

CDW: A Conceptual Overview 2017

by Margaret Gonsoulin, PhD

March 29, 2017



Thanks to:

- Richard Pham, BISL/CDW for his mentorship
- Heidi Scheuter and Hira Khan for organizing this session

Poll #1: Your CDW experience

- How would you describe your level of experience with CDW data?
 - 1- Not worked with it at all
 - 2
 - 3
 - 4
 - 5- Very experienced with CDW data

Agenda for Today

- Get to the bottom of all of those acronyms!
- Learn to think in “relational data” terms
- Become familiar with the components of CDW
 - Production and Raw Domains
 - Fact and Dimension tables/views
- Understand how to create an analytic dataset
 - Primary and Foreign Keys
 - Joining tables/views

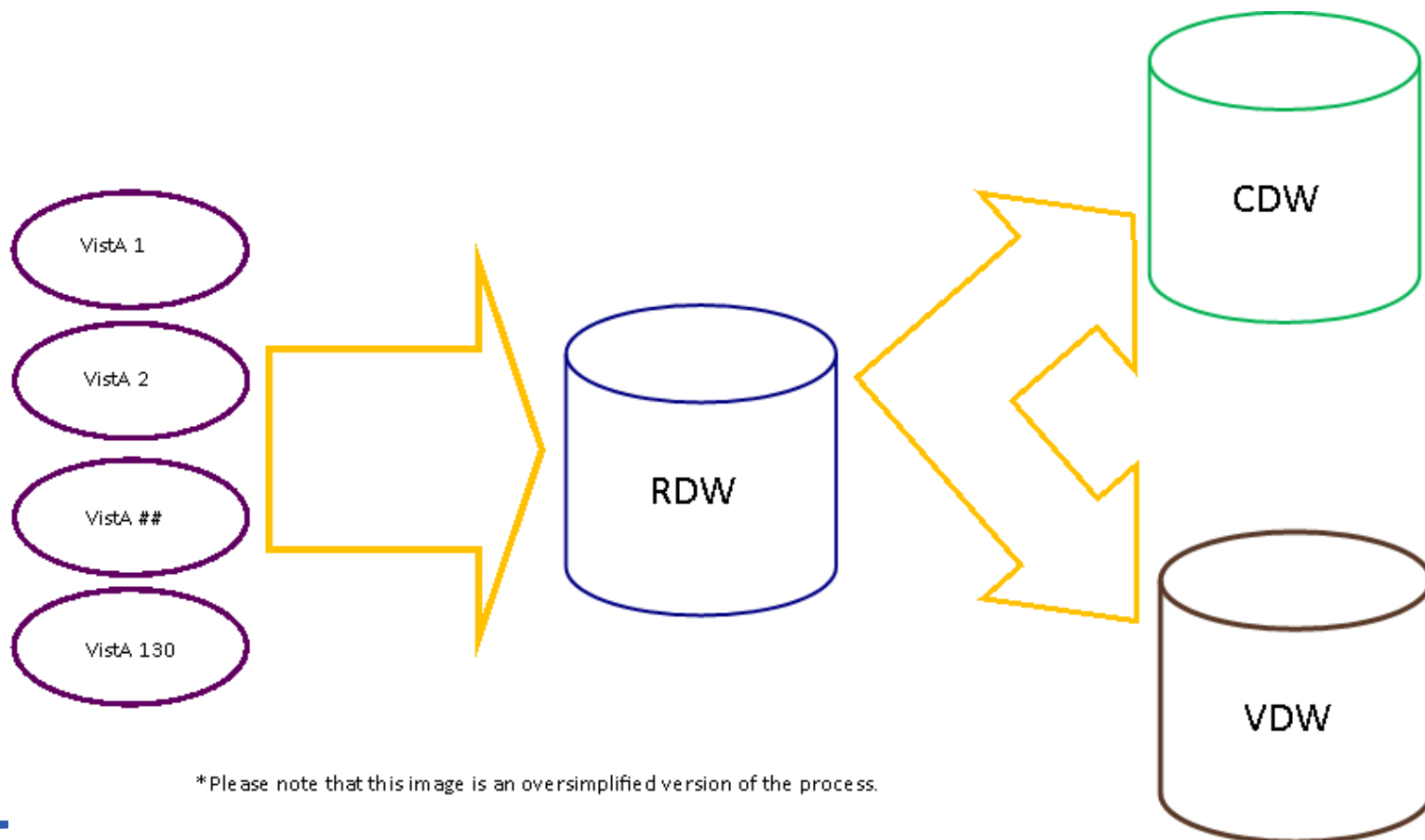
Agenda for Today

- **Get to the bottom of all of those acronyms!**
- Learn to think in “relational data” terms
- Become familiar with the components of CDW
 - Production and Raw Domains
 - Fact and Dimension tables/views
- Creating an analytic dataset
 - Primary and Foreign Keys
 - Joining tables/views

“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.

