Research Data Management Implementation Utilizing Dataverse

Mercè Crosas, Ph.D.
Director of Data Science

IQSS
The Institute for Quantitative Social Science
HARVARD UNIVERSITY

Jonathan Crabtree
Assistant Director of Computing
and Archival Research

UNC
THE ODUM INSTITUTE

Science in the Digital Age

Raw Data
(Collect, Store
& Analyze)
Research computing
facilities,
Domain specific
databases

Research
Results

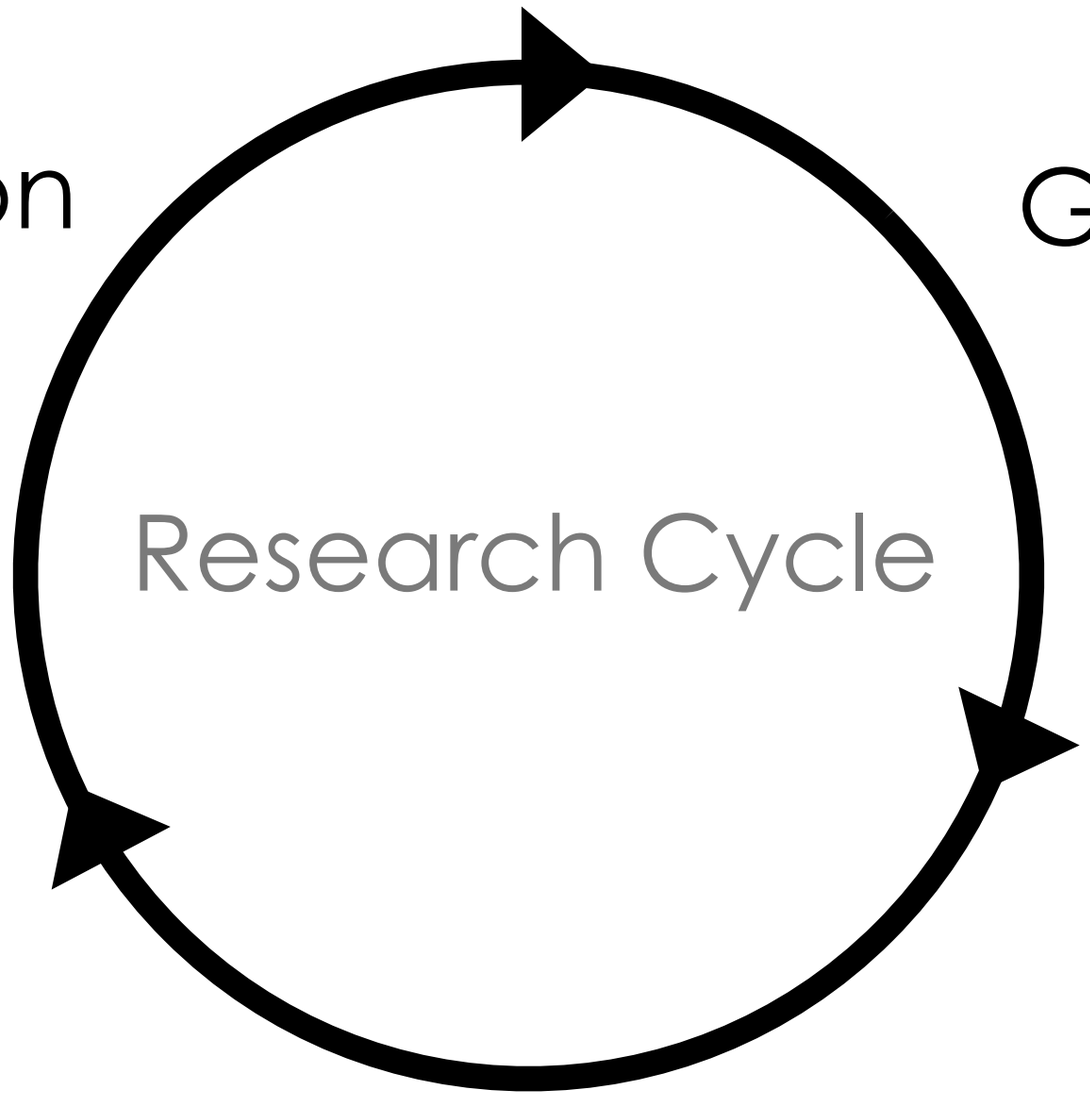
Publications,
Presentations,
& Grant Results
Journals, Books,
Reports, Web Sites

Data citation

Final Data, Code
& other Materials
(Share, Cite &
Archive)
**The Dataverse
Network**

Project Idea
Grant Proposal

Publication



Data Collection,
Processing and
Analysis

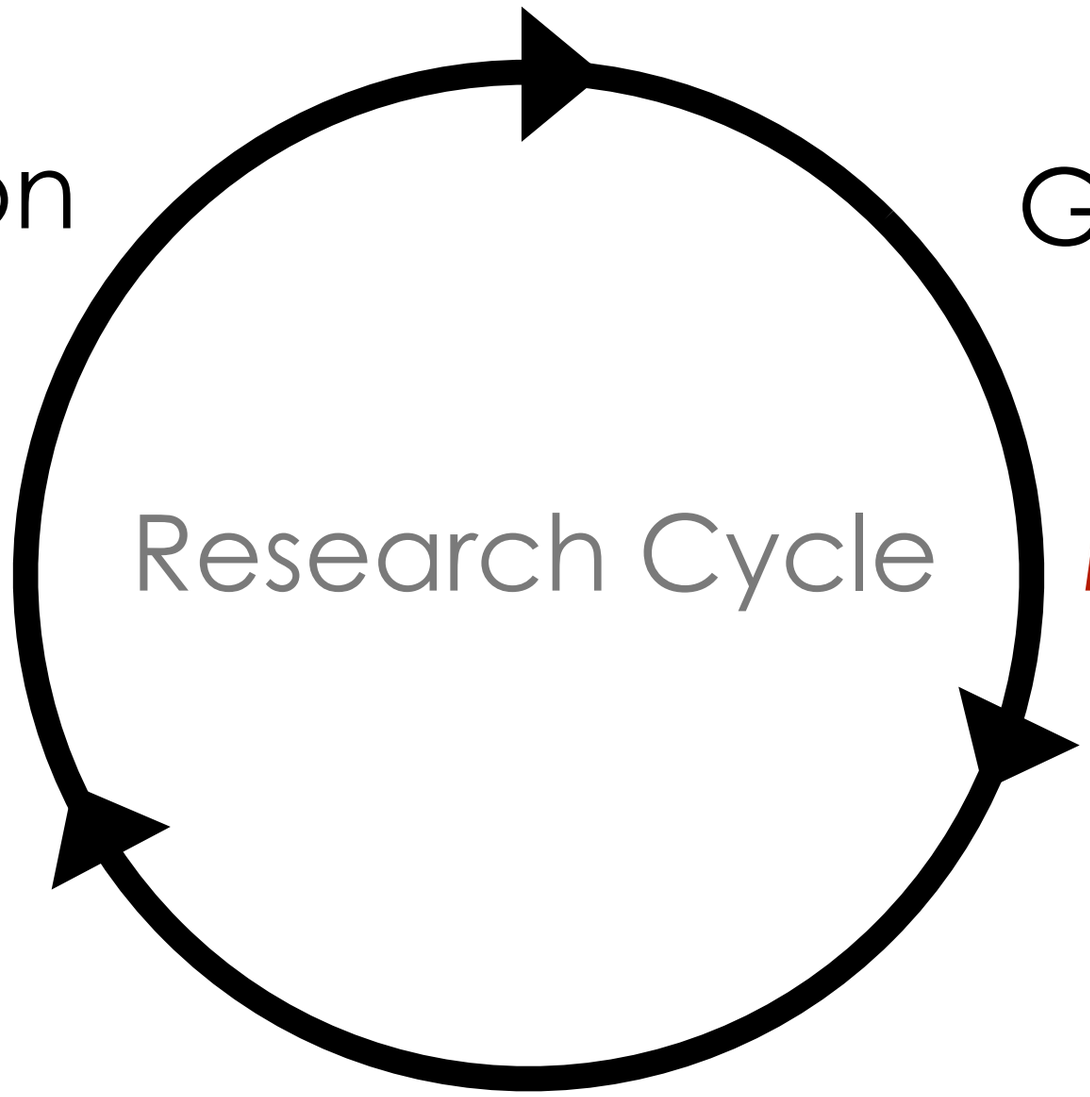
Project Idea
Grant Proposal

Data
Management
Plan!

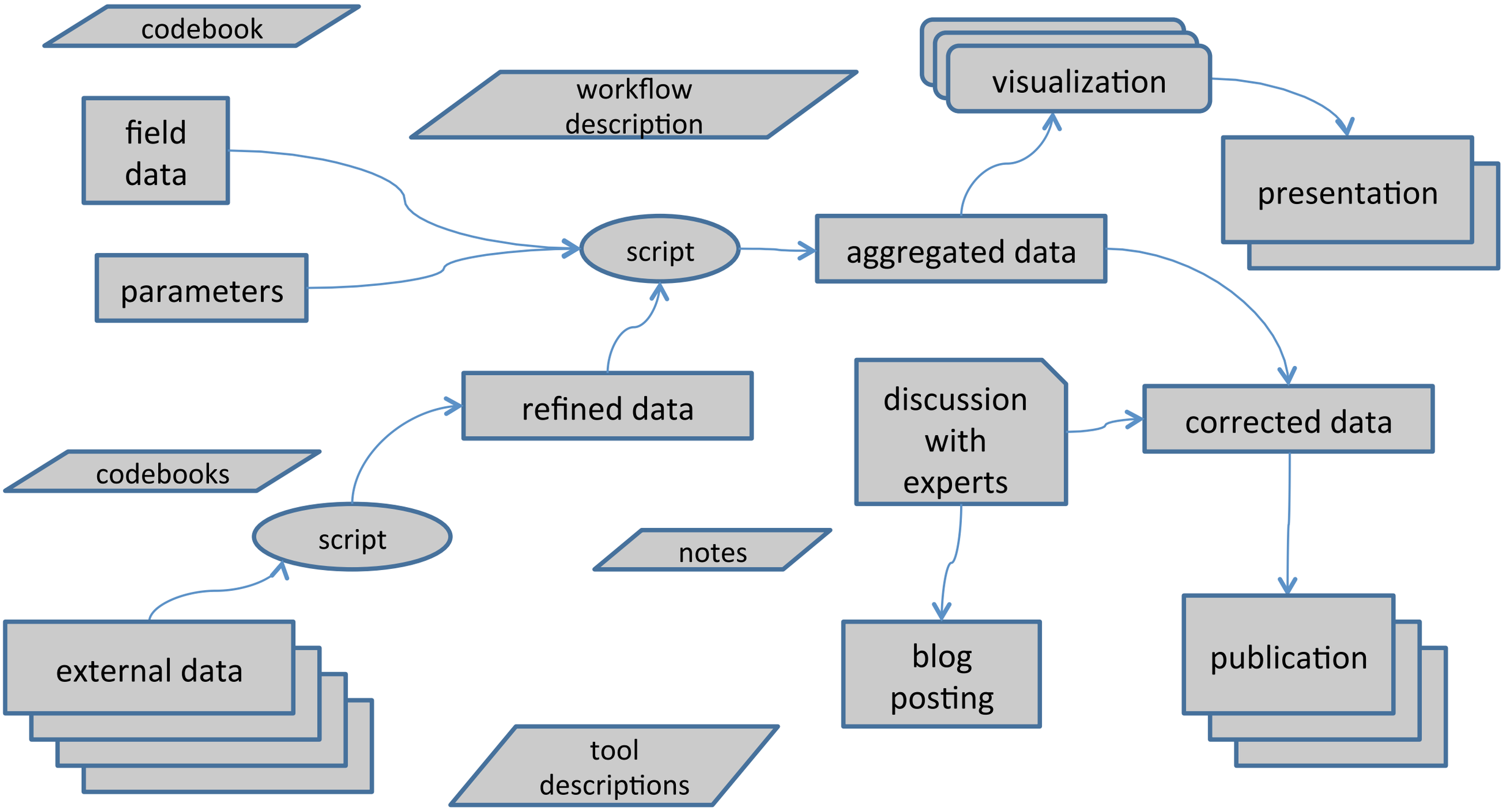
Data Collection,
Processing and
Analysis

