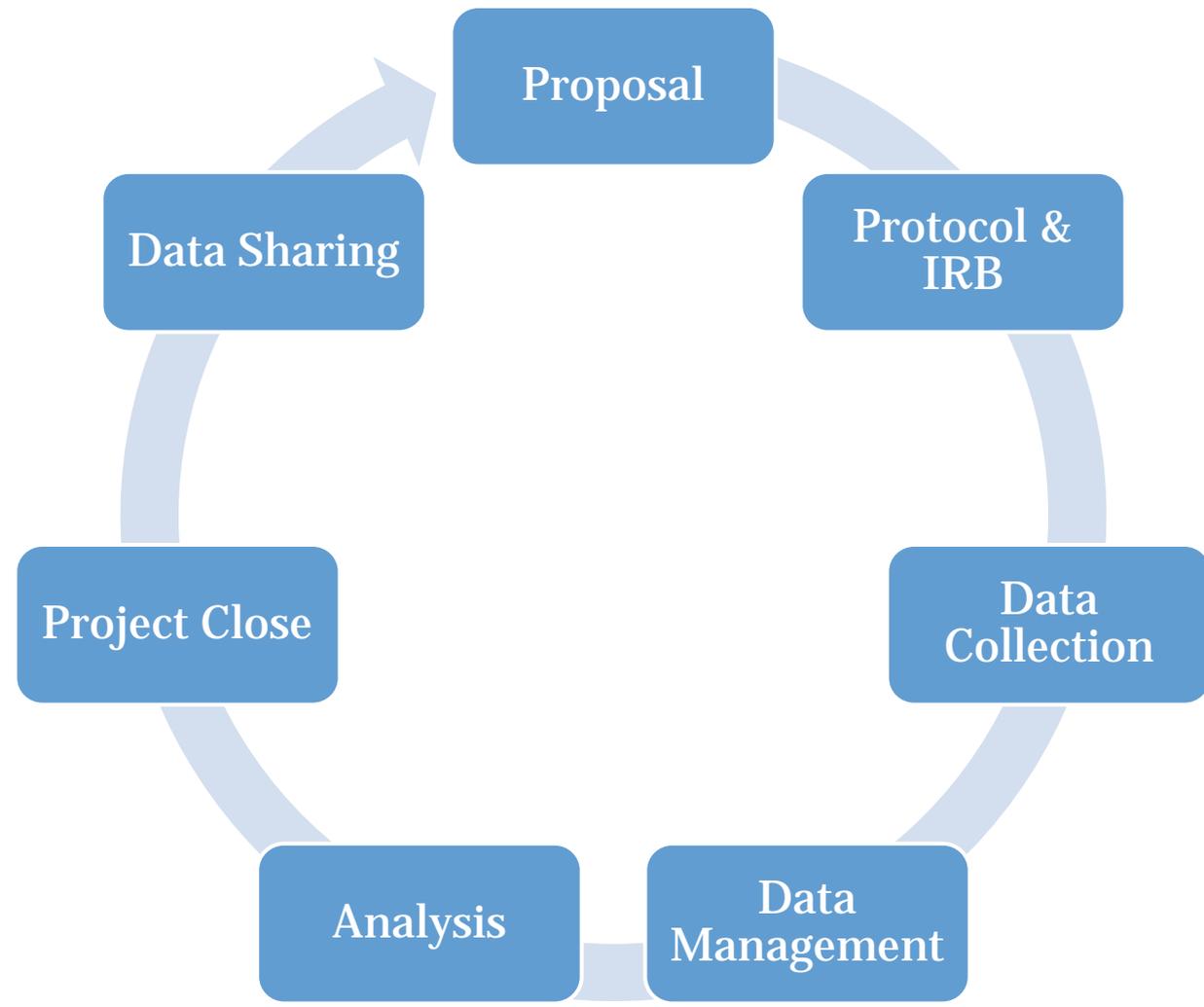


GDP 2014 Acknowledgements

- **Laurel Copeland, San Antonio VA**
- **Brian C. Sauer, Salt Lake City VA**
- **Kevin Stroupe, Hines VA**
- **Linda Williams, Indianapolis VA**
- **Brenda Cuccherini, ORD**
- **Denise Hynes, VIREC**
- **Arika Owens, VIREC**
- **Maria Souden, VIREC**

Research Life Cycle



Good Data Practices Series Overview

May 8

- *The Best Laid Plans: Plan Well, Plan Early* - Jennifer Garvin

May 15

- *"The Living Protocol:" Managing Documentation While Managing Data* - Matt Maciejewski

May 22

- *Controlled Chaos: Tracking Decisions During an Evolving Analysis* - Pete Groeneveld

May 29

- *Reduce, Reuse, Recycle: Planning for Data Sharing* - Linda Kok

Poll #1: About you

- What is your role in research and level of experience?
 - Research investigator
 - New? Experienced?
 - Data manager/analyst
 - New? Experienced?
 - Project coordinator
 - New? Experienced?
 - Other – please describe via the Q&A function

Poll #2: GDP Experience

- How many GDP sessions have you attended or viewed online this month, including today's?
 - 1
 - 2
 - 3
 - 4

Session 4: Reduce, Reuse, Recycle: Planning for Data Sharing

Linda Kok, MA

VIReC

Poll Question #3

- Are you working on or planning a project that will produce data that might be shared for re-use?
 - Yes, for re-use by me (other protocol, same PI)
 - Yes, for re-use by others (other protocol, different PI)
 - Not at this time

Session 4: *Reduce, Reuse, Recycle:* *Planning for Data Sharing*

- Traditional project close activities
- Why re-use research data?
- Project close activities if sharing data
- What is a research data repository?
- Requirements for a research data repository
- Creating a research data repository

Traditional project close activities

Traditional project close - protocol

- Notify R&D Committee
- Notify IRB



Traditional project close - data

- Identify the datasets to be retained
- Arrange for secure storage until scheduled in VHA's Records Control Schedule (RCS 10-1)
 - Check with local facility for procedures
 - Non-VA data use agreements may require data destruction – refer to the agreement

Why re-use research data?