Flow of data into the warehouse

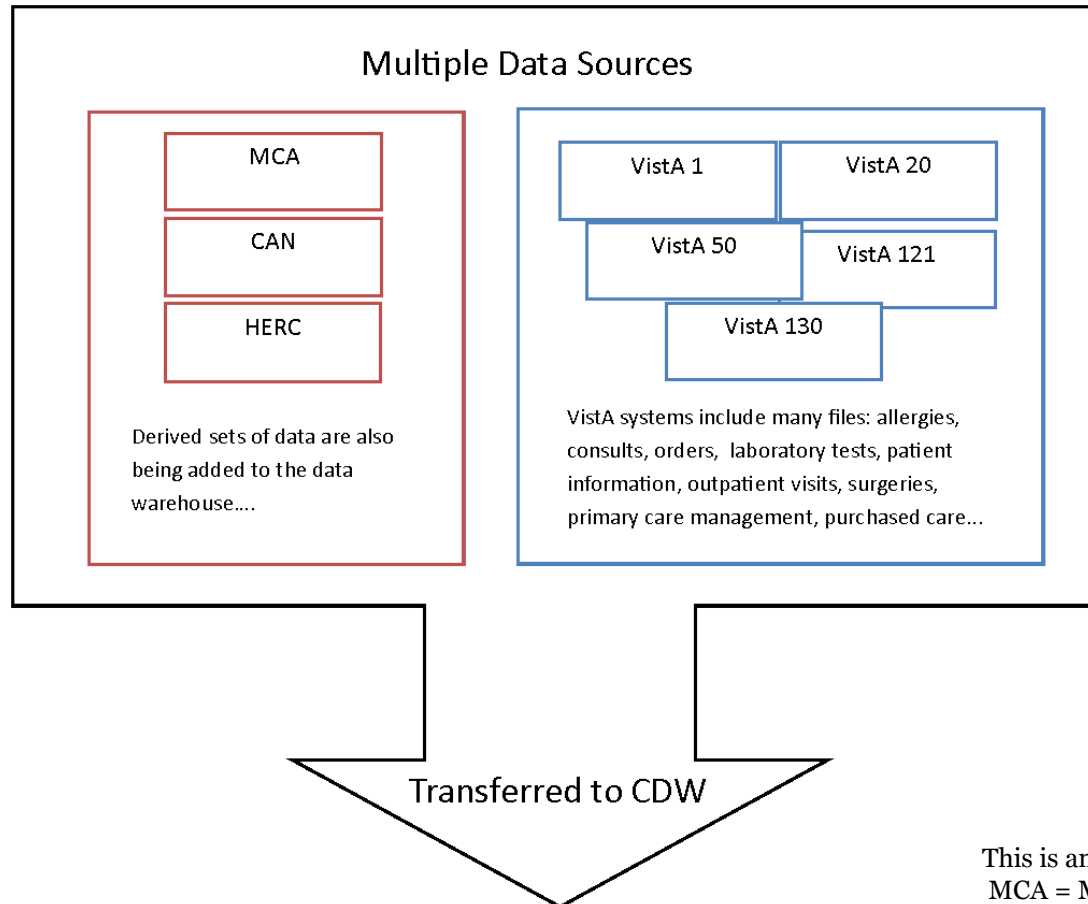


*Please note that this image is an oversimplified version of the process.

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.”

What type of data get stored here?



notes

This is an oversimplified description.

MCA = Managerial Cost Accounting

CAN = Care Assessment Need

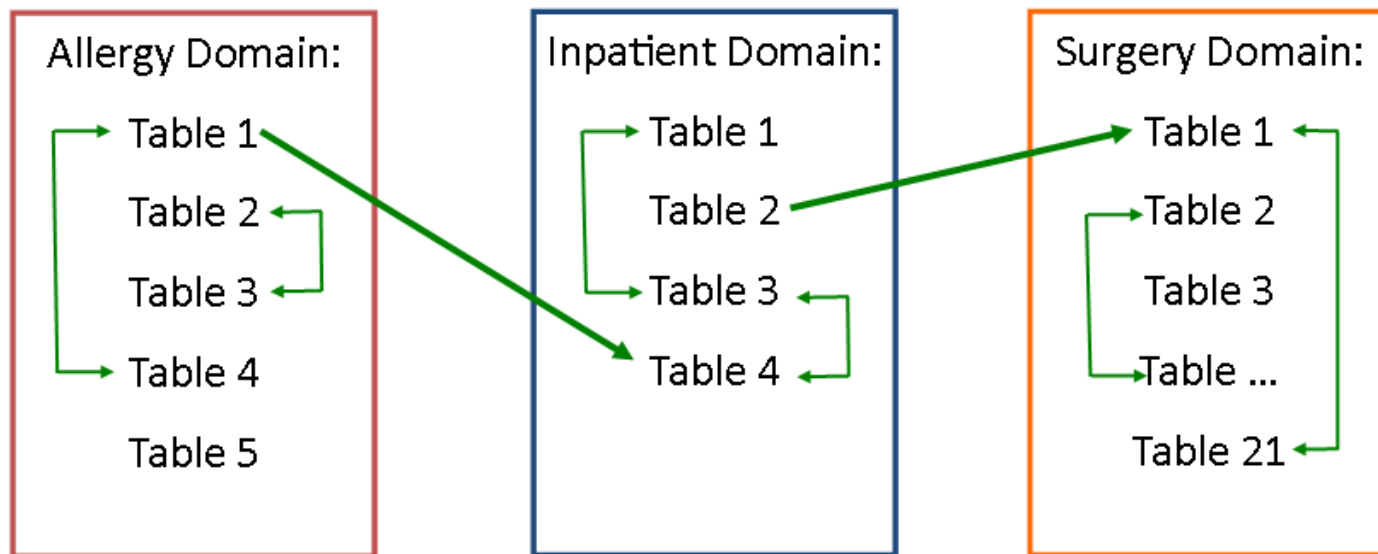
HERC = Health Economics Resource Center

How the CDW organized?

