How to Manage Your Data

Lisa Spiro

June 2019

This workshop draws heavily on materials from the <u>University of Minnesota Libraries</u>, <u>New</u> <u>England Collaborative Data Management Curriculum</u>, <u>MIT Libraries</u> & <u>DataOne</u>.

Quick Poll: Raise Your Hand If You Have Ever...

- Forgotten what you called a file and/or where you put it
- Discovered unnecessary duplicates, then struggled over which to keep
- Lost data due to hardware failure, lost devices, etc.

Objectives for This Session

- 1. Understand the importance of managing data.
- 2. Learn how to create a good data management plan.
- 3. Name and organize your files effectively.
- 4. Manage versions of files.
- 5. Create tidy data.
- 6. Document your data.
- 7. Know options for storing, backing up and archiving your data.

1.Why Managing Your Data Matters

What Is Research Data?

"recorded factual materials that are commonly accepted in the scientific community as necessary to validate research finding" (<u>NSF</u>)

Examples:

- Text, e.g. interview transcriptions & field notes
- Audio & video recordings
- Images
- Numbers, e.g. measurements

What is data management?

The process of storing, organizing, describing, preserving, and sharing data so that research results can be validated, data can be understood, and future use is facilitated.



https://www.dataone.org/best-practices

Why Is Managing Your Data Important?

- Keep track of your data, working more efficiently.
- Prevent data loss.
- Uphold standards of research integrity.
- Make it easier to share and re-use data.
- Meet funder, <u>university</u> & increasingly journal requirements.
- Be kind to Future You and your collaborators.

If the data you need still exists; If you found the data you need; If you understand the data you found; If you trust the data you understand; If you can use the data you trust; Someone did a good job of data management.

Rex Sanders, USGS

2. Plan



Typical Components of Data Management Plan (NSF)

- 1. the **types of data** and other materials to be produced in the course of the project;
- 2. the **standards** to be used for data and metadata format and content;
- 3. policies for **access & sharing** including provisions for appropriate protection of privacy, security, IP, etc.;
- 4. policies and provisions for **re-use**, **re-distribution**, and the production of derivatives; and
- 5. plans for **archiving** data, samples, and other research products, and for **preservation** of access to them.

Create a Data Management Plan Using DMP Tool

	Lea	m + Lisa Spiro + Languag
ice University		🔄 Lisa Spiro (Fondren Libi
y Dashboard Create plan Admin Features -		
OC demo		
roject Details Plan overview Write Plan Share Download		
expand all collapse all 0/5 answered		
+ Roles and responsibilities (0 / 1)		
+ Expected data (0 / 1)		
+ Period of data retention (0 / 1)		
+ Data format and dissemination (0 / 1)		
- Data storage and preservation of access (0 / 1)		-
The Data Management Plan should describe physical and cyber resources and facilities that will be used for the effective preservation and storage of research data. These can include third party facilities and repositories.	Guidance	Comments
B I ⊞ · I≣ - ∂ ⊞·	NSF	
	The DMP should descrit resources and facilities I effectively preserve and These can include third- repositories.	be physical and cyber hat will be used to store research data. party facilities and
Save	Consider the following:	
	 What is the long-t maintaining, cural data? Which archive/rep you identified as a 	erm strategy for Ing, and archiving the pository/database have a place to deposit data?

https://dmptool.org/

Key Principles for Data Management Planning

- 1. Investing time in organizing your data now will save you time later.
- 2. Be clear and consistent.
- 3. Work out your data management procedures with collaborators.
- 4. Document your procedures.
- 5. Understand that there is no one right way; it's what works for you and your collaborators.

3. Organize Your Data



How to Create a Hierarchical File System

- 1. Organize your files in a predictable, easy-to-sort way.
- 2. Use relevant categories to organize folders (e.g. Project/Site/Date).
- 3. Select a meaningful naming convention for folders.

