

The Data Management Plan with Dataverse

Mercè Crosas, Ph.D.
Director of Product Development



The
Dataverse

The Data
Management
Plan

The Data
Management
Plan
with Dataverse

The
Dataverse

The Data
Management
Plan

The Data
Management
Plan
with Dataverse

The Intellectual Origin of the Dataverse Network

"The replication standard holds that sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party can replicate the results without any additional information from the author."

- ▶ King, Gary. 1995 "Replication, Replication"
- ▶ Altman, Micah, King, Gary. 2007 "A Proposed Standard for the Scholarly Citation of Quantitative Data"

What You Need to Make it Work

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)
- ✓ Conversion to archival formats
- ✓ Extraction of Metadata from data sets
- ✓ Metadata standards (DDI, Dublin Core)
- ✓ Inter-operability (OAI, APIs)

Sidney Verba
Carl H. Pforzheimer University Professor
Department of Government, Harvard University (email)

BIO PUBLICATIONS DATA

CDD-Hewlett Data Center
A Center Committed To The Provision Of Objective, Verifiable Data For Socio-Economic Development Of Ghana

ABOUT US PEOPLE ANNOUNCEMENTS DATA CALENDAR CONTACT

Ghana Center for Democratic Development Dataverse

Search Studies [Go] Advanced Search Tips

Sort By: [v] Studies: 23 [K] [v] [1] [2] [3] [v] [v]

April 2012

1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 1 2 3 4 5

Ghana: Local Public Finance and Decentralization, 1994-2004
by Tewodaj Mogues
Abstract: The purpose of this data compilation effort was to use the

hdi:1902.1/17846
79 downloads + analyses
Last Released: Mar 19, 2012

Sidney Verba Dataverse

Search Studies [Go] Advanced Search Tips

Sort By: [v] Studies: 11 [K] [v] [1] [2] [v] [v]

Political Participation in America, 1967
by Verba, Sidney; Nie, Norman H.
Abstract: The focus of this study is individual political participation, both at the local and national levels. Respondents were asked to describe the most important recent changes and problems within their communities, how ...

hdi:1902.2/7015
1 download/analysis
Last Released: Feb 9, 2011

Public Opinion and the War in Vietnam Study, 1966
by Verba, Sidney; Brody, Richard et al.
Abstract: This study, conducted by the National Opinion Research Center,

hdi:1902.2/7295
0 downloads + analyses
Last Released: Feb 9, 2011

James M. Snyder, Jr.
Professor of Government

James Snyder Dataverse

Search Studies [Go] Advanced Search Tips

Sort By: [v] Studies: [v]

Replication data for: Electoral Selection with Parties and Primaries
by James M. Snyder, Jr.; Michael M. Ting
Abstract: This data collection contains information on state legislative election returns from 1967 through 2003. Each row represents an individual candidate who ran for state legislative ...

hdi:1902.1/15624
114 downloads + analyses
Last Released: Apr 11, 2011

State Legislative Election Returns, 1967-2003
by Carsey, Thomas M.; Berry, William D.; Miami, Richard G.; Powell, Lynda W.; Snyder, James M.
Abstract: This data collection contains information on state legislative election returns from 1967 through 2003. Each row represents an individual candidate who ran for state legislative ...

hdi:1902.2/21480
10 downloads
Last Released: Feb 8, 2011

An Empirical Investigation of the Dynamics of PAC Contributors
by Romer, Thomas; Snyder, James M. Jr.
Abstract: These data and computer programs are part of ICPSR's Publication-Related Archive and are distributed exactly as they arrived from the data depositor. ICPSR has not checked or processed this material. Users ...

hdi:1902.2/1092
0 downloads
Last Released: Feb 8, 2011

Dataverse Network
(centralized repository)

