

First Time Research Users' Guide to CDW:

Getting Started with this
Relational Database

by Margaret Gonsoulin, PhD



Thanks to:

- **Richard Pham, BISL/CDW**
- **Lucy Zhang, VIREC**
- **Linda Kok, VIREC**
- **Kristin de Groot, VIREC**
- **Maria Souden, VIREC**
- **Denise Hynes, VIREC**

Agenda for Today

- **Define terminology used with data warehouses and relational databases**
- **Provide an orientation to the way these terms are utilized in the context of CDW**
- **Identify the resources that can be used to continue learning about the content and structure of CDW**

By the end of this talk,

We hope that a new CDW user will:

- Understand the basic organization of CDW
- Learn some of the jargon associated with relational databases
- Know where to go to learn more about CDW data
- Be able to communicate more effectively with CDW and VINCI when you are preparing to make a data request, but this talk does not cover the request process or access issues.

Poll #1: About you

- What is your role in VA research?
 - Research investigator/PI
 - Data manager/analyst
 - Project coordinator
 - Other – please describe via the Q&A function

Poll #2: Your CDW experience

- What is your level of experience with CDW data?
 - 1- Not worked with it at all
 - 2
 - 3
 - 4
 - 5- Very experienced with CDW data

The Organization of CDW



Corporate Data Warehouse (CDW)

Production Domains
(e.g., Allergy, Appointments)

Tables/Views
(e.g., Dim.AllergicReaction
and Allergy.Allergy)

Raw Domains
(e.g., Surgery, DSS)

Tables/Views
(e.g., Dim.SignsSymptoms
and Dss.RAD)

C “DW”

- The “DW” in CDW stands for “Data Warehouse.”
- Data Warehouse = a data delivery system intended to give users the information they need to support their business decisions.
- In ordinary terms, it is a large storage facility for “big data.”

“C” DW, “R” DW & “V” DW

- Users will see documentation referring to xDW.
- The “x” is a variable waiting to be filled in with either:
 - “V” for VISN,
 - “R” for region or
 - “C” for corporate (meaning “national VHA”)
- Each organizational level of the VA has its own data warehouse focusing on its own population.
- This talk focuses on CDW only.

C“D”W- The data.

- Data stored in this warehouse are in a “relational” format.
- In ordinary terms, this is data that has been separated out into multiple tables that look like excel spreadsheets.
- Linking keys are added into these tables so that users may reassemble the tables for analytic use.
- SQL programming language can be used to reassemble the tables.

Simple Example, Flat File

LastName	FirstName	Address	CityState	Zipcode	Item1	Item2	Date
Jones	Marianna	123 Oak St	Bee, AR	70788	Shirt	Pants	1/2/2014
Frank	Josie	11 Pine Ave	Flip, OK	30032	Skirt	Sweater	2/2/2014
Plank	Bill	230 5 th St	Miner, TX	11201	Pants	Pants	3/3/2014

There are three main types of information in this “flat file”

1. Customer information (name and address)
2. Items for Sale (shirt, sweater, pants, skirt)
3. Purchases Made (two items and date of purchase)

Simple Example, Relational Data

CustomerTable

CustomerKey	LastName	FirstName	Address	CityState	Zipcode
1	Jones	Marianna	123 Oak St	Bee, AR	70788
2	Frank	Josie	11 Pine Ave	Flip, OK	30032
3	Plank	Bill	230 5 th St	Miner, TX	11201

ItemTable

ItemKey	Item
1	Shirt
2	Pants
3	Skirt
4	Sweater

PurchasesTable

PurchaseKey	ItemSoldKey	CustomerKey	Date
1	1	1	1/1/2014
2	2	1	1/1/2014
3	3	2	2/2/2014
4	4	2	2/2/2014
5	1	3	3/3/2104
6	1	3	3/3/2014

Example MedSAS to CDW

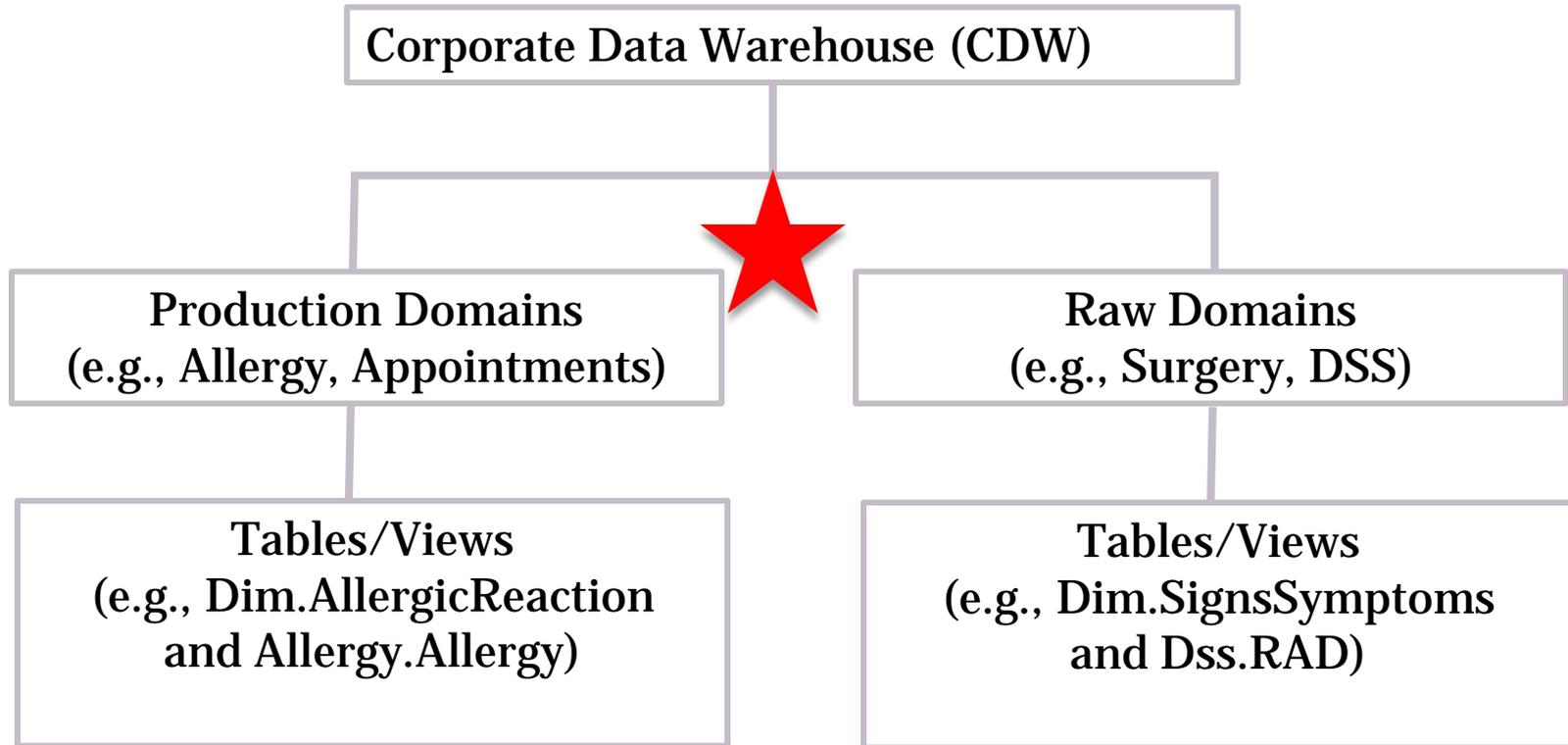
The Original MedSAS Inpatient Encounter File breaks into the following tables:

1. Patient.Patient
2. SPatient.SPatient
3. Outpat.Visit
4. Outpat.VProcedure
5. Outpat.VDiagnosis
6. Outpat.VProvider
7. PatSub.PatientServicePeriod
8. PatSub.PatientRace
9. PatSub.PatientEthnicity
10. Dim.Race
11. Dim.Ethnicity
12. Dim.MaritalStatus
13. Dim.Religion
14. Dim.Institution
15. Dim.Location
16. Dim.Eligibility
17. Dim.CPT
18. Dim.ICD
19. Dim.VistaSite
20. Dim.MeansTestCategory
21. Dim.Division

CDW → MedSAS

- A user would have to join together those 21 tables to approximate the MedSAS IE file
- There would be an extra set of variables that were not in MedSAS
- And, a set of derived variables that are in MedSAS would be missing

The Organization of CDW



“Domains”

- Groups of tables based on the subject matter found within the group of tables
- For example, adverse reaction information is grouped into the domain called *Allergy 1.0*