Example of a Directory Structure

		1-Proposals
		2-Finance
	1-ProjectManagement	3-Reports
2-EthicsGovernance		1-EthicsApproval
	2-EthicsGovernance	2-ConsentForms
ProjectFolder		1-Inputs
ProjectFolder	3-ExperimentOne	2-Data
		3-DataAnalysis
		4-Outputs
	4-Dissemination	1-Presentations
		2-Publications
		3 Publicity

Nikola Vukovic

What to Avoid...







Overlapping categories



Another Option: Tags

Pros:

More flexible

Cons:

Harder to be consistent.

Social Science / Sociology / Urban spatial history Sponsored stokenewington Temperance toread tufnellpark UK news UNITED States urban geography urban history Urban Renewal Urbanisation victorian Victorian period Waste Water World news

• May lose tags in moving to different computers.

Systems that support tags:

-<u>Evernote</u> -<u>Windows 10</u> -<u>Gmail</u> -<u>Mac OS</u> -Zotero -Box

The Problem of File Names

A STORY TOLD IN FILE NAMES:			
Location: 😂 C:\user\research\data			~
Filename 🔺	Date Modified	Size	Туре
 data_2010.05.28_test.dat data_2010.05.28_re-test.dat data_2010.05.28_re-re-test.dat data_2010.05.28_calibrate.dat data_2010.05.28_huh??.dat data_2010.05.29_more.dat data_2010.05.29_#\$@*&!!.dat data_2010.05.29_rap.dat data_2010.05.29_notbad.dat data_2010.05.29_woohoo!!.dat data_2010.05.29_USETHISONE.dat analysis_graphs.xls ThesisOutline!.doc Notes_Meeting_with_ProfSmith.txt DUNK 	3:37 PM 5/28/2010 4:29 PM 5/28/2010 5:43 PM 5/28/2010 7:17 PM 5/28/2010 7:20 PM 5/28/2010 9:58 PM 5/28/2010 12:37 AM 5/29/2010 3:22 AM 5/29/2010 4:16 AM 5/29/2010 4:47 AM 5/29/2010 5:08 AM 5/29/2010 7:13 AM 5/29/2010 7:26 AM 5/29/2010 11:38 AM 5/29/2010 2:45 PM 5/29/2010	420 KB 421 KB 420 KB 1,256 KB 30 KB 30 KB 30 KB 437 KB 670 KB 1,349 KB 2,894 KB 455 KB 38 KB 1,673 KB	DAT file DAT file
🚦 data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file
<			>
Type: Ph.D Thesis Modified: too many times	Copyright: Jorge Cham	www.phdo	comics.com

http://phdcomics.com/comics.php?f=1323

Principles for Effective File Naming

• Files are **distinguishable** from each other within their containing folder.

• Files are easy to locate, browse and sort.

• If files are moved to another storage platform, their names will retain **useful context**.

(EDINA and Data Library, n.d.) | RDMRose

File Naming Best Practices

- Be descriptive: Use shared, meaningful terminology. Incorporate relevant terms such as project name, place, date, experiment, instrument, subject, etc. Example: AirQual_Lufkin_Sensor1_201709007
- Be consistent: Use the same structure and terms across projects so that files fall into a useful order (for sorting) and you can easily identify them.
 Example: AvSAT_Ric_2017 AvSAT_Ric_2016 AvSAT_UTx 2017

Guidelines for File Naming

Guideline	Example
Avoid special characters, like / , . #?	Exp01a.xls, NOT Exp#1.a.xls
Don't use blank spaces. Use CamelCharacters or _ to link together keywords.	Site01_Sensor002, NOT Site1 Sensor 2
Use yyyymmdd for dates	200180617, NOT 0617218
Use leading zeroes , e.g. 0001, 001, etc	Experiment002.xls, NOT Experiment2.xls

Which file naming scheme works the best?

- A. bridgedata1bridgedata2bridgedata3
- B. bridge1_sensor2_02142013 bridge1_sensor2_02152013 bridge1_sensor2_02162013
- C. madisonavebridge_sensor2_20130214 madisonavebridge_sensor2_20130215 madisonavebridge_sensor2_20130216
- D. madisonavebridge_sensor2_feb142013 madisonavebridge_sensor2_02152013 madbridge_s2_feb162013



University of Minnesota Libraries



Instructions: Review the handout, then partner with 2-3 people to decide on a file naming system in order to **sort by interviewee name**.