Reduce, Reuse, Recycle!

- Reduce redundant (& expensive) data preparation
- Reuse existing research data
- Recycle subsets of research data



Why make research-generated data available for re-use?

“Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone.”

August, 2012

IBM, Bringing Big Data to the Enterprise



Open Data/Open Science



“To fuel entrepreneurship, innovation, and scientific discovery”

- Exec. Order May 2013

Forbes - New Posts Most Popular America's Richest Counties Lists Hip-Hop's Richest

We've doubled our commitment to businesses grow internationally Find out more >>

Matthew Herper, Forbes Staff
I cover science and medicine, and believe this is biology's century.

PHARMA & HEALTHCARE | 1/30/2014 @ 7:05AM | 24,934 views

In Stunning Win For Open Science, Johnson & Johnson Decides To Release Its Clinical Trial Data To Researchers

5 comments, 1 called-out + Comment Now + Follow Comments

Drug companies tend to be secretive, to say the least, about studies of their medicines. For years, negative trials would not even be published. Except for the U.S. Food and Drug Administration, nobody got to look at the raw information behind those studies. The medical data behind important drugs, devices, and other products was kept shrouded.

Today, Johnson & Johnson is taking a major step toward changing that, not only for drugs like the blood thinner Xarelto or prostate cancer pill Zytiga but also for the artificial hips and knees made for its orthopedics division or even consumer products. "You want to know about Listerine trials? They'll have it."

“If science is to be progressive and self-correcting it is critical for multiple groups to look at the data and draw their own conclusions and put the results in public view.”

Harlan Krumholz, MD
Yale School of
Medicine

NIH data sharing goals



Data should be

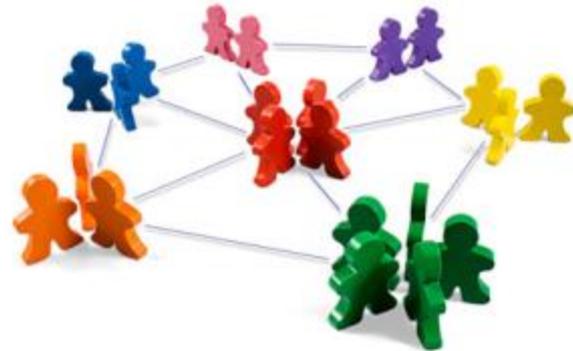
- Widely & freely available
- With timely release and sharing
- While protecting patient/subject privacy

NIH: Reasons to share data

- Expedite translation of research into knowledge, products & procedures
- Facilitate the education of new researchers
- Permit testing of new alternative hypotheses & methods
- Support methods and measurement studies
- Reinforce open scientific inquiry

Why make your data available for re-use?

- To re-use the data yourself
- Save data prep time & \$\$\$ on your next project
- Promote your research
- Enable new discoveries with your data



Is your project data ready for sharing?

- **Importance of documenting as you go**
 - **Accurate, systematic record of the research process**
 - **Capture decisions and the reasoning behind them**
 - **Must haves**
 - Project description
 - Data description
 - Methodology description

Do you have authority to share?

- HIPAA compliant patient authorizations
- IRB approval of waiver of HIPAA authorization
- VHA data owners/steward permission
- Agreements with non-VHA data sources



Project close if sharing data

Traditional project close

- Notify R&D Committee
- Notify IRB
- Identify and secure the data to be retained



Project close if sharing data

- Notify R&D Committee
- Get IRB approval for a VA research data repository
- Identify data to be shared
- And more...



What is a research data repository?

Research data repository (RDR)

- A dataset or collection of datasets produced and managed under an IRB approved research protocol to provide for re-use for subsequent research
- Permits re-use of research-generated data within the VHA for IRB approved research protocols
- Data collected for use exclusively for one specific research protocol do not constitute a research data repository

Research data repository content

- Data generated during the course of one research protocol to be used by other subsequent research protocols
- Data collected specifically for use by more than one research protocol

Sources of data in RDRs

- **Directly from research subjects or from review of the subject's administrative, medical, or other records**
- **Indirectly from existing databases**
 - **Research data**
 - **Non-research sources**
 - **VA or non-VA sources**

Two management approaches

1. PI creates a new IRB approved research data repository to self-manage
2. PI delegates management of sharing the dataset(s) to an existing VA IRB approved research data repository

Note: May need to amend the existing protocol either way

Summary: When is an RDR required?