Metadata Data Files

ajps American Journal of Political Science

American Journal of Political Science (AJPS) Dataverse

Search Studies [Go] Advanced Search Tips

Sort By: [v] Studies: 1

Replication data for: Poverty and Support for Militant Politics: Evidence from Pakistan
by Graeme Blair; C. Christine Fair; Neil Malhotra; Jacob N. Shapiro
Abstract: Policy debates on strategies to end extremist violence frequently cite poverty as a root cause of support for the perpetrating groups. There is little evidence to support this contention, particularly in the Pakistani case. Pakistan's urba...Continue [+]

hdi:1902.1/17042
42 downloads
Last Released: Apr 5, 2012

Henry A. Murray Research Archive
at Harvard University
HOME POLICIES CONTACT US

Murray Research Archive Original Collection Dataverse

Search Studies [Go] Advanced Search Tips

Sort By: [v] Studies: 440 [K] [v] [1] [2] [3] [4] [v] [v]

Henry A. Murray Research Archive Original Collection

Economic Theory and Demography

Terman Life Cycle Study of Children with High Ability, 1922-1986
by Louis M. Terman; Robert R. Sears; Lee Cronbach; Pauline S. Sears; Albert Hastorf
Abstract: This study began by comparing a group of children with high intelligence quotients with groups of children typical of the general population, to discover similarities and differences. Research was continued ...

hdi:1902.1/00882
188 downloads + analyses
Last Released: Mar 21, 2012

HARVARD UNIVERSITY DEPARTMENT OF GOVERNMENT

Government Department Dissertations Dataverse

Search Studies [Go] Advanced Search Tips

Sort By: [v] Studies: 34 [K] [v] [1] [2] [3] [4] [v] [v]

Replication data for: Just Leave it to the Courts: How, When, and Why Congress Abdicates the Legislative Power
by Nancy Ellen Billica
Abstract: What happens to a bill once signed into law? While many details of the legislative process have been subject to exhaustive study, scholars of American politics have done surprisingly little systematic tracking of statutes beyond enactment...Continue [+]

hdi:1902.1/Z/GFMDJGCB
59 downloads + analyses
Last Released: Feb 13, 2012

Replication data for: The Micropolitics of Realism: A General Equilibrium Approach to Security Maximization
by William Kent Griffiths
Abstract: The argument extends the neo-realist rethinking of realism, with particular attention to the nature of the actors in world politics, the economic substructure of power and security, and the rigor with which assumptions are specified and co...Continue [+]