Data
Sharing
and
Archiving

Publication

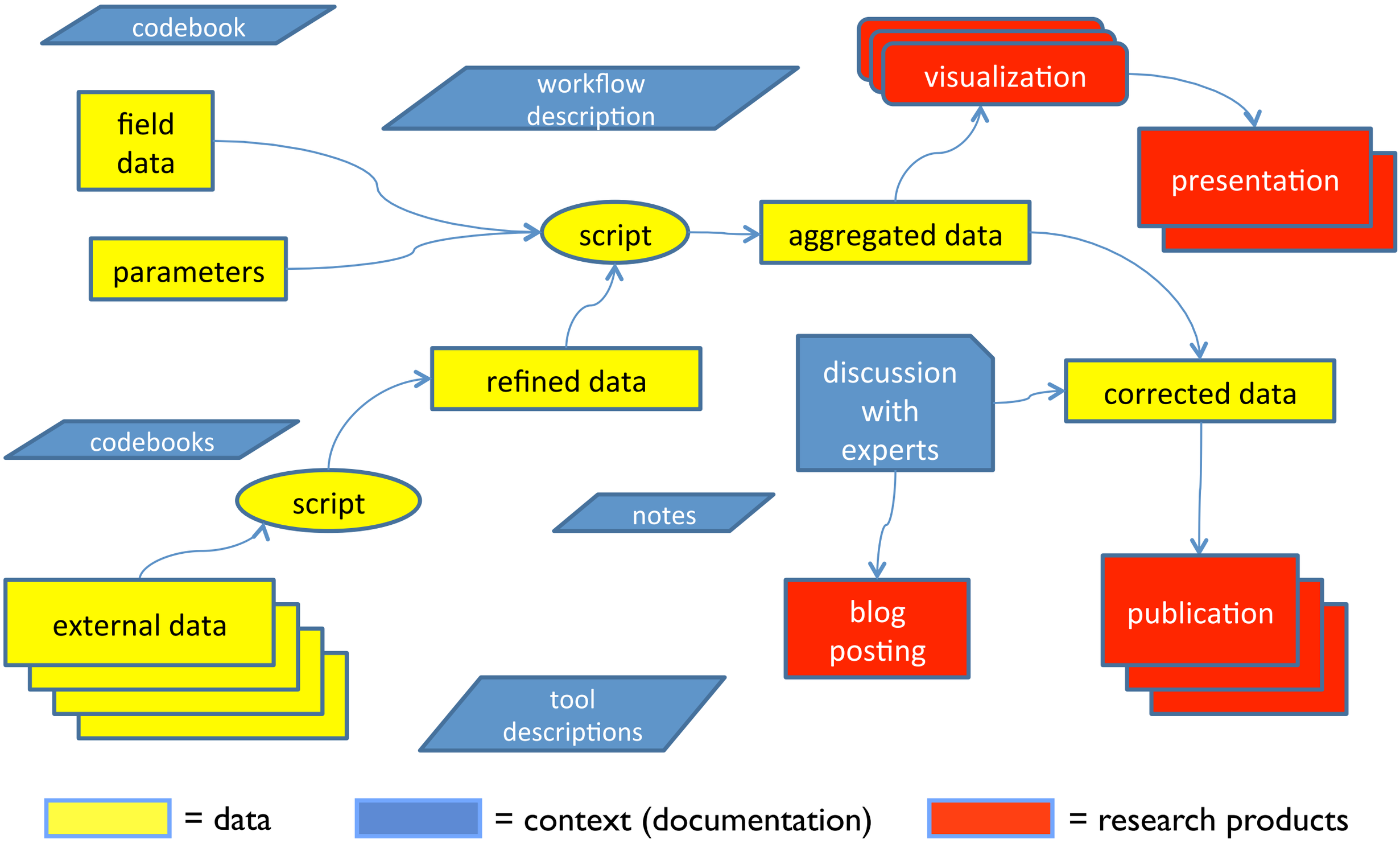


Which Data?



Slide acknowledgment: Andrea Goethals

Which Data?



Slide acknowledgment: Andrea Goethals

Assortment of Data Types?

- Survey data
- Weather measurements
- Computer models
- Simulation outputs
- Software
- Input data
- Federated queries
- Geospatial data
- Enumerations
- Vital statistics
- Interviews
- Transcripts
- Medical and genetic data
- Scripts
- Administrative records
- Field data
- Bio images
- Experimental data
- Spreadsheets
- Visualizations
- Instrument outputs
- Structured data products
- Database
- Quantitative tabular data
- Qualitative data
- Recordings

The Dataverse Network

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

Researcher

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

Researcher

- ✓ Deposits data and enters metadata

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

Researcher

- ✓ Deposits data and enters metadata
- ✓ Gets data citation (handle, UNF)

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

Researcher

- ✓ Deposits data and enters metadata
- ✓ Gets data citation (handle, UNF)
- ✓ Displays data on own web site

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

Researcher

- ✓ Deposits data and enters metadata
- ✓ Gets data citation (handle, UNF)
- ✓ Displays data on own web site
- ✓ Manages data permissions

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

Researcher

- ✓ Deposits data and enters metadata
- ✓ Gets data citation (handle, UNF)
- ✓ Displays data on own web site
- ✓ Manages data permissions
- ✓ Updates new versions

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

Researcher ⟷ Centralized Data Repository

- ✓ Deposits data and enters metadata
- ✓ Gets data citation (handle, UNF)
- ✓ Displays data on own web site
- ✓ Manages data permissions
- ✓ Updates new versions

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

Researcher ⟷ Centralized Data Repository

- ✓ Deposits data and enters metadata
- ✓ Gets data citation (handle, UNF)
- ✓ Displays data on own web site
- ✓ Manages data permissions
- ✓ Updates new versions

- ✓ Backups and replication of data in different locations (LOCKSS)

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

Researcher

- ✓Deposits data and enters metadata
- ✓Gets data citation (handle, UNF)
- ✓Displays data on own web site
- ✓Manages data permissions
- ✓Updates new versions

Centralized Data Repository

- ✓Backups and replication of data in different locations (LOCKSS)
- ✓Conversion to archival formats

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

Researcher

- ✓Deposits data and enters metadata
- ✓Gets data citation (handle, UNF)
- ✓Displays data on own web site
- ✓Manages data permissions
- ✓Updates new versions

Centralized Data Repository

- ✓Backups and replication of data in different locations (LOCKSS)
- ✓Conversion to archival formats
- ✓Extraction of Metadata from data sets

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

Researcher

- ✓Deposits data and enters metadata
- ✓Gets data citation (handle, UNF)
- ✓Displays data on own web site
- ✓Manages data permissions
- ✓Updates new versions

Centralized Data Repository

- ✓Backups and replication of data in different locations (LOCKSS)
- ✓Conversion to archival formats
- ✓Extraction of Metadata from data sets
- ✓Metadata standards (DDI, Dublin Core)

The Dataverse Network

A repository for research data that takes care of long term preservation and good archival practices, while the researcher keeps control of and gets recognition for his data

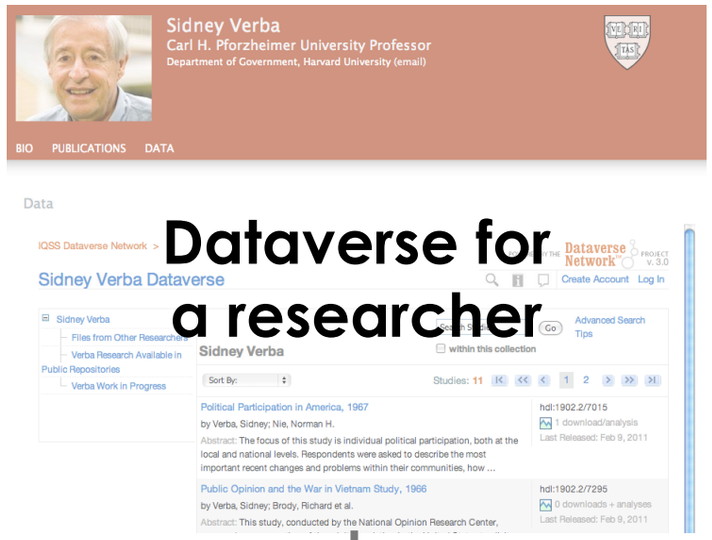
Researcher

- ✓Deposits data and enters metadata
- ✓Gets data citation (handle, UNF)
- ✓Displays data on own web site
- ✓Manages data permissions
- ✓Updates new versions

Centralized Data Repository

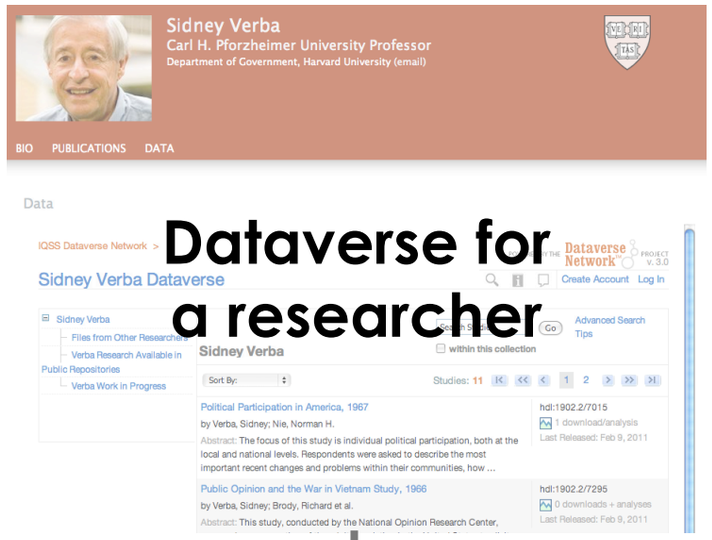
- ✓Backups and replication of data in different locations (LOCKSS)
- ✓Conversion to archival formats
- ✓Extraction of Metadata from data sets
- ✓Metadata standards (DDI, Dublin Core)
- ✓Inter-operability (OAI, APIs)