3 minutes to discuss

University of Minnesota Libraries

4. Manage versions

Versioning: Which one is authoritative?

DataAnalysis.xls DataAnalysis2.xls DataAnalysisSept2017.xls DataAnalysisFinal.xls DataAnalysisFinalFINAL.xls

Manual Options for Managing Versions

- Retain original, raw files and significant iterations.
- Use careful file naming: record major changes via whole numbers (v01), minor via an additional number (v02_01)
- Put older versions in an archive folder.
- Create a <u>version control table:</u>

Version Number	Author	Purpose/Change	Date
0-1	Jackie Wilson, Project Manager	Initial draft – to line manager	12/07/2011
0-2	Jackie Wilson, Project Manager	Consultation draft – to working group	21/08/2011
0-3	Jackie Wilson, Project Manager	Second consultation draft – to working group	08/10/2011
1-0	Jackie Wilson, Project Manager	Final version – approved by Project Board	18/11/2011

Software for Managing Versions

Accessing multiple versions:

• <u>Box, Google Drive</u> & other storage services

Version control software:

• <u>GitHub</u>: <u>Researchers</u> and educators can receive GitHub Team (unlimited repositories) for free.

Accessing Version History on Box.com



https://community.box.com/t5/Organizing-and-Tracking-Content/Accessing-Version-History/ta-p/50452

Version Control

"Version control is a system that records changes to a file or set of files over time so that you can recall specific versions later." (Pro Git)

- See who does what.
- Access any version of file.
- Roll back changes.
- Enable new branches of project.

Manage and Access Versions of Files with Git(Hub)

	h / git4phi			Watch +	3 .	★ Star	7	¥ Fork
pdate master	README.md	https://github.con	n/rzach/git4	<u>lphi</u>		Browse	e files	
rzach	committed on Jul 4	1 parent	0a9437b commit f8ct	a8b8ec50331f6a2	2d5e3ad	777d870e	e10bae	259
Showing	1 changed file with 1 addition a	nd 1 deletion.				Unified	Sp	olit
	README.md				\diamond		View	1
串	@@ -5,7 +5,7 @@ Git for PM	flosophers						
登 5 5 5 6 6 7 7	A basic introduction to t collaborate on document of	ilosophers he revision control system Git for non-pr miting.	ogrammers, specifical	ly for using Gi	it as a	way to		
5 5 5 5 7 7 8	 -5,7 +5,7 @ Git for PP A basic introduction to t collaborate on document v -The guide is written in N /blob/master/git4phi.md). +The guide is written in N /blob/master/git4phi.md). 	ilosophers the revision control system Git for non-pr miting. Markdown, the file is git4phi.md, and [car You can download the latest release, ind //git4phi/releases)	ogrammers, specifical be read here](https: be read here](https: luding a printable PC	ly for using Gi //github.com/rz //github.com/rz F version, [her	it as a zach/gi1 zach/gi1 re]	way to t4phi t4phi		

<u>Researchers</u> and educators can receive <u>GitHub</u> Team (unlimited repositories) for free.

5. Create tidy data.



The Problems with Messy Data

- Difficult to analyze
- Requires time to clean
- Confusing to other users- and to Future You
- Raises questions about your credibility

Keep Your Data Tidy

- Make each variable a column & each observation a row
- Make column headers variable names
- Atomize your data; put only a single piece of information in each cell (e.g. city, state, country)
- Be consistent in how you will handle empty values (e.g. NULL, NA, leave blank)