Primary Data Collection?	Existing Data Sources?	Identifying Information?	To be used for more than one protocol?	Data Repository Required?
Yes	No	De-identified	No	No
No	Yes	De-identified	No	No
Yes	Yes	De-identified	No	No
Yes	No	IIHI	No	No
No	Yes	IIHI	No	No
Yes	Yes	IIHI	No	No
Yes	No	De-identified	Yes	Yes
No	Yes	De-identified	Yes	Yes
Yes	Yes	De-identified	Yes	Yes
Yes	No	IIHI	Yes	Yes
No	Yes	IIHI	Yes	Yes
Yes	Yes	IIHI	Yes	Yes

Requirements for RDR

Scientific/ethical oversight committee(s)

- To provide scientific and ethical advice for repositories that contain a number of different databases or provide data broker services
- May be the IRB of record
- Consult with ORD about the requirement

Administration of RDR

- VA investigator responsible for all activities of the data repository – (not WOC or IPA)
 - Name an administrator
 - Does not have to be the PI
 - A data repository specialist
 - May also be the administrator
 - IT support available for granting data access

Administrative Stability

- IRB and R&D Committee must approve changes to an RDR
 - Changes to the repository administrator
 - Combining administration of data with another research data repository

RDR policies and procedures

- **Criteria for releasing data**
- **Request review process**
 - Procedures for verifying R&D Committee approval
 - Procedures for verifying IRB approval - if identifiable data
 - Record of requests & approvals
- **Data privacy protection**
- **Data security**

VHA data sharing requirements



VA | Defining
HEALTH CARE | **EXCELLENCE**
in the 21st Century

- VHA Handbook 1200.12 *“Use of Data and Data Repositories in VHA Research”*
- Contact Dr. Brenda Cuccherini at ORD for guidance

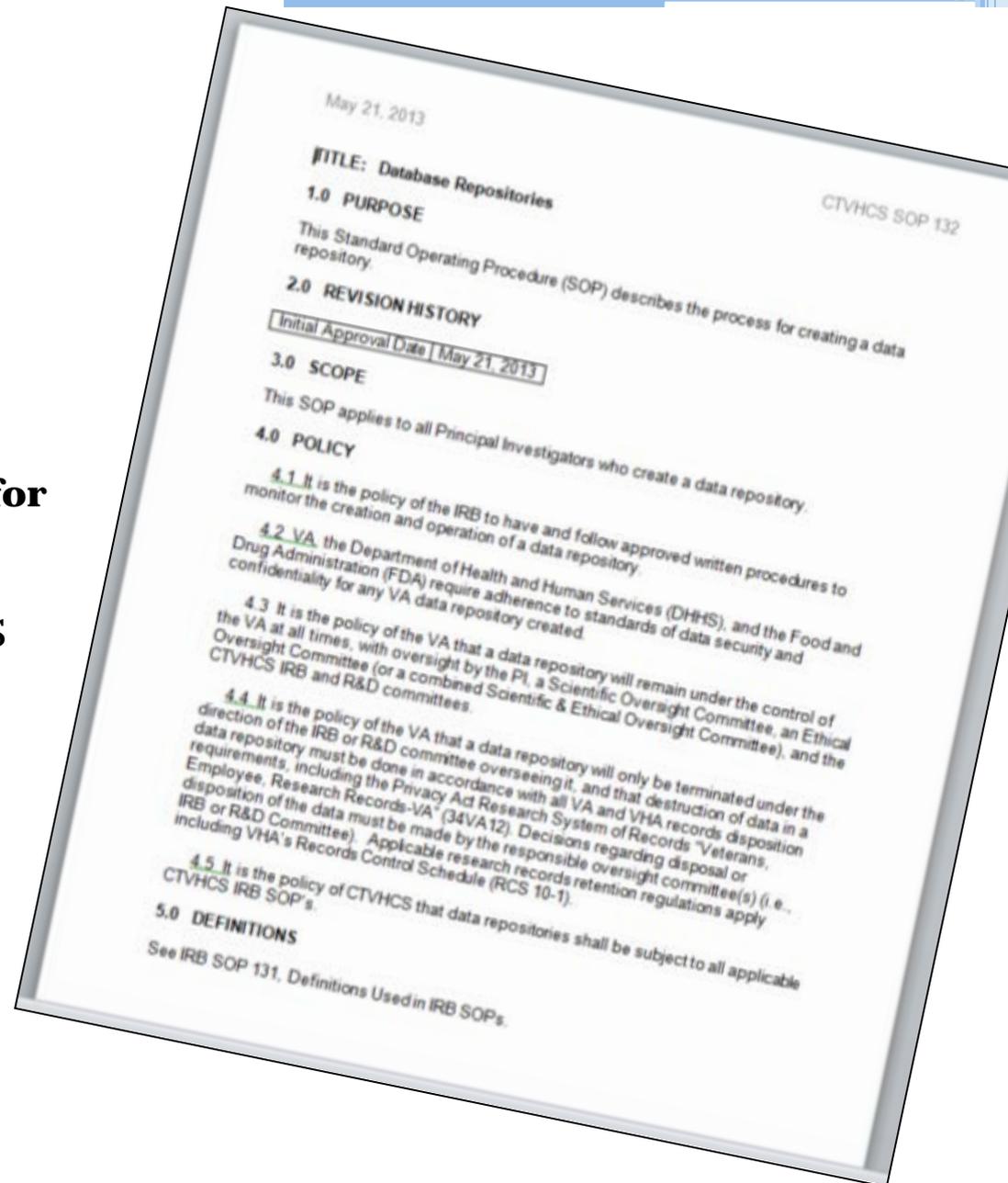
Creating a VA RDR

RDR Example:

Policy & Procedures

IIR 09-335 Surgical Treatment Outcomes for Patients with Psychiatric Disorders (STOPP)