Dataverse Network (centralized repository)

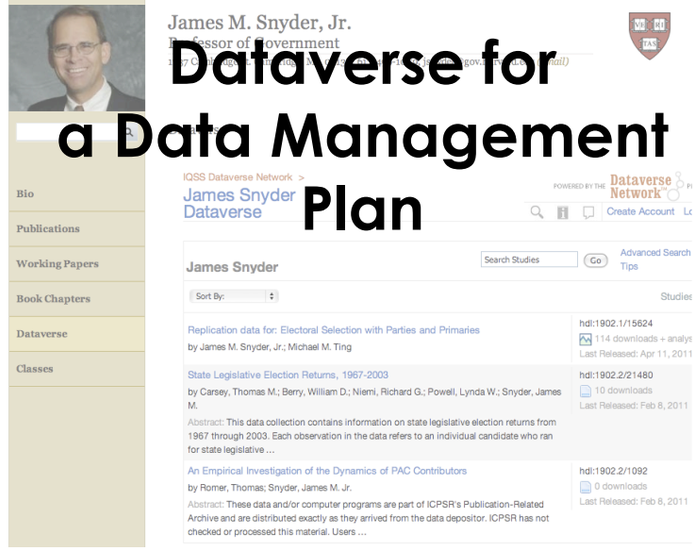


Dataverse for a researcher

Dataverse Network (centralized repository)

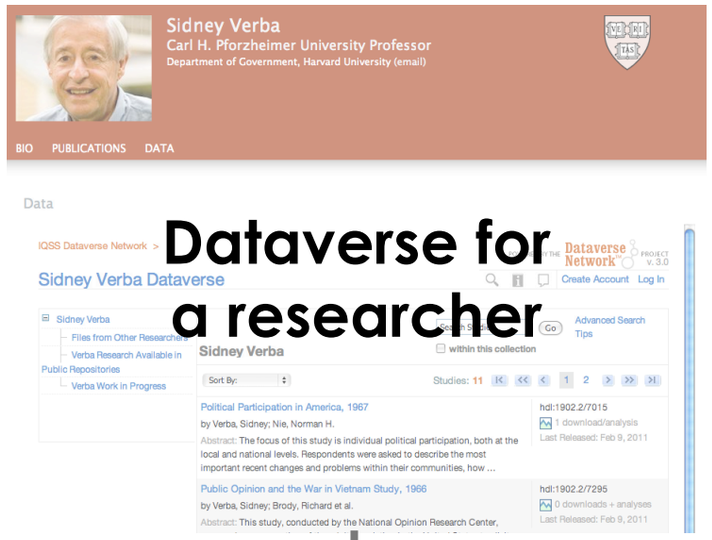


Dataverse for a researcher



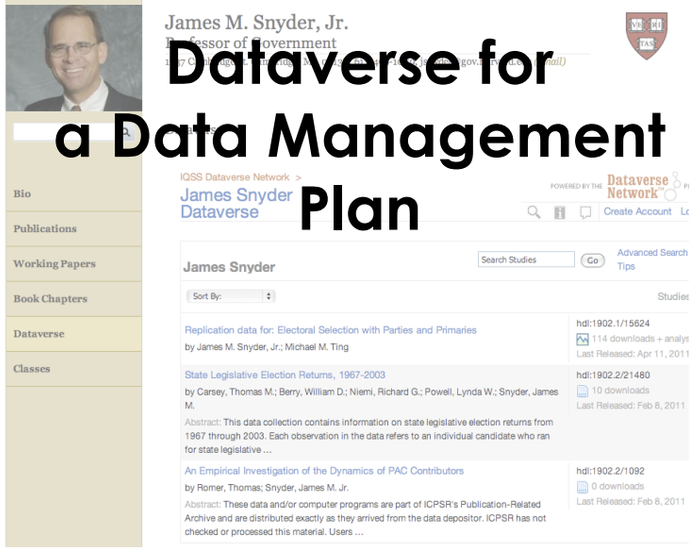
Dataverse for a Data Management Plan

Dataverse Network (centralized repository)

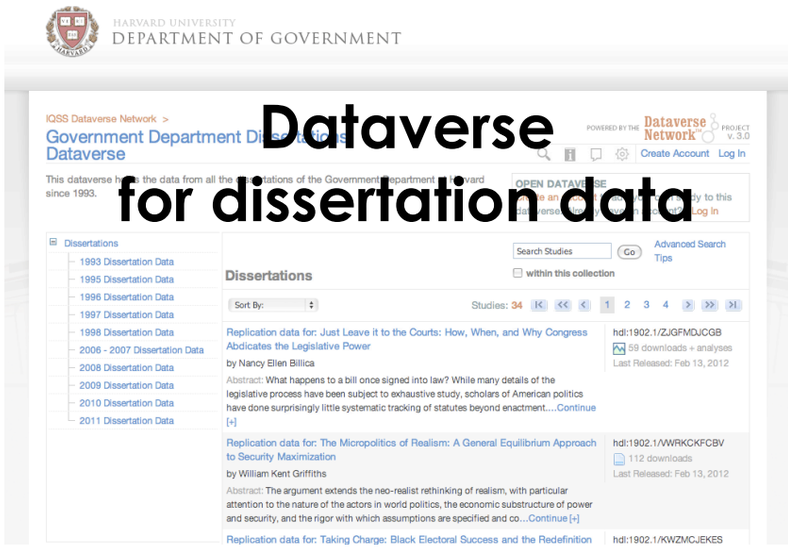


Dataverse for a researcher

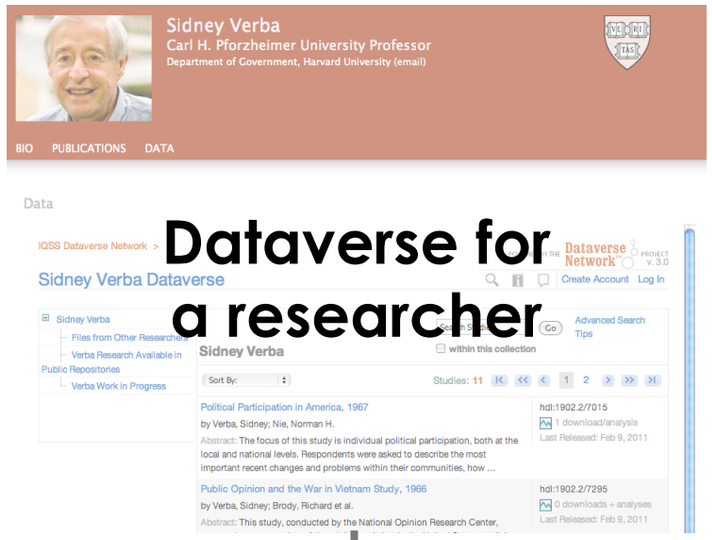
Dataverse Network (centralized repository)