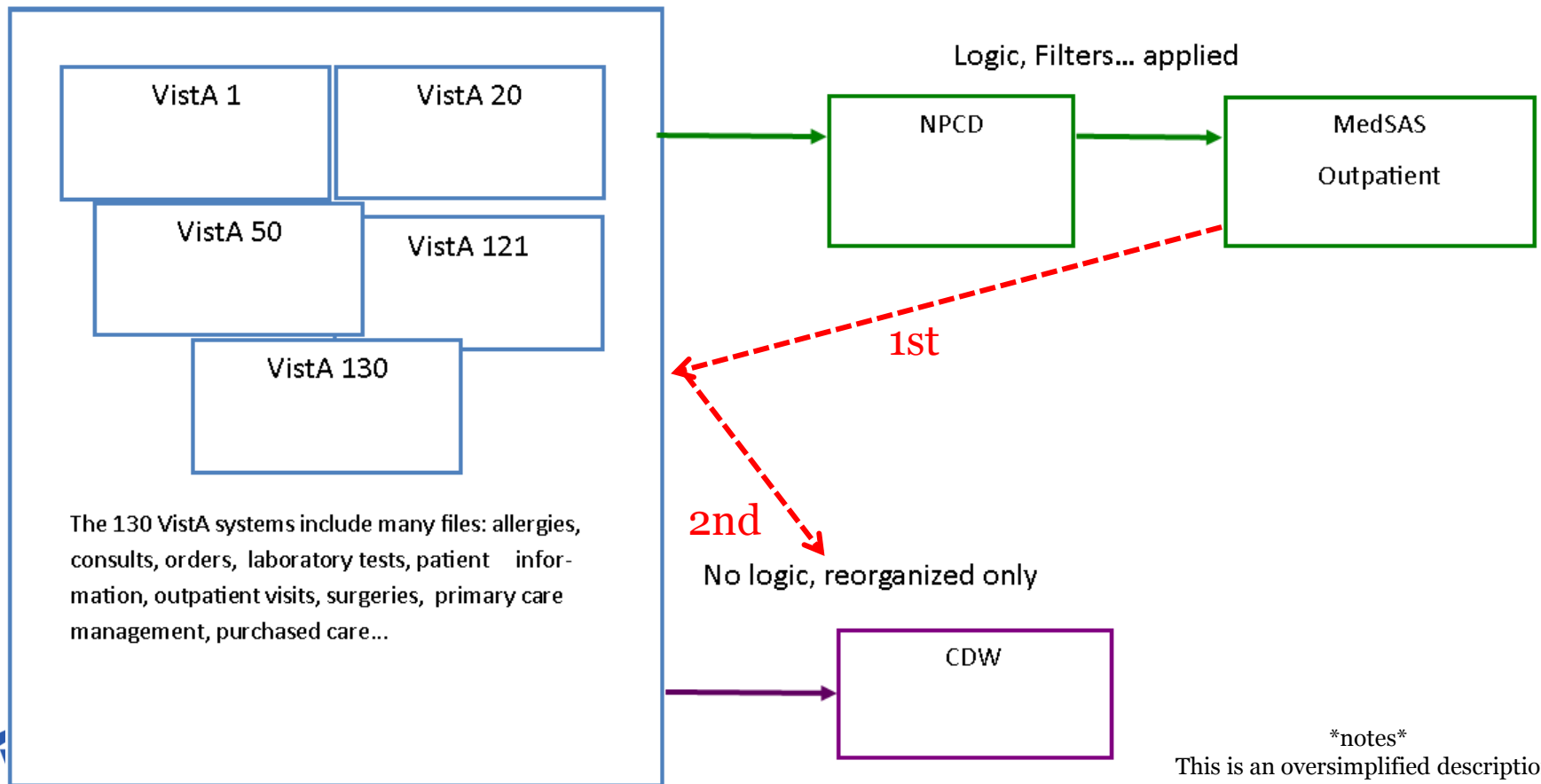
CDW organizes data into data domains

with multiple tables in each domain

and builds linking keys to connect tables and domains:



How does CDW fit into VA Data?



notes

This is an oversimplified description.
 NPCD = National Patient Care Database
 MedSAS = Medical SAS

Agenda for Today

- Get to the bottom of all of those acronyms!
- **Learn to think in “relational data” terms**
- Become familiar with the components of CDW
 - Production and Raw Domains
 - Fact and Dimension tables/views
- Creating an analytic dataset
 - Primary and Foreign Keys
 - Joining tables/views

C“D”W- The data.