Laurel Anne Copeland PhD MPH BS
Central Texas Veterans Health Care
System, Temple, TX



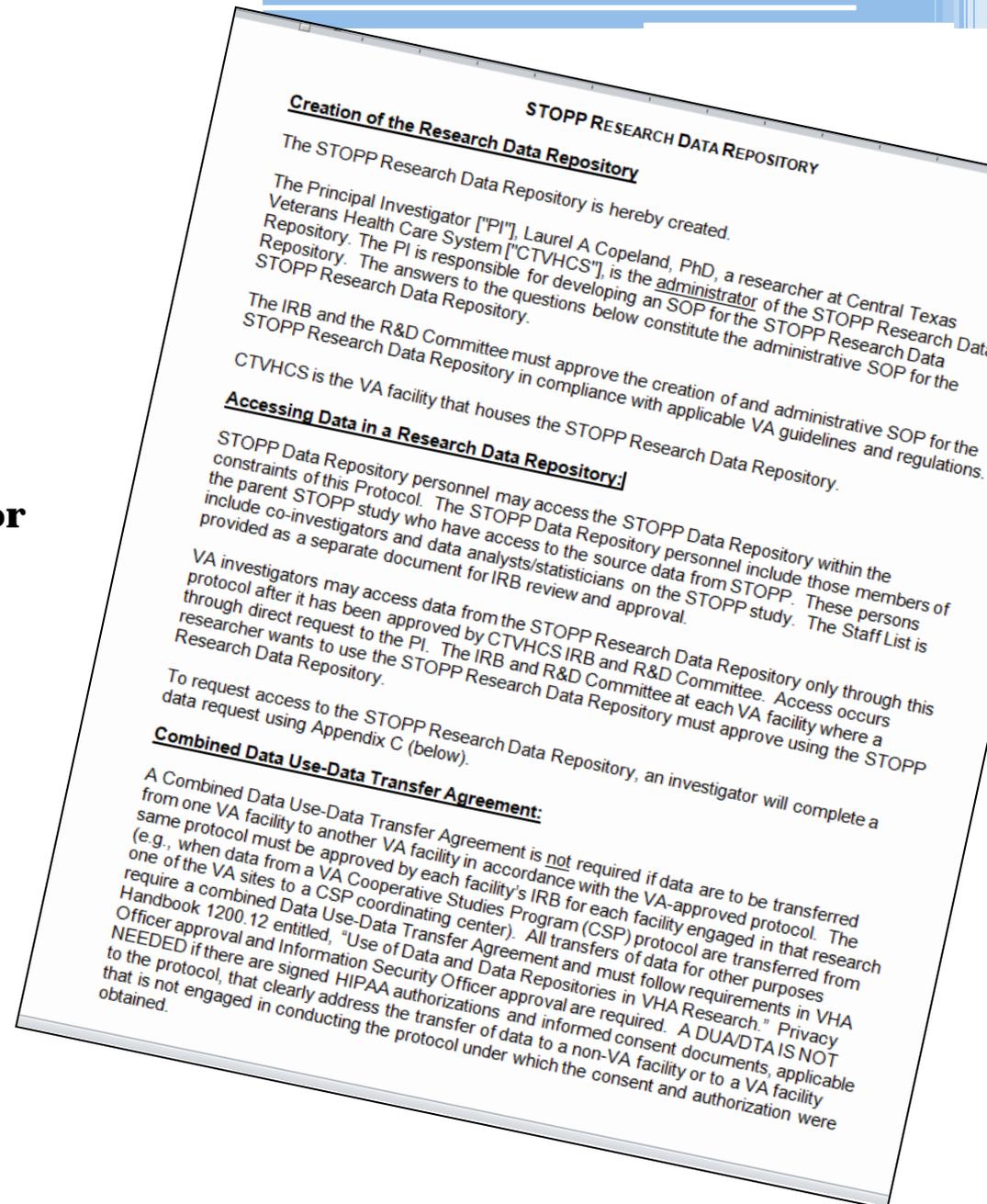
RDR Example:

Request review process

IIR 09-335

Surgical Treatment Outcomes for Patients with Psychiatric Disorders (STOPP)

Laurel Anne Copeland PhD MPH BS
Central Texas Veterans Health Care
System, Temple, TX



Consideration for a data repository host

- Data security
- Scheduled back-ups
- File recovery system
- Adequate volume capacity
- Compatible data formats
- Long-term data retention capacity
- Data access provisioning capability



Workspace - VHA Data Portal > Home
 https://vhacdrdget.vha.med.va.gov/RDWeb/Pages/en-US/Default.aspx