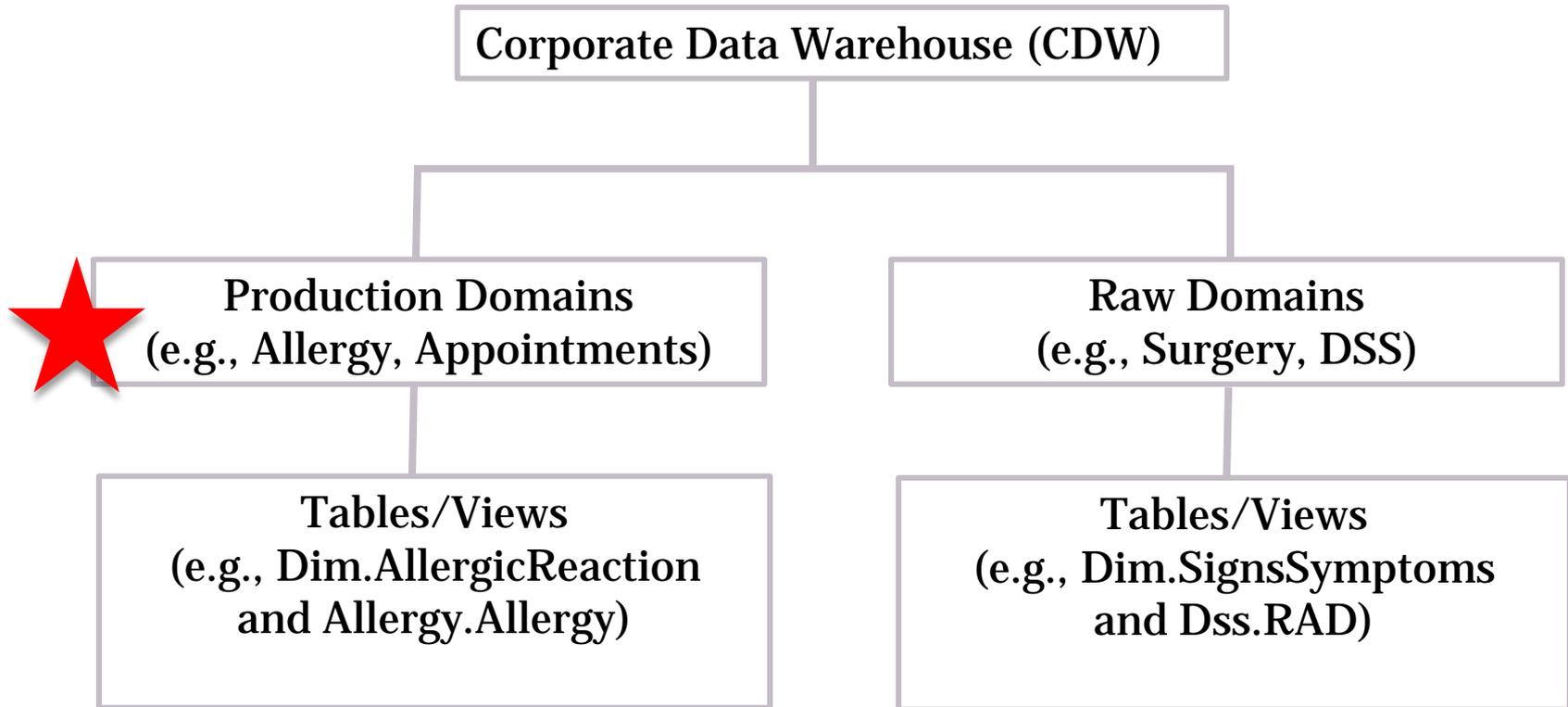
Tables in Allergy 1.0 Domain (4/2014)

Allergy.AllergicReaction	Dim.Reactant
Allergy.AllergicReactionAssessment	Dim.ReactantDrugClass
Allergy.Allergy	Dim.ReactantDrugIngredient
Allergy.AllergyComment	Dim.ReactantSynonym
Allergy.AllergyDrugClass	Dim.Reaction
Allergy.AllergyDrugIngredient	Dim.ReactionEffectiveDate
	Dim.ReactionSynonym

2 Types of Domains

1. **Production** – contains tables that have been structured by BISL to support re-joining of tables.
 2. **Raw** – contains tables that are direct extracts from the source system (e.g., VistA) that are simply being housed with no editing performed on them.
- ❖ **Note** - These domains are works-in-progress, and some raw domains will eventually become production domains.

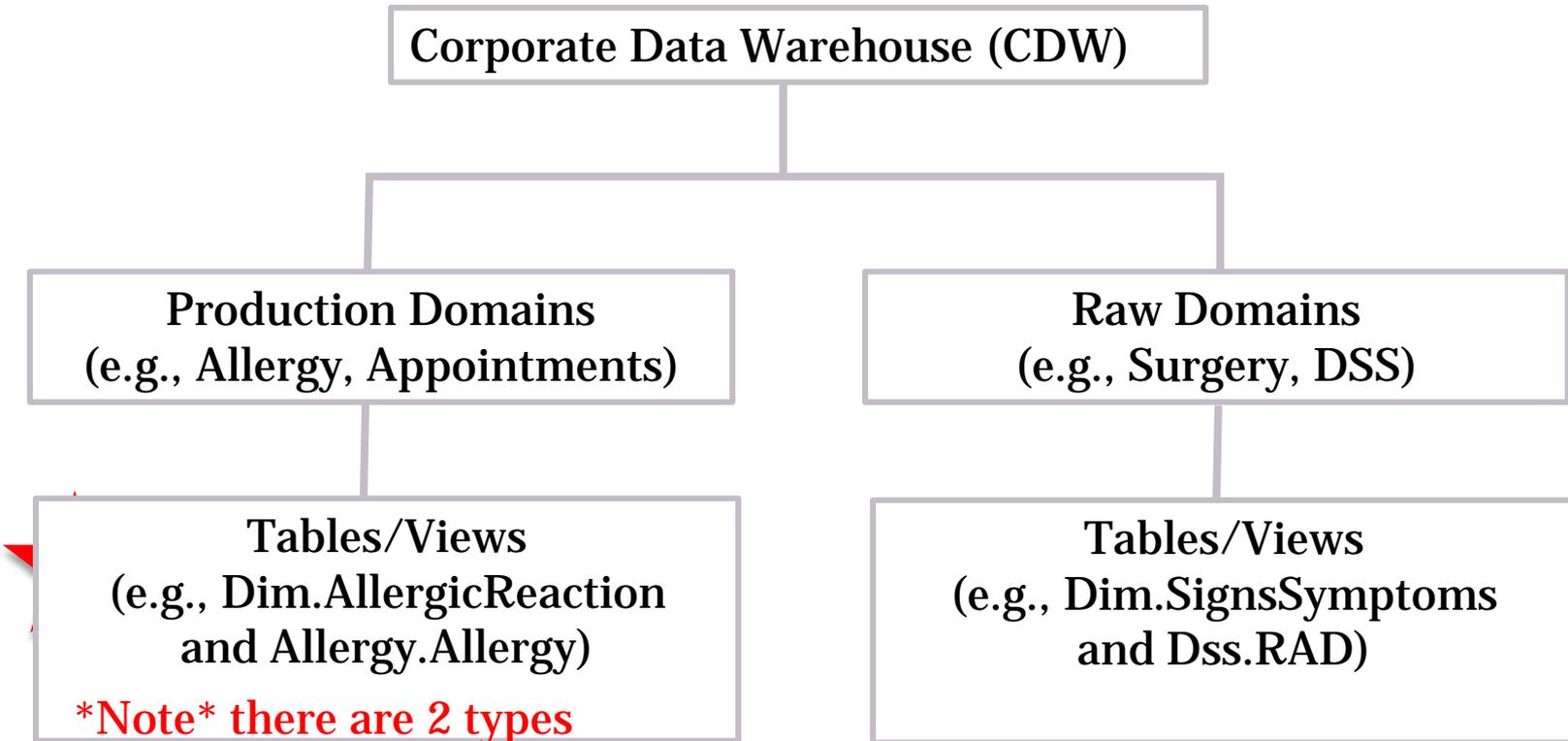
The Organization of CDW



Production Domains

- **Examples of Production Domains:**
 - **Consult 2.0**
 - **Lab Microbiology 1.0**
 - **Mental Health 1.0**
 - **Patient 2.0**
 - **Vital Sign 1.0**

The Organization of CDW



Tables v. Views

- A table is a set of columns and rows that contain data elements
- A view is the result of a procedure that pulls information out of a database into a virtual table; in simple terms it is a “virtual” table.

