

VINCI data services

HSR&D Cyber Seminar

February 27, 2011



VA Informatics
and Computing
Infrastructure

Agenda

- What is VINCI (quick review)
- VINCI data
- VINCI data services
- Planned enhancements/new features
- Questions?

Quick VINCI Review

- Secure analytical workspace for VA researchers
- Dedicated high performance hardware
- Software tools, custom applications
- Access to data analysis software tools
- RDP connection - Cloud computing
- Analysis and development workspaces
- Most complete collection of VA data
- Staff to assist you

VINCI Data Center

- Located within AITC
- Many large, high performance servers
- Hundreds of terra bytes of high performance SAN
- Newly improved networking
- Dedicated system administrators
- Regular data backup and archiving

VINCI Software

- SQL Server
- SAS, including high performance SAS grid
- Stata, PASW(SPSS), R
- MS Office tools
- Other publishing software
- NLP tools
- Other software can be purchased

VINCI Data

- CDW production data - Vitals, Pharmacy, Lab, many more
- CDW Raw data - TIU, Radiology, Allergies
- DSS NDEs - SAS and SQL format
- Med SAS - SAS and SQL format
- Vital Status

VINCI Data Services

- Data needs assessment
- Preparatory to research work
- Cohort selection
- Data extraction
- Formatting, indexing & security
- Raw data cleaning & standardizing
- Data description documents
- Project tracking system

Data Needs Assessment

- Consultation with your study team to prepare you for making a DART request
- Query development and to find estimated cohort size
- Multiple simple to moderately complex aggregations provided
- We can assist with determining whether a sufficient sample size is available for your study
- Check the data description docs first

Preparatory to Research Work

- All the services of the Data needs assessment plus more
- Primarily used to check study feasibility
- More complex and thorough aggregations are provided
- Your own work database in VINCI
- Use of the VINCI Workspace and all the tools
- Access to individual data records, as needed