Welcome to VINCI Workspace

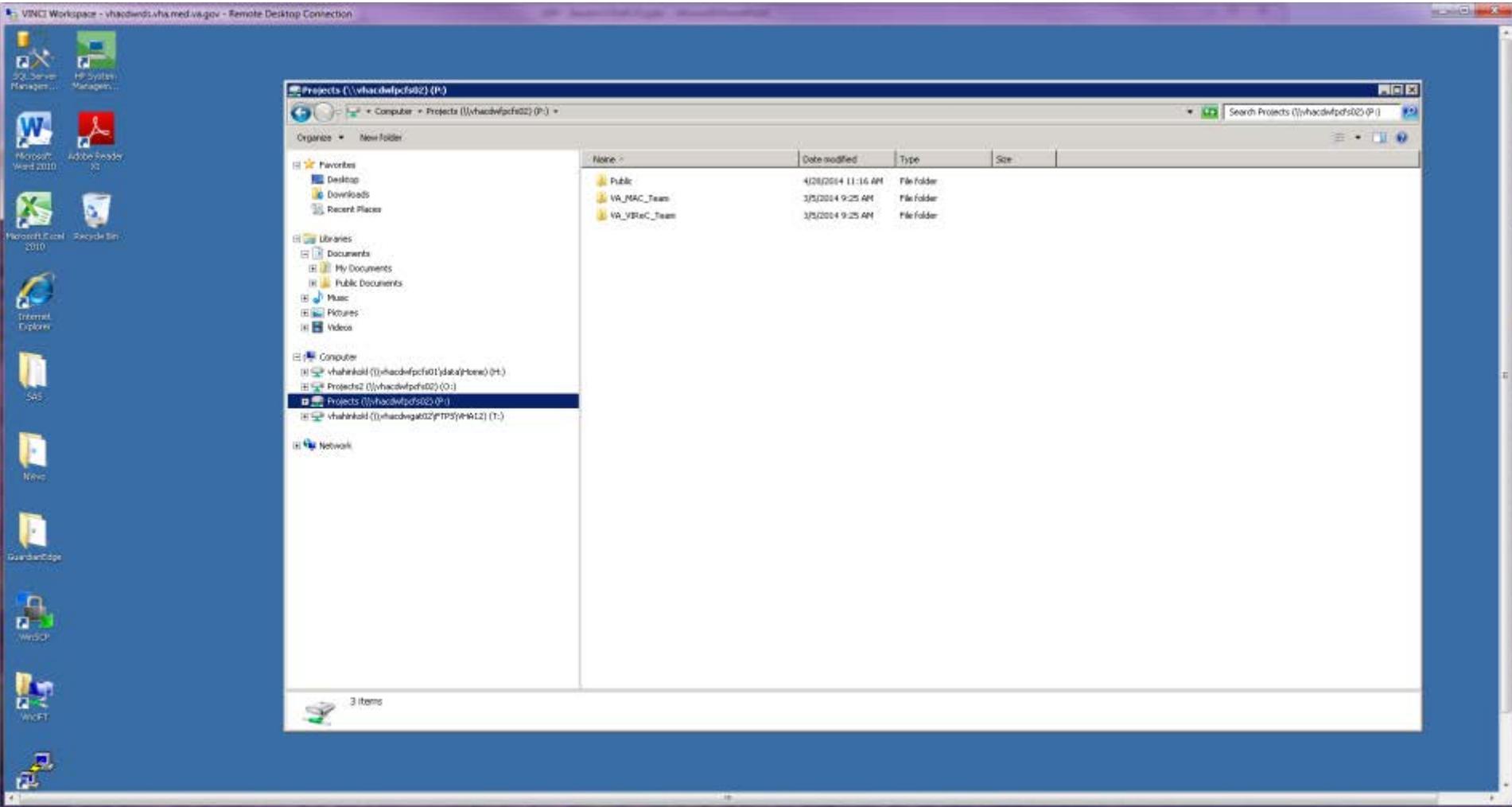
Standard and Development Workspaces

Standard Workspace | Development Workspace
Help

Important Instructions
 Click once on any icon below to start the application.
 When prompted to enter your login, please enter it in the form of **DomainName\Login**.
 This site is only supported on the Internet Explorer browser with Java Script enabled.
 If you are not able to see the application icons below, [click here](#)  to access the VINCI Standard Workspace Desktop to access all applications.

 Acrobat Distiller 9	 Adobe Acrobat 9 Pro	 Adobe Reader X	 Altova XMLSpy	 ArcMap 10	 EndNote	 Explorer	 GATE 5.2.1 GUI	 GuardianEdge Removable Storage	 MATLAB R2011b
 Microsoft Access 2010	 Microsoft Excel 2010	 Microsoft InfoPath Designer 2010	 Microsoft InfoPath Filler 2010	 Microsoft OneNote 2010	 Microsoft PowerPoint 2010	 Microsoft Word 2010	 MySQL Query Browser	 notepad++	 NVivo 8
 Protégé 3.3.1 (Knowtator)	 Protégé 3.4.8 (Knowtator)	 R x64 2.15.2	 R x64 2.15.3	 R x64 3.0.1	 SAS & SAS Grid 9.2	 SAS & SAS Grid 9.3	 SAS Enterprise Guide 4.3	 SAS Enterprise Guide 5.1	 SAS Grid Enterprise Miner
 SAS Power and Sample Size 3.1	 SAS Universal Viewer	 SAS XML Mapper	 SPSS Modeler 15	 SPSS Statistics 21	 SQL Server Management Studio 2012	 StataMP 11 (64-bit)	 StataMP 12 (64-bit)	 StatTransfer 9	 TextPad
 v3nlpClient	 v3nlpDictionary	 VinciFT	 WinDiff	 WinZip 10.0					

My VINCI Workspace



Poll Question #4

- If there were a central research data repository available in the VHA, how likely would you be to share data from one of your research projects for re-use?

5 = Very likely

4 = Likely

3 = Maybe

2 = Unlikely

1 = Never

Poll Question #5

- How likely would you be to use data from someone else's research data repository?