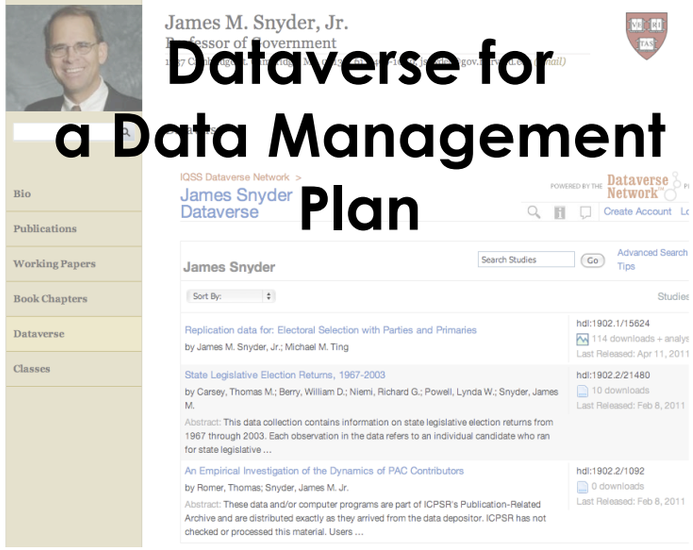
Dataverse for a Data Management Plan



Dataverse for dissertation data

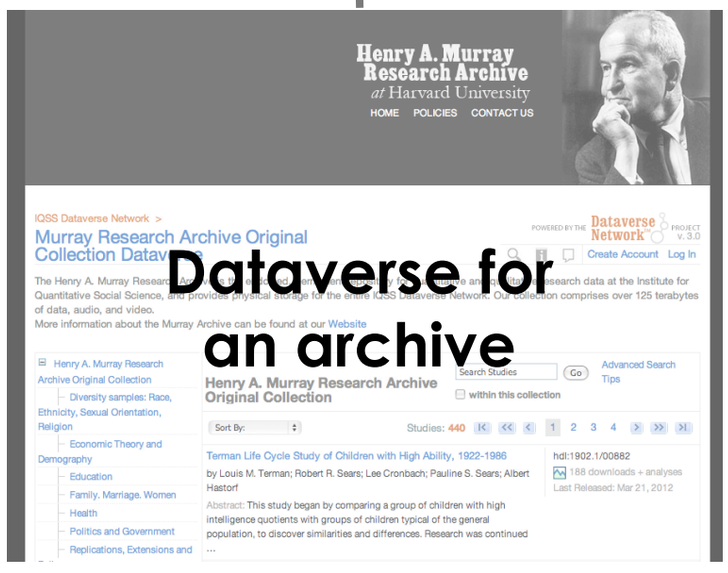


Dataverse for a researcher

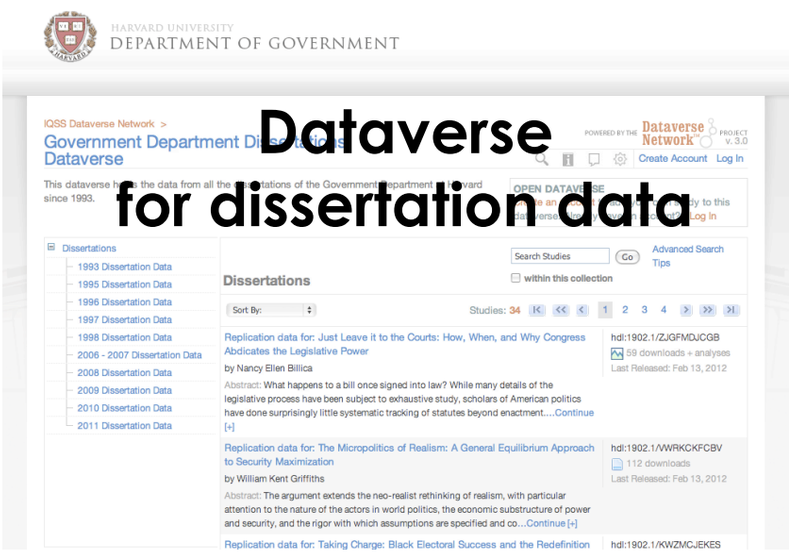


Dataverse for a Data Management Plan

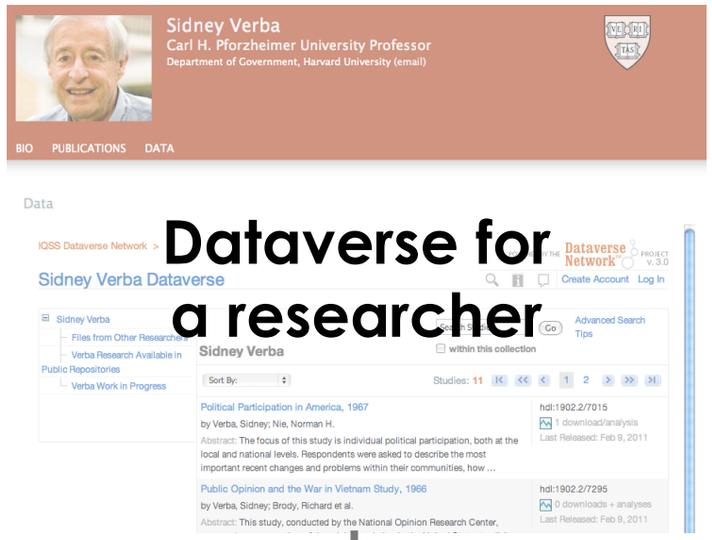
Dataverse Network (centralized repository)



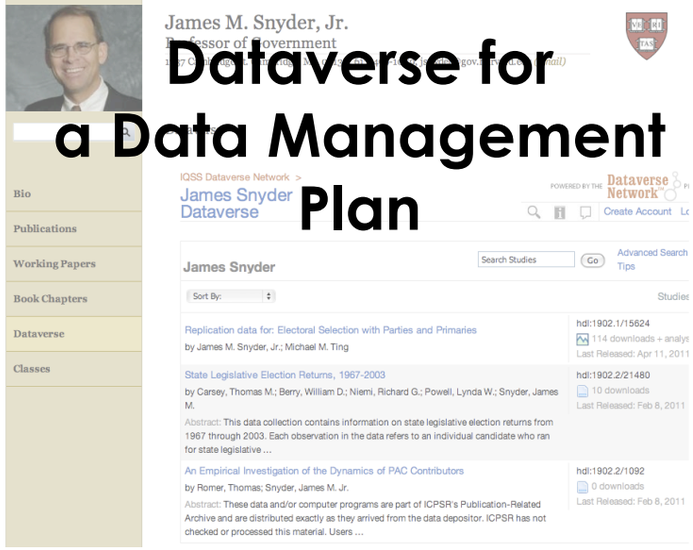
Dataverse for an archive



Dataverse for dissertation data

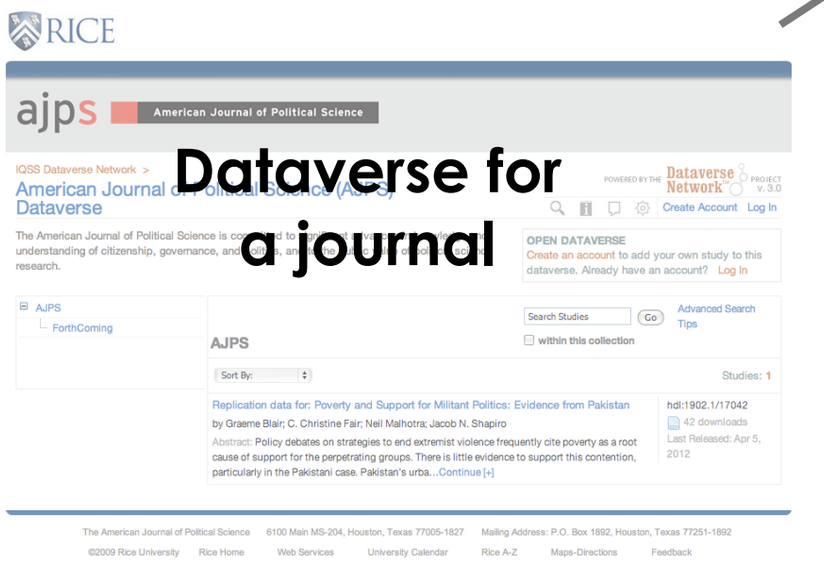


Dataverse for a researcher

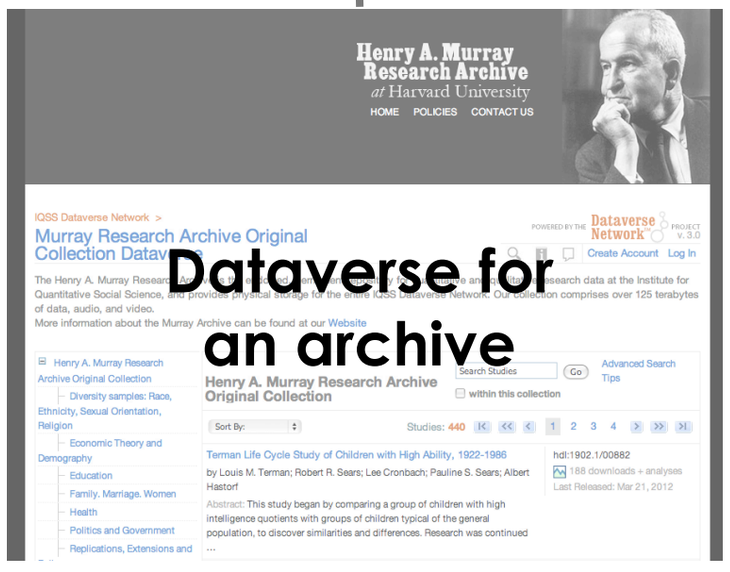


Dataverse for a Data Management Plan

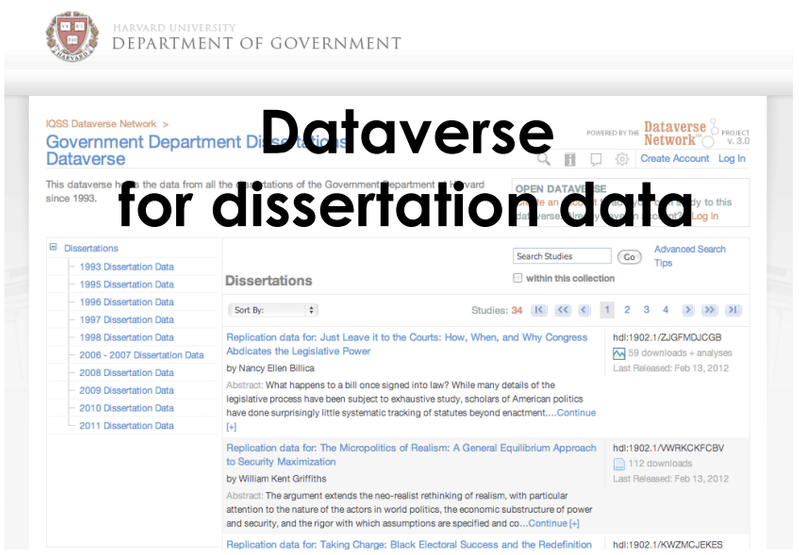
Dataverse Network (centralized repository)



Dataverse for a journal



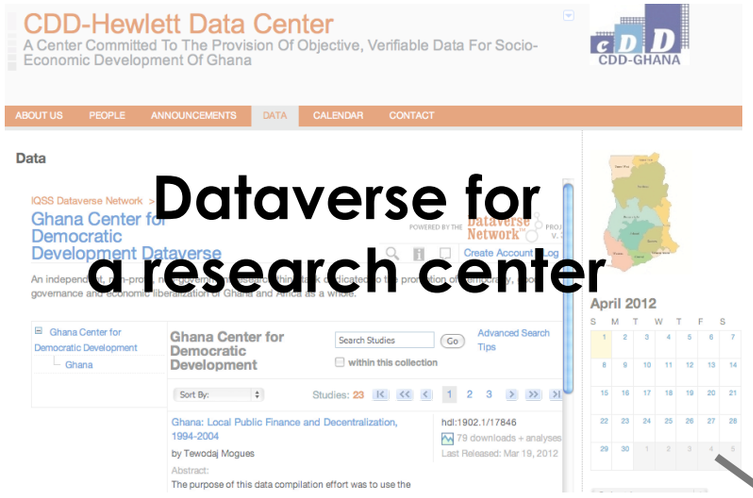
Dataverse for an archive



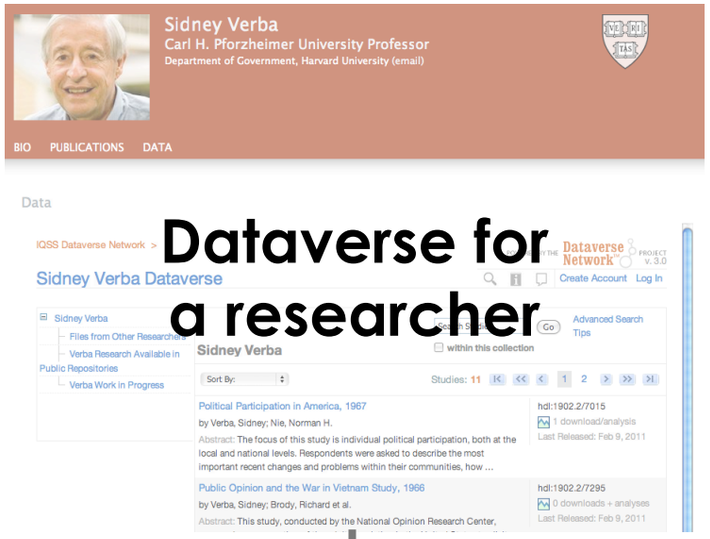
Dataverse for dissertation data

Dataverse Network (centralized repository)