Sta3n	State	County	Country	PeriodOfService	MaritalStatus	InsuranceCoverageFlag	Religion
676	WISCONSIN	MILWAUKEE	UNITED STATES	PERSIAN GULF WAR	NEVER MARRIED	N	LUTHERAN
629	ARKANSAS	JEFFERSON	UNITED STATES	VIETNAM ERA	DIVORCED	U	BAPTIST
586	MISSISSIPPI	TIPPAH	UNITED STATES	PERSIAN GULF WAR	MARRIED	U	UNKNOWN/NO PREFERENCE
629	TEXAS	SMITH	UNITED STATES	POST-VIETNAM	SEPARATED	U	BAPTIST
636	NEBRASKA	LINCOLN	UNITED STATES	POST-KOREAN	*Missing*	U	*Missing*
623	OKLAHOMA	TULSA	UNITED STATES	PERSIAN GULF WAR	NEVER MARRIED	Y	UNKNOWN/NO PREFERENCE
676	PENNSYLVANIA	LANCASTER	UNITED STATES	POST-VIETNAM	NEVER MARRIED	NULL	ROMAN CATHOLIC CHURCH
623	OKLAHOMA	TULSA	UNITED STATES	VIETNAM ERA	MARRIED	Y	BAPTIST
676	WISCONSIN	MONROE	UNITED STATES	PERSIAN GULF WAR	MARRIED	Y	BAPTIST
636	CALIFORNIA	CALAVERAS	UNITED STATES	PERSIAN GULF WAR	MARRIED	U	CHRISTIAN (NON-SPECIFIC)
676	WISCONSIN	PORTAGE	UNITED STATES	OTHER NON-VETERANS	*Missing*	NULL	*Missing*
674	TEXAS	ANDERSON	UNITED STATES	WORLD WAR II	WIDOWED	Y	UNKNOWN/NO PREFERENCE
695	WISCONSIN	WAUKESHA	UNITED STATES	VIETNAM ERA	MARRIED	Y	*Missing*
674	TEXAS	HARRIS	UNITED STATES	VIETNAM ERA	DIVORCED	U	PROTESTANT, NO DENOMINATION

1. Dimension Tables

- These tables are used as supporting tables; each one holds a specific type of information that is meant to be accessed repeatedly
- These tables
 - do not contain patient information
 - can be viewed with basic read access
 - are relatively small in size

2. Fact Tables

- These tables hold measurements
- They tend to be very large (up to billions of records)
- They have patient and staff identifiers in them
- Therefore,
 - It is necessary to request permission to view subsets of these tables
 - It is important to plan your cohort to reduce the size of your request

Simple Example, Fact & Dimension

CustomerTable (Fact)

CustomerKey	LastName	FirstName	Address	CityState	Zipcode
1	Jones	Marianna	123 Oak St	Bee, AR	70788
2	Frank	Josie	11 Pine Ave	Flip, OK	30032
3	Plank	Bill	230 5 th St.	Miner, TX	11201

ItemTable (Dimension)

ItemKey	Item
1	Shirt
2	Pants
3	Skirt
4	Sweater

PurchasesTable (Fact)

PurchaseKey	ItemSoldKey	CustomerKey	Date
1	1	1	1/1/2014
2	2	1	1/1/2014
3	3	2	2/2/2014
4	4	2	2/2/2014
5	1	3	3/3/2104
6	1	3	3/3/2014

In CDW, Prefixes or Schema

- Each table name is preceded by a prefix or schema used to indicate whether or not it is a:
 - A fact (e.g., Schema.TableName)
 - A dimension table (always a “dim.” schema)

Fact Tables	Dimension Tables
Dental.DentalAlerts	Dim.OrderableItem
CPRSOrder.OrderedItem	Dim.AdmitSource
Immun.Immunization	Dim.Ethnicity
Patient.Patient	Dim.Antibiotic

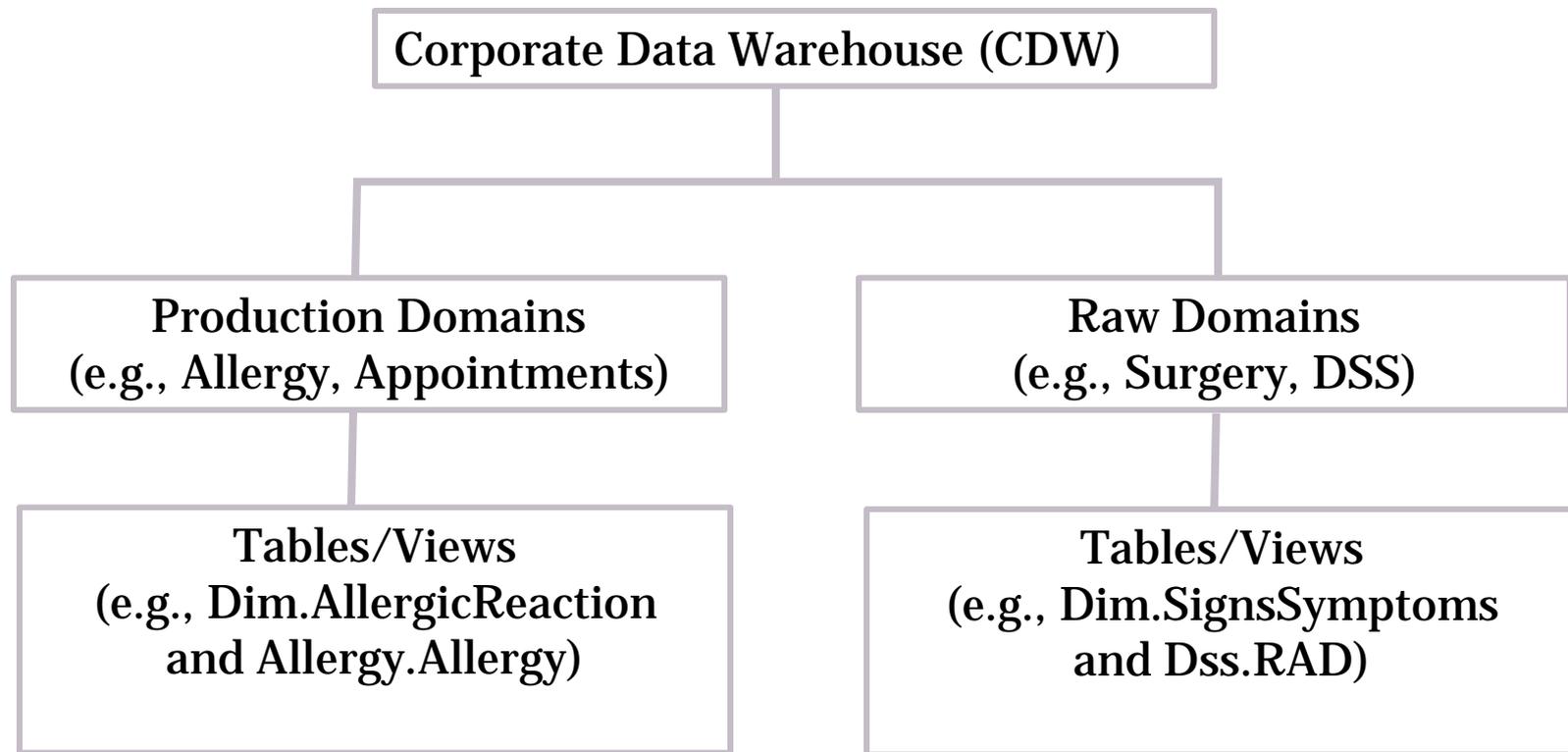
Examples of Dimension Tables

- The **Dim.ICD9** table from the Inpatient 2.0 domain contains:
 - the International Classifications of Diseases (ICD) Codes
 - a description of the diagnosis
 - a diagnostic related group (DRG) identifier
 - the major diagnostic code (MDC)
- The **Dim.Race** table in the Patient 2.0 domain contains:
 - a list of racial categories
 - a list of abbreviations

Examples of Fact Tables

- The **MH.SurveyResult** table in the Mental Health domain contains:
 - the name of the survey, instrument or test
 - a patient identification number
 - the date the test was given to the patient
 - the final raw score
 - any applicable transformed scores for the test
- The **Con.Consult** table from the Consult 2.0 domain contains:
 - the provisional diagnosis that prompted the request
 - time/date
 - the type of request
 - any procedures that are being requested
 - the urgency
 - the location
 - identifiers for the patient and the staff members involved

The Organization of CDW



The “joining keys” used to link tables in the production domain.