5 = Very likely

4 = Likely

3 = Maybe

2 = Unlikely

1 = Never

Resources

Massachusetts Institute of Technology (MIT) Libraries

Help Yourself : Subject Guides

MIT Libraries

Data Management and Publishing

[Home](#)

[Why Manage Your Data?](#)

[Data Planning Checklist](#)

[What is Data?](#)

[Evaluate Your Data Needs](#)

[Funder and Journal Requirements](#)

[Data Management Plans](#)

[Writing an NSF Data Management Plan](#)

[Documentation and Metadata](#)

[File Formats](#)

[Organizing Your Files](#)

[Backups and Security](#)

[Sharing Your Data](#)

[Citing Data](#)

[Data Integration](#)

Manage Your Data

The MIT Libraries supports the MIT community in the management and curation of research data by providing the following services:

Data Management Guide

This Data Management and Publishing Guide is a practical self-help guide to the management and curation of research data throughout its life cycle. It provides guidance on a range of topics, including: [planning for data management, documentation/metadata, file formats, data organization, data security and backup, citing data, data integration, funder requirements, ethical and legal issues, and sharing and archiving data.](#)

Assistance with Creating Data Management Plans

Many [funders](#), such as the National Science Foundation, have requirements for data sharing and [data management plans](#). We can help you to put together such a plan, assess the data management needs of your particular project, and assist in identifying solutions for data management and archiving.

Workshops

Our [workshops](#) teach you how to manage data more efficiently for your own use and help you to effectively share your data with others.

Individual Consultation and Collaboration with Researchers

We are available for individual consultation on data management issues, and can provide expertise in areas such as data organization and preservation, connect you to a network of data management services, and advocate for your needs. We can help you in understanding your data management needs and recommending optimal practices for keeping your data usable, now and into the future.

Faculty Successes:

"I've had thousands of downloads of my published data--I am impressed that it's been so useful to others!"

Esther Duflo, Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics, MIT

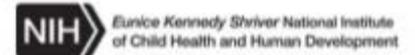
For advice on a data management project, contact:

data-management@mit.edu

Inter-university Consortium for Political and Social Research (ICPSR)



A data archive for demography and population sciences



[Log In/Create Account](#)
[Announcements](#)
[Contact Us](#)

[Search Holdings](#) [Deposit Data](#) [Restricted Data](#) [NICHD Funded Studies](#) [Publications](#) [NICHD Pop. Centers](#) [About Us](#)

[Why Deposit Data with ICPSR?](#)
[What Should My Deposit Include?](#)
[What Happens to My Data After the Deposit?](#)

Deposit Your Data

ICPSR welcomes and encourages deposits of digital data. Deposits are made using a secure form to describe the data collection and upload content.

Users must have or create a MyData, Facebook, or Google account to sign in to ICPSR and deposit data.

We strongly recommend that users consult the [DSDR Pre-Deposit Guide](#) (pdf) before depositing data with DSDR. For a more comprehensive overview regarding the data management steps that should be taken before depositing data with DSDR, please refer to the [ICPSR guide](#).

[Data Deposit Form](#)

For deposits that involve physical materials, please email deposit@icpsr.umich.edu

Our Partners

Population Center

PSC ICPSR

UNC
CAROLINA
POPULATION
CENTER

MPC

RAND

[Home](#) | [Privacy Policy](#)

© 2012 The Regents of the University of Michigan



<http://www.icpsr.umich.edu/icpsrweb/content/DSDR/deposit.html>

Contact Information

Linda Kok, MA

Hines VA Hospital

VIReC@ va.gov

708-202-2413

VIReC Resources

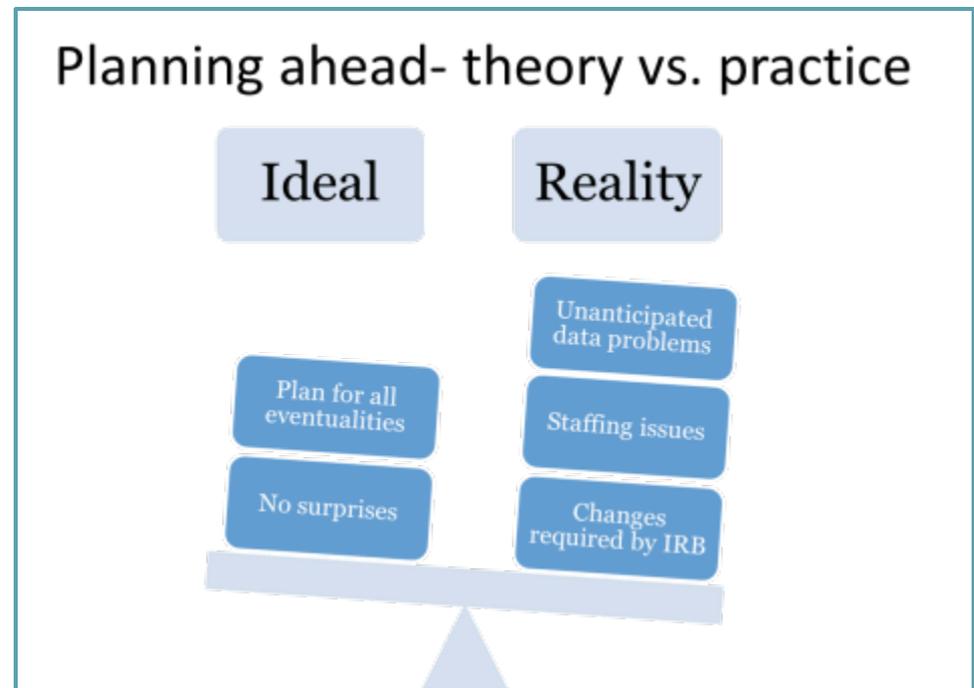
- **VIReC Help Desk**
 - VIReC staff will answer your question and/or direct you to available resources on topics
 - VIReC@va.gov
- **HSRData Listserv**
 - Join at the VIReC Intranet web site
 - Discussion among 800+ data stewards, managers, and users
 - Past messages archived on intranet
- www.virec.research.va.gov