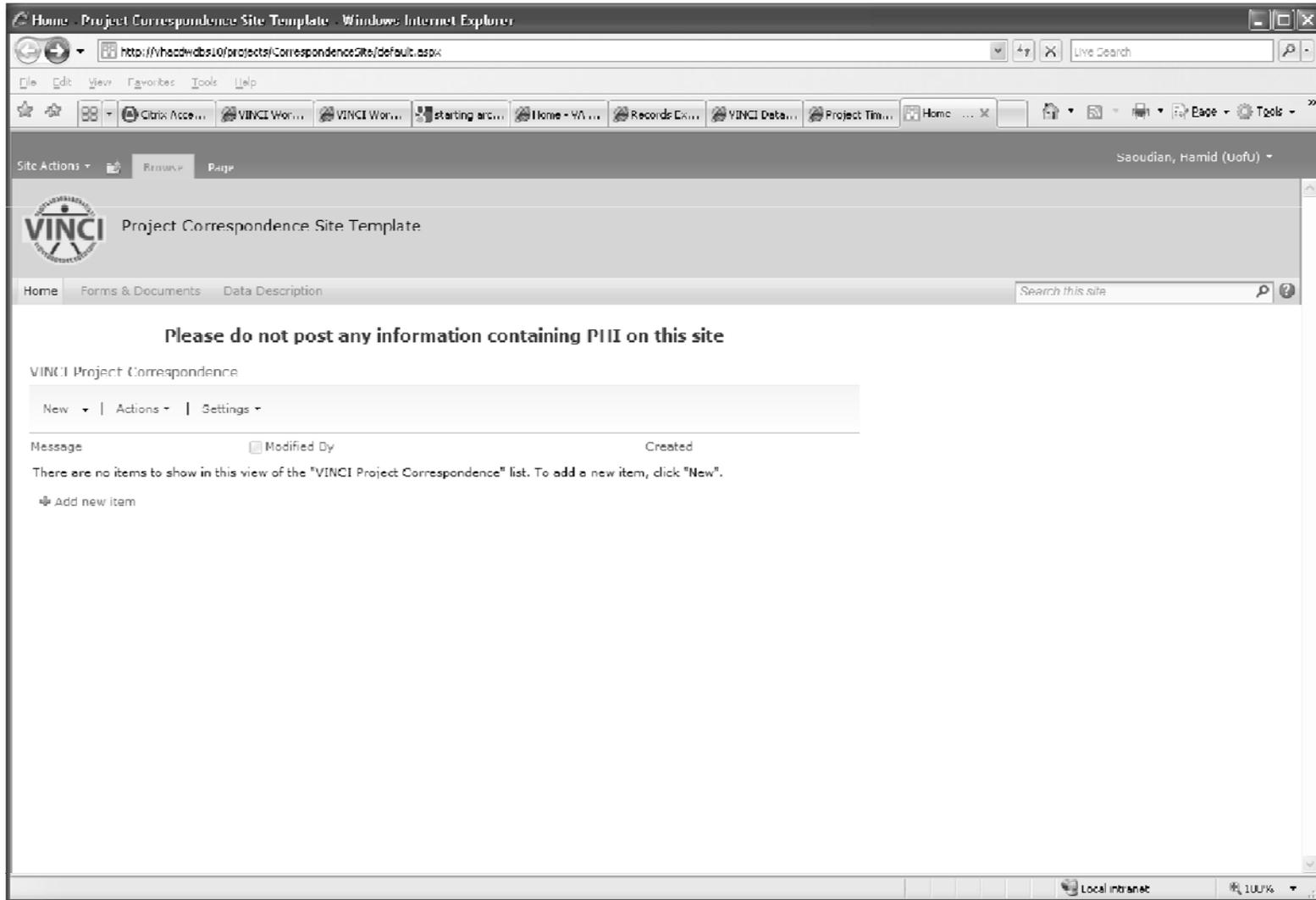
Cohort Selection

- VINCI data manager creates patient cohort based on research users' inclusion and exclusion criteria
- Collaboratively define moderately complex rules for cohort creation, filtering strategies
- Custom scripting of rules into optimized T-SQL code
- Cohorts can be created containing indicator column for each rule or combination of rules, using multiple domains

Data Extraction

- Research study team receives project approval from NDS
- Data access group created including research study team participants
 - Used for granting access to data & workspace
- Research study team receives alerts that the correspondence site has been created
 - All communication takes place on the Site
 - Correspondence Site is easy to use

Data Extraction



Data Extraction

- Research study team uploads cohort to a secure site or VINCI creates cohort per requirements
- Research study team completes the data selection forms on correspondence site
- NDS approved data domains are extracted and provided in your own high performance SQL Server database
- Analysis performed by research project team staff

Formatting, indexing & security

- Data can be transformed to multiple formats as needed
- Databases can be accessed from analysis tools
- Consultation regarding query tuning
- Adding indexes to your data
- Only your study and the VINCI data managers will have access to your database

Raw Data Cleaning & Standardizing

- Performed on all CDW Raw domains, TIU, Radiology, Allergies, etc.
- Data type optimization
- Naming standardization
- Data content standardization
- Indexes added
- Keys added for easier joins

Data Description Documents

- Information on domains, including the variable and variable descriptions contained within each table, available on VINCI's website
- Includes CDW Production and RAW data
- Repackaged VistA FileMan descriptions of tables and columns
- Lineage from VistA to CDW to your database

Data Description Documents

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

The screenshot shows a Windows Internet Explorer browser window displaying the VINCI Data Description page. The address bar shows the URL: <http://vaww.vinci.med.va.gov/vincicentral/data.aspx>. The page title is "Data Description". The navigation menu includes: VINCI Data Home, Analyze Data, Upcoming Data, **Data Descriptions**, Metadata, Reports, Sessions, Feedback, and Links. The main content area is titled "VINCI Data Home" and contains the following text:

The following is a list of the three categories of data available through VINCI.

CDW
The Veterans Administration Corporate Data Warehouse (CDW) is a national repository comprising data from VA clinical and administrative systems. The objective of CDW is to provide data and tools to support management decision making, performance measurement and research objectives. Its premise is that face analysis at the enterprise level. The CDW operates within the VA Office of Information & Technology Field Operations. Multiple VA data sources are being merged so that cohorts will be defined by time period. CDW data are stored in a relational database. Multiple VA data sources are being merged so that cohorts will be defined by time period. CDW data are stored in a relational database. Multiple VA data sources are being merged so that cohorts will be defined by time period. These data are kept current by frequent updates with new data from the source databases so timely data are available.

RAW
The purpose of CDW-Raw is to hold domains that have not been obtained through the normal CDW process. The Raw extractor allows us to obtain data quickly and to make changes as subject area experts refine their requirements. Certain domains will be promoted to CDW Prod, and others will not. The Raw extraction data are not as refined as the CDW primary dataset. They are extracted directly from the Vista sites through a process as a data snapshot, rather than continuously as does CDW. The Raw extractions are therefore not kept current. However, some are updated monthly depending on the domain. Extractions may be repeated as needed.

SAS DATASETS
The SAS Datasets are national administrative data for VHA-provided health care utilized primarily by veterans, but also by some non-veterans (e.g., employees, research participants). The datasets are provided in SAS format by fiscal year (Oct. 1 - Sept. 30). These data are extracted from the National Patient Care Database (NPCD) maintained by the VHA Office of Information at the Austin Information Technology Center (AITC), the central repository for VA data. Having historically been maintained by AITC mainframe, VINCI has acquired a complete copy of those Datasets. In all of the SAS Datasets, each patient has a unique identifier referred to as the Scribed SSN, which is a formula-based encryption of the Social Security Number. The identifier is consistent for a given patient across datasets and fiscal years.