- Data stored in this warehouse are in a “relational” format.
- In ordinary terms, these are data that have been separated out into multiple tables that look like 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.

A common database, a “flat file”

LastName	FirstName	Address	CityState	Zipcode	ICD9	Sta3n	VisitDate
Jones	Marianna	123 Oak St	Bee, AR	70788	110.6	578	1/2/2014
Frank	Josie	11 Pine Ave	Flip, OK	30032	377.75	358	2/2/2014
Plank	Bill	230 5 th St	Miner, TX	11201	202.05	402	3/3/2014

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

1. Patient information (name and address)
2. Diagnosis the led to the visit (ICD9 code)
3. Information about the visit (station number and date)

Transformed into a Relational Database

PatientTable

PatientKey	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

DiagnosisTable

DxKey	ICD9
1	110.6
2	202.05
3	280.8
4	377.75

VisitTable

VisitKey	DxKey	PatientKey	Sta3n	Date
1	1	1	578	1/1/2014
2	4	2	358	2/2/2014
3	2	3	402	3/3/2014

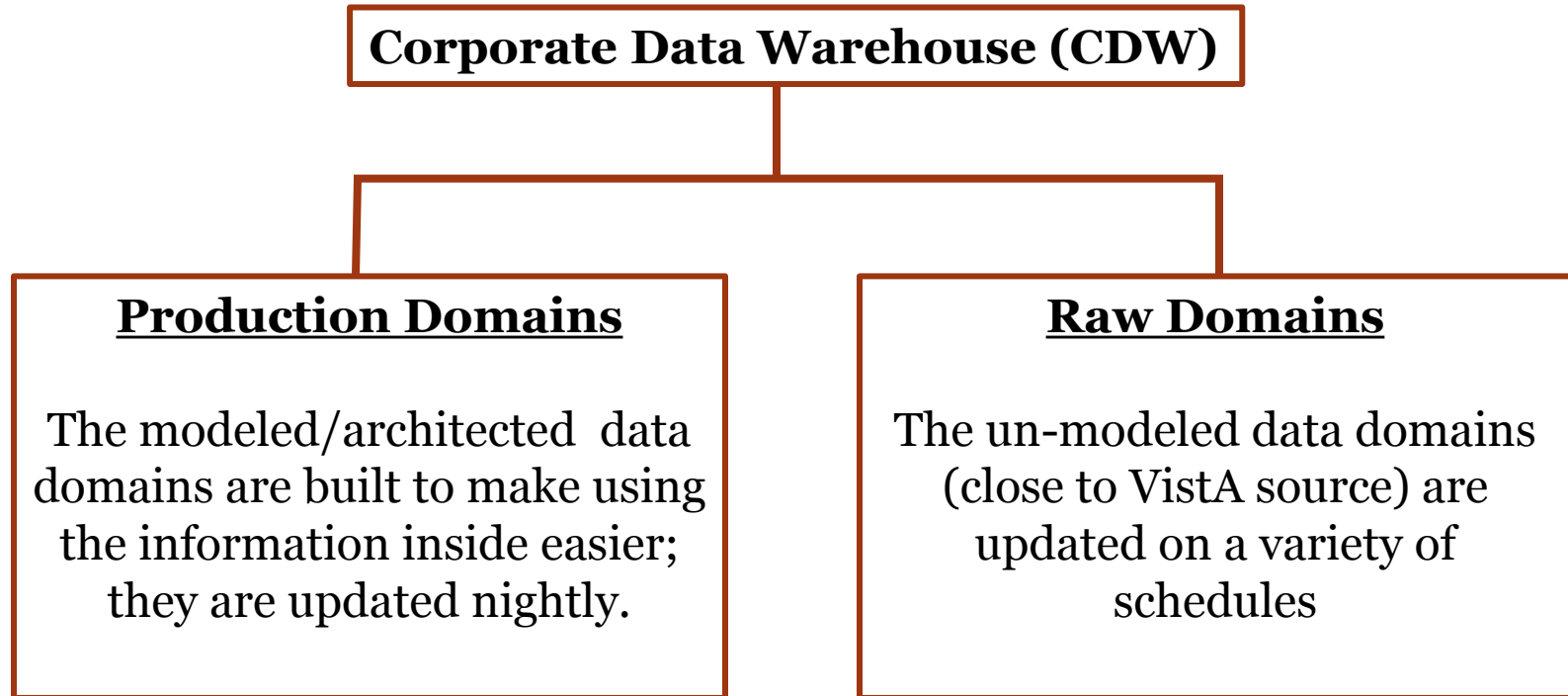
Poll #2: About you

- Which of the following best describes your role in the VA? (check all that apply)
 - Operations/Quality Improvement only
 - Research investigator/PI
 - Research data manager/analyst
 - Project coordinator
 - Both operations & research work
 - Other

Agenda for Today

- Get to the bottom of all of those acronyms!
- Learn to think in “relational data” terms
- **Become familiar with the components of CDW**
 - Production and Raw Domains
 - Fact and Dimension tables/views
- Creating an analytic dataset
 - Primary and Foreign Keys
 - Joining tables/views