Recap

- **Why did we do this series?**

Session1 *The Best Laid Plans: Plan well, plan early*

- **Benefits of early data planning**
- **Planning for data privacy & security**
- **Feasibility testing**
- **Planning for data re-use after the research**



Session 2 *The Living Protocol: Managing Documentation While Managing Data*

- Living Protocol
- Example: Managing secondary data
- Example: Managing linkage of primary & secondary data
- Conclusions: The value of documentation

“Objective: Share examples of conducting these tasks in timely manner after....”

- Trial and error
- Begging, borrowing and stealing best practices from other investigators”

Session 3 *Controlled Chaos: Tracking Decisions During an Evolving Analysis*

- Challenges of documentation
- Why good documentation?
- A schematic for document organization
- “Good practices:” Documents, communications & presentations
- Why organize?

“Explain everything clearly to your “future self” who will have to decipher these documents while writing a scientific manuscript in the distant future.”

Whiteboard:

- Things you learned?
- Things you will use in your research?
- Additional topics for future Good Data Practices sessions

Questions?

Bonus slides

Data planning checklist - Handout to be developed

- **Formalizing Your Data Management Plan**
 - Description of the project
 - Description of the data to be collected
 - Standards to be applied for formats, metadata, etc.
 - Plans for short-term storage and data management: e.g., file formats, local storage and back up procedures, and security
 - Description of legal and ethical issues: e.g., intellectual property, confidentiality of study participants
 - Access policies and provisions: i.e., how will you make data available to others, any restrictions to data reuse, etc.
 - Provisions for long-term archiving and preservation
 - Assigned data management responsibilities: i.e., which persons will actually be responsible for ensuring data management; compliance monitoring over time

Importance of documenting as you go

- Accurate, systematic record of the research process
- Captures decisions and the reasoning behind them
- Reduces mistakes, confusion and wasted time
- Provides the basis for re-using data with confidence

It's all in the details...



Documentation checklist (handout)

- Project Description
- Data description
- Methodology description
- Study Citation
- Study Abstract

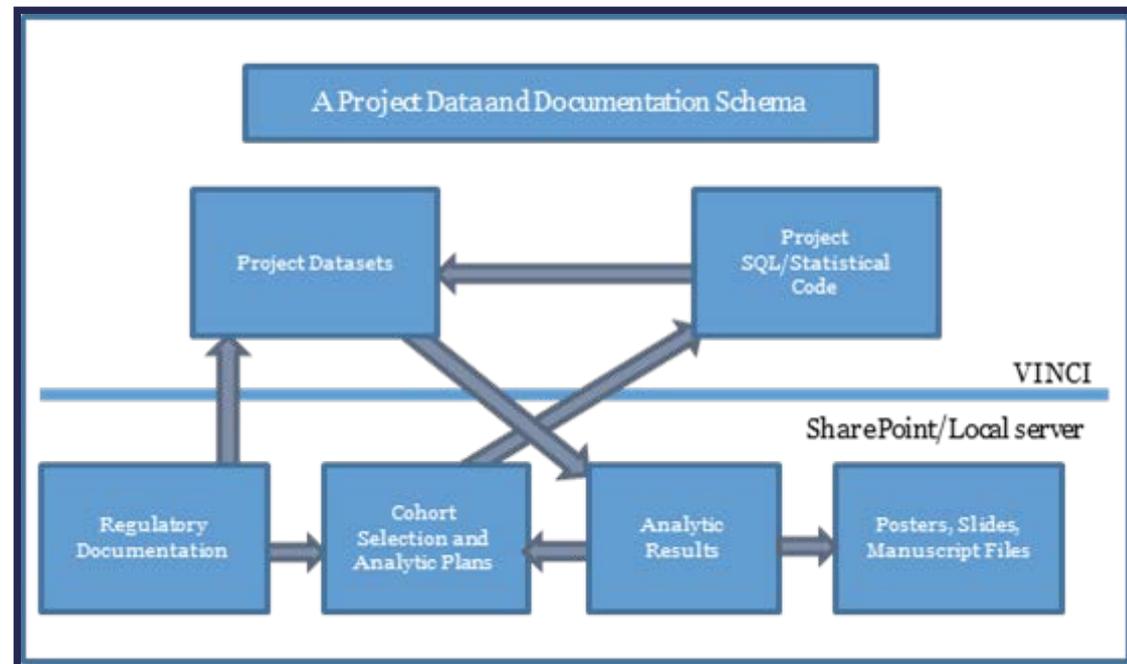


Project description

- **A brief narrative & summary information about the data collection**
 - **Creator (PI)**
 - **Level of analysis**
 - **Identifier used**
 - **Time frame for the data collected or used**
 - **Funder**
 - **Where you did your study**