Joining Keys

- **Primary Key** –A column in every table that uniquely identifies each row.
 - For example, each drug listed in Dim.LocalDrug has a primary key associated with it.
- **Foreign Key** –These are column(s) in a table that correspond to or reference a primary key in another table.
 - Values of the foreign key may repeat
 - Names of foreign keys may differ from names of their matching primary keys

Simple Example, Joining Keys

CustomerTable (Fact)

CustomerKey (PK)	LastName	FirstName	Address	CityState	Zipcode
1	Jones	Marianna	123 Oak St	Bee, AR	70788
2	Frank	Josie	11 Pine Ave	Flip, OK	30032
3	Plank	Bill	230 5 th St.	Miner, TX	11201

ItemTable (Dimension)

ItemKey (PK)	Item
1	Shirt
2	Pants
3	Skirt
4	Sweater

PurchasesTable (Fact)

PurchaseKey (PK)	ItemSoldKey (FK)	CustomerKey (FK)	Date
1	1	1	1/1/2014
2	2	1	1/1/2014
3	3	2	2/2/2014
4	4	2	2/2/2014
5	1	3	3/3/2104
6	1	3	3/3/2014

Appt.Appointment	
AppointmentSID (PK)	bigint
PatientEN	varchar(50)
AppointmentEN	varchar(50)
Sta3n	smallint
PatientSID (FK)	int
AppointmentDateTime	datetime2(0)
AppointmentVistaErrorDate	varchar(50)
AppointmentDateTimeTransformSID	bigint
AppointmentDateSID (FK)	int
AppointmentMadeDateTime	datetime2(0)
AppointmentMadeVistaErrorDate	varchar(50)
AppointmentMadeDateTimeTransformSID	bigint
AppointmentMadeDateSID (FK)	int
AppointmentTypeSID (FK)	smallint
CancelNoShowCode	varchar(50)
VisitSID (FK)	bigint
LocationSID (FK)	int
DesiredAppointmentDate	date
DesiredAppointmentVistaErrorDate	varchar(50)
DesiredAppointmentDateTransformSID	bigint
PurposeOfVisit	varchar(50)
DataEntryStaffSID (FK)	int
AutoRebookDateTime	datetime2(0)
AutoRebookVistaErrorDate	varchar(50)
AutoRebookDateTimeTransformSID	bigint
SchedulingRequestType	varchar(50)
NextAvailableAppointment	varchar(50)
FollowUpVisitFlag	char(1)
CancellationReasonSID (FK)	int
CancellationRemarks	varchar(255)
CancelDateTime	datetime2(0)
CancelVistaErrorDate	varchar(50)
CancelDateTimeTransformSID	bigint
CancelDateSID (FK)	int
CancelByStaffSID (FK)	int
LengthOfAppointment	decimal(9,4)
OverbookFlag	char(1)
EligibilitySID (FK)	int
AppointmentCancelledFlag	char(1)
AppointmentMadeFromLocationDate	date
AppointmentMadeFromLocationVistaErrorDate	varchar(50)
AppointmentMadeFromLocationDateTransformSID	bigint
ConsultSID (FK)	bigint
CheckInStaffSID (FK)	int
CheckInDateTime	datetime2(0)
CheckInVistaErrorDate	varchar(50)
CheckInDateTimeTransformSID	bigint
CheckInEnteredDateTime	datetime2(0)
CheckInEnteredVistaErrorDate	varchar(50)
CheckInEnteredDateTimeTransformSID	bigint
CheckOutStaffSID (FK)	int
CheckOutDateTime	datetime2(0)
CheckOutVistaErrorDate	varchar(50)
CheckOutDateTimeTransformSID	bigint
CheckOutEnteredDateTime	datetime2(0)
CheckOutEnteredVistaErrorDate	varchar(50)
CheckOutEnteredDateTimeTransformSID	bigint
FileManFileNumber: 2.98	
FileManFileName: APPOINTMENT	

Appointment

Entity Relationship (ER) Diagrams are drawings of the relationships between the tables in the domain.

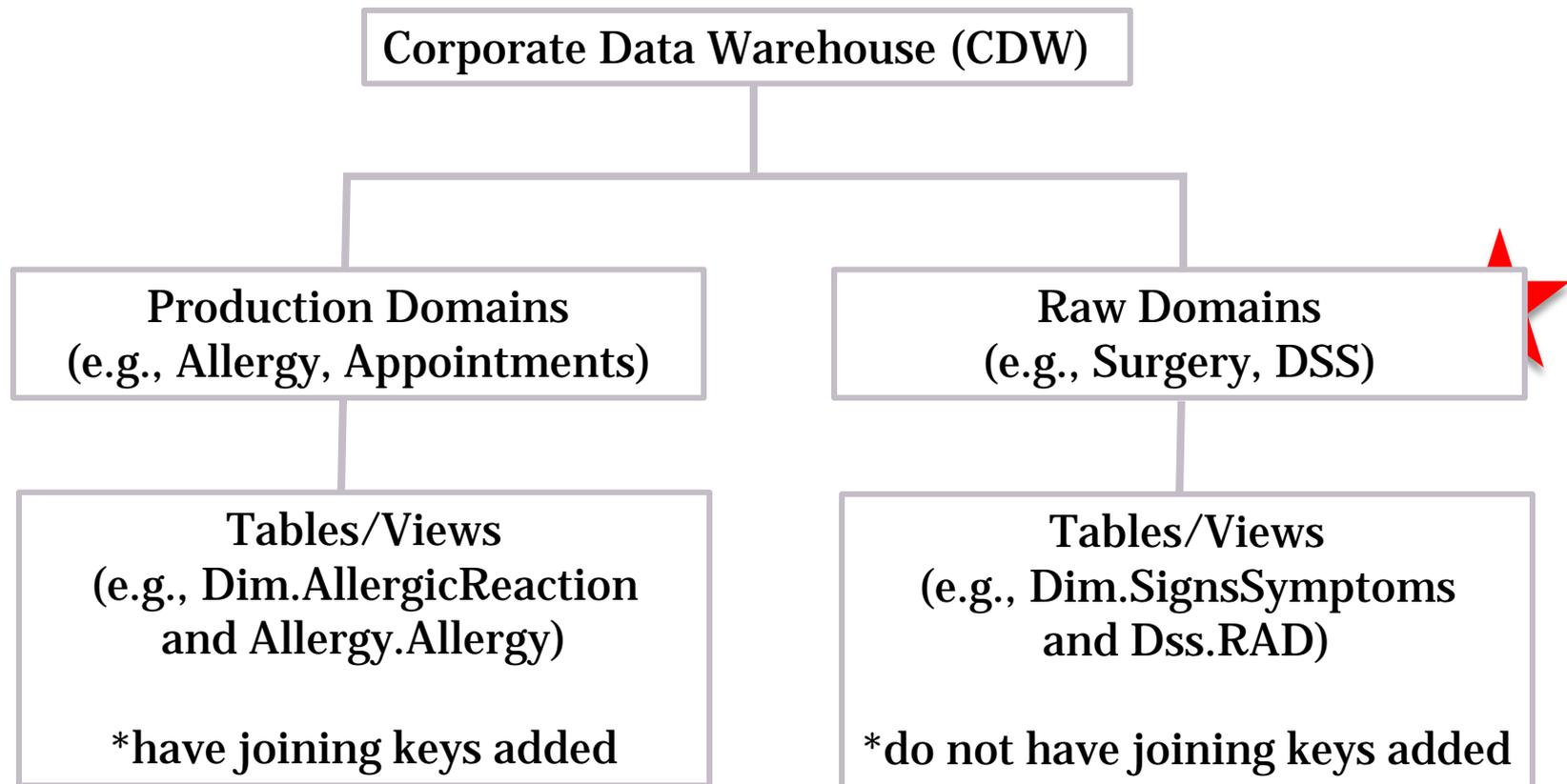
(PK) indicates a Primary Key

(FK) indicates a Foreign Key

These two join on *CancellationReasonSID*

Dim.CancellationReason	
CancellationReasonSID (PK)	int
CancellationReasonEN	varchar(50)
Sta3n	smallint
CancellationReason	varchar(50)
CancellationReasonType	varchar(50)
CancellationReasonSynonym	varchar(50)
CancellationReasonInactiveFlag	char(1)
Dimension Type: Fixed	
FileManFileNumber: 409.2	
FileManFileName: CANCELLATION REASONS	

The Organization of CDW



Raw Domains

Examples of Raw Domains

DSS

Oncology
Prosthetics
Radiology
Surgery

TIU

A couple of notes:

- The **Decision Support System (DSS)** domain is the same database that had been housed on the Austin mainframe.
- **TIU (Text Integrated Utilities) notes** are here while the titles of these text notes are in the production domains.
- A deeper discussion of these domains is beyond the scope of this seminar.

II. Places to Learn about the Data

- There are 6 main sites housing information about the data:
 1. VHA Data Portal
 2. VA Information Resource Center (VIReC)
 3. CDW SharePoint
 4. Data Architecture Repository (DAR)
 5. VA Informatics and Computing Infrastructure (VINCI)
 6. Managerial Cost Accounting Office (MCAO) National Data Reporting Page

VHA Data Portal

- **This site provides:**
 - **Links to all of the other sites mentioned in this talk**
 - **Information about how to apply for access to CDW data**
- **This is good place to find the links to the documentation you may need**

VHA Data Portal

This site guides users to all of the other sites mentioned in this talk

Quick View
CDW is an evolving repository of national VHA data. It is a physical implementation of a logical data model at the enterprise level for VHA.

CDW
Overview
Content
Structure
Access
Resources

VHA Corporate Data Warehouse (CDW)

Overview

VHA's Corporate Data Warehouse (CDW) is a national repository comprising data from several VHA clinical and administrative systems. The objective of CDW is to facilitate reporting and data analysis at the enterprise level by incorporating data from multiple data sets throughout the VHA into one standard database structure. CDW provides data and tools to support management decision making, performance measurement, and research objectives.

