

# Introduction to VINCI data services

---

HSR&D Cyber Seminar

October 11, 2011



VA Informatics  
and Computing  
Infrastructure

# Agenda

---

- What is VINCI (review)
- Available data
- VINCI data services
- VINCI data access/analysis tools
- Planned enhancements/new features
- Questions?

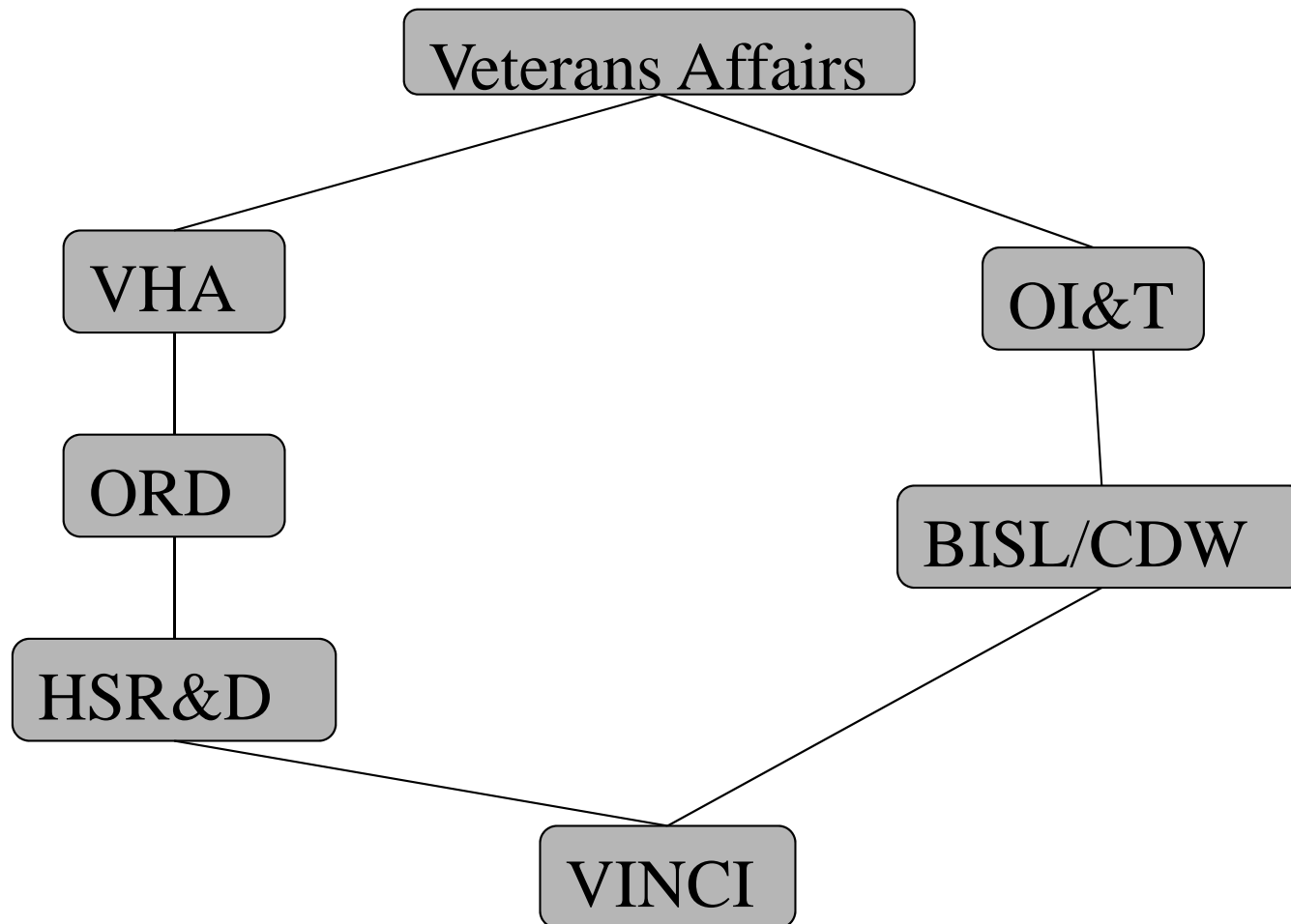
# What is VINCI?