2 Parts of CDW (Production and Raw)



Domains

Groups of tables based on the subject matter found within the group of tables

- Production Domains:
 - Outpatient 2.1
 - Lab Microbiology 2.0
 - Mental Health 1.2
 - Patient 2.0
 - Vital Sign 1.0
 - Patient Insurance 1.0
- Raw Domains
 - Intravenous Meds
 - Oncology
 - Prosthetics
 - Bill Claims
 - Radiology

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 trillions of records)
- They have patient and staff identifiers in them
- Therefore,
 - It is necessary to request permission to view these tables
 - It is important to plan your cohort to reduce the size of your request

Fact tables and Dimension tables

PatientTable (Fact)

PatientKey	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

DiagnosisTable (Dim)

DxKey	ICD9
1	110.6
2	202.05
3	280.8
4	377.75

VisitTable (Fact)

VisitKey	DxKey	PatientKey	Sta3n	Date
1	1	1	578	1/1/2014
2	4	2	358	2/2/2014
3	2	3	402	3/3/2014

Prefixes or Schema / Type of table

- Each table name is preceded by a prefix or schema used to indicate whether or not it is a:
 - A fact table (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

Two types of tables/views

Health Factor 2.1 Domain

HF.HealthFactor

Dim.HealthFactorType

Fact table

Dimension table

HF and Dim tables in Health Factors

HF.HealthFactor Content

- Patient identifiers
- Staff identifiers
- Descriptions of the health factor type (such as “colonoscopy” and “flu vaccine done elsewhere”)
- Comments related to the health factor (such as “biopsy of polyp - benign” and “normal”).

Dim.HealthFactorType Content

- Whether the health factor is a “factor” or a “category”
- Whether or not the health factor is gender specific
- Text describing the factors in words (e.g., “current smoker,” “weight management,” or “abuse/neglect”)

Agenda for Today

- Get to the bottom of all of those acronyms!
- Learn to think in “relational data” terms
- Become familiar with the components of CDW
 - Production and Raw Domains
 - Fact and Dimension tables/views
- **Creating an analytic dataset**
 - Primary and Foreign Keys
 - Joining tables/views

Joining Keys

- Primary Key –A column in every table that uniquely identifies each row.
- 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

Joining Keys

PatientTable (Fact)

PatientKey (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

DiagnosisTable (Dim)

DxKey (PK)	ICD9
1	110.6
2	202.05
3	280.8
4	377.75

VisitTable (Fact)

VisitKey (PK)	DxKey (FK)	PatientKey (FK)	Sta3n	Date
1	1	1	578	1/1/2014
2	4	2	358	2/2/2014
3	2	3	402	3/3/2014

2 methods of identifying joining keys

The screenshot shows the BSL CDW Home page. At the top, there are navigation tabs for BSL, CDW, and VISNs. Below the navigation, the BSL logo is on the left, and the text 'CDW Home' is on the right. A dropdown menu is open for 'Community', showing 'MetaData' as an option, which is circled in red. An arrow points from this 'MetaData' option to the 'CDW Metadata' link in the 'NEW TO CDW?' section, which is also circled in red. The 'NEW TO CDW?' section contains a heading 'NEW TO CDW?' and a sub-heading 'Are you getting started with the Corporate Data Warehouse (CDW)?'. Below this are three bullet points: 'Intro and Policies', 'CDW Support', and 'CDW Metadata'. The 'WHAT'S IN THE WORKS?' section contains a heading 'WHAT'S IN THE WORKS?' and three bullet points: 'General Announcements', 'Training Announcements', and 'CDW Domain Status and Priority (Excel)'. The 'EXTERNAL LINKS' section contains a heading 'EXTERNAL LINKS' and five bullet points: 'HSR&D Listserv', 'VHA Data Portal', 'DAR', 'NDS', and 'System Status & Planned Outages'. The 'DATA ACCESS' section contains a heading 'DATA ACCESS' and a paragraph: 'Do you want to get connected to CDW? There are many types of access to CDW. The best thing to do before you start is to define what kind of access you need.' Below this are three bullet points: 'Local Data Access (LSV)', 'Data Access information', and 'CDW Customers and Workgroups'.

BSL CDW VISNs

CDW Home CDW Support Community MetaData

BSL CDW Home

Site Contents

NEW TO CDW?

Are you getting started with the Corporate Data Warehouse (CDW)?

- Intro and Policies
- CDW Support
- CDW Metadata

WHAT'S IN THE WORKS?

- General Announcements
- Training Announcements
- CDW Domain Status and Priority (Excel)

EXTERNAL LINKS

- HSR&D Listserv
- VHA Data Portal
- DAR
- NDS
- System Status & Planned Outages

DATA ACCESS


Do you want to get connected to CDW? There are many types of access to CDW. The best thing to do before you start is to define what kind of access you need.