Content

CDW is an evolving repository of national VHA data containing clinical, enrollment, financial, administrative, utilization, Veteran benefits information and more. Data are available from October 1999 to present. Visit the [CDW SharePoint site](#) and [VINCI Data Descriptions](#) page for more information about CDW data.

Structure

CDW uses multiple servers to store data. Data on these servers are stored in relational databases organized into data domains. Each domain comprises logically or conceptually related sets of data tables. Domains generally indicate the VistA application from which most of the data elements originate (e.g., Vital Signs or Mental Health Assessment).

CDW SQL databases are categorized as production or raw.

- **CDW Production (CDW-Prod)** – Data on the CDW-prod servers are mostly uploaded from VistA with no filtering of records, editing, or business rules applied. These data are modeled and indexed to promote efficient querying.
- **CDW Raw (CDW-RAW)** – CDW-Raw refers to the test and evaluation database. CDW-RAW data are extracted directly from the data source (e.g., VistA). These data are not modeled, standardized, or indexed and include DSS NDEs in SQL tables and other SAS-to-SQL tables. When the domain is developed for production it is removed from Raw.

CDW also includes copies of [Austin Information Technology Center \(AITC\) Mainframe SAS](#) data sets, including legacy DSS NDE SAS data sets.

Access

The process for requesting access to CDW data depends on the data access category. Select a category from the list below for detailed information on requesting access to these data:

- [Operations/ Non-Research](#)
- [Preparatory to Research](#)
- [Research](#)

Visit the [Data Access](#) page for more information about accessing VHA data.

Resources

- The [CDW SharePoint site](#) contains information on newly released and missing data, an up-to-date list of data domains, structural documentation, and more.
- VIREC provides the following resource on CDW:
 - [VIREC Resource Guide: VA Corporate Data Warehouse \(CDW\)](#) provides a general overview of the CDW including CDW structure, content, and utility.
 - [VIREC Summary Documentation](#) on CDW datasets includes SAS Proc Contents and record level frequencies.
 - [VIREC's CDW Cyber Seminar Mini-Series](#) provides information on using the CDW.
- Visit the [VINCI Workspace](#) page to learn more about using the VINCI Workspace to store and analyze CDW data.

All Rights Reserved Department of Veteran Affairs | [Sitemap](#) | [Section 508 Accessibility](#) | [Privacy Policy](#) | [OIA Homepage](#) | [HSR&D Homepage](#) | [Contact Us](#)

From the home page, you can click:

**-Data Sources
-CDW**

**Then, expand
Corporate Data
Warehouse and
select the CDW
link.**

Direct Link- <http://vaww.vhadataportal.med.va.gov/DataSources/CDW.aspx>

VIReC Products

- VIReC has three main products available:
 1. Archived Cyberseminars
 1. Metadata
 2. Lab Chemistry Domain
 3. Workload Calculations
 2. A CDW Resource Guide
 3. A set of products that allow a glimpse into CDW data
- Come here for summary information and a tour of the data before you have access

VA INFORMATION RESOURCE CENTER (VIREC)

- [VIREC Home](#)
- [VA/CMS Home](#)
- [About Us](#)
- [New Users of VA Data](#)
- [Data Transition to CDW](#)
- [News & Updates](#)
- [FAQs](#)
- [Acronyms](#)
- [HelpDesk](#)



[Data Issues Brief, May 2014](#)

[Upcoming Cyberseminar](#)

[New: Risk Adjustment Tutorial](#)

[VIREC News & Updates](#)

[Data Transition to CDW](#)

At a Glance

[How VIREC Can Help](#): Introduction to resources VIREC develops for researchers using VA data.

[Data Issues Brief](#): News and updates about VA data and information systems.

[HSRData-L Listserv](#): Virtual community of VA researchers sharing knowledge about VA data and database use.

[VA/CMS Data for Research Project](#): VIREC serves as the data custodian for Centers for Medicare and Medicaid Services (CMS) data for research use in the VA.

[Journal Supplements](#): Collaborations with peer reviewed journals on VA data and informatics research topics.

Requesting & Accessing Data

[Data Access Overview](#): VIREC provides guidance on accessing and requesting data through our collaboration on the VHA Data Portal.

[Data Tools](#): Information about tools used to work with VA data.

[Preparatory to Research](#): Overview of using data during protocol development.

[DART](#): Introduction to the Data Access Request Tracker.

Resources for Researchers

[Data Sources & Data Topics](#): Select a specific data source or data topic described by VIREC.

[Research User Guides](#): Detailed descriptions of select VA data sources, including variable descriptions.

[Data Reports](#): Technical reports, data investigations, and data quality updates.

[Summary Information](#): Current and historical information about dataset variables, counts, frequencies, and more.

[Publications](#): A complete list of peer reviewed articles and journal supplements published by VIREC.

[VHA Data Portal](#): A one-stop-shop for information about VHA data including details on requesting data access.

Education

[VIREC Cyberseminars](#): Key issues in clinical informatics and database use.

[Toolkit for New Users](#): A starting point for learning about VA data.

[Tutorials](#): Step-by-step instruction on accessing and using select VA data and information systems.

[Presentations](#): Learn about topics relevant to VA research from experts in the field.

Click here for the CDW Resource Guide & A glimpse of the data

Click here for Cyberseminars

<http://vaww.virec.research.va.gov/CDW/Overview.htm>

The Resource Guide



VA INFORMATION RESOURCE CENTER (VIREC)

- VIREC Home
- VA/CMS Home
- About Us
- New Users of VA Data
- Data Transition to CDW
- News & Updates
- FAQs
- Acronyms
- HelpDesk

VHA Corporate Data Warehouse (CDW)

Overview

The Veterans Health Administration (VHA) Corporate Data Warehouse (CDW) is a national repository of data from VISTA and several other VHA clinical and administrative systems. The CDW is physically located at the Austin Information Technology Center.

The CDW is a relational database organized into a collection of data domains. Domains represent logically or conceptually related sets of data tables. Domain themes generally indicate the application in the VistA electronic health record system from which most of the data elements in the domain come (e.g., Vital Signs or Mental Health Assessment).

Available Data

Data are available from October 1999 - present. Availability varies by domain. CDW's SharePoint site provides an up-to-date list of available domains with some structural documentation in the Metadata Reports section.

CDW data domains are classified as either production or raw. VistA is the primary source of CDW data. No filtering of records, editing of content or business rules have been applied to these data. Raw data domains reflect the source system structure (VistA) and are not modeled, standardized, or indexed. Tables in production domains have been structured to support flexibility of queries.

Access

The process for requesting access depends on how the data will be used. VIREC provides data access information through our collaboration on the VHA Data Portal Intranet site. Visit the Portal for details on accessing CDW data by selecting from the following data use categories:

- Research
- Preparatory to Research
- Operations

Resources for Researchers

- [VIREC Resource Guide: VA Corporate Data Warehouse \(CDW\)](#) provides a general overview of the CDW including CDW structure, content and utility.
- [Summary Documentation](#) on the CDW datasets includes SAS Proc Contents.
- Medical SAS Outpatient variables have been mapped to CDW data fields. Visit the ["NPCD CDW Transition"](#) page on CDW's SharePoint site to access the map and for information on differences in variable formatting, data transformations and calculated variables only in the Medical SAS files.
- VINCI's Intranet site, [VINCI Central](#), provides in-depth information on VINCI including data sources, support, training, and guides for using VINCI.
- The [CDW SharePoint site](#) has an announcements section that provides up-to-date information on topics such as newly released data, missing data and training available.



Expand "Data Sources"

Select "CDW"

Find resource guide here



<http://vaww.virec.research.va.gov/CDW/Overview.htm>

VIReC - CDW Documentation

**List of columns
& their formats**

20 sample records

CDW Documentation

Overview
VIReC documentation for the Corporate Data Warehouse (CDW) includes domain layouts, data contents, sample records, and frequencies. This summary information is available by domain or schema.

Domain Layout
A list of schemas and tables for each CDW domain.
[View Domain Layout](#)

Data Contents
A list of data elements for each CDW domain.
[View Data Contents](#)

Sample Records
Randomly selected records for each CDW domains.
[View Sample Records](#)

Frequencies & Counts
Record counts, null counts, and selected discrete value frequencies by schema.
[View Frequencies & Counts](#)

CDW

- Overview
- Data Transition to CDW
- Documentation

General Resources

- Data Access
- Data Sources
- Data Tools
- Data Topics
- Products & Services

VIReC Home | VA/CMS Home | VA R&D Home | VA HSR&D Service | Viewers | VIReC Sitemap | Contact Us

VA Intranet Home | About VA | Organizations | Locations | Section 508 Accessibility | No Fear Act
Reviewed/Updated Date: January 30, 2014

List of domains & tables

**Frequencies for
discrete values &
counts of null values**

- <http://vaww.virec.research.va.gov/Intro/Documentation-Index.htm>

- These products are meant to provide a preliminary look at resources found inside CDW (especially the Data Profiling tables and searchable metadata tables) that aren't viewable until after the researcher applies for access.