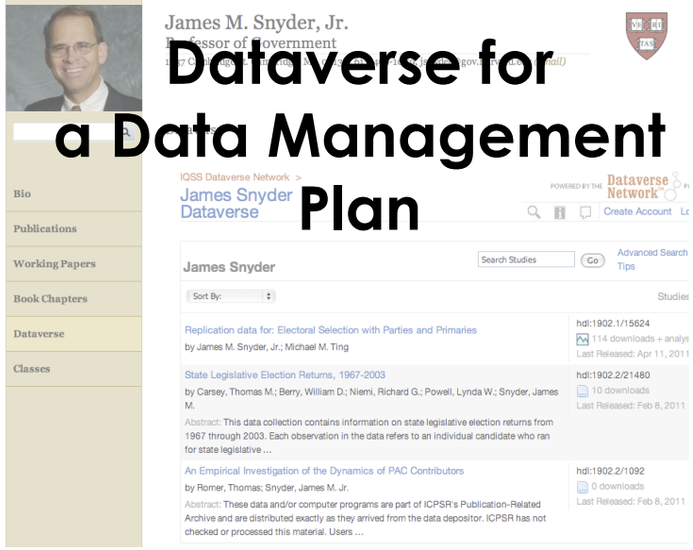
Dataverse for a research center



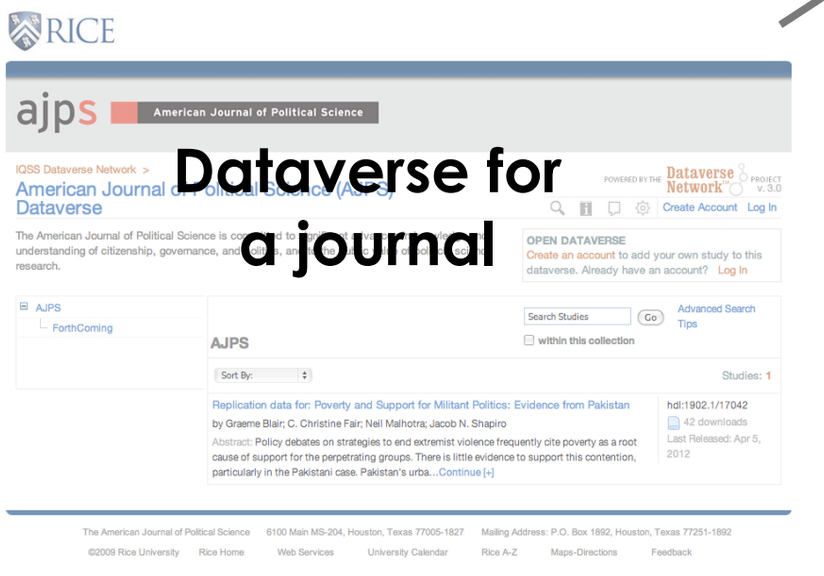
Dataverse for a researcher



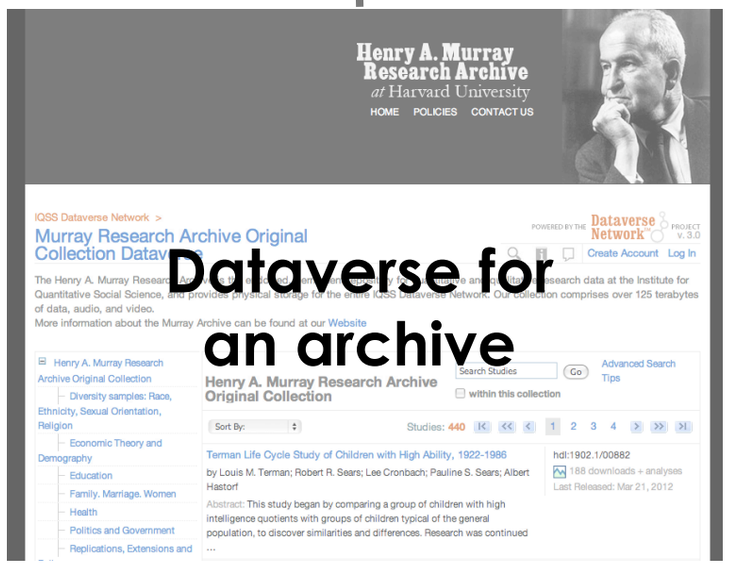
Dataverse for a Data Management Plan



Dataverse for a journal



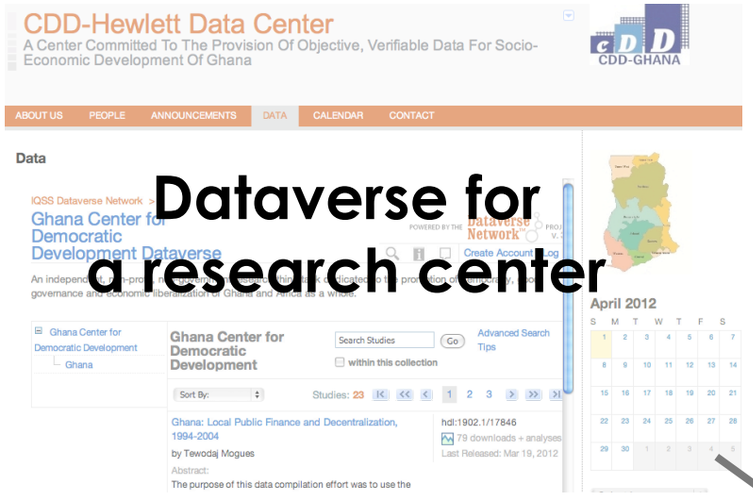
Dataverse for an archive



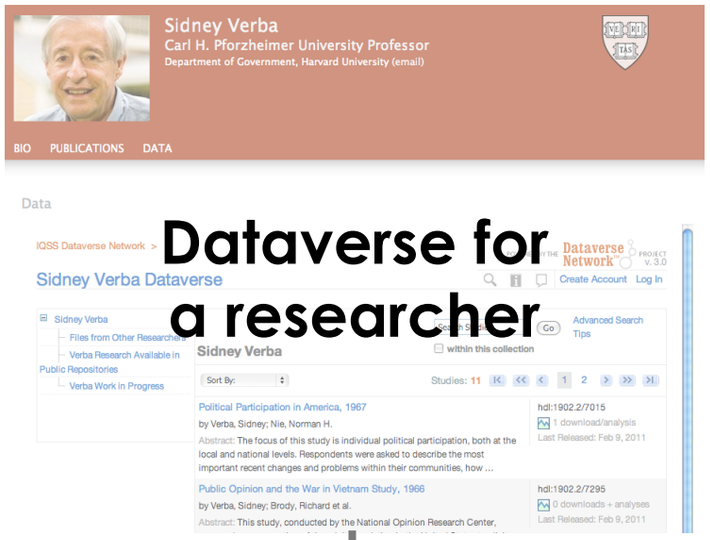
Dataverse for dissertation data

Dataverse Network (centralized repository)