- Local Data Access (LSV)
- Data Access information
- CDW Customers and Workgroups

Click “CDW metadata report”

BISL ▾
CDW ▾
VISNs ▾

BROWSE PAGE



[CDW Home](#)
CDW Support ▾
Community ▾
MetaData ▾

MetaData Home

Libraries

- [CDW Metadata Wiki](#)
- [Metadata Documents](#)
- [Reports - SSRS](#)
- [Data Sources](#)

Reports

- [Schemas](#)
- [VINCI Central](#)

Social

- [VA Pulse](#)


Site Contents










[CDW Home](#) > [MetaData](#)







CDW MetaData

Welcome to the CDW MetaData site, the site that contains information about the metadata for the CDW Data Domains in production. To begin, launch the [CDW Metadata Report](#). Information pertaining to the *Meta* and *BaseCamp* views in a CDW database may be found in the [CDW Metadata Wiki](#).

Metadata Documents

Current View ... 

✓ 	Name	Modified	Modified By
	Active Directory 1.0	... July 22, 2016	 Dean, Mark A.
	Allergy 1.0	... February 19, 2014	 Anderson, Stephen M. (BISL)
	Appointment 2.0	... February 19, 2014	 Anderson, Stephen M. (BISL)
	Consult 2.1	... October 1, 2014	 Dean, Mark A.

Actions |  |  |  | 1 of 1 |  |  |  | Find

CDW Metadata

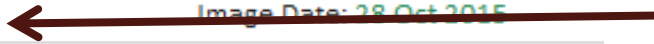
Contains a grouped list of available CDW ER Diagrams and members.

ImageDescription

Active Directory 1.0	Image Date: 20 Jul 2016
Allergy 1.0	Image Date: 01 Feb 2014
Appointment 2.2	Image Date: 25 Apr 2016
Consult 2.1	Image Date: 24 Sep 2015
CPRSOrder 1.0	Image Date: 11 Aug 2014
Data Profiling 1.0	Image Date: 21 Feb 2014
Dental 1.0 Diagram 1 of 2	Image Date: 28 Oct 2015
Dental 1.0 Diagram 2 of 2 for Analytics	Image Date: 20 Apr 2016
Dimensions A Through D 7/8/2015	Image Date: 08 Jul 2015
Dimensions E Through K 7/8/2015	Image Date: 08 Jul 2015
Dimensions L Through O 5/4/2016	Image Date: 04 May 2016
Dimensions P Through R 3/7/2016	Image Date: 07 Mar 2016
Dimensions S Through Z 3/7/2016	Image Date: 07 Mar 2016
Dimensions, MRSA	Image Date: 31 Dec 2015
Dimensions, Place	Image Date: 30 Mar 2016
EDIS 1.0	Image Date: 13 Jan 2017

The metadata page lists each domain alphabetically

Click the name of the domain to open the “ER Diagram”



Pharmacy Patient Domain

Patient.PharmacyPatient	
PharmacyPatientSID (PK)	int
PatientIEN	varchar(50)
PatientSID (FK)	int
Sta3n	smallint
MismatchedPatientInternalEntryNumber	varchar(50)
SafetyCapFlag	char(1)
MailPreference	varchar(50)
DialysisPatientFlag	char(1)
MailStatusExpirationDate	date
MailStatusExpirationVistaErrorDate	varchar(50)
MailStatusExpirationDateTransformSID	bigint
FirstServiceDate	date
FirstServiceVistaErrorDate	varchar(50)
FirstServiceDateTransformSID	bigint
ActualHistorical	varchar(50)
Narrative	varchar(255)
PharmacyPatientStatusSID (FK)	int
CommunityNursingHomeFlag	char(1)
NursingHomeContractFlag	char(1)
ContractLastDate	date
ContractLastVistaErrorDate	varchar(50)
ContractLastDateTransformSID	bigint
RespitePatientStartDate	date
RespitePatientStartVistaErrorDate	varchar(50)
RespitePatientStartDateTransformSID	bigint
RespitePatientEndDate	date
RespitePatientEndVistaErrorDate	varchar(50)
RespitePatientEndDateTransformSID	bigint
ClozapineRegistrationNumber	varchar(50)
ClozapineStatus	varchar(50)
LastClozapinePrescriptionDate	date
LastClozapinePrescriptionVistaErrorDate	varchar(50)
LastClozapinePrescriptionDateTransformSID	bigint
DemographicsSentFlag	char(1)
ClozapineResponsibleProviderSID	int
RegistrationDate	date

Dim.PharmacyPatientStatus	
PharmacyPatientStatusSID (PK)	int
PharmacyPatientStatusIEN	varchar(50)
Sta3n	smallint
PharmacyPatientStatus	varchar(50)
PharmacyPatientStatusAbbreviation	varchar(50)
DaysSupplySetting	decimal(9,4)
RefillsSetting	decimal(9,4)
RenewableFlag	char(1)
AMISCategory	varchar(50)
ExemptFromCopaymentFlag	char(1)
ExemptFromCHAMPUSBillingFlag	char(1)
FileManFileNumber: 53	
FileManFileName: RX PATIENT STATUS	

Patient.PharmacyPatientScriptalk	
PharmacyPatientScriptalkSID (PK)	int
PatientSID (FK)	int
PatientIEN	varchar(50)
PharmacyPatientScriptalkIEN	varchar(50)
Sta3n	smallint
ScriptalkEnrollmentDateTime	datetime2(0)
ScriptalkEnrollmentVistaErrorDate	varchar(50)
ScriptalkEnrollmentDateTimeTransformSID	bigint
ScriptalkPatientFlag	char(1)
IndicationStatus	varchar(50)
StaffSID (FK)	int
FileManFileNumber: 55.0108	
FileManFileName: SCRIPTALK ENROLLMENT ACTIVITY	