---

- Secure analytical workspace
- Protect PHI/PII
- Regulated data access for research
- Access to data analysis software tools
- Access to high performance servers
- Large high speed data storage
- Custom software
- RDP connection - Cloud computing
- Staff to assist you

# VINCI Organization

---



# What is VINCI

---

- Data Center in AITC (Austin, Texas)
- Staff mostly in Salt Lake City field office
- VINCI works closely with NDS for data access approval
- Serving both research and operations users
- CDW as data provider, VINCI providing analysis service - working in partnership
- VINCI includes custom software development work for research use

# Poll Question

---

- Have you already used VINCI data/services?
  
- Yes
- No
- Did not know about VINCI

# Available Data

---

- CDW Production Data
- CDW Raw Data
- SAS Datasets

# CDW Production Data

---

- Data warehouse modeled data
- Hosted in SQL Server 2008-R2
- Fact and dimension tables
- Extracted through CDW's journaling process from shadow servers
- Unique CDW identifiers
- Indexed for fast queries
- Updated nightly



# CDW Production Data

---

- Patient Demographics
- Vital Signs
- Outpatient Pharmacy
- Consults
- Health Factors
- Outpatient Encounters
- Immunization
- Lab Chemistry
- PCMM

# CDW Production Data

---

- Orders
- Appointments
- More domains coming soon

# CDW Production Data

---

- For updated production domain information see:

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

<http://vaww.dwh.cdw.portal.va.gov/Pages/welcome.aspx>

# CDW Raw Data

---

- Extracted directly from Vista Cache databases
- Snapshot data
- Same structure as source, not modeled
- Data types and column names standardized by VINCI - indexes added
- Updated weekly, monthly, as needed
- Can be joined with prod. and other data
- Additional data domains can be requested

# CDW Raw Data

---

- Bill Claims
- Fee Basis
- Non-VA Meds
- MyHealthVet
- Patient Treatment File
- Prosthetic Surgery
- Inventory
- Travel

# CDW Raw Data

---

- Intravenous Meds
- TIU
- Allergies/Adverse Events
- Radiology

# Text Integration Utilities (TIU)

---

- Medical document information such as clinical notes and nurses notes
- Available for all regions (working on last 7 stations)
- Full Text Indexed - Very fast searches
- Contains real SSN-access regulated by NDS
- Updated as needed
- Does not include radiology and pathology report

# Allergies/Adverse Event (Raw)

---

- Contains patient allergies records and adverse reaction events
- Includes: allergies recorded dates, locations, allergy reactant, such as food, drug, and others
- Includes all stations
- Update monthly



# Radiology (Raw)

---

- Include report notes, impression text and additional clinical history
- Radiology procedures includes: MRI, CT scan, bone scan, ultrasound, echocardiogram
- Includes all stations
- Updated monthly

# SAS Datasets

---

- Uploaded from AITC mainframe
- Regularly updated with mainframe data
- Available in both SAS and SQL Server format
- Combined in VINCI for faster & easier access

# DSS NDE (SAS)

---

- All NDEs, include LAB, LAR, OUT, DISCH, PHA, and RAD
- FY 2005-current combined and provided by DSS team
- Data prior to FY 2005 was transformed, combined and loaded to SQL Server by VINCI staff
- Monthly or quarterly update

# MedSAS (SAS)

---

- Available in VINCI in both SAS and SQL Server table format
  - Transformed & loaded to SQL Server by VINCI
  - FY 2000 - 2010
  - Combined over multiple years in SQL Server
- Includes:
  - Outpatient SE/SF files
  - Inpatient Encounter (SE)
  - Inpatient datasets
  - Inpatient census datasets

# Vital Status

---

- includes date of death, Medicare yearly enrollment indicators, date of last activity, and Veteran status flags, etc.
- Transformed and loaded from SAS to SQL Server tables by VINCI
- Updated quarterly

# Other Research Data

---

- PBM
- CMS
- HERC
- Active duty personnel data
- Registry data
- Birth/death certificate data, or cause of death.

