI Central About Using VINCI Data Guides Support VINCI Workspace

Home

VINCI
VA Informatics and Computing Infrastructure
The mission of VINCI is to improve the healthcare of Veterans by providing researchers access to integrated national data sets and tools for analysis in a secure, high-performance computing environment. The vision of VINCI is to become VA researchers' preferred data source and data-processing environment.

Learn About VINCI
About VINCI Workspace Software Available Data

I Want to Use VINCI
Get a VINCI Workspace How to use VINCI

I Am a Current User
Support Training VINCI Guides SAS Guides FAQs

Launch the VINCI Workspace
[UPLOAD files to Workspace](#) [DOWNLOAD files from Workspace](#)

Did you know?
VINCI has 5,708 VINCI Workspace users as well as 2,113 research and 430 operations project data folders in the VINCI Workspace.

Applications
Data Access Request Tracker
VINCI File Upload Utility
VINCI File Download Utility

Shortcuts
My Collaboration Sites
VINCI Workspace
VHA Data Portal
HSR&D
VIREc
HSRData-L Listserv

Just Ask a VINCI Concierge

CRISP Security Scan Schedule
(scans slow down the VINCI system)
Nov. 2-8, 2015
Nov. 30-Dec. 6, 2015
Jan. 4-10, 2016
Feb. 1-7, 2016
Feb. 29-Mar. 6, 2016

Upload existing cohort



VA
Informatics and
Computing
Infrastructure

Faler, Allen (LongView) Welcome to VINCI's File Upload Site

**Important Note: Files containing patient information can only be uploaded to project folders.
Maximum file size limit is 2 GB.
For uploads exceeding 2 GB, please email vinci@va.gov.**

Please compress (Zip) your files before uploading.

Select a file to upload:

 Browse...

Upload to:

- H:\Upload
- P:\SAS_Users\Upload
- P:\ORD_Percipio\Upload
- P:\MAVERIC\Upload
- P:\CDW_SStaff\Upload



Upload existing cohort

- Load
- Test → left pad
- Duplicates
- Not found
- Test patients

Standard Cohort Tables

- [Src].[CohortCrosswalk]
- [Src].[CohortPatientSID]
- [Src].[CohortScrSSN]
- [Src].[CohortPatientICN]

VINCI Cohort Creation

- Most popular
- Specific criteria required
- Cohort creation script

VINCI Cohort Creation

- Domain/Table Name
- Column Name
- Actual codes, dates, names

VINCI Cohort Creation

- Outpatient, Inpatient, Purchased Care (Fee)
- ICDCode Column from [Dim].[ICD9] joined to Outpatient, Inpatient, and Fee
- ICDCode IN ('115.90', '115.92', '115.99', '130.2', '360.40', '360.41', '360.89', '362.16', '362.41', '362.56', '362.70', '371.00', '371.20', '379.00', '446.5', '871.9', '996.51')

VINCI Cohort Creation

- Outpatient VDiagnosis Table
- [VDiagnosisDateTime] Column
- [VDiagnosisDateTime] >= '2010-01-01'
AND [VDiagnosisDateTime] < '2013-01-01'

VINCI Cohort Creation

- Pharmacy Outpatient
- [LocalDrugNameWithDose] Column from [Dim].[LocalDrug] joined to Pharmacy Outpatient
- [LocalDrugNameWithDose] LIKE '%LORAZEPAM%'

VINCI Cohort Creation

- Outpatient, Inpatient, Purchased Care (Fee)
- Sta3n (VA Station) Column
- Sta3n = 508

VINCI Cohort Creation

- Outpatient, Inpatient, Purchased Care (Fee)
- ICDCode Column from [Dim].[ICD9] joined to Outpatient, Inpatient, and Fee
- ICDCode IN ('250.51', '250.52', '250.53', '250.6', '250.61', '250.62', '250.63', '250.7', '250.71', '250.72', '250.73', '250.8', '250.81', '250.82', '250.83', '250.9', '250.91', '250.92', '250.93')