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

(PK) indicates a Primary Key

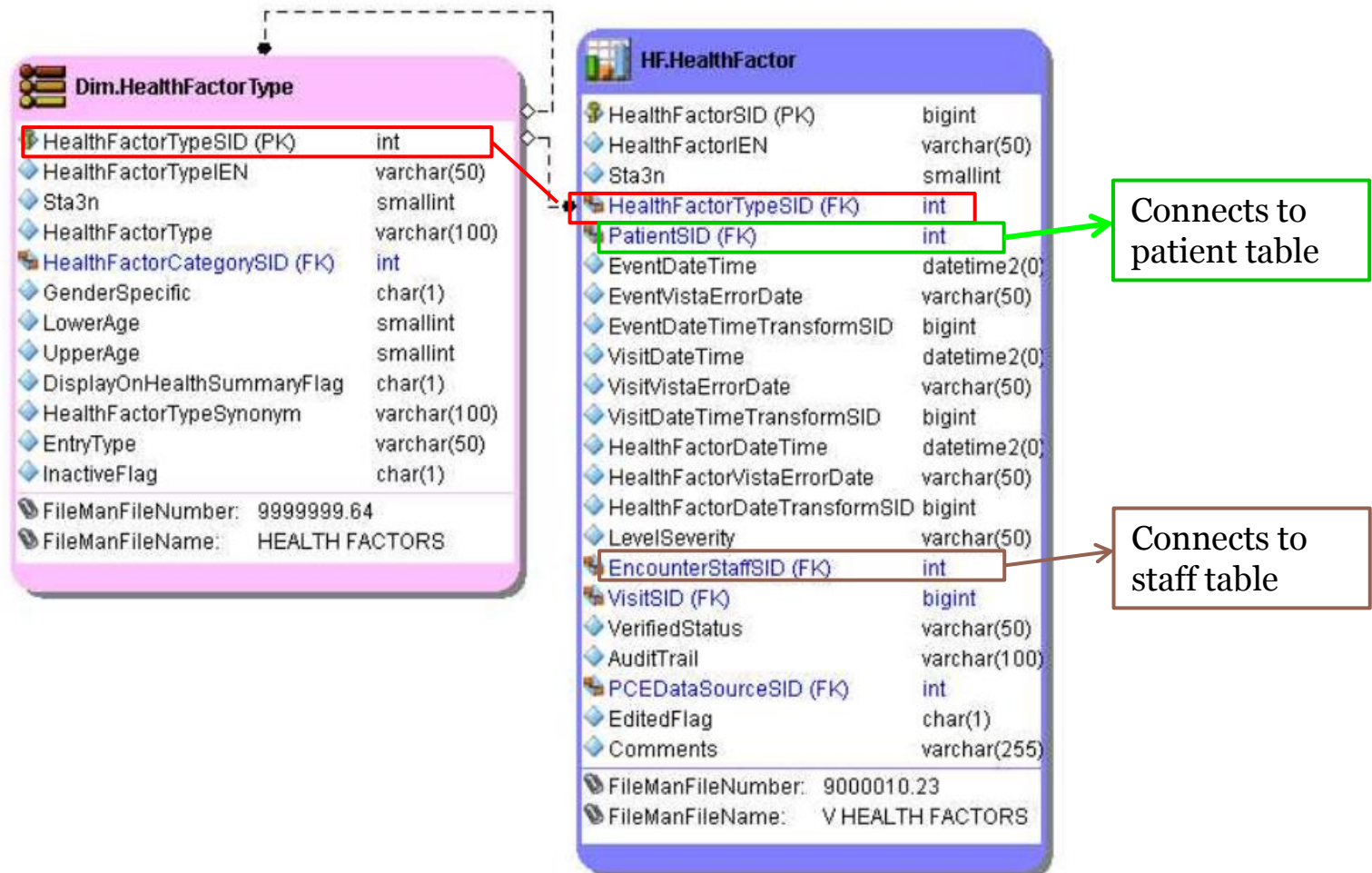
(FK) indicates a Foreign Key







These two join on *PharmacyPatientStatusSID*

The Patient Domain (connections within a domain)



The Health Factors Domain (connections inside & outside of domain)






Actions |  |  |  1 of 1  |  |  | Find

CDW Metadata

Contains a grouped list of available CDW ER Diagrams and members.