The screenshot also shows a dropdown menu for "Data Descriptions" with the following items: Allergies, Appointment, Contacts, Demographics, Dimensions, Encounter, Health Factor, Immunization, Lab/Chem, Mental Health, Outpatient Pharmacy, Radiology, TB, and Vital Signs.



Project tracking system

- VINCI data team uses a project tracking system to manage your project
- Project duration, time waiting for VINCI, hours spent on your project by VINCI, notes written by VINCI, domains extracted, number of records
- Information can be made available to project team
- Develop progress and management reports

Project tracking system

VINCI Data Project Tracking - Windows Internet Explorer

Search/Filter Data Projects

Assigned To:

Current Status:

Project Type:

Search Clear Filter

Project Name	Protocol	NDS/Tracking Number	Cohort Size	% Complete	Type	Complexity	Status	Assigned To
SPRINT_20120204			100	100	CDW Extract	Medium	Completed	Loft, Dan
Bernal_20110210D	Women Veterans Cohort Study	20110210D	74056	100	CDW Extract	Medium	Completed	Fair, Allen
WVet_20110101D	WVet Cohort	20110101D	7	100	WVet Extract	High	In Progress	Loft, Dan
Brenner_20110705D	Traumatic Brain Injury among Homeless Veterans	20110705D	1010	100	VINCI Extract	Medium	Completed	Fair, Allen
Steady_20110303D	The effects of inpatient care consultation on post-hospital utilization and costs of	20110303D	40	100	CDW Extract	Medium	Waiting...	Bedharshinan, Sundaragandhi
Bruce_20101010D	Reemerging Native Veterans: Service Utilization	2010 04 013 D 021	38493	100	CDW Extract	Medium	Waiting...	Yan, Yimin
Hogan_20090101D	Organizational Complexities and Processes in Medication	20090101D	100	100	Organizational	Medium	Completed	Medley, Daniel
Dwyer_20080204	Organizational Complexities of Adherence to Medication	20080204	4330	100	CDW Extract	Medium	Completed	Loft, Dan
WVet_20110401D	WVet Cohort among veterans' inspection for a new area	20110401D	1910881	100	WVet Extract	High	Completed	Loft, Dan
Durata_20110401D			100	100	CDW Extract	Medium	Completed	Sebastian, Hristi
Rut_20111004D	Electronically Retrieved Cohort of Hospital & Discharged Veterans	20111004D	444213	100	CDW Extract	Medium	Completed	Fair, Allen
Coffey_20110101D	Comparative Effectiveness of Anticoagulants for the Treatment of Myocardial Infa	20110101D	34587	100	CDW Extract	Low	Completed	Fair, Allen
Cannon_20100901D	Function and Analysis of a Genetically-Identified Phenotype Encoder	2010 04 020 D 020	100000	100	CDW Extract	Medium	Scheduled	Yan, Yimin
Cannon_20110504D	US Department of Agriculture (USDA) Registry	2011 05 004 D 010	1914	85	CDW Extract	Medium	Waiting...	Yan, Yimin
Cannon_20110504D	Genetic Phenotype Search	2011 05 004 D 010	1914	85	CDW Extract	Low	Scheduled	Yan, Yimin
Cannon_20100601D	Genetic Phenotype Search	2010 06 01D 010	1914	85	CDW Extract	Medium	Completed	Loft, Dan
Cannon_20111010D	Race, obesity, and triglyceride use as predictors of clinical outcomes in multi	2011 10 10D 0	5206	80	CDW Extract	Medium	In Progress	Hacker, Ryan
Cannon_20111010D	Use of electronic veterans health administration (VHA) services for discharged an	2011 10 10D 0	19403	100	CDW Extract	Medium	Completed	Hacker, Ryan
Cannon_20111010D	Influence of obesity, African-American race, and advanced age on clinical outcom	2011 10 10D 0	4371	100	CDW Extract	Low	Completed	Hacker, Ryan
CDW - Raw Data Cleaning/Tr			30	100	ETL Work	High	In Progress	Yan, Yimin

VINCI Data Project Tracking - Windows Internet Explorer

Search/Filter Data Projects

Assigned To:

Current Status:

Project Type:

Search Clear Filter

Complexity: Medium

Status: Completed

Type: CDW Extract

Assigned To: Fair, Allen

% Complete: 100

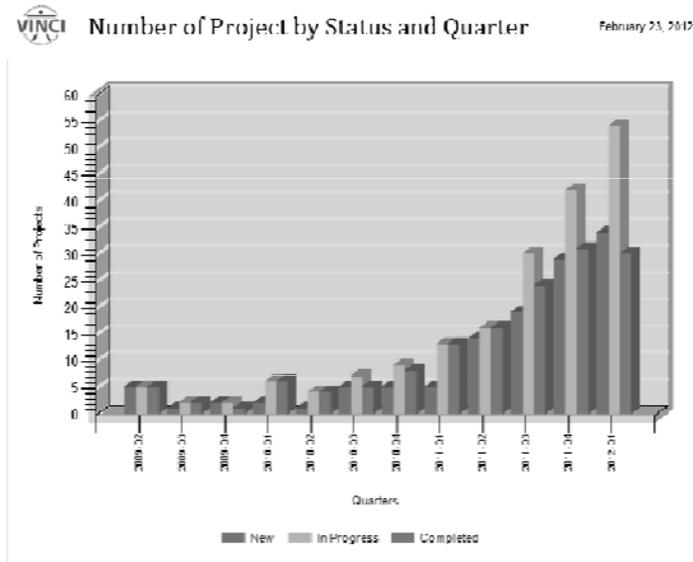
Project Name	Protocol	NDS/Tracking Number	Cohort Size	% Complete	Type	Complexity	Status	Assigned To
Bernal_20110210D	Women Veterans Cohort Study	20110210D	74056	100	CDW Extract	Medium	Completed	Fair, Allen

Data Project Details

Domain	Record Count	Submission Date	Rate
CDW - CDW	182793	5/12/2011	
Health Factors	2067670	5/12/2011	
Lab Chemistry	799287	5/12/2011	
Immunology Tests	2802294	5/12/2011	
Medical Health Assessment	1324008	5/12/2011	
Trauma/Injury	1261288	5/12/2011	
Veterans	182793	5/12/2011	



Project tracking system



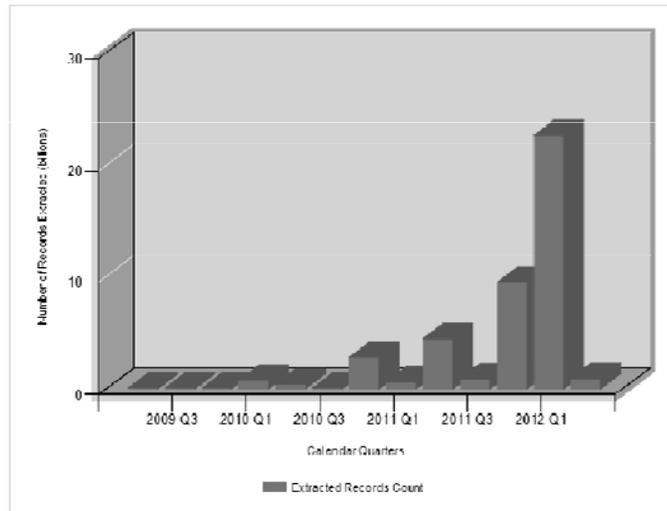
Cal Year	Cal Quarter	Fiscal Year	Fiscal Quarter	New Projects	In Progress	Completed
2009	2	2009	3	5	5	5
2009	3	2009	4	1	2	2
2009	4	2010	1	2	2	1
2010	1	2010	2	2	6	6
2010	2	2010	3	1	4	4
2010	3	2010	4	5	7	5
2010	4	2011	1	5	9	8
2011	1	2011	2	5	13	13
2011	2	2011	3	14	16	16
2011	3	2011	4	10	30	24
2011	4	2012	1	29	42	31
2012	1	2012	2	34	54	30

Project tracking system



Extracted Records Counts

February 25, 2012



Cal Year	Cal Quarter	Fiscal Year	Fiscal Quarter	Records Extracted
2009	2	2009	3	73,383,256
2009	3	2009	4	8,323,490
2009	4	2010	1	25,455,825
2010	1	2010	2	778,042,649
2010	2	2010	3	116,066,278
2010	3	2010	4	106,510,212
2010	4	2011	1	2,837,775,133
2011	1	2011	2	645,226,879
2011	2	2011	3	1,522,010,885
2011	3	2011	4	928,148,527
2011	4	2012	1	9,498,617,869
2012	1	2012	2	22,673,097,156
2012	4	2013	1	939,660,529



Enhancements/New Features

- Knowledgebase System
- Cleaning and standardizing data
- Creation of “Less Identifiable” data domains
- Metadata & data profile reports
- Data analysis cubes
- Aggregate reports
- Cohort selection tools

Poll Question

- Which new features do you want to be made available in VINCI data services?
 - A. Knowledge base
 - B. Cleaning and standardizing data
 - C. “Less identifiable” data domains
 - D. More metadata & data profile info
 - E. Analysis cubes & aggregate reports
 - F. Cohort selection tools
 - G. Other (Please specify using Q/A option)

Questions?

Contact Information:

vinci@va.gov

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

Q: For data that is available through the SAS files in Austin, is the data available in the same format? For data not available in SAS format through Austin, are there guides to the data formatting?