Messy vs. Tidy Data

country	TOOT	column	oncoc	country	Imor	POT	000	00000	•
country	year	corumn	Cases	country	year	sex	age	cases	
AD	2000	m014	0	AD	2000	m	0 - 14	0	
AD	2000	m1524	0	AD	2000	m	15 - 24	0	
AD	2000	m2534	1	AD	2000	m	25 - 34	1	
AD	2000	m3544	0	AD	2000	m	35-44	0	
AD	2000	m4554	0	AD	2000	m	45-54	0	
AD	2000	m5564	0	AD	2000	m	55-64	0	
AD	2000	m65	0	AD	2000	m	65 +	0	
AE	2000	m014	2	AE	2000	m	0-14	2	
AE	2000	m1524	4	AE	2000	m	15 - 24	4	
AE	2000	m2534	4	AE	2000	m	25-34	4	
AE	2000	m3544	6	AE	2000	m	35 - 44	6	
AE	2000	m4554	5	AE	2000	m	45-54	5	
AE	2000	m5564	12	AE	2000	m	55-64	12	
AE	2000	m65	10	AE	2000	m	65 +	10	
AE	2000	f014	3	AE	2000	f	0-14	3	
	(a) Molt	en data			(b) 7	lidy da	ata		

Table 10: Tidying the TB dataset requires first melting, and then splitting the column column into two variables: sex and age.



	A	В	C	D	E
1	Date	ID	Plasmid	Primer	Results
2	970910	E1 5411	MDM970905E1	MSAF5411	Unreadable
3	970911	J1 5411	MDM970905J1	MSAF5411	unteadable
4	070047	E5411	MDM970905E	MSAF5411	T173A, HA iss present
5	970917	J5411	MDM970905J	MSAF5411	STOTA, HA tag present
6	971104	A4	AH971022A4	MSAF8259	GST clone - wrong, no GST!
7	074040	A6	AH971204A6	pUC19SP2	U.S.E clone wrong
8	9/1216	C9	AH971216C9	pUC19SP2	U.S.E clone wrong
9		A15	AH971230A15	pUC19SP2	R261A, L283A
10	980114	AS	AH971230A5	pUC19SP2	WT
11		D9	AH971230D8	MSAF1818	N-terminal HA tag present
12	960313	AH2	AH971116A7	MSAF1818	HA tag present
13	980330	A2	AH980325A2	MSAF1818	R261A, L263A, R269A, F271A
14		C1	AH980325C1	MSAF8259	R261A, L263A
15		C2	AH980325C2	MSAF8259	unreadable
16	980402	C3	AH980325C3	MSAF8259	R261A, L263A
17	1000	C4	AH980325C4	MSAF8259	R261A, L263A
18		C5	AH980325C5	MSAF8259	no mutation
19	980424	E8	AH980325E8	MSAF8259	L263A ranky
20	980504	H1B	random mut. H1B	MSAF8259	221-284 no mutation
21	DEDEDT	430A1	AH980430A1	MSAF8259	WT no R269A, F271A
22	960201	430A2	AH960430A2	MSAF8259	WT no R269A, F271A
25		325E20	AH980325E20	MSAF8259	L263A only
24		325E21	AH980325E21	MSAF8259	correct, R261A, L260A
25	000544	325E22	AH980325E22	MSAF8259	L263A only
26	960211	325E26	AH980325E26	MSAF8259	WT
27		325E28	AH980325E28	MSAF8259	L263A only
28		325E30	AH980325E30	MSAF8259	WT
29	000340	B12REV	AH980707B12	reverse	215-284 3xHA correct
30	960/16	C1REV	AH980707C1	reverse	226-284 3xHA correct
31		A1REV	AH980717A1	reverse	not close enough to primer
32	960722	A3REV	AH980717A3	reverse	WT (incorrect)
33	(*******)	A7REV	AH980717A7	reverse	unreadable
34	980902	A23REV	AH980707A23	reverse	221-284 3xHA correct
35	001004	A11	AH981015A11	1818	R288A, F271P
36	961021	A4	AH981015A4	1818	RZMAA, F271A
37	and the second second	A11	AH981015A11	1818	R289A F271A

What issues do you see with this spreadsheet?



6. Document your data.

What information would you want to know about this file?



Why Document Data?