CDW SharePoint

- **CDW has four main sources of metadata:**
 1. Metadata for production domains
 2. Metadata for raw domains
 3. SAS & VistA - CDW Crosswalks (XWalk)
 4. Four user guides:
 - Introduction and Policies
 - Projects and Development
 - Query Best Practices
 - Querying CDW Raw
- Here a user can find information about VHA data origins, content of the tables, linking keys, data formats and help with SQL

Metadata for Production Domains

The screenshot shows the VA Metadata portal interface. The browser address bar displays <https://vaww.dwh.cdw.portal.va.gov/metadata/default.aspx>. The page header includes the United States Department of Veterans Affairs logo and the text "United States Department of Veterans Affairs Metadata". A navigation menu contains "CDW Home", "CDW Support", "Community", "Domain Teams", "Metadata", and "TechTeam". A left sidebar lists "MetaData Home", "Libraries", "Reports", "Metadata Documents", "Data Sources", "Lists", and "Discussions". The main content area features a link to "Execute the [Metadata Report](#)" and a sub-note: "-- ER Diagrams: click on the SubModel name in the [Metadata Report](#)". Below this is a table of "Metadata Documents".

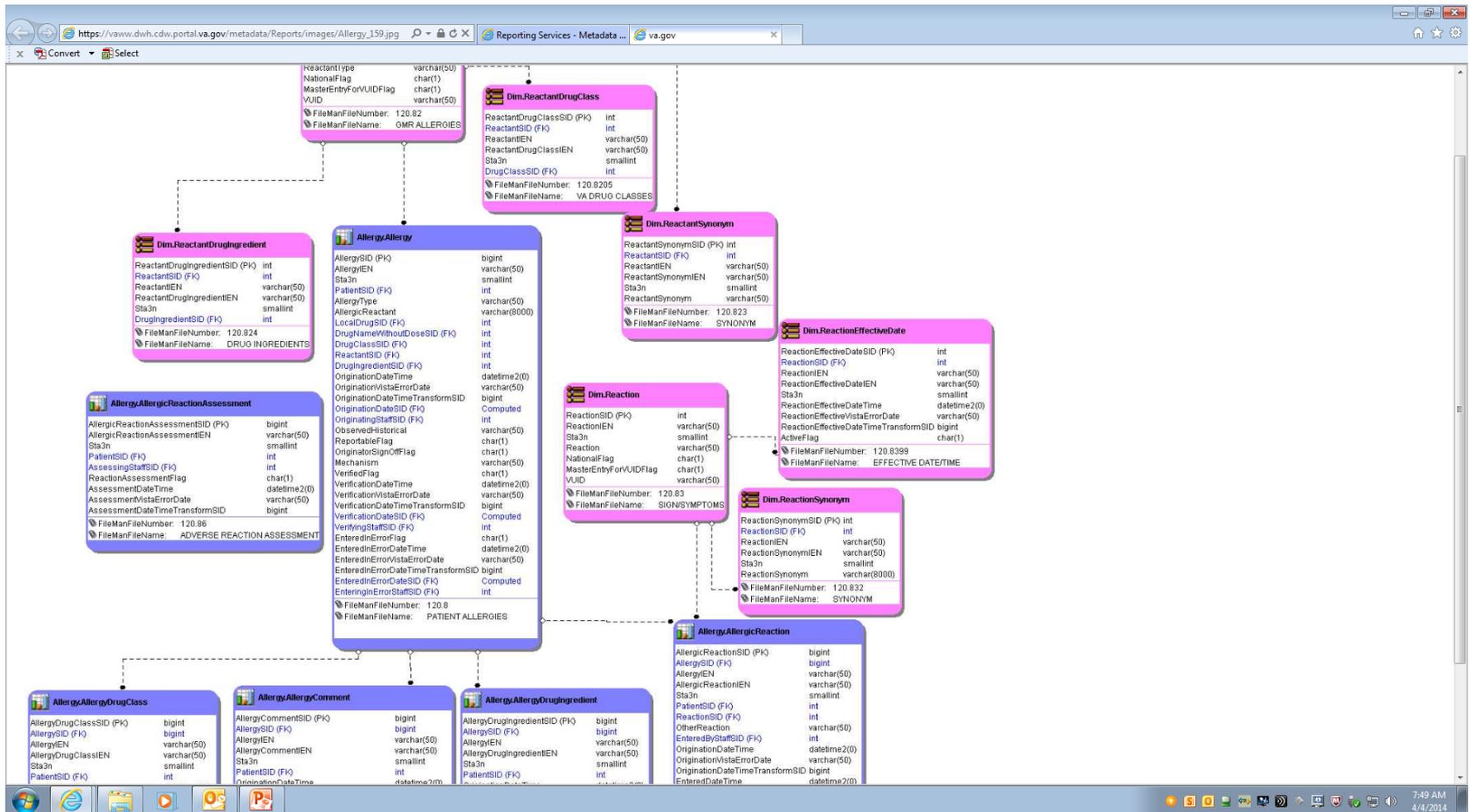
Type	Name	Modified	Modified By
Folder	Allergy 1.0	2/19/2014 6:15 PM	Anderson, Stephen M. (CDW)
Folder	Appointment 2.0	2/19/2014 4:25 PM	Anderson, Stephen M. (CDW)
Folder	Consult 2.0	2/19/2014 4:27 PM	Anderson, Stephen M. (CDW)
Folder	Dental 1.0	2/19/2014 5:52 PM	Anderson, Stephen M. (CDW)
Folder	DSS	2/19/2014 5:59 PM	Anderson, Stephen M. (CDW)
Folder	ICD-10	2/19/2014 5:51 PM	Anderson, Stephen M. (CDW)
Folder	Inpatient 1.0	2/19/2014 4:22 PM	Anderson, Stephen M. (CDW)
Folder	Inpatient 2.0	2/19/2014 4:22 PM	Anderson, Stephen M. (CDW)
Folder	Lab Chem 2.0	2/19/2014 5:52 PM	Anderson, Stephen M. (CDW)
Folder	Microbiology 1.0	2/19/2014 5:54 PM	Anderson, Stephen M. (CDW)
Folder	Patient 3.0	4/2/2014 12:24 PM	Nielsen, Hans C.
Folder	Pharmacy	2/19/2014 5:48 PM	Anderson, Stephen M. (CDW)
Folder	Purchased Care (Fee) 1.0	2/19/2014 6:01 PM	Anderson, Stephen M. (CDW)
Folder	Race Ethnicity	2/19/2014 6:01 PM	Anderson, Stephen M. (CDW)
Folder	Text Patient Logic	2/19/2014 5:54 PM	Anderson, Stephen M. (CDW)
Folder	TIU (Text Document Titles) 1.0	2/19/2014 5:51 PM	Anderson, Stephen M. (CDW)
File	DateTimeTransformations	10/11/2012 6:32 PM	Anderson, Stephen M. (CDW)
File	DRG to MDC Explanation	1/2/2013 12:32 PM	Pham, Richard H. VHAREN
File	Linking Patient Data in the CDW	4/25/2013 12:29 PM	Kennedy, Lauren
File	Medication Image Library	3/16/2013 7:24 AM	Pham, Richard H. VHAREN
File	NPCD Business Rules_042012	5/15/2012 12:29 PM	Kennedy, Lauren
File	studentsT	12/20/2012 7:35 AM	Hickey, Susan
File	Tier 2 Definition	12/8/2011 9:51 PM	Anderson, Stephen M. (CDW)
File	Using The Meta Schema To View The CDW Translation	12/20/2012 1:17 PM	Pham, Richard H. VHAREN

Diagrams & Table Level Metadata

Submodel (links to diagrams)	Schema	Table Name	Vista File	Key Date Column	Table Description	Table Description Author	Description Date	Table Description Technical
Allergy 1.0	Allergy	AllergicReaction	REACTIONS	OriginationDateTime			3/16/2013 9:29:00 AM	This table will have no cutoff date. All rec
	Allergy	AllergicReactionAssessment	ADVERSE REACTION ASSESSMENT	AssessmentDateTime			10/9/2013 2:44:00 PM	There will be no cutoff field for this table.
	Allergy	Allergy	PATIENT ALLERGIES	OriginationDateTime		Pham 4/2013	2/25/2011 12:03:00 AM	Cutoff logic - Take all allergies. There sho
	Allergy	AllergyComment	COMMENTS	OriginationDateTime			8/3/2012 9:53:00 PM	No cutoff date. Take all records.
	Allergy	AllergyDrugClass	DRUG CLASSES	OriginationDateTime			4/1/2013 6:05:00 AM	There is no cutoff date for this table. Take
	Allergy	AllergyDrugIngredient	DRUG INGREDIENTS	OriginationDateTime			4/1/2013 6:05:00 AM	There is no cutoff date for this table. Take
	Dim	Reactant	GMR ALLERGIES				5/3/2012 11:58:00 PM	
	Dim	ReactantDrugClass	VA DRUG CLASSES				3/17/2013 11:04:00 PM	
	Dim	ReactantDrugIngredient	DRUG INGREDIENTS				5/22/2013 9:45:00 AM	
	Dim	ReactantSynonym	SYNONYM				3/20/2013 7:57:00 PM	
Appointment 2.0	Appt	Appointment	APPOINTMENT	AppointmentDateTime			4/14/2011 4:02:00 PM	The cutoff field for Appointment is Appoin
	Appt	AppointmentMultiple	PATIENT	AppointmentMadeDate			10/9/2012 4:26:00 PM	The cutoff field for AppointmentMultiple is
	Dim	CancellationReason	CANCELLATION REASONS				10/9/2012 9:07:00 PM	Records that the AppointmentMadeDate are not processed

Diagrams are intended to be displays of the content of and links between the tables.