The data in CDW are generally in SQL format. VINCI data managers can provide most of the data in either SAS or SQL formats. The "Austin SAS files" generally refers to the DSS or MedSAS data. Both these sources are originally SAS format, but can be provided in SQL also.

Q: Is it possible to get a copy of the naming standardization you have in place?

Standardization is really a VHA-wide activity. The group responsible is Standardization and Terminology. There is a subgroup for Lab standardization and a pharmacy group. The best source of information is VIREC - the VA Information Resource Center.

Q: Please elaborate in more detail regarding data standardization - we were told that help was NOT available with this. We want only plasma or serum glucose levels, for example.

The VINCI data management group will try to assist with identification of the correct data for a study. But standardization, especially of Lab data, is a VHA-wide effort. The best source of information is VIREC - the VA Information Resource Center.

Q: When researchers create projects, they don't necessarily want nor need involvement by the VINCI data managers, i.e., what is to prevent "rogue" VINCI data managers from corrupting or damaging research stored on the servers? In many research studies also, many IRBs DO NOT accept external and non-approved personnel from having access to a researcher's work product.

The permission to access data from the CDW and some other data sources is granted through National Data Systems (NDS). There are other sources of data and other data stewards in the VA. The VINCI data managers have been given rights to provide data for research projects as directed by NDS. Researchers do not have access to the CDW. If the IRB requires data to be accessed only by those named on the IRB, the project would be limited to local or regional data sources, if available, and as approved by those data stewards.

Q: Are the data provided 21 CFR Part 11 compliance? If can be made so?

Q: Will the RAW data be eventually productionized so that it will have the same availability as the production dataset?

Yes, the CDW RAW files will be modeled and moved to CDW Production.

Q: How is missing data handled?

Data quality and data cleaning are still largely the responsibility of the user. Efforts throughout the VA are underway to improve data quality, but there is no systemic source of "clean" data. VIREC (VA Information Resource Center) is your best source of information about data and the quality of the data.

Q: You said that researchers can bring data into VINCI, is there some sort of data use agreement (DUA) in place to protect the researcher's primary data? Does the data become property of VINCI?

VINCI is not a data "owner". According to our mission, we prevent the download of data that can be identified with a patient (PII/PHI) without written permission of Dr. Seth Eisen, head of HSR&D. If one

imports/uploads data to the VINCI servers for analysis, one may keep a copy of the data. Downloading of aggregate/summary data or results/graphs/figures/manuscripts is always permitted.

That said, VINCI is a Research Data Repository, and can make available derived data with the proper approvals. Derived data may be datasets created by combining other data, or data that is produced by research activity, an example of which would be annotated medical records used to train NLP tools. These are the only data that would be specific to VINCI as a repository.

Q: Would it be possible to have our own 'honest data brokers' to access CDW data and abstract the data ourselves, without help from VINCI staff?

Data access permissions for the CDW are determined by NDS. At this time, researchers are not allowed access to the CDW. It is still possible to access certain data on the mainframe in Austin. Those permissions are also handled by NDS.

Q: Are Inpatient medications available in sql format through VINCI?

In-patient and outpatient pharmacy are available from CDW (SQL format). The PBM database is available for research with permissions directly from PBM.

Q: ...and are radiology notes subject

??

Q: Can we have access to microbiology (lab) data?

There is currently available a database of microbiology data (organism, sample, antibiotic, resistance/sensitivity). These data are not part of CDW, but can be used for research with oversight by NDS. Please contact Tori Barrett (victoria.barrett@va.gov) for more information.

Q: Do WOC appointees get the same level of access or priority to VINCI services?

Yes.

Q: What does the data look like in VINCI? For data that is available through the SAS files, is it similar in format to the SAS files?

Most data are available in both SAS and SQL format.

Q: HI. Do you have any ongoing telepatholgy projects i.e. whole slide imaging, etc

No image data are currently available in the CDW.

Q: How do we upload our own data to VINCI workspace?

Data are uploaded by ftp (file transfer protocol). Instructions are available on the VINCI Website – vaww.vinci.med.va.gov/vincicentral.

Q: How do we delete a study workspace on VINCI?

There are rules in the VA about deleting/maintaining data. The VINCI staff will assist with the archiving of data that must be retained. Otherwise, a workspace can be deleted simply by notifying VINCI staff at vinci@va.gov.

Q: How frequently is the data backed up? If uploaded by user

Data are backed up nightly and archived to tape weekly. Data, whether materialized by VINCI staff or uploaded by a project's staff are backed up in Project folders nightly.

Q: How frequent the data is updated to date through VINCI?

CDW data are updated nightly. There are many other sources of data that are updated on specific schedules. For example, the TIU (note) data are a VistA extract and are currently extracted annually.

Q: How is VINCI working with other VA or VHA offices in terms of things like data standardization, metadata development, etc?