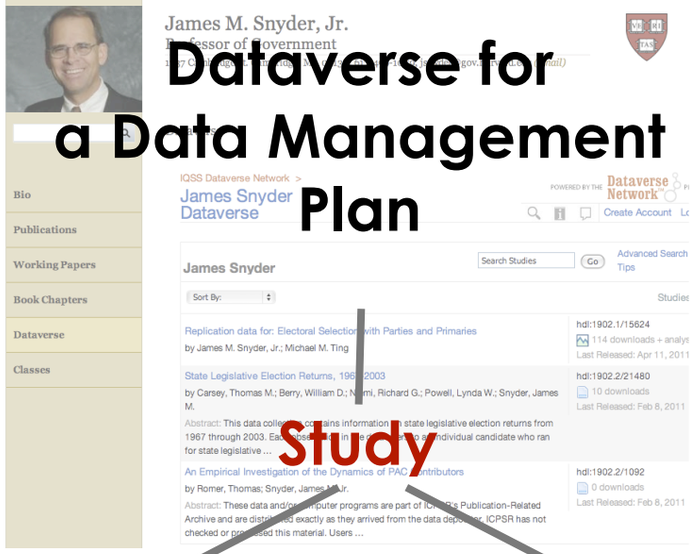
Dataverse for a research center



Dataverse for a researcher



Dataverse for a Data Management Plan

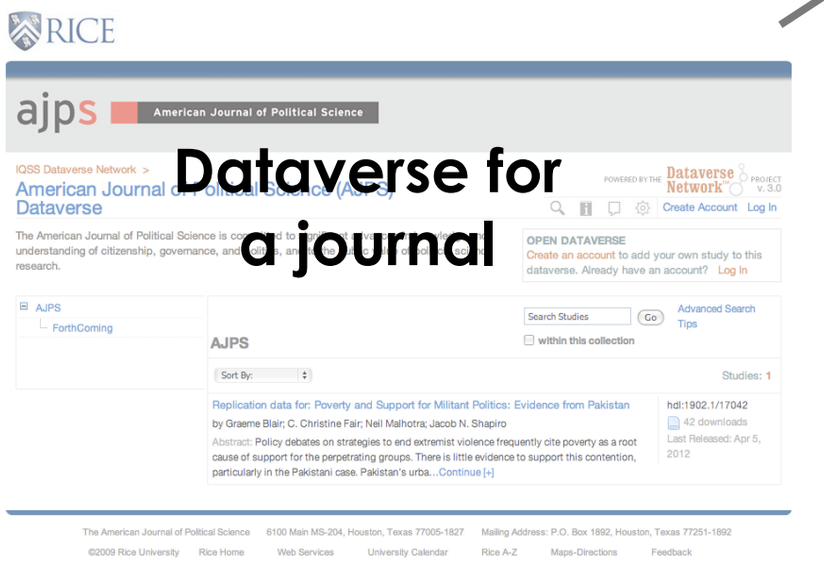


Study

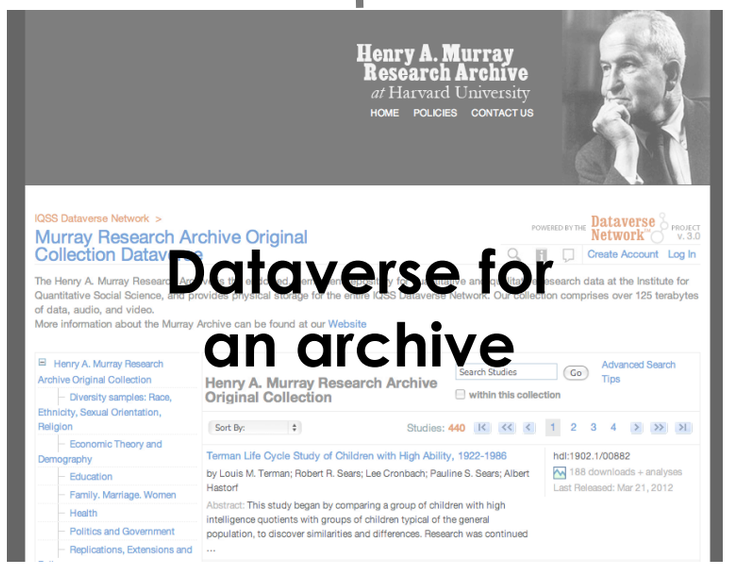
Metadata

Data Files

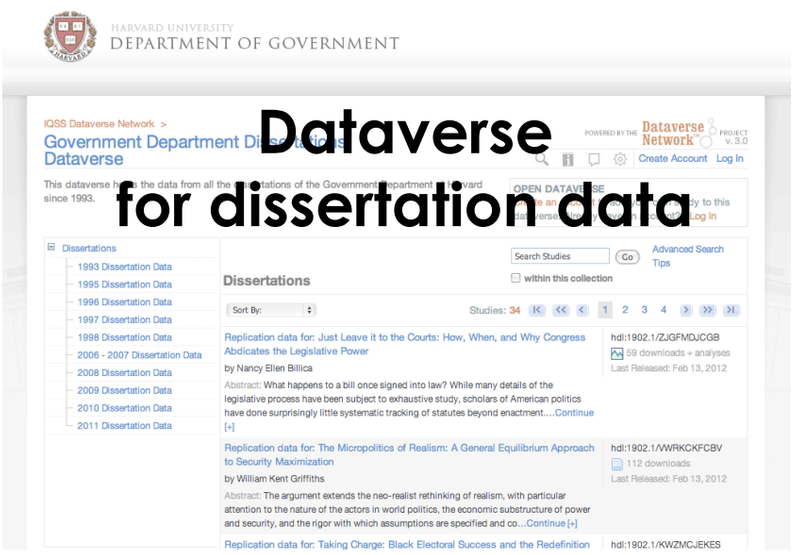
Dataverse for a journal

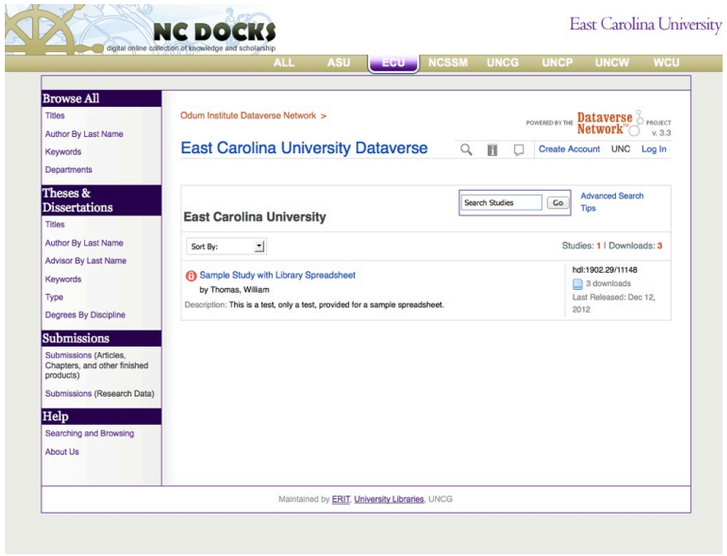


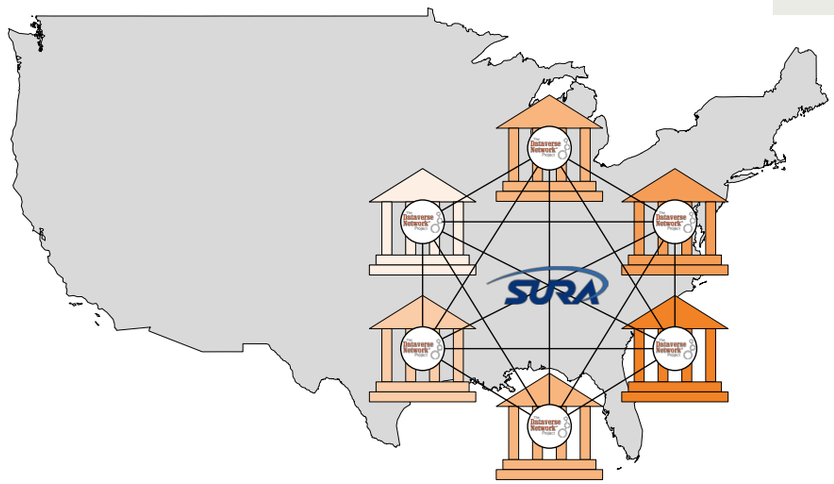
Dataverse for an archive

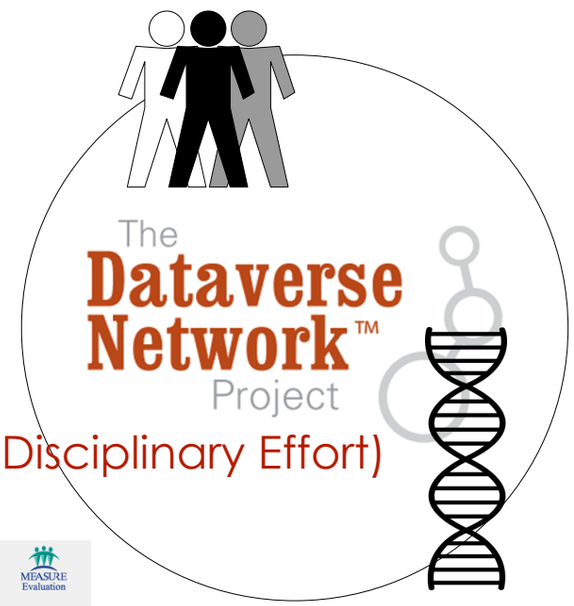


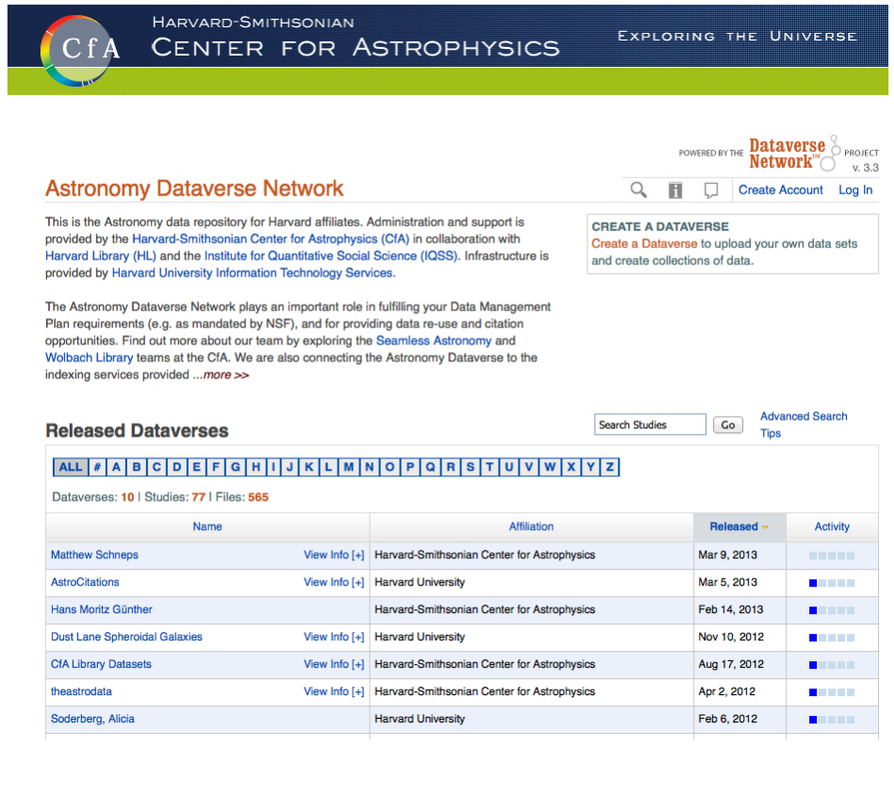
Dataverse for dissertation data

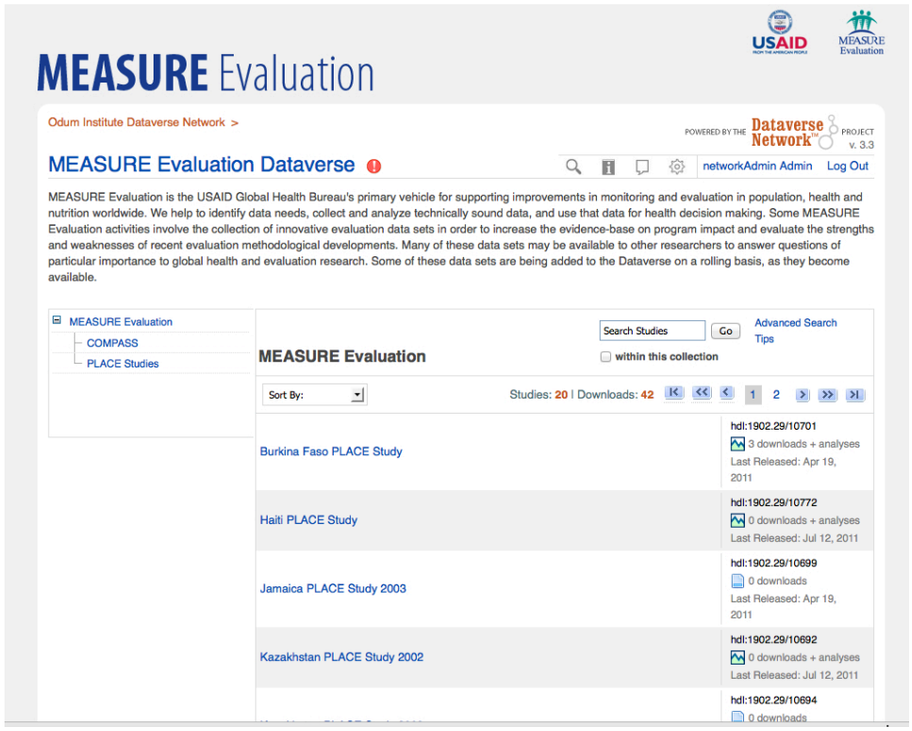


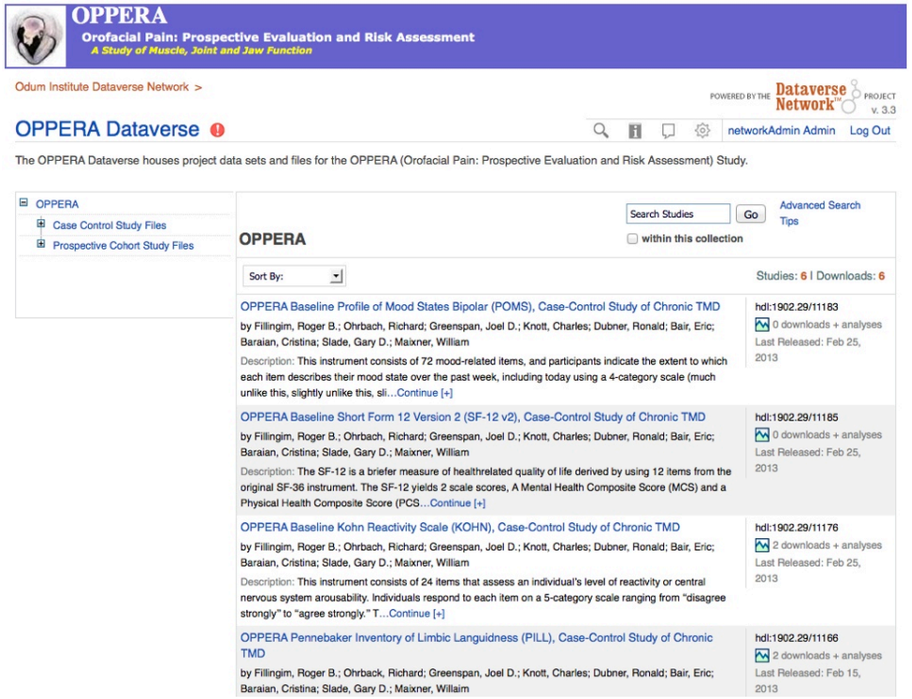

NC DOCKS
digital online collection of knowledge and scholarship
East Carolina University
ALL ASU ECU NCSSM UNCG UNCP UNCW WCU
Browse All
Titles
Author By Last Name
Keywords
Departments
Theses & Dissertations
Titles
Author By Last Name
Advisor By Last Name
Keywords
Type
Degrees By Discipline
Submissions
Submissions (Articles, Chapters, and other finished products)
Submissions (Research Data)
Help
Searching and Browsing
About Us
Odum Institute Dataverse Network >
East Carolina University Dataverse
Create Account UNC Log In
East Carolina University
Search Studies Go
Advanced Search Tips
Sort By:
Studies: 1 | Downloads: 3
Sample Study with Library Spreadsheet
by Thomas, William
Description: This is a test, only a test, provided for a sample spreadsheet.
hdl:1902.29/11148
3 downloads
Last Released: Dec 12, 2012
Maintained by ERIT, University Libraries, UNCG