Table descriptions define the types, meanings and sources of the measurements.



- Blue boxes represent fact tables & pink represent dimension tables.
- The lines represent links between tables.
- Measurements used to link (keys) are labeled in the image with (PK) or (FK).

Metadata for Raw Domains

United States Department of Veterans Affairs Database - Raw Extractor
CDW Database Team

Community

Recently Modified
CDW Database
Raw Extractor
Master Server List
Team Pictures

CDW RAW
Libraries
Site Pages
Shared Documents
Reports
Pages
Database Team Wiki
Weekly Schedule Review

Lists
Tasks
DBA Team Tasks
DBA Team Calendar

Discussions
Team Discussion

Database Team Blog

Site Content
Announcements
Database Team
DBA Team Calendar
DBA Team Task
Extraction Schedules
Links
List
Pages
Reports
Shared Document
Site Assets

Name	Schedule	Last Extract	Next Extract	Time	Records/Space	NDS Privilege	Project Notes
Allergies v1 Schema: Allergies since Dec 2010	Weekly Monday	MAR 31 2014	APR 07 2014	12 Hours	158 M recs 8 Gb	Basic (CDW_Full)	6/20/13 - Allergies will be promoted to CDW-Prod toward 2014 Extracts performed on same schedule as Non-VA Meds. POC: Mike Wrobel
Beneficiary Travel (CB0) v2 Schema: Travel since Dec 2010	Bi-Weekly Payday Wednesday	MAR 20 2014	APR 09 2014	10 Hours	105 M recs 32 Gb	Basic (CDW_Full)	POC: Kristina Ebersole Change extraction day from Wednesday to Monday
Bill Claims (CB0) v6 AKA: Integrated Billing Schema: BillClaims since Oct 2010	Quarterly	JAN 17 2014	APR 04 2014	One weekend+ cleanup	1.6B+ recs 214 Gb Date ranged extracts	Privileged (CDW_SPatient)	3/18/13 - requires Privileged Read Access (SPatient) Q4 extract included these tables, back to 10/1/2011: CDWWork.BillClaims_v3_ar_transaction_233 CDWWork.BillClaims_v3_patient_enrollment_27_11 CDWWork.BillClaims_v3_explanation_of_benefits_361_1 CDWWork.BillClaims_v3_billiclaims_399 CDWWork.BillClaims_v3_integrated_billing_action_350 CDWWork.BillClaims_v3_claims_tracking_type_356_8 CDWWork.BillClaims_v3_claims_tracking_nonbilla_356_8 CDWWork.BillClaims_v3_billiclaims_prescrip_362_2 CDWWork.BillClaims_v3_nonbillable_clinic_stop_codes_352_3 CDWWork.BillClaims_v3_nonbillable_clinic_352_4 CDWWork.BillClaims_v3_sub_procedures_399_0504 CDWWork.BillClaims_v3_claims_tracking_356 CDWWork.BillClaims_v3_accounts_receivable_430 CDWWork.BillClaims_v3_ar_debtor_340 CDWWork.BillClaims_v3_ar_insurance_type_2_352 CDWWork.BillClaims_v3_insurance_company_36 Ernesto Castro's team decides what data to extract next, the date ranges, but they move data out of the CDW to their more comprehensive billing database. View the link on the left for more detailed information.
CAPRI Audit Trail							POC: Peter Change Peter.Cheng1@va.gov Team also uses CBP data This is not a standard VistA database. Rather, it is a stand-alone database in Falling Waters. so the CDW data comes from a single database. Each file extracted requires a

Under
"Community"
& "CDW Raw"

- Contains status reports on the raw domains
- <https://vaww.dwh.cdw.portal.va.gov/TechTeam/DBA/SitePages/Raw%20Extractor.aspx>

SAS & VistA Crosswalks

Department of Veterans Affairs
Corporate Data Warehouse

NDS » Home

CDW Home CDW Support Community Data Teams MetaData TechTeam

Home
Dashboards
Monitor
Nagios
Network
Perceiver

Projects
3-Tier Support
CDW CRISP Audit
Domain Teams Portal
NPCD CDW Transition
Open Data Initiative
Standardization

Reports
Acronyms/Terms
Applications
CDW Customers and Projects
LSV Customers and Stations
CDW FAQs
Common Values
CPT-ICD Xwalk
ICD-9 Categories
Facilities
FIPS Codes
SAS Datasets
SAS XWalk
Schemas
VistA-CDW Xwalk

Toolbox
CDCO Monitor
Energy Units
HIPAA Rules
Section 508
Shared Documents
SharePoint Quirks

Community Home
CDW Governance Board
SharePoint Governance Board
VA GIS User Home
RDW
NDS
IEHR Mapping
CDW RAW
Data Access Committee

Systems
reports
development
it status and artifacts
tools
sports and Toolbox items open in a new tab/window
links open in the same window

Useful links - gentle reminder to right-click link to open elsewhere

- ANR, Automated Notification Reports list
- ARC, Allocation Resource Center
- DAR, Data Architecture Repository
- DSS, Decision Support System reports
- DSS program documents including stop code identifiers
- HIG, Health Information Governance
- NDS, NDS intranet organizational website
- NPCD, National Patient Care Database
- OIA, Office of Informatics and Analytics
- PSSG, Planning Systems Support Group
- VINC, Veterans Information and Computing Infrastructure
- VIREC, VA Information Research Center
- VSSC, VHA Support Service Center
- VHA Data Access
 - VHA Data Consortium
 - VHA Data Portal
 - VHA Forms and Templates
 - VHA VACO
- Wiki, VA portal

Click “Community”
Then select “NDS” from drop down menu
Then click Xwalk of choice in left banner

These links will open a search that will allow users to enter either the name of a measurement in CDW or the name of a measurement from VistA or SAS (depending on choice) and find out what it is called in the other source.

<https://vaww.dwh.cdw.portal.va.gov/Regions/NDS/SitePages/Home.aspx>

Data Architecture Repository (DAR)

- This site serves as a repository for metadata from around the VA.
- It includes metadata for VHA information and VistA information.
- VistA is one of the primary sources of CDW data.
- Referencing these files can help clarify the meaning of CDW data.

DAR's VistA Metadata

the Metadata Repository

Metadata Repository provides a means to catalog, search, report, and manage metadata via a web-accessible portal. Metadata is data about data; data that defines and describes the characteristics of data. Availability of an enterprise metadata repository enables data management and governance; promotes data understanding, data quality, data standardization, standardization, and data integration. Below, use the tabs above to view all Metadata, Value Domain, or Thesaurus content, or click on a specific portal within Metadata Portal below.