# Poll Question

---

- Which data analysis tools do you use?
- SQL Server
- SAS
- STATA
- PASW (SPSS)
- R
- Others

# VINCI Data Services

---

- Data description documents
- Preparatory consulting
- Cohort selection
- Data extraction
- Data formats
- Data access and security
- External data



# 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, Available 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 vital signs like blood pressure, weight and height, within a target 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 right side of the page shows a partial view of a dropdown menu with the following items: Allergies, Appointment, Contacts, Demographics, Dimensions, Encounter, Health Factor, Immunization, Lab/Chem, Mental Health, Outpatient Pharmacy, Radiology, TB, and Vital Signs.



# Preparatory Consulting

---

- Data needs assessment
- Work with VINCI prior to NDS approval to solidify requirements
- Contact: [vinci@va.gov](mailto:vinci@va.gov)

# Cohort Selection

---

- If the study does not have a cohort, we will help to create one
  - Research Study Team provides specific criteria
  - VINCI provides data necessary for calculations (if any)
  - Research Study Team completes cohort calculations (if any)

# 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

# Correspondence Site

Home - Name\_2011001010 - Windows Internet Explorer

http://vhacdwbst10/projects/CorrespondenceSite/Name\_2011001010/default.aspx

Site Address: Name\_2011001010

X-Fox: http Previous Next Options

Site Address: Name\_2011001010

Site Address: Name\_2011001010

Please do not post any information containing PHI on this site

VINCI Project Correspondence

New | Actions | Settings

Message

Modified By: Created

Faler, Allen L. 10/9/2011 3:46 PM

**Welcome to VINCI,**

CDW/VINCI has received your NDS approval and we look forward to working with you. NDS has approved the \_\_\_\_\_ Data Set(s) for this study.

Please post a reply to this message by clicking on "New" above and then "Project Message". In your reply, please indicate whether you have a cohort file ready for upload. Once we receive your reply, we will commence work on the study.

**Purpose of This Site**

This site has been created to facilitate correspondence between the VINCI/CDW Data Team and your project team. Please post messages to the home page of this site to communicate with the VINCI data team. You may post a new message by clicking on "New" above and then "Project Message". All correspondence 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. That message will contain a link to this site.

**Data Extraction Process for Researchers**

It seems to really help Researchers if they know the steps that must be taken, which ones involve them, and when to expect each step. Not knowing the process seems to cause dissatisfaction with the process because every step is a surprise and it is unclear how to help things go smoothly. Please review the process found at the following link: <http://vaww.vinci.med.va.gov/vincicentral/Library/VINCI%20Data%20Materialization%20Process%20for%20Researchers.pdf>

Add new item



# 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
- Analysis performed by research project team staff

# Data Formats

---

- SQL Server database
- SAS files
- PASW (SPSS)
- STATA
- Excel
- Flat files
- Others
- Tools available for data conversion

# Data Access & Security

---

- Access groups created based on IRB and NDS approved research team
- Only research team members have access to the data
- Data stored on secure VINCI servers
- Regular data backup and archiving
- Workspace vs. collaboration site
- Project work can be performed in VINCI
- Export final result & publication



# External Data

---

- Research Study Team may upload other data for analysis into project database or workspace
- Secure data upload process
- Optional direct database upload
- VINCI data managers work on behalf of research team to upload data from other data providers

# Data Processing/Analysis tools

---

- SQL Server as primary data store
- Multiple high performance servers
- Most data queries performed in SQL Server
- Accessible by all analysis software
- SSIS, SSRS, SSAS
- High speed intra-server network will allow distributed queries

# SAS

---

- High performance SAS server
  - 2 TB of RAM, 64 cores, 2 TB SSD
  - Launch grid jobs
- SAS grid - very large data analysis work
  - 10 high performance servers
  - Most advanced SAS implementation in VA
  - Dedicated SAN
  - Additional SAS modules
  - SAS knowledge base SharePoint site
- Dedicated SAS administrator

# Other Data Analysis Tools

---

- STATA
- SPSS
- R
- Annotation & NLP tools
- Terminology databases
- Other software can be purchased as needed

# Enhancements/New Features

---

- New CDW data domains
- Additional Raw data & standardization
- Metadata & data profile reports
- Cubes
- Aggregate reports
- Knowledge base
- DOD research data sharing
- Cohort selection tools
- I2B2 pilot project

# Poll Question

---

- Do you plan on using VINCI data/services in the next:
  - 6 months
  - Year
  - Not sure
  - Not planning on using VINCI

# Questions?

---

Contact Information:

[vinci@va.gov](mailto:vinci@va.gov)

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