VINCI Cohort Creation

- Spatient Table
- [Age] Column
- [Age] > 65

Standard Cohort Tables

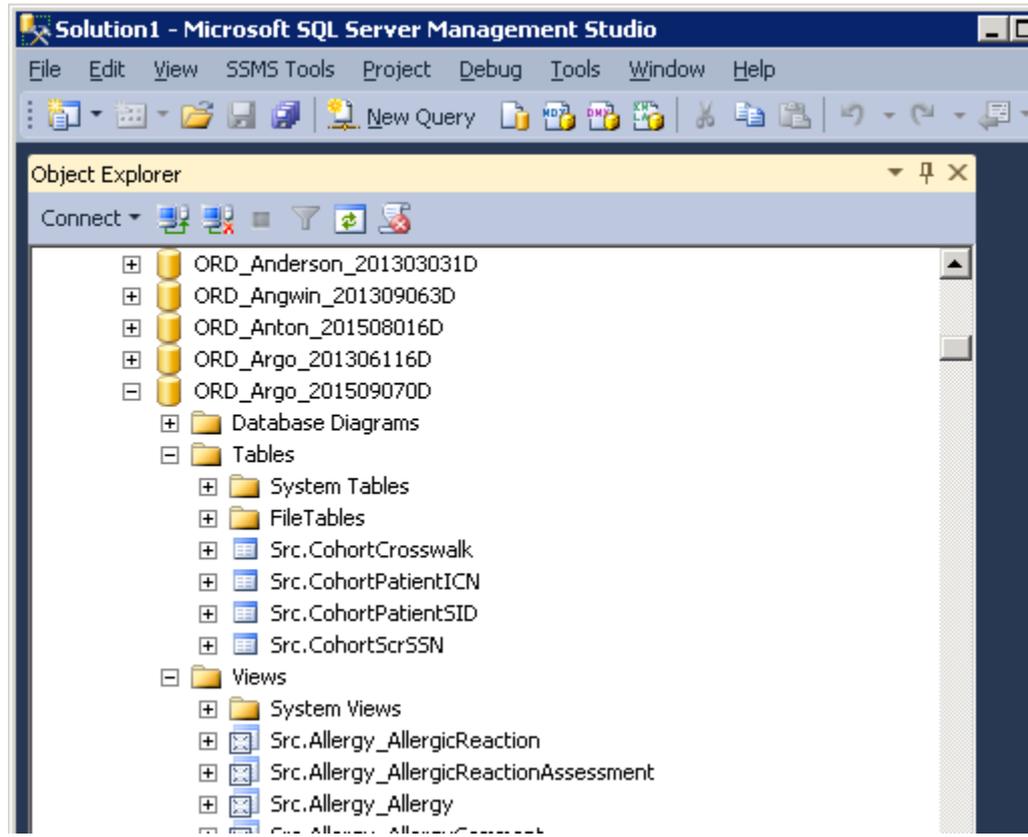
- [Src].[CohortCrosswalk]
- [Src].[CohortPatientSID]
- [Src].[CohortScrSSN]
- [Src].[CohortPatientICN]
- [CohortName] Column

Study Cohort Creation

- Complex, lengthy criteria
- Umbrella cohort
- [VINCI1] Database

Data Provisioning

■ Data Domain Views



Database Connection

- VINCI Workspace/Download
- User database connection guides
 - SAS
 - SPSS
 - R
 - STATA
 - MS Access
 - MS Excel

Recommendations

- Data Analyst
- [Dflt] schema
- Other schemas
- IRBs in DART

Third-party data stewards

- Datasets we do not have (not all-inclusive)
 - CCR
 - NDI
 - VSSC (direct upload into database)
 - CMS
 - PBM
 - OEF_OIF
- Project coordinates with third-party
 - VHA Data Portal