Search For: Search IN: Metadata Name Only WITH: Any of these words
 Identifying Fields Only All of these words
 All Attribute Fields These words as a phrase

Use Thesaurus

Use the options below to apply filter(s) to search results, returning only matches on selected criteria

Metadata Source	Metadata Category	Metadata Type
<input checked="" type="checkbox"/> Any	<input checked="" type="checkbox"/> Any	<input checked="" type="checkbox"/> Any
<input type="checkbox"/> DSS NDE	<input type="checkbox"/> Application	<input type="checkbox"/> Application
<input type="checkbox"/> NCA AMAS	<input type="checkbox"/> Business Rule	<input type="checkbox"/> BIRLS Column
<input type="checkbox"/> NCA BOSS	<input type="checkbox"/> Data Element	<input type="checkbox"/> CSP Element
<input type="checkbox"/> NCA FNOD	<input type="checkbox"/> Data Model	<input type="checkbox"/> Child Table
<input type="checkbox"/> VA FMS	<input type="checkbox"/> Data Store	<input type="checkbox"/> OSS NDE
<input type="checkbox"/> VA NCVAS	<input type="checkbox"/> File	<input type="checkbox"/> Foreign Key
<input type="checkbox"/> VA OALC	<input type="checkbox"/> Information Domain	<input type="checkbox"/> Schema
<input type="checkbox"/> VA VADS	<input type="checkbox"/> Schema	<input type="checkbox"/> USVETS Field
<input type="checkbox"/> VA VIM		<input type="checkbox"/> VHIM Attribute
<input type="checkbox"/> VBA BIRLS		<input type="checkbox"/> VistA Cross-reference
<input type="checkbox"/> VBA CH33 LTS		<input type="checkbox"/> VistA Index
<input type="checkbox"/> VBA Corporate		<input type="checkbox"/> BIRLS Table
<input type="checkbox"/> VBA VADIR		<input type="checkbox"/> CSP Study
<input type="checkbox"/> VHA ADR		<input type="checkbox"/> Column
<input type="checkbox"/> VHA CSP		<input type="checkbox"/> Enterprise Information Domain
<input type="checkbox"/> VHA Corporate Database Monograph		<input type="checkbox"/> Monograph Database
<input type="checkbox"/> VHA Data Stewardship		<input type="checkbox"/> Table
<input type="checkbox"/> VHA VistA Platinum		<input type="checkbox"/> Unique Key
		<input type="checkbox"/> VHIM Classifier/Element
		<input type="checkbox"/> VHIM Domain/Package
		<input type="checkbox"/> VistA Field
		<input type="checkbox"/> VistA Package
		<input type="checkbox"/> VistA File

VA

- FMS - Financial Management System is a corporate system that is the official system of record for the General Ledger (GL) and other financial management information for the Department of Veterans Affairs. FMS is a mainframe based application that provides batch oriented processing for transactional updates and report generation.
- NCVAS - The NCVAS (National Center for Veterans Analysis and Statistics) Analysis System is a dimensional database to be populated with the critical Veteran data and information necessary for the Department of Veterans Affairs to provide a comprehensive view of the Department's Veterans. NCVAS is used to create sophisticated datasets such as Unique Veterans Lists of VA, custom by Role, and based on data for "thematic" analysis.

**Select “VHA” from the left banner
Then select “VistA” from the menu**

**Once inside, you will want to
search by file name or number
or field name or number. The
files and fields are listed in the
CDW Metadata.**

- <http://vaausdarmul81/pls/apex/f?p=2000:1:246212839773488>

VINCI's Documentation

VINCI Central

Home > Data > Available Data

Available Data

The following is a list of the three categories of data available through VINCI. To see a description of the tables and columns contained within a domain, select it from the Data Descriptions on the right.

- See CDW MetaData Reports Page on CDW site.
- See CDW-Raw Domains Page on CDW site.

CDW

The Veterans Administration's Corporate Data Warehouse (CDW) is a national repository comprising data from several Veterans Health Administration (VHA) 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 incorporating data from multiple differing data sets throughout the VHA into one standard database structure will facilitate reporting and data analysis at the enterprise level. The CDW operates within the VA Office of Information & Technology Field Operations Business Intelligence Service Line.

CDW data are stored in a relational database. Multiple VA data sources are being merged so that cohorts will be definable by attributes such as ICD-9 codes and CPT codes from both inpatient and outpatient encounters or from abnormal values of vital signs like blood pressure, weight and height, within a target time period. Data are kept current by refreshing on a nightly basis and is available beginning October 1, 1999. CDW is considered to be the best model data.

CDW-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 data set. They are extracted directly from the VistA sites through a process as data snapshots, 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 Data Sets

The SAS data sets 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 data sets are provided in SAS format by fiscal year (Oct. 1 - Sept. 30). These data are extracted from the National Patient Care Database (NPCCD) maintained by the VHA Office of Information at the Austin Information Technology Center (AITC), the central repository for VA data. Having historically been monitored by AITC mainframe, VINCI has acquired a complete copy of those data sets.

In all of the SAS data sets, each patient has a unique identifier referred to as the Scrambled SSN, which is a formula-based encryption of the Social Security Number. The identifier is consistent for a given patient across data sets and fiscal years.

Please note that VINCI consolidated the pre-2005 DSS data from the mainframe SAS files that were available at Austin (AITC). Since VINCI is not the data owner, no attempt was made by VINCI to modify, enhance, or change the data other than consolidating the original individual SAS files and making the data available as SQL Server tables. Please visit the **VIReC site** to find more information DSS data.

Data Descriptions

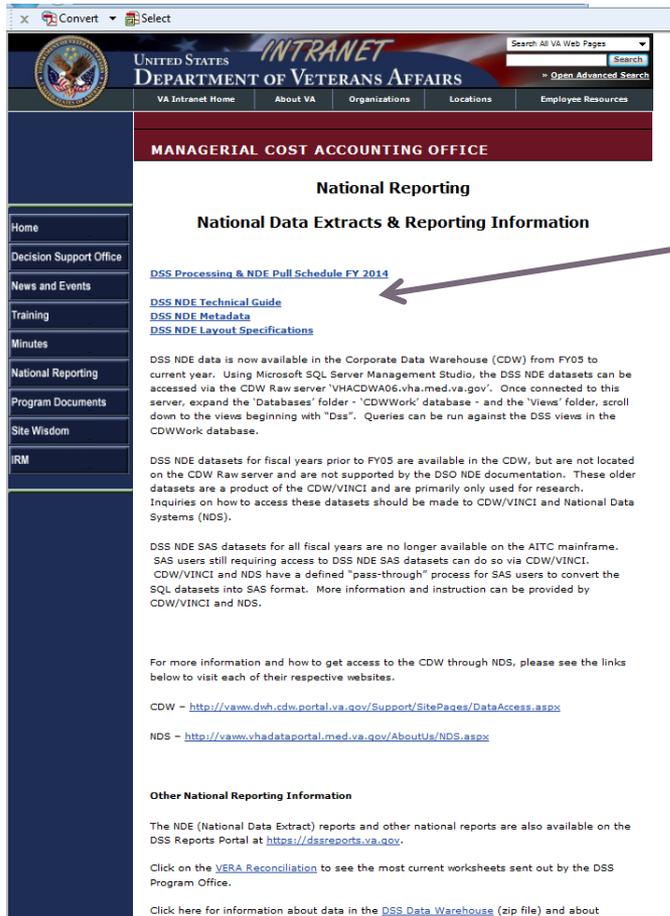
- Allergies
- Appointment
- BCMA
- Beneficiary Travel
- Bill Claims
- Compensation & Pension Exam
- Consult
- Dimensions (CDW)
- Dimensions (CDW-Raw)
- DSS
- Echo
- Electronic Time Attendance
- Emergency Dept. Inc. Software
- Equipment Inventory
- Fee
- Health Factors
- Immunization
- Inpatient Encounters
- Inpatient Pharmacy, IV
- Inpatient Pharmacy, Unit Dose
- Lab Chemistry
- MedSAS
- Mental Health Assessment
- Microbiology
- My Health Vet
- Orders (CPRS)
- Outpatient Pharmacy
- Outpatient Visits
- Patient
- PCMM
- Pharmacy, Non-VA Meds
- Prostheses
- Radiology
- SPatient
- Staff
- Surgery
- TIU Documents
- Vital Signs
- Vital Status

Here one will find additional documentation related to various CDW domains.

The documentation is organized by domain. It contains some descriptions of meaning of content found in tables and columns.

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

MCAO National Data Extracts (NDE)



The screenshot shows the VA Intranet website for the Managerial Cost Accounting Office (MCAO). The page is titled "National Reporting" and "National Data Extracts & Reporting Information". A red arrow points to the link "DSS NDE Technical Guide" in the left-hand navigation menu. The main content area contains several paragraphs of text, including a section titled "DSS Processing & NDE Pull Schedule FY 2014" and a section titled "DSS NDE data is now available in the Corporate Data Warehouse (CDW) from FY05 to current year." The page also includes a search bar at the top right and a navigation menu on the left.

Here a user can find a technical guide, metadata and layout specifications for Decision Support Services (DSS) National Data Extracts (NDE).

As mentioned above, DSS data is currently in CDW Raw.

Summary/Conclusion

- **Being comfortable with terms like domains, fact/dimension tables, and primary/foreign keys will be an important part of working with CDW databases.**
- **Using resources and metadata from the above-named sources will be essential to defining a cohort in a research request for data access.**

Contact Information

Margaret Gonsoulin, PhD

VIREC@va.gov

708-202-2413

Questions?

A decorative graphic consisting of several horizontal blue lines of varying thickness and length, extending from the left side of the slide towards the right, positioned below the word 'Questions?'. The lines are arranged in a stepped fashion, with some being longer than others, creating a modern, layered effect.

“Whiteboard” poll:

- What was most useful to you in today’s session?
- What topics would you like to see in future CDW cyberseminars?
- What other resources (manuals, guides, documentation, etc.) would be helpful in using CDW?