- Makes it easier for you to interpret your own data
- Facilitates collaboration, sharing, and reuse
- Promotes successful long-term preservation of data

New England Collaborative Data Management Curriculum

Create a Readme File to Document a File or Directory

Typical contents:

- What: title & description
- When: date of data collection
- Who: name & contact info of creator
- Where: location where data was captured
- How:
 - Method of data collection, creation or processing
 - Restrictions on accessing files

Files to replicate Sean Bolks and Richard J. Stoll, <u>"The Arms Acquisition Process</u>: The Effect of Internal and External Constraints on Arms Race Dynamics," *The Journal of Conflict Resolution* 44, no. 5 (October 1, 2000): 580–603.

FileContenttable1.dtaStata data file with data for Table 1table1.doStata .do file with commands to replicate Table 1table2.dtaStata data file with data for Table 2table2.doStata .do file with commands to replicate Table

Simple Example of a ReadMe File

Create a Codebook to Describe the Contents of Data Files

"A codebook is an essential document that informs the data user about the study, data file(s), variables, categories, etc., that make up a complete dataset. The codebook may include a dataset's record layout, list of variable names and labels, concepts, categories, cases, missing value codes, frequency counts, notes, universe statements, and so on." http://www.ddialliance.org/training/getting-started-new-content/create-a-codebook

Codebook Example



RP COOPERATIVE INSTITUTIONAL RESEARCH PROGRAM at the HIGHER EDUCATION RESEARCH INSTITUTE AT LICLA

للاقطاب والأواط

2017 CIRP Freshman Survey (Codebook)

#	Variable Name	Variable Descripion	
	ACE	College I.D.	
	SUBJID	Subject I.D.	
	STUID	Student I.D. as entered on form	
	GRPA	Group Code A	
	GRPB	Group Code B	
1	SEX	Your sex:	
		1 = Male	
		2 = Female	
2	TRANSGENDER	Do you identify as transgender?	
		1=No	
	CONTRACTOR OF	2=Yes	E I
3	YRGRADHS	In what year did you graduate from high school?	
1000		1=2017	
		2=2016	
		3=2015	
		4=2014 or earlier	
		5=Did not graduate but passed G.E.D. test	
		6=Never completed high school	

.

Exercise

Think through creating a readme file for one of your datasets (real or imagined) or the "Dr. Psi" data using this template from <u>Cornell</u>.

See "Guidelines for writing 'readme' style metadata" http://data.research.cornell.edu/sites/default/files/SciMD ReadMe_Guidelines_v4_1_0.pdf

7. Store, Backup and Archive Data

Data Storage Definition

- The media (optical or magnetic) to which you save your data files and software.
- All storage media are vulnerable to risk and obsolescence.
- Storage media should be evaluated and updated every 2-5 years.

Data Backup Definition

- Allows you to *restore* your data if original data is lost or damaged due to:
 - Hardware or software malfunction
 - Environmental disaster (fire, flood)
 - Theft
 - Unauthorized access

>> TEST YOUR BACKUP PLAN!

New England Collaborative Data Management Curriculum

3-2-1 Backup Rule



Save 3 copies of your data.



Use 2 types of storage.



Keep 1 remote copy.

Overview of Data Storage, Backup and Sharing Options at Rice

- **Network or Cloud Storage**
 - **storage.rice.edu** U: drive, departmental shares
 - Research Data Facility (RDF) larger scale storage for research Rice Box: unlimited cloud storage
- **Backup Options**
 - **storage.rice.edu** backups/snapshots
 - Crash Plan for Rice workstations
- Data Sharing- Globus Connect

Options for faculty/ staff: <u>https://kb.rice.edu/page.php?id=70762</u>

Options for students: <u>https://kb.rice.edu/page.php?id=65636</u>

storage.rice.edu

- Location: Networked
- Storage quotas
 - Undergraduates: 2 GB
 - Graduates, Staff, Faculty: 5 GB
 - Colleges, Depts, Centers, Institutes: 40 GB
- Performance Subject to network
- Accessibility
 - NetID folder: Private, not shared
 - Groups: Any Rice NetID holder by request