The major data standardization efforts are through the Standardization and Terminology office. VIREC is the best source of information about the quality of data for research.

Q: How long does it now take to go from first contact with VINCI to project approval and to having data ready for analysis?

Access to data is granted by approval from NDS. That process depends on the nature of the request, the complexity of the data and the number of other applications that are in process at the same time. A complete and accurate data access application certainly speeds up the process. Once a project has approval for a set of data, the VINCI data managers will contact the research team within 2 business days. The time it takes to have data ready for analysis depends on the complexity of the request.

Please remember that access to the Austin mainframe is also granted by permission from NDS. Once a group/person has permission to access the mainframe, their acquisition of the data proceeds at their own pace.

Q: If we create a Microsoft document in VINCI how do we move it our local computer? Email? Or...

Files are downloaded from VINCI through a utility called File Transfer. We use this utility in order to copy and audit all files that are downloaded. The VINCI policy states that no files containing PHI/PII may be downloaded. Otherwise, any aggregated/summary/results data or figure/manuscripts may be freely removed using the File Transfer utility.

Q: How much storage is available for individual researcher's data tables?

There is currently more than 500 TB of storage available. We do not limit individual project storage.

Q: How should our IRB and VINCI interact?

If a project will use the VINCI environment (servers, software, storage), the IRB needs to specify VINCI as the location of the data. Suggested wording for the IRB can be found on VINCI Central (vaww.vinci.med.va.gov/vincicentral) under the library tab.

Q: How to access TIU? What format is the data for TIU?

TIU are text documents. One requests permission for TIU through NDS just as other CDW data.

Q: How to get approval for Preparatory to research work?

Instructions for Preparatory to Research permissions are on the NDS website.

Q: Are Pathology reports available in CDW Raw?

Pathology reports are not currently available in CDW.

Q: I have formed a work group in VINCI, but am I supposed to have a VINCI member on my study team?

No. Please do make sure that you have VINCI Workspace access and not just a Study Registration collaboration site.

Q: How will VINCI intersect with the work of the open source communities?

The VINCI group that is doing custom software development for Natural Language Processing (NLP), Annotation/Chart Reader and Information Extraction uses all open-source tools. All VINCI custom software is open-source.

Q: Would you say that standardizations of the core legacy applications such as lab, pharmacy is a key dependency on pulling national data thru VINCI?

The office of Standardization and Terminology is charged with the standardization of data throughout VHA. VINCI is not part of this office.

Q: How does data validation occur in Vinci?

VINCI staff is not responsible for data validation. Please refer to VIREC for information about data quality in VHA.

Q: If we do all the work ourselves, create cohort, pull the data. Do we need to submit DRAT application as well?

All access to CDW, and some other data sources, is handled through NDS. It is the responsibility of the data steward to grant access to data for research. Researchers do not have access to the CDW so cannot "pull data". If a project is accessing local sources of data, or other sources not controlled by NDS, one must contact the steward for those data for approval. VIREC provides information about sources of data available for research and the process to access those data.

Q: is it possible to create a cohort using data from both the production and RAW datasets (linked by ssn or some other Unique Identifier)?

Yes. There are several ways to link data from CDW and other sources. The CDW uses an IEN unique identifier but it can be cross-walked to other identifiers used in other sources.

Q: Do the TIU snapshots contain full text from all CPRS documentation notes?

Most, but not all, text data in VistA are stored in TIU. Each station has peculiarities. One cannot guarantee that all notes are in TIU.

Q: do we need IRB approval prior to applying for VINCI access?

It is possible to have a VINCI Workspace (access to servers, software, storage) without an IRB. If a study is getting data from CDW (with permissions from NDS), we will generally create a workspace for your project when your data access request is approved. If you are bringing in your own data or simply want to "explore", you may set up a workspace in advance. Please see VINCI Central for instructions. (vaww.vinci.med.va.gov/vincicentral)

Q: Do researchers need to be able to write perl, sql, etc in order to perform queries on the provided dataset?

The data can be provided in SAS or SQL format. VINCI staff do not do any analyses, they can, though provide data in Access or some other formats if requested.

Q: Is the idea that the data analysis cubes, aggregate reports, less identifiable domains, etc would not require NDS permission/approval?

No. All CDW data access currently required NDS approval. Please remember there are many sources of data in VHA and other sources may have their own data steward responsible for approving access for research.

Q: Could explain "knowledge base" briefly again? I apologize...I did not hear that description.

A compilation of information regarding VA research data and other related concepts and issues. The information in the knowledgebase is accessible to all VA research staff, VINCI staff and others with interest in VA research data. All knowledgebase users can contribute and add information to it. It will be maintained and moderated by VINCI staff.

Q: Is tumor registry in CDW too?

CDW does not have the tumor registry data. The VINCI data managers can help you link those data with CDW data but you must work with the data steward for the tumor data to gain access for your project.