hdi:1902.1/AWRKCKFCBV
112 downloads
Last Released: Feb 13, 2012

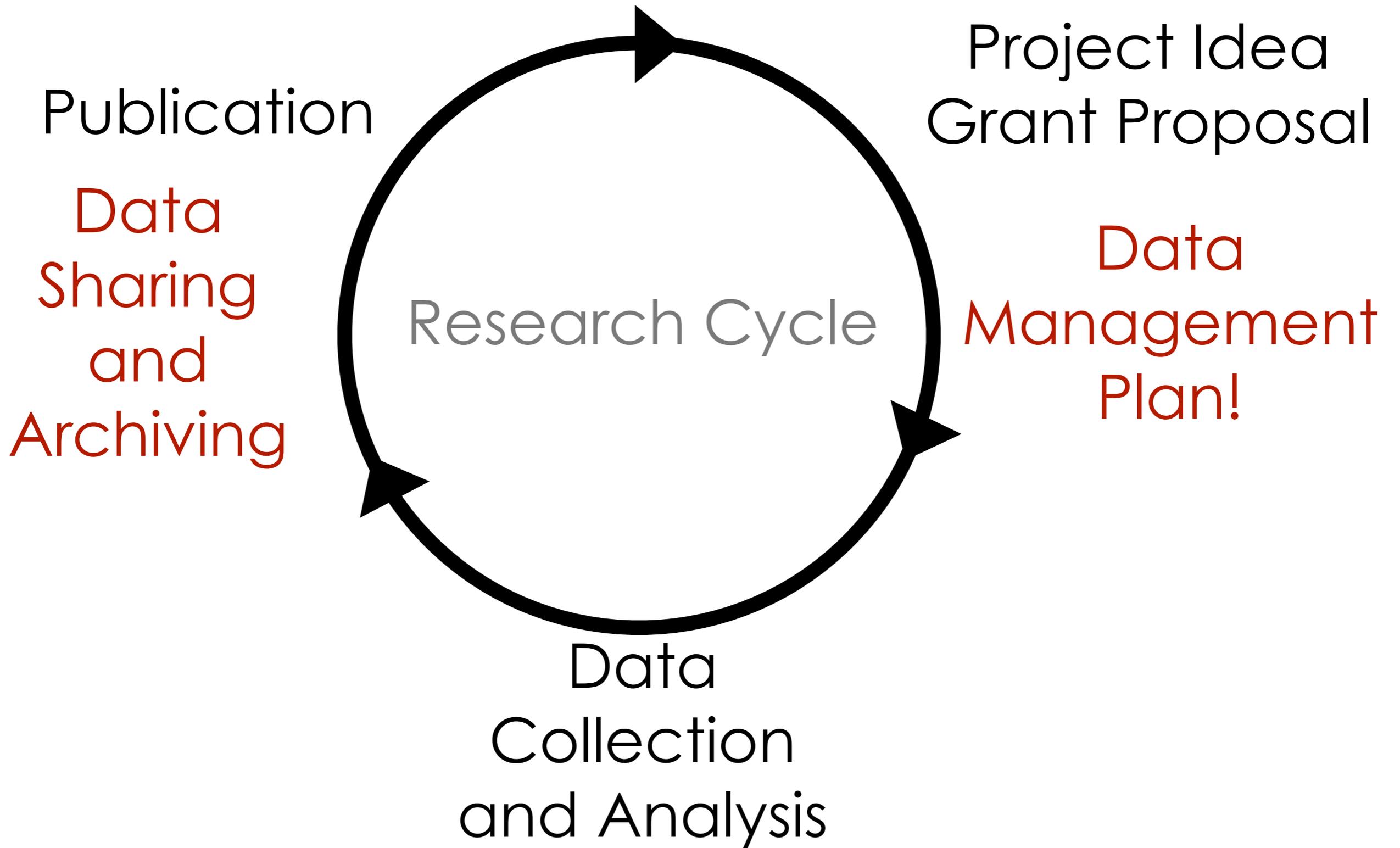
Replication data for: Taking Charge: Black Electoral Success and the Redefinition

hdi:1902.1/KWZMCJEKES

The
Dataverse

The Data
Management
Plan

The Data
Management
Plan
with Dataverse



Use Case: James Snyder's NSF proposal



James M. Snyder, Jr. Professor of Government

1737 Cambridge St. Cambridge, MA 02138, 617-496-1089, jsnyder@gov.harvard.edu (*email*)



Bio

Professor Snyder's primary research and teaching interests are in American politics, with a focus on political representation. He has written on a variety of topics, including elections, campaign finance, legislative behavior and institutions, interest groups, direct democracy, the media, and corruption. He is a Research Associate at the National Bureau of Economic Research, and a Fellow of the American Academy of Arts and Sciences.

Class

His articles have appeared in the *American Political Science Review*, the *American Journal of Political Science*, the *Journal of Politics*, the *American Economic Review*, the *Journal of Political Economy*, *Econometrica*, and many other journals and edited volumes. He is co-author of *The End of Inequality: One Person, One Vote and the Transformation of American Politics*. Professor Snyder taught for six years in the Department of Economics at the University of Chicago, and for eighteen years in the Departments of Political Science and Economics at the Massachusetts Institute of Technology.

Publications

Working Papers

Book Chapters

Dataverse

 [Admin](#)

Checklist for generic NSF Data Management Plan

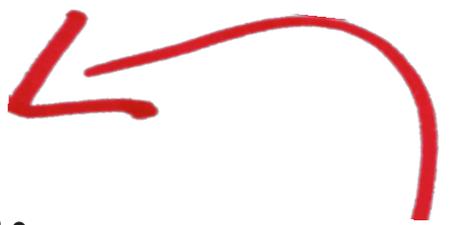
Data Description	What data will be generated? How will you create the data? (simulated, observed, experimental, software, physical collections)
Existing Data	Will you be using existing data? Relationship between the data you are collecting and existing data.
Audience	Who will potentially use the data?
Access and Sharing	How will data files be shared? How will others access them?
Formats	What data formats will you be creating?
Metadata and Documentation	What documentation will you provide to describe the data? Metadata formats and standards?
Storage, backup, replication, versioning	Are the data files backed up regularly? Are there replicas in different locations? Are older versions of the data kept?
Security	Are the system and storage that will be used secure?
Budget	Any costs for preparing the data? Costs for storage and long-term access?
Privacy, Intellectual Property	Does the data contain private or confidential information? Any copyrights?
Archiving, Preservation, Long-term Access	What plans do you have to archive the data and other research products? Will it have long-term accessibility?
Adherence	How will you check for adherence of this plan?