Accessing your backups on storage

Name ^	Date modified	Туре
길 2015-03-23_1917-0500.UJ-p_daily	3/21/2015 12:04 AM	File folder
퉬 2015-03-24_1917-0500.UJ-p_daily	3/21/2015 12:04 AM	File folder
퉬 2015-03-25_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder
퉬 2015-03-26_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder
퉬 2015-03-27_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder
퉬 2015-03-28_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder
퉬 2015-03-29_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder
퉬 2015-03-30_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder
퉬 2015-03-31_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder
퉬 2015-04-01_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder
퉬 2015-04-02_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder
퉬 2015-04-03_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder
퉬 2015-04-04_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder
퉬 2015-04-05_1917-0500.UJ-p_daily	3/25/2015 12:08 AM	File folder

\\storage.rice.edu\?-home\~snapshot

Research Data Facility

- Location: On Site network data shares
- Storage quotas
 - 500GB per researcher
 - Additional storage available with cost recovery
- Performance Subject to network
- Accessibility
 - Based on NetID and ADRICE security groups
 - Can be shared to multiple users in a research group

Storage Features of Rice Box

"enterprise cloud-based storage and collaboration service"

- Use <u>unlimited storage</u> (max file size = 15 GB)
- Access prior <u>versions</u> (up to 100)
- <u>Sync</u> files and download for offline use
- Files automatically <u>backed up at</u>
 multiple data centers
- Control file/folder permissions

hare 'BoxTest'						
vit	vite People					
ΛH	d names or email addres are					
vit	e as Editor 🔺					
	Co-owner					
	Manage security, upload, download, preview, share, edit, and delete					
~	Editor					
	Upload, download, preview, share, edit, and delete					
	Viewer Uploader					
	Upload, download, preview, share, and edit					
	Previewer Uploader					
	Upload and preview					
	Viewer					
	Download praview and chara					

Collaboration Features of Box

- <u>Share</u> files, links, notes, etc. with collaborators at Rice & beyond
- Integrate <u>Google Suite</u>
- Take <u>Box Notes</u> and share (can't currently export)
- Use <u>tags</u> to sort and search files (can't currently export)



Security and Rice Box

"Box encrypts all data it stores, allow for granular access controls, and facilitates access monitoring. Folders must be properly configured to take full advantage of these security enhancements."

See <u>Box User Guidelines - Rice Data</u> Includes recommended settings for sharing, tagging, notifications, and more. Contact Help Desk about high risk & confidential data.

Backup: CrashPlan

- Availability: Rice-owned computers
- Cost: \$82.56/year/person (up to 4 devices)
- Location: Off-site cloud storage
- Procedure: Incremental
- Frequency: Adjustable up to every minute
- Retention: Adjustable up to forever

https://kb.rice.edu/page.php?id=72955

Backup frequency	r.	
New version		every 15 minutes
Additional version	s to keep from:	
Last week		every 15 minutes
Last 90 days	-0	every day
Last year		every week
Previous years		every month
Remove deleted f	iles	ever
Defaults		Cancel

CrashPlan PROe or crashplan.rice.edu

Sharing: Globus Connect

- Widely used service for large data exchange between participating institutions
- Can be used in our HPC environment or from your desktop with Globus Connect Personal
- Accessibility
 - Contact Center for Research Computing to be added to license
 - $\circ~$ Arrange for access to peer institution end points

Confidential and Regulated Data (High Risk)

- Health Information, including Protected Health Information
- Health Insurance policy ID numbers
- Social Security Numbers
- Credit card numbers
- Financial account numbers
- Export controlled information under U.S. laws
- Driver's license, passport and visa numbers
- Donor contact information and non-public gift information
- Rice data classified as confidential under policy <u>808</u>, e.g. birth dates

Securing Your Data

Provide appropriate <u>security</u> for data (e.g. antivirus protection, access control, encryption, deidentification of data).