Q: What's the sequence for approval? Which one comes first? DART or VINCI. I got an VINCI access approval for a project, but still waiting for the DART approval for 6 weeks already.

It is possible to have a VINCI Workspace (access to servers, software, storage) without an IRB. If a study is getting data from CDW (with permissions from NDS), we will generally create a workspace for your project when your data access request is approved. If you are bringing in your own data or simply want to “explore”, you may set up a workspace in advance. Please see VINCI Central for instructions. (vaww.vinci.med.va.gov/vincicentral)

Just a polite reminder, the NDS process is independent from VINCI. DART is the workflow application.

Q: Is VINCI only available to researchers? If we need data sources for non research projects can we still utilize VINCI?

VINCI is the analytical platform (servers, software, storage). It is available for use by Researchers and non-researchers.

Q: Do you include SPSS, in addition to SAS?

SPSS is available as are more than 50 other commercial and custom software applications. Please see VINCI Central (vaww.vinci.med.va.gov/vincicentral) for a complete list.

Q: please provide link to virec?

The VIREC site can be reached at: <http://www.virec.research.va.gov/>

Q: This is my first time joining. I am unfamiliar with this service. Is there a website I can go to learn more specifics?

Please start with VINCI Central vaww.vinci.med.va.gov/vincicentral

Q: through the presentation, VINCI sounds serving for 'research' purpose. Can we get data through VINCI for operational project, e.g., quality improvement project? For instance, can we get adverse event incidence report owned by National Center of Patient Safety through VINCI to work with local patient safety managers to improve the safety of the care?

VINCI is the analytical platform (servers, software, storage). It is available for use by Researchers and non-researchers. There are many sources of data throughout VHA. One must work with the steward of the specific source of data to obtain permission for use. This is generally true for both research and operations. Please work with the steward for the National Center of Patient Safety to obtain permission for use of the data. VINCI data managers may be able to assist with materialization of those data once approved.

Q: What is the process for getting approval to get data?

There are many sources of data throughout VHA. One must work with the steward of the specific source of data to obtain permission for use. The data steward for CDW and a number of other commonly used data sources is NDS. Please see the NDS website for instructions.

Q: You encourage storage of data; where?

VINCI is an analytical platform (servers, software, storage). If using the VINCI Workspace, your data will be stored and backed up by the VINCI systems administrators. If you use the VINCI Workspace for a research project your IRB needs to identify VINCI is as the location of the data.

Q: Has a privacy officer evaluated whether the VINCI data has been de-identified with respect to HIPAA regulations?

There is no source of de-identified data in the VA. CDW data are fully identified.

Q: Will we have direct access to the Med SAS, vital status, DSS NDE data? Right now, we go directly to those files (mainframe) to develop our own cohorts. Will we be able to do that, or will we have to go through a VINCI analyst to get the data for us. We are concerned about the extra time and complexity involved in adding another layer of people to our projects.

All access to CDW, and some other data sources, is handled through NDS. It is the responsibility of the data steward to grant access to data for research. Researchers do not have access to the CDW so cannot "pull data". If a project is accessing local sources of data, or other sources not controlled by NDS, one must contact the steward for those data for approval. VIREC provides information about sources of data available for research and the process to access those data. At this time, researchers are not allowed access to the CDW. It is still possible to access certain data on the mainframe in Austin. Those permissions are also handled by NDS.

Q: What application do we need to submit in order to use VINCI to store and analyze our own data (not VINCI data)?

If one would like to use the VINCI Workspace (servers, software, storage) just see VINCI Central (vaww.vinci.med.va.gov/vincicentral) for instructions. It is a simple ftp to bring in data for which a project has permission.

Q: If we load our own data into VINCI, are we allowed to download the data back to our personal computers?

If one imports/uploads data to the VINCI servers for analysis, one may keep a copy of the data. Downloading of aggregate/summary data or results/graphs/figures/manuscripts is always permitted.

Q: Can the presenter eventually review a 'typical timeline' for getting data from VINCI?

Access to data is granted by approval from NDS. That process depends on the nature of the request, the complexity of the data and the number of other applications that are in process at the same time. A complete and accurate data access application certainly speeds up the process. Once a project has approval for a set of data, the VINCI data managers will contact the research team within 2 business days. The time it takes to have data ready for analysis depends on the complexity of the request.

Please remember that access to the Austin mainframe is also granted by permission from NDS. Once a group/person has permission to access the mainframe, their acquisition of the data proceeds at their own pace.

Q: If I don't have access to the VA intranet, can I still see the data description?

No.

Q: In the cohort identification process, is it possible to reconstruct a consort figure?

We need more information to be able to answer – please contact us at VINCI@va.gov

Q: Would we be allowed to download CDW data from VINCI into our computers if we originally requested to have CDW data downloaded into our computer?