Data Description	What data will be generated? How will you create the data? (simulated, observed, experimental, software, physical collections)
Existing Data	Will you be using existing data? Relationship between the data you are collecting and existing data.
Audience	Who will potentially use the data?
Access and Sharing	How will the data be shared? How will others access them?
Formats	What data formats will you be creating?
Metadata and Documentation	What documentation will you provide to describe the data? Metadata formats and standards?
Storage, backup, replication, versioning	Are the data files backed up regularly? Are there replicas in different locations? Are older versions of the data kept?
Security	Are the system and storage that will be used secure?
Budget	Any costs for preparing the data? Costs for storage and long-term access?
Privacy, Intellectual Property	Does the data contain private or confidential information? Any copyrights?
Archiving, Preservation, Long-term Access	What plans do you have to archive the data and other research products? Will it have long-term accessibility?
Adherence	How will you check for adherence of this plan?

The Political Economy of U.S. State Courts: The Influence of Media and Selection Systems

James M. Snyder, Jr., Harvard University

Claire S. H. Lim, *Stanford University*



Grant
Proposal

Data Collected by this project will include:

- 1) collection of detailed national scale data set on press coverage of the U.S. state courts,
- 2) collection of data on election of judges in trial, appellate, and supreme courts in all states, and
- 3) construction of data set on county composition of judicial districts of trial courts in all states.

Data Description	What data will be generated? How will you create the data? (simulated, observed, experimental, software, physical collections)
Existing Data	Will you be using existing data? Relationship between the data you are collecting and existing data.
Audience	Who will potentially use the data?
Access and Sharing	How will data files be shared? How will others access them?
Formats	What data formats will you be creating?
Metadata and Documentation	What documentation will you provide to describe the data? Metadata formats and standards?
Storage, backup, replication, versioning	Are the data files backed up regularly? Are there replicas in different locations? Are older versions of the data kept?
Security	Are the system and storage that will be used secure?
Budget	Any costs for preparing the data? Costs for storage and long-term access?
Privacy, Intellectual Property	Does the data contain private or confidential information? Any copyrights?
Archiving, Preservation, Long-term Access	What plans do you have to archive the data and other research products? Will it have long-term accessibility?
Adherence	How will you check for adherence of this plan?



James M. Snyder, Jr.
Professor of Government

1737 Cambridge St. Cambridge, MA 02138, 617-496-1089, jsnyder@gov.harvard.edu (*email*)



IQSS Dataverse Network >

James Snyder
Dataverse

POWERED BY THE **Dataverse Network™** PI

Bio

Publications

James Snyder

[Advanced Search Tips](#)

- All data collected or generated will be deposited in the researcher's Dataverse.
- The Dataverse allows researchers to deposit data in an organized, well curated and citable network... ultimately facilitating access and sharing.

Data Description	What data will be generated? How will you create the data? (simulated, observed, experimental, software, physical collections)
Existing Data	Will you be using existing data? Relationship between the data you are collecting and existing data.
Audience	Who will potentially use the data?
Access and Sharing	How will data files be shared? How will others access them?
Formats	What data formats will you be creating?
Metadata and Documentation	What documentation will you provide to describe the data? Metadata formats and standards?
Storage, backup, replication, versioning	Are the data files backed up regularly? Are there replicas in different locations? Are older versions of the data kept?
Security	Are the system and storage that will be used secure?
Budget	Any costs for preparing the data? Costs for storage and long-term access?
Privacy, Intellectual Property	Does the data contain private or confidential information? Any copyrights?
Archiving, Preservation, Long-term Access	What plans do you have to archive the data and other research products? Will it have long-term accessibility?
Adherence	How will you check for adherence of this plan?