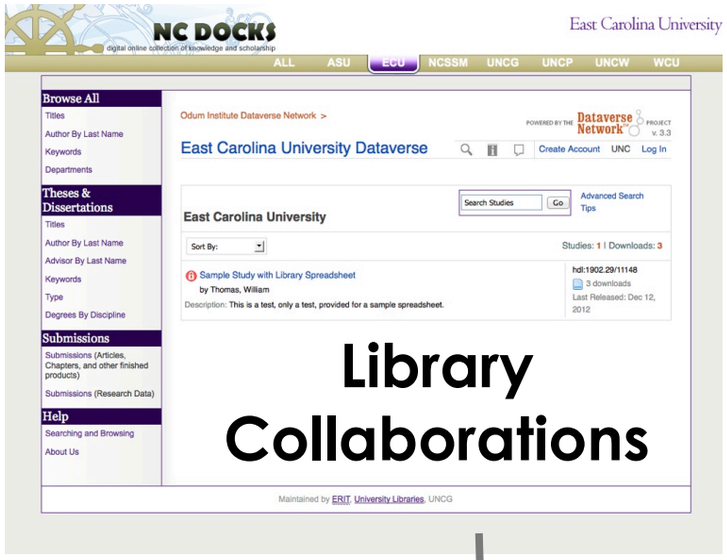

SURA


The
Dataverse
Network™
Project
(Multi-Disciplinary Effort)


Harvard-Smithsonian
CfA
Center for Astrophysics
Exploring the Universe
POWERED BY THE Dataverse Network™ PROJECT v. 3.3
Astronomy Dataverse Network
Create Account Log In
This is the Astronomy data repository for Harvard affiliates. Administration and support is provided by the Harvard-Smithsonian Center for Astrophysics (CfA) in collaboration with Harvard Library (HL) and the Institute for Quantitative Social Science (IQSS). Infrastructure is provided by Harvard University Information Technology Services.
The Astronomy Dataverse Network plays an important role in fulfilling your Data Management Plan requirements (e.g. as mandated by NSF), and for providing data re-use and citation opportunities. Find out more about our team by exploring the Seamless Astronomy and Wolbach Library teams at the CfA. We are also connecting the Astronomy Dataverse to the indexing services provided ...more >>
CREATE A DATAVERSE
Create a Dataverse to upload your own data sets and create collections of data.
Released Dataverses
Search Studies Go
Advanced Search Tips
ALL # A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Dataverses: 10 | Studies: 77 | Files: 565
Name
Affiliation
Released
Activity
Matthew Schneps View Info [+] Harvard-Smithsonian Center for Astrophysics Mar 9, 2013
AstroCitations View Info [+] Harvard University Mar 5, 2013
Hans Moritz Günther Harvard-Smithsonian Center for Astrophysics Feb 14, 2013
Dust Lane Spheroidal Galaxies View Info [+] Harvard University Nov 10, 2012
CfA Library Datasets View Info [+] Harvard-Smithsonian Center for Astrophysics Aug 17, 2012
theastrodata View Info [+] Harvard-Smithsonian Center for Astrophysics Apr 2, 2012
Soderberg, Alicia Harvard University Feb 6, 2012

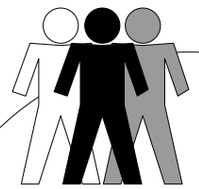

USAID
MEASURE Evaluation
MEASURE Evaluation
Odum Institute Dataverse Network >
POWERED BY THE Dataverse Network™ PROJECT v. 3.3
MEASURE Evaluation Dataverse
networkAdmin Admin Log Out
MEASURE Evaluation is the USAID Global Health Bureau's primary vehicle for supporting improvements in monitoring and evaluation in population, health and nutrition worldwide. We help to identify data needs, collect and analyze technically sound data, and use that data for health decision making. Some MEASURE Evaluation activities involve the collection of innovative evaluation data sets in order to increase the evidence-base on program impact and evaluate the strengths and weaknesses of recent evaluation methodological developments. Many of these data sets may be available to other researchers to answer questions of particular importance to global health and evaluation research. Some of these data sets are being added to the Dataverse on a rolling basis, as they become available.
MEASURE Evaluation
COMPASS
PLACE Studies
MEASURE Evaluation
Search Studies Go
Advanced Search Tips
within this collection
Sort By:
Studies: 20 | Downloads: 42
1 2
Burkina Faso PLACE Study
hdl:1902.29/10701
3 downloads + analyses
Last Released: Apr 19, 2011
Haiti PLACE Study
hdl:1902.29/10772
0 downloads + analyses
Last Released: Jul 12, 2011
Jamaica PLACE Study 2003
hdl:1902.29/10699
0 downloads
Last Released: Apr 19, 2011
Kazakhstan PLACE Study 2002
hdl:1902.29/10692
0 downloads + analyses
Last Released: Jul 12, 2011
hdl:1902.29/10694
0 downloads


OPPERA
Orofacial Pain: Prospective Evaluation and Risk Assessment
A Study of Muscle, Joint and Jaw Function
Odum Institute Dataverse Network >
POWERED BY THE Dataverse Network™ PROJECT v. 3.3
OPPERA Dataverse
networkAdmin Admin Log Out
The OPPERA Dataverse houses project data sets and files for the OPPERA (Orofacial Pain: Prospective Evaluation and Risk Assessment) Study.
OPPERA
Case Control Study Files
Prospective Cohort Study Files
OPPERA
Search Studies Go
Advanced Search Tips
within this collection
Sort By:
Studies: 6 | Downloads: 6
OPPERA Baseline Profile of Mood States Bipolar (POMS), Case-Control Study of Chronic TMD
by Fillingim, Roger B.; Ohrbach, Richard; Greenspan, Joel D.; Knott, Charles; Dubner, Ronald; Bair, Eric; Baraian, Cristina; Slade, Gary D.; Maixner, William
Description: This instrument consists of 72 mood-related items, and participants indicate the extent to which each item describes their mood state over the past week, including today using a 4-category scale (much unlike this, slightly unlike this, sli...Continue [+]
hdl:1902.29/11183
0 downloads + analyses
Last Released: Feb 25, 2013
OPPERA Baseline Short Form 12 Version 2 (SF-12 v2), Case-Control Study of Chronic TMD
by Fillingim, Roger B.; Ohrbach, Richard; Greenspan, Joel D.; Knott, Charles; Dubner, Ronald; Bair, Eric; Baraian, Cristina; Slade, Gary D.; Maixner, William
Description: The SF-12 is a briefer measure of healthrelated quality of life derived by using 12 items from the original SF-36 instrument. The SF-12 yields 2 scale scores, A Mental Health Composite Score (MCS) and a Physical Health Composite Score (PCS...Continue [+]
hdl:1902.29/11185
0 downloads + analyses
Last Released: Feb 25, 2013
OPPERA Baseline Kohn Reactivity Scale (KOHN), Case-Control Study of Chronic TMD
by Fillingim, Roger B.; Ohrbach, Richard; Greenspan, Joel D.; Knott, Charles; Dubner, Ronald; Bair, Eric; Baraian, Cristina; Slade, Gary D.; Maixner, William
Description: This instrument consists of 24 items that assess an individual's level of reactivity or central nervous system arousability. Individuals respond to each item on a 5-category scale ranging from "disagree strongly" to "agree strongly." T...Continue [+]
hdl:1902.29/11176
2 downloads + analyses
Last Released: Feb 25, 2013
OPPERA Pennebaker Inventory of Limbic Languidness (PILL), Case-Control Study of Chronic TMD
by Fillingim, Roger B.; Ohrback, Richard; Greenspan, Joel D.; Knott, Charles; Dubner, Ronald; Bair, Eric; Baraian, Cristina; Slade, Gary D.; Maixner, Willaim
hdl:1902.29/11166
2 downloads + analyses
Last Released: Feb 15, 2013