NDS approves requests for access to data. The request must specify where the data are to be stored/used. It is possible to obtain permission for access to data on local research servers.

Q: Who are VINCI data managers? How can we contact them?

Please see VINCI Central for contact information. vawww.vinci.med.va.gov/vicnicentral

Q: Are you radiology reports and radiology orders available in the CDW?

Yes.

Q: How frequently the data is backed up?

Data are backed up nightly and archived to tape weekly. Data, whether materialized by VINCI staff or uploaded by a project's staff are backed up in Project folders nightly.

Q: Will your infrastructure match that of OnLive Desktop for Apple Ipad? They have 1-gigabit Internet and can download 20 MB file in a second.

No

Q: There is a big problem in using the lab data. Lab values and names exist in the database just as they were created in individual VA facilities. Researchers may want to study glucose levels, for example. But there are 2000 different names of glucose tests (we've obtained a list).

The VINCI data management group will try to assist with identification of the correct data for a study. But standardization, especially of Lab data, is a VHA-wide effort. The best source of information is VIREC - the VA Information Resource Center.

Q: Do you have endnote software?

Yes, but VINCI is not open to the internet so some functionality of Endnote is not available.

Q: Can we have access to Freeware (free software) through VINCI? What is the process?

It is possible to bring in your own software. When you request access to VINCI Workspace, the staff will ask you about the software you need for your work. If you need to bring in specialized software we will create an isolated workspace for your use so that your software cannot affect the system.

Q: How do we request software purchases? What is the timeline for these purchases?

VINCI is funded to provide software on request, but each request is evaluated in light of budget, utility to others and operability. The procurement process in the VA is long and convoluted.

Q: If a researcher has access to VINCI data, does he still need to request access to mainframe account separately? Thanks.

VINCI is the analytical platform (servers, storage, software). If you have been given access to CDW data for a project, the VINCI staff will help materialize those data either in the VINCI Workspace or to your local server as instructed by NDS. Most data that can be obtained from the mainframe are available from CDW.

Q: medSAS and DSS data are the same with that in AITC?

Yes, except that they can be provided from CDW in SQL format as well.

Q: Please provide more information about help available with data cleaning - this is a big problem for us.

The VINCI data managers have experience with the data used for research so that they can provide some suggestions about the quality of the data. Please see VIREC for more information about data quality.

Q: Does Vinci currently support neuroimaging analysis (e.g., structural or functional MRI)?

CDW does not have any image data.

Q: How long does it take to get permission from DART in general?

Access to data is granted by approval from NDS. That process depends on the nature of the request, the complexity of the data and the number of other applications that are in process at the same time. A complete and accurate data access application certainly speeds up the process. Once a project has approval for a set of data, the VINCI data managers will contact the research team within 2 business days. The time it takes to have data ready for analysis depends on the complexity of the request.

Please remember that access to the Austin mainframe is also granted by permission from NDS. Once a group/person has permission to access the mainframe, their acquisition of the data proceeds at their own pace.

Q: What about researchers that only need a codebook and access to VINCI? In other words, what provisions are being made to accommodate advanced researchers?

All access to CDW, and some other data sources, is handled through NDS. It is the responsibility of the data steward to grant access to data for research. Researchers do not have access to the CDW so cannot "pull data". If a project is accessing local sources of data, or other sources not controlled by NDS, one must contact the steward for those data for approval. VIREC provides information about sources of data available for research and the process to access those data. At this time, researchers are not allowed access to the CDW. It is still possible to access certain data on the mainframe in Austin. Those permissions are also handled by NDS.

Q: What is needed is for VINCI staff to develop ways to make identification of commonly used tests more convenient for researchers. Is this possible, and if so, who should we work with? This would help lots of investigators, so each group doesn't have to "reinvent the wheel" with this.

2/27/2012 12:00pm
Saoudian, Hamid

VA Informatics and Computing Infrastructure
VINCI Data Services

Standardization is really a VHA-wide activity. The group responsible is Standardization and Terminology. There is a subgroup for Lab standardization and a pharmacy group. The best source of information is VIREC - the VA Information Resource Center.

Q: Can NLP be used on radiology notes?

Yes. VINCI can provide some custom NLP tools.

Q: Will the data manager extract the cohort data in SAS format or SQL format?

Both formats are possible.

Q: Does a project need to have a well-defined cohort criteria from the beginning of the data abstraction? Can cohort be modified over time?

A cohort can be modified over time and VINCI data managers can help with cohort identification.

Q: Is it possible to have extracted data provided in another format other than SQL Server database (e.g. SAS or STATA)?

All SQL data can be provided in SAS format. May "native" SAS datasets are available. In general, SAS data sets can be used with most other analytical software programs such as Stata.

Q: What are the steps for getting NDS approval...is there a website with forms, etc.?

Please see the National Data Systems (NDS) website.