James M. Snyder, Jr.

Professor of Government

1737 Cambridge St. Cambridge, MA 02138, 617-496-1089, jsnyder@gov.harvard.edu (email)



- Quantitative data will be deposited either in SPSS, Stata, CSV, Tab delimited, or GraphML.
- Images in JPEG-2000 or TIFF. [good practice]
- Audio in MP3 or WAVE. [good practice]
- Dataverse accepts all data formats to accommodate the flexibility researchers need.

Choose a Data Type

- ✓ Tabular Data
 - SPSS/POR
 - SPSS/SAV
 - Stata
 - CSV (w/SPSS card)
 - TAB (w/DDI)
- Network Data
 - GraphML
- Other

Tabular and Network Data files can be subset and analyzed using the Dataverse Network analysis tools. These files will take longer to upload and you'll get a notification once the upload is completed. Tabular files will also get Universal Numerical Fingerprint (UNF). All other files types will be available for download only.

Category	File Name	Description	Size (bytes)	Remove
----------	-----------	-------------	--------------	--------

Save Cancel

Data Description	What data will be generated? How will you create the data? (simulated, observed, experimental, software, physical collections)
Existing Data	Will you be using existing data? Relationship between the data you are collecting and existing data.
Audience	Who will potentially use the data?
Access and Sharing	How will data files be shared? How will others access them?
Formats	What data formats will you be creating?
Metadata and Documentation	What documentation will you provide to describe the data? Metadata formats and standards?
Storage, backup, replication, versioning	Are the data files backed up regularly? Are there replicas in different locations? Are older versions of the data kept?
Security	Are the system and storage that will be used secure?
Budget	Any costs for preparing the data? Costs for storage and long-term access?
Privacy, Intellectual Property	Does the data contain private or confidential information? Any copyrights?
Archiving, Preservation, Long-term Access	What plans do you have to archive the data and other research products? Will it have long-term accessibility?
Adherence	How will you check for adherence of this plan?



- Study metadata will be entered in the Dataverse Cataloging Information page which provides ~ 100 fields to choose from, plus custom fields.
- Basic metadata fields are: **Title, identifier, year, author, abstract, keywords.**
- Additional documentation will be uploaded in PDF or plain text formats. Code can be uploaded too.
- A formal Data Citation will be generated automatically.
- Metadata will be exported into XML (DDI, DC).

Michael M. Ting
(FirstName LastName)

Department of Political Science and SIPA Columbi.

Producer *

Producer

Data Description	What data will be generated? How will you create the data? (simulated, observed, experimental, software, physical collections)
Existing Data	Will you be using existing data? Relationship between the data you are collecting and existing data.
Audience	Who will potentially use the data?
Access and Sharing	How will data files be shared? How will others access them?
Formats	What data formats will you be creating?
Metadata and Documentation	What documentation will you provide to describe the data? Metadata formats and standards?
Storage, backup, replication, versioning	Are the data files backed up regularly? Are there replicas in different locations? Are older versions of the data kept?
Security	Are the system and storage that will be used secure?
Budget	Any costs for preparing the data? Costs for storage and long-term access?
Privacy, Intellectual Property	Does the data contain private or confidential information? Any copyrights?
Archiving, Preservation, Long-term Access	What plans do you have to archive the data and other research products? Will it have long-term accessibility?
Adherence	How will you check for adherence of this plan?



James M. Snyder, Jr.

- The Dataverse will keep multiple versions of the data.
- Deposited data will never be destroyed (unless legally required).
- In addition, the Dataverse Network at IQSS provides system backups in a daily basis.