ImageDescription

 Active Directory 1.0	Image Date: 20 Jul 2016
 Allergy 1.0	Image Date: 01 Feb 2014
 Appointment 2.2	Image Date: 25 Apr 2016
 Consult 2.1	Image Date: 24 Sep 2015
 CRSOrder 1.0	Image Date: 11 Aug 2014
 Data Profiling 1.0	Image Date: 21 Feb 2014
 Dental 1.0 Diagram 1 of 2	Image Date: 28 Oct 2015
 Dental 1.0 Diagram 2 of 2 for Analytics	Image Date: 20 Apr 2016
 Dimensions A Through D 7/8/2015	Image Date: 08 Jul 2015
 Dimensions E Through K 7/8/2015	Image Date: 08 Jul 2015
 Dimensions L Through O 5/4/2016	Image Date: 04 May 2016
 Dimensions P Through R 3/7/2016	Image Date: 07 Mar 2016
 Dimensions S Through Z 3/7/2016	Image Date: 07 Mar 2016
 Dimensions, MRSA	Image Date: 31 Dec 2015
 Dimensions, Place	Image Date: 30 Mar 2016
 EMIS 1.0	Image Date: 13 Jan 2017

A second way of looking at linking keys:

expand the domain information to see table level metadata reports

Getting a list of joining keys

Dimensions S Through Z 4/17/2015	Image Date: 17 Apr 2015					
Dimensions_Place	Image Date: 09 Oct 2015					
Encounter 1.0	Image Date: 29 Oct 2013					
Health Factor 2.0	Image Date: 11 Mar 2015					
Health Factor 2.1						
DWViewName	Field Count	FileMan File Data Source	View Version	Relevant Dates	Relationships	
Dim_HealthFactorType	12	HEALTH FACTORS (9999999.64)	DWViewDeployed: xDWWork View Version: 3			
HF_HealthFactor	21	V HEALTH FACTORS (9000010.23)	DWViewDeployed: SPVNext View Version: 34	Partition Key: HealthFactorDateTime		
ICD-9-CM and ICD-10-CM	Image Date: 24 Sep 2015					
ICD-9-PCS and ICD-10-PCS	Image Date: 10 Aug 2015					
Immunization 2.1	Image Date: 03 Jun 2015					

Reading the “CDW foreign keys” table...

CDW Foreign Keys

Provides a listing of the foreign and primary keys for CDW views.

FKSchemaName	FKViewName	FKViewFieldName	FKViewVersion	PKSchemaName	PKViewName	PKViewFieldName
HF	HealthFactor	EncounterStaffSID	9	Staff	Staff	StaffSID
HF	HealthFactor	EncounterStaffSID	34	Staff	Staff	StaffSID
HF	HealthFactor	HealthFactorDateSID	34	Dim	Date	DateSID
HF	HealthFactor	HealthFactorTypeSID	9	Dim	HealthFactorType	HealthFactorTypeSID
HF	HealthFactor	HealthFactorTypeSID	34	Dim	HealthFactorType	HealthFactorTypeSID
HF	HealthFactor	PatientSID	9	Patient	Patient	PatientSID
HF	HealthFactor	PatientSID	34	Patient	Patient	PatientSID
HF	HealthFactor	PCEDDataSourceSID	9	Dim	PCEDDataSource	PCEDDataSourceSID
HF	HealthFactor	PCEDDataSourceSID	34	Dim	PCEDDataSource	PCEDDataSourceSID
HF	HealthFactor	VisitSID	9	Outpat	Visit	VisitSID
HF	HealthFactor	VisitSID	34	Outpat	Visit	VisitSID

Report Date: 11/18/2015 10:08:17 AM

The foreign key EncounterStaffSID in HF.HealthFactor connects to the primary key StaffSID in Staff.Staff.

Connecting the HF to a patient

CDW Foreign Keys

Provides a listing of the foreign and primary keys for CDW views.

FKSchemaName	FKViewName	FKViewFieldName	FKViewVersion	PKSchemaName	PKViewName	PKViewFieldName	PKViewVersion
HF	HealthFactor	EncounterStaffSID	9	Staff	Staff	StaffSID	9
HF	HealthFactor	EncounterStaffSID	34	Staff	Staff	StaffSID	34
HF	HealthFactor	HealthFactorDateSID	34	Dim	Date	DateSID	34
HF	HealthFactor	HealthFactorTypeSID	9	Dim	HealthFactorType	HealthFactorTypeSID	9
HF	HealthFactor	HealthFactorTypeSID	34	Dim	HealthFactorType	HealthFactorTypeSID	34
HF	HealthFactor	PatientSID	9	Patient	Patient	PatientSID	9
HF	HealthFactor	PatientSID	34	Patient	Patient	PatientSID	34
HF	HealthFactor	PCEDataSourceSID	9	Dim	PCEDataSource	PCEDataSourceSID	9
HF	HealthFactor	PCEDataSourceSID	34	Dim	PCEDataSource	PCEDataSourceSID	34
HF	HealthFactor	VisitSID	9	Outpat	Visit	VisitSID	9
HF	HealthFactor	VisitSID	34	Outpat	Visit	VisitSID	34

Report Date: 11/18/2015 10:08:17 AM

The foreign key PatientSID in HF.HealthFactor connects to the primary key PatientSID in Patient.Patient

Summary/Conclusion

- CDW contains a variety of data.
- Currently, most of it comes from VistA.
- It is broken into many domains, fact tables and dimension tables that must be brought back together to create an analytic set.
- SQL and linking keys are necessary tools for creating that analytic set of data.

Questions?

Margaret Gonsoulin, PhD

VIReC@va.gov

708-202-2413