Consult IT regarding data security

Approved Services

This table indicates which classifications of data are allowed on a selection of commonly used Rice IT Services.



https://vpit.rice.edu/it-security/resources/risk-classifications/approved-services

Data Archiving Definition

- Provides a final version of your data.
- Stored for the long-term.
- May be shared publicly through a data archive.

Why Archive Your Research Data with a Data Repository?

- Conform to publisher or funder requirements
- Get cited
 - "studies that made [gene expression microarray] data available in a public repository received 9% ... more citations than similar studies for which the data was not made available." (Piowowar & Vision, 2013)
- Promote future research

Data Archiving Options

Public Repositories:

- <u>Discipline based repository (e.g. GenBank or PANGEA)</u>
- General data repository (e.g. FigShare or Dataverse)
- Institutional repository (e.g. Rice Digital Scholarship Archive)

Private Approaches:

• Long-term storage

Rice Data Sharing Option: Rice Digital Scholarship Archive



FA

🕈 Rice Scholarship Home / Faculty & Staff Research / Rice Research Data / View Item

The Acceptability of War and Support for Defense Spending: Evidence from Fourteen Democracies, 2004–2013 [Replication Data]

_	_		1
_		-	
	-	-	
_		-	20
-	_	-	

	_		
-	-		
-	-	-	
-	-		
-	_	-	
_			t.
-	_	-	

Name:	esbulld.zip
Size:	3.011Mb
Format:	application/zip
Description:	Original data files

Name: esbuildNonproprietary.zip Size: 2.651Mb Format: application/zip Description: Nonproprietary data files View/Open

Vlew/Open

https://scholarship.rice.edu/

Data Archiving Caveats

- Do not share confidential data (unless it has been de-identified and approved through IRB).
- Consult with your collaborators before publishing data.
- It may be possible to embargo data so that it is not available until the related publication is released.

What Does Research Data Services Offer?

https://library.rice.edu/research-data-services

- Workshops on R, Python, Excel, etc. (including upcoming 2 day workshops from Software Carpentry)
- Consulting on finding, analyzing, managing, and visualizing data, including during Friday office hours
- Publishing and preserving data through the Rice Digital Scholarship Archive; providing DOIs
- Reviewing data management plans

Tools and Resources

Use OSF to Manage Your Research Workflow and Collaborate

				Factor (Carton (
🎇 Demo Project			Private Make Public	A+ C 0
ntributors: Courtney Soderberg Illaced Institutions: Center For Open Science te created: 2017-01-31 10:31 AM Last Updated: 201 tegory: Project C scription: roject created to learn how to use the OSF: ense: No license	7-01-31 10:31 AM			
Wiki 🛛 🖄		Citation	osf.io/d5ntj ~	
		Components	Add Component	Link Projects
Files	Ø	No components have been added to this project.		
Click on a storage provider or drag and drop to uple	bad			
	Q Filter i	Tags		
Name A Modified A Modi		add a tag		
		Recent Activity		
		Courtney Soderberg added C	enter For Open Science affiliation b	o Demo

https://osf.io/

- Organize files in one place
- Share with collaborators
- Control files access
- Integrate with tools like Box
- Track versions
- Make work citable
- Facilitate reproducibility
- Free & open source

Resources

Borer, Elizabeth T., et al "Some Simple Guidelines for Effective Data Management."

Bulletin of the Ecological Society of America (2009): 205–14.

DataOne Primer on Data Management,

https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf

Dataverse, Data Management Plans, <u>http://best-practices.dataverse.org/data-</u> <u>management/</u>

ICPSR Guide to Social Science Data Preparation and Archiving, http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/

Svend Juul et al, "Take good care of your data,"

http://www.epidata.dk/downloads/takecare.pdf

UK Data Archive, *Managing and Sharing Data: Best Practices for Researchers*, <u>http://www.data-archive.ac.uk/media/2894/managingsharing.pdf</u>

Thanks!

Please contact <u>researchdata@rice.edu</u> with any questions.

Visit us online at <u>http://researchdata.rice.edu/</u>.

Help us shape future workshops! Please complete this evaluation:

http://library.rice.edu/requests/course-evaluationform