The screenshot shows the Dataverse interface with a sidebar on the left labeled "Dataverse" and a main content area with tabs for "Cataloging Information", "Documentation, Data and Analysis", "User Comments", and "Versions". The "Versions" tab is active, displaying a "Version History" table with the following data:

<input type="checkbox"/>	Version	Status	Comments	Released	Contributors
<input type="checkbox"/>	3	Released		Mon Apr 11 15:12:13 EDT 2011	IQSS Admin
<input type="checkbox"/>	2	Archived		Tue Apr 05 18:29:06 EDT 2011	IQSS Admin
<input type="checkbox"/>	1	Archived		Tue Mar 08 09:11:00 EST 2011	Helen Lewis

Data Description	What data will be generated? How will you create the data? (simulated, observed, experimental, software, physical collections)
Existing Data	Will you be using existing data? Relationship between the data you are collecting and existing data.
Audience	Who will potentially use the data?
Access and Sharing	How will data files be shared? How will others access them?
Formats	What data formats will you be creating?
Metadata and Documentation	What documentation will you provide to describe the data? Metadata formats and standards?
Storage, backup, replication, versioning	Are the data files backed up regularly? Are there replicas in different locations? Are older versions of the data kept?
Security	Are the system and storage that will be used secure?
Budget	Any costs for preparing the data? Costs for storage and long-term access?
Privacy, Intellectual Property	Does the data contain private or confidential information? Any copyrights?
Archiving, Preservation, Long-term Access	What plans do you have to archive the data and other research products? Will it have long-term accessibility?
Adherence	How will you check for adherence of this plan?



developing
SECURE
technology
SOLUTIONS



ENTERPRISE SECURITY POLICY

RESEARCH DATA SECURITY POLICY

Welcome

- The Dataverse Network at IQSS follows “good computer use practices” set by the Security & Privacy group at Harvard.

CONTACT US

Jump

Go

requirements and policies that support these commitments and works to communicate the policies to the University community. Each Harvard School is responsible for implementing these University guidelines and for developing local policies, where needed, to facilitate a secure environment that is consistent with University requirements.

Harvard community members-- student, staff, and faculty-- encounter sensitive information every day, information such as student grades and evaluations, staff evaluations, credit card numbers, bank accounts, salaries, and personal information including home addresses for example. This information is considered confidential by the University and by the person the information is about.

Data Description	What data will be generated? How will you create the data? (simulated, observed, experimental, software, physical collections)
Existing Data	Will you be using existing data? Relationship between the data you are collecting and existing data.
Audience	Who will potentially use the data?
Access and Sharing	How will data files be shared? How will others access them?
Formats	What data formats will you be creating?
Metadata and Documentation	What documentation will you provide to describe the data? Metadata formats and standards?
Storage, backup, replication, versioning	Are the data files backed up regularly? Are there replicas in different locations? Are older versions of the data kept?
Security	Are the system and storage that will be used secure?
Budget	Any costs for preparing the data? Costs for storage and long-term access?
Privacy, Intellectual Property	Does the data contain private or confidential information? Any copyrights?
Archiving, Preservation, Long-term Access	What plans do you have to archive the data and other research products? Will it have long-term accessibility?
Adherence	How will you check for adherence of this plan?

IQSS Dataverse Network

Access the world's largest collection of social science research data here by searching across or browsing through one of the virtual data archives - called "dataverses" - listed

CREATE A DATAVERSE

Create a Dataverse to upload your own data sets

- The Dataverse Network at IQSS is free and open to all social science research data.
- Restrictions are 2GB per file, with no limit in the number of files.

(In the future, a fee might apply to archive very large collections - currently under review)

Research Projects
Scholars

University of the Thai Chamber of Commerce
[View Info \[+\]](#)

University

Apr 9, 2012



American Journal of Political Science (AJPS)

Rice University

Apr 4, 2012



Data Description	What data will be generated? How will you create the data? (simulated, observed, experimental, software, physical collections)
Existing Data	Will you be using existing data? Relationship between the data you are collecting and existing data.
Audience	Who will potentially use the data?
Access and Sharing	How will data files be shared? How will others access them?
Formats	What data formats will you be creating?
Metadata and Documentation	What documentation will you provide to describe the data? Metadata formats and standards?
Storage, backup, replication, versioning	Are the data files backed up regularly? Are there replicas in different locations? Are older versions of the data kept?
Security	Are the system and storage that will be used secure?
Budget	Any costs for preparing the data? Costs for storage and long-term access?
Privacy, Intellectual Property	Does the data contain private or confidential information? Any copyrights?
Archiving, Preservation, Long-term Access	What plans do you have to archive the data and other research products? Will it have long-term accessibility?
Adherence	How will you check for adherence of this plan?