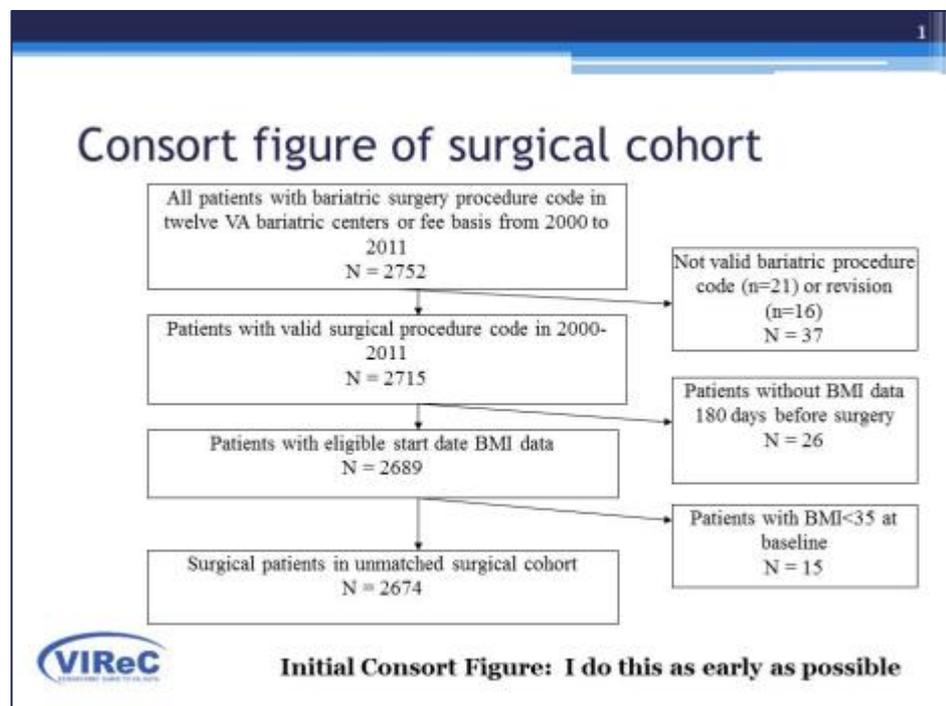
Data description

- Technical information describing each data set
 - File format and structure
 - Variable list - Name, description, format, label, missing data...
- Session 3, slide 25



Methodology description

- **The Cohort definition**
 - with decisions & reasoning
 - a consort chart – (see Session 2, slide 40)



Sources of data

- Data collection instrument
- Existing data sources
 - See table from Session 2, slide 34
- Process description for creating the analytic dataset.

Plan for Pulling Multiple VA Claims Datasets

Dataset	Aim	Outcome To Create	Covariates To Create (in brief)	Years We Have on Cases	Years We Have on Controls
VASQIP	All	--	Surgery type	2000-2011	Not applicable
Fee Basis	All	--	Surgery type	2000-2011	Not applicable
Mini-Vitals	All, 2	Death	Age, gender	Most current	Most current
HERC	3	Cost		2000-2011	??
DSS LAR	All	Lab results	Baseline values for A1c, LDL	2000-2011	2000-2011
PBM	1	Disease control	Medications at baseline	2000-2011	2000-2011
CDW	All	BP, Weight Δ	Baseline value of BP, BMI	2000-2011	2000-2011
OPC	All	Utilization, complications	Race, marital status, Dx-based covars (comorbidity)	2000-2011	In process (covariate & exclusions)
PTF	All	Utilization, complications	Dx-based covars	2000-2011	In process (covariate & exclusions)
DCG	All	--	DCG risk score	2000-2011	2000-2011
Enrollment	All	--	Copay status (from Priority St)	2000-2011	2000-2011

Documentation of derived analytic variable

Note:

- Description
- Source
- SAS code

From: “Impact of Medicare Drug Benefit on VA Drug Use, Healthcare Use and Cost”;
 Funding Agency: Department of Veterans Affairs Health Services Research and Development;
 Study Number: IIR 07-165-2 - Kevin Stroupe, PI

VARIABLE: COPAY

DESCRIPTION:
 Three co_pay groups(no copay, some copay, all copay) based on veteran's priority category

Data Type: Numeric

Label: 'if PRI01_8 =1 then copay=0 if 2 <= PRI01_8 <= 6 then copay=1 if PRI01_8>6 then copay=2'

Program Location: hsrfilesS:\PartD\7 Programs\Jenny_huo\demog.sas

SOURCES:

Variable Name	File	Description
pr01_8	Name:VA Enrollment file(PSSG) Type: SAS	Three co_pay groups(no copay, some copay, all copay) based on veteran's priority category
ENRLPRI0	Location: hsrdataS:\PartD\Original_data\AA_C_DATA\pssg04-10	
ELIG		

SELECTION CRITERIA or CALCULATION or CODES:

```
To determine priority category:
if pr01_8=' ' then PRIORITY= pr01_8;
else if pr01_8=' ' and ENRLPRI0=' ' then PRIORITY=ENRLPRI0;

if PRIORITY='1 ' then copay=0;
else if '2 '<= PRIORITY <= '6 ' then copay=1;
else if PRIORITY>'6 ' and priority not in ('99' '') then copay=2;
if priority in ('99' '') and elig=' ' then do;
if substr(elig,1,1)='A' then copay=0;
else if substr(elig,1,1)='G' then copay=1;
else copay=2;
end;
```

If listed as multiple priorities, assign the most frequently occurring

Study Citation & Abstract

- The study citation others must use to cite your work in developing the data.
- Abstract of the study
 - Describe theoretical framework
 - Research questions addressed
 - Specific hypotheses tested