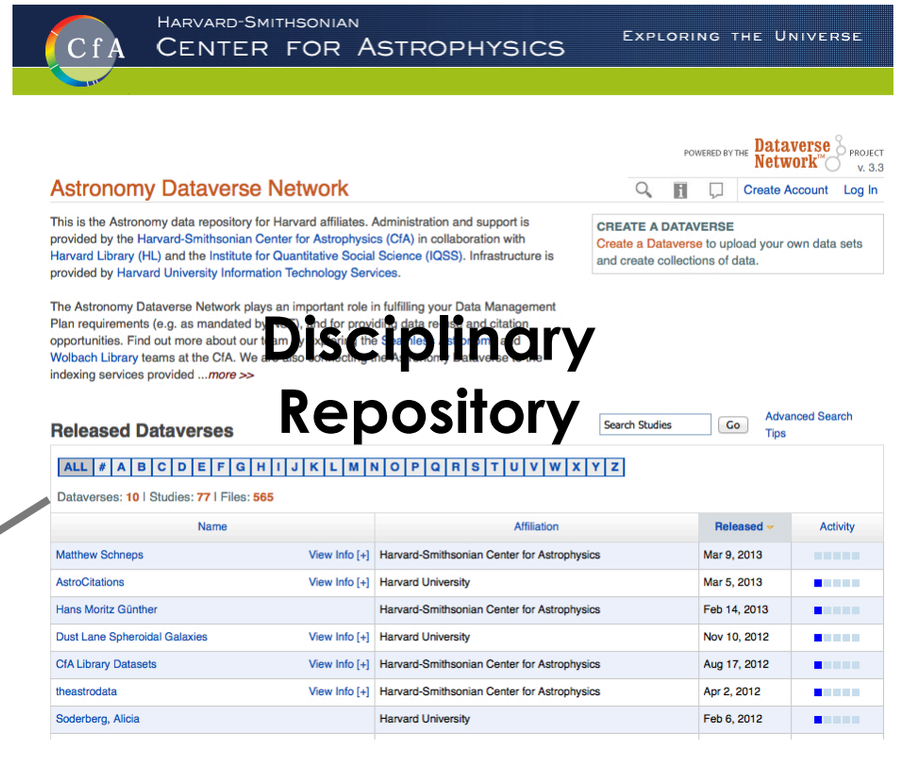
Library Collaborations

Research Collaborations

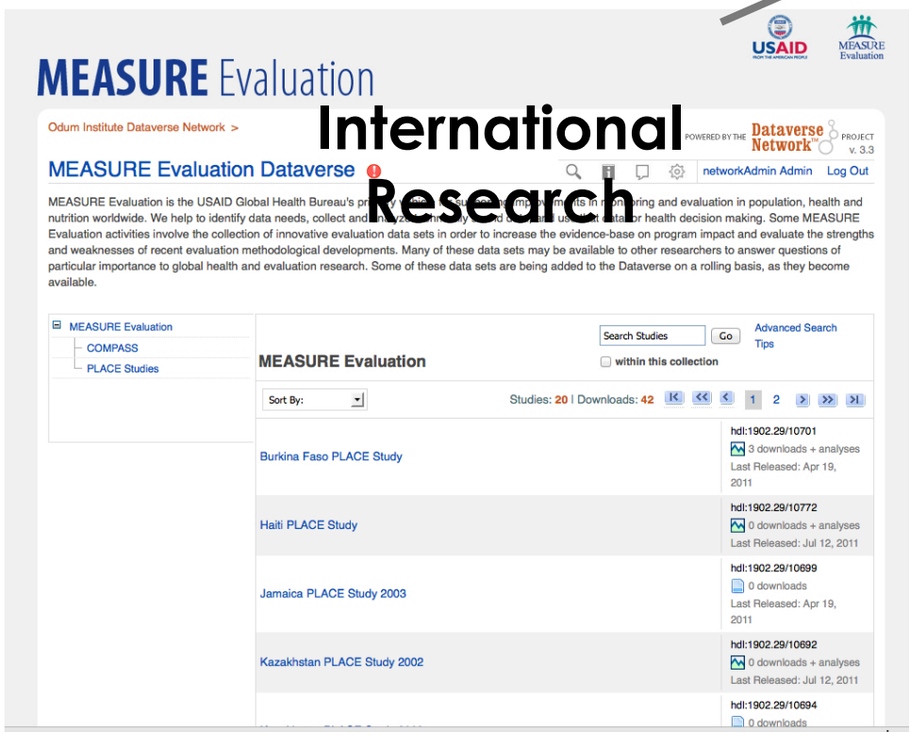


The Dataverse Network™ Project

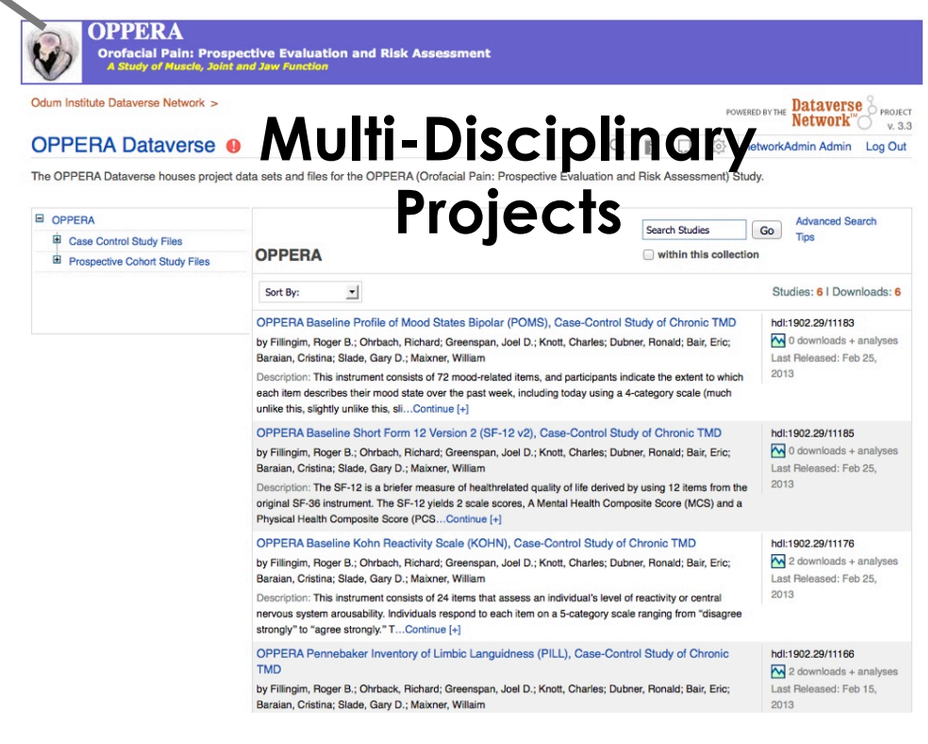
(Multi-Disciplinary Effort)



Disciplinary Repository



International Research



Multi-Disciplinary Projects

Why Dataverse?

Online Analysis

Discoverability

Ingest Tools

Distributed Preservation

Enable Data Sharing

Data Curation

Preservation Auditing

Standardization

Next Steps

Integration with CyberInfrastructure

eirods-consortium.org

DFC DataNet FEDERATION CONSORTIUM

NSF

Big Data: DataBridge

A Sociometric System for Long-Tail Science Data Collections

Secure Data Management and Auditing

Auditing Data Management Plans

SafeArchive

DMPTool
Guidance and Resources for your Data Management Plan

Secure Data Management and Delivery

Synthetic data

i·R·O·D·S

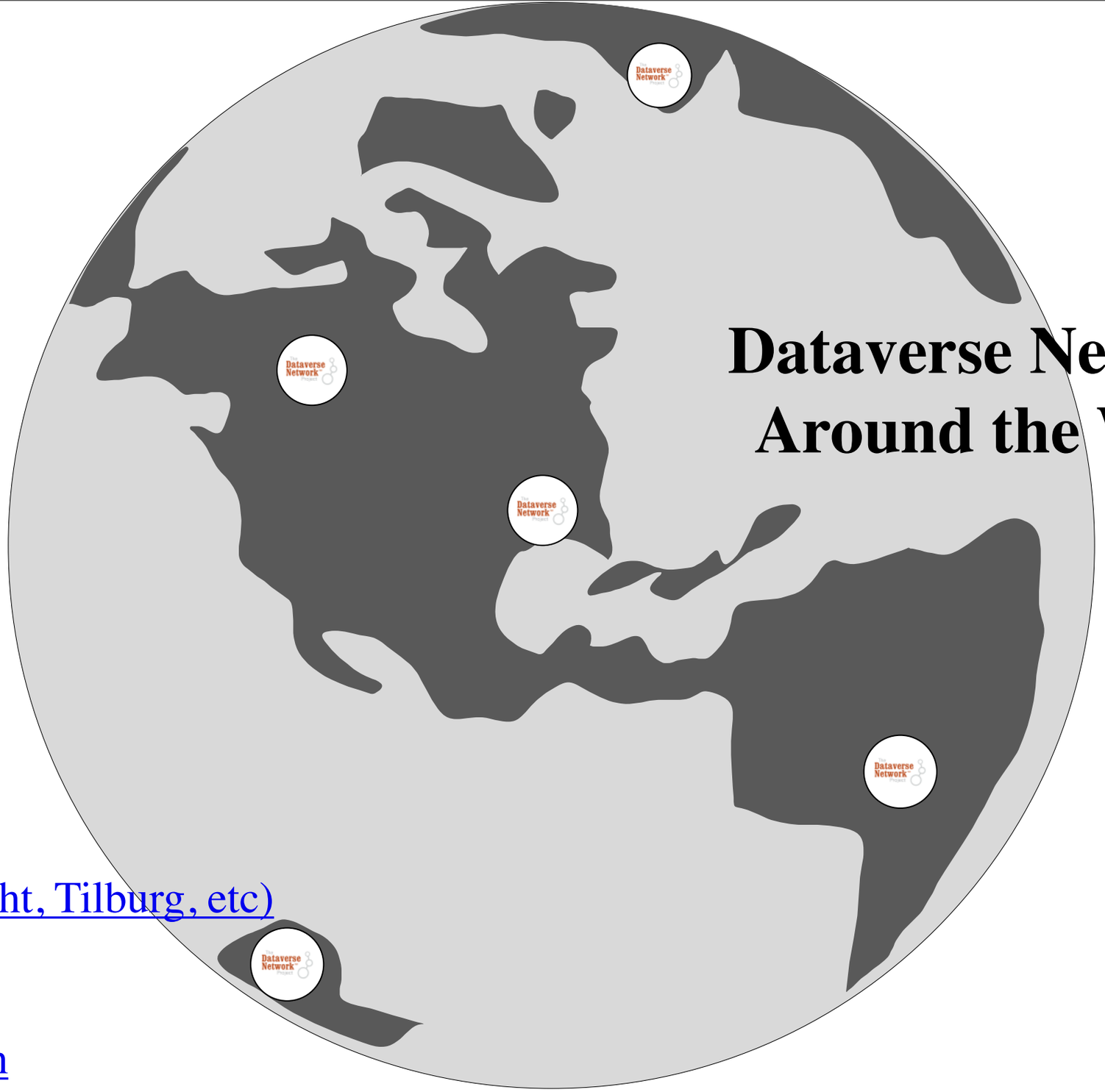
Differential privacy

Dataverse Networks at Harvard (collaboration between IQSS, Harvard Library and University IT):

- IQSS Dataverse Network:
- Astronomy Dataverse Network:

Dataverse Networks at UNC Chapel Hill:

- Odum Institute Dataverse Network:
- NC DOCKS Dataverse
- SURA Pilot Project Dataverse



Dataverse Networks Around the World

- Odum Institute
- Dutch Universities (Utrecht, Tilburg, etc)
- UC Davis
- UC San Diego
- University of Copenhagen
- Department of Cross Cultural and Regional Studies, University of Copenhagen
- European Union Democracy Observatory (Beta version)
- Ontario Council of University Libraries
- Thailand

DATA MANAGEMENT CHALLENGES

- Funding agencies and journal publishers are placing additional demands on researchers for data sharing and access
- Data-intensive science has enabled collaborative research that crosses disciplinary and institutional boundaries
- Researchers do not have access to or are unaware of tools and services to help manage their data
- Discipline specific customization to tools

Thank You

Jonathan_Crabtree@unc.edu

mcrosas@iq.harvard.edu

http://thedata.org

http://www.odum.unc.edu