James M. Snyder, Jr.

- Data uploaded to the Dataverse cannot contain confidential information.
- Researcher agrees to the terms of use upon uploading the data, which states that:
 - You give permission and any required licenses to IQSS and the Archive to store and backup the materials
 - The Materials do not infringe upon the copyrights or other intellectual property rights, ..
 - If human subjects were studied in the collection of the Materials, you collected the Materials with IRB approval
 - The Materials do not contain high-risk confidential information ...
 - You give permission and any required licenses to IQSS to make the Content available for archiving, preservation and access, within the Data Preservation Alliance for the Social Sciences ("Data-PASS")

Data Description	What data will be generated? How will you create the data? (simulated, observed, experimental, software, physical collections)
Existing Data	Will you be using existing data? Relationship between the data you are collecting and existing data.
Audience	Who will potentially use the data?
Access and Sharing	How will data files be shared? How will others access them?
Formats	What data formats will you be creating?
Metadata and Documentation	What documentation will you provide to describe the data? Metadata formats and standards?
Storage, backup, replication, versioning	Are the data files backed up regularly? Are there replicas in different locations? Are older versions of the data kept?
Security	Are the system and storage that will be used secure?
Budget	Any costs for preparing the data? Costs for storage and long-term access?
Privacy, Intellectual Property	Does the data contain private or confidential information? Any copyrights?
Archiving, Preservation, Long-term Access	What plans do you have to archive the data and other research products? Will it have long-term accessibility?
Adherence	How will you check for adherence of this plan?



- The IQSS Dataverse Network commits to good archival practices:
 - Metadata is exported to XML
 - Data files are re-formatted for long term access
 - All versions are kept
 - Metadata and data are replicated to multiple locations through LOCKSS

prompted to save a single archive file. Study files that have restricted access will not be downloaded.

Select all files

Download All Selected Files



replication.codebook.pdf

Adobe PDF - 33 KB - 20 downloads

Download

This codebook describes the variables in the following datasets:
tables 1 2 replication.dta table 3 replication.dta

table_3_replication.tab

Tab Separated - 601 KB
- 9 downloads + analyses



Download as...
✓ Tab Delimited
Original File
Splus
Stata
R

replication data tables

TABULAR DATA 16156 Cases 9 Variables

View Data Citation [+]

tables_1_2_replication.tab

Tab Separated - 2 KB - 6 downloads + analyses



Download as...

replication data tables

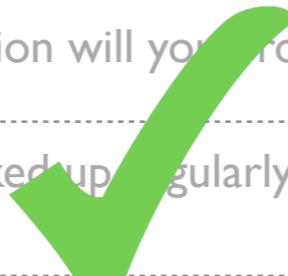
TABULAR DATA 48 Cases 6 Variables



Access Subset/Analysis

View Data Citation [+]

Data Description	What data will be generated? How will you create the data? (simulated, observed, experimental, software, physical collections)
Existing Data	Will you be using existing data? Relationship between the data you are collecting and existing data.
Audience	Who will potentially use the data?
Access and Sharing	How will data files be shared? How will others access them?
Formats	What data formats will you be creating?
Metadata and Documentation	What documentation will you provide to describe the data? Metadata formats and standards?
Storage, backup, replication, versioning	Are the data files backed up regularly? Are there replicas in different locations? Are older versions of the data kept?
Security	Are the system and storage that will be used secure?
Budget	Any costs for preparing the data? Costs for storage and long-term access?
Privacy, Intellectual Property	Does the data contain private or confidential information? Any copyrights?
Archiving, Preservation, Long-term Access	What plans do you have to archive the data and other research products? Will it have long-term accessibility?
Adherence	How will you check for adherence of this plan?



<http://thedata.org>

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

- IQSS Dataverse Network: Open to all Social Science research data
- Astronomy Dataverse Network: Open to the Harvard-Smithsonian Center for Astrophysics

Dataverse Networks in other institutions:

- The software is open-source and it's free to install in any institutions.

Thanks