

**Some advice from a reproducible
researcher about how some advice
from research data repositories to
irreproducible researchers about
reproducibility and repositories
might help researchers, repositories,
and reproducibility**

Thomas J. Leeper

Department of Government
London School of Economics and Political Science

Why reproducibility?

- Journal requirements
- Funding agency requirements
- Institutional requirements
- The coming revolution

Why reproducibility?

- Journal requirements
- Funding agency requirements
- Institutional requirements
- The coming revolution

How can we shift thinking from *extrinsic* motivations to intrinsic motivations?

Getting intrinsic



Open Science
@openscience



Following

"Reproducibility is collaboration with people you don't know, incl. yourself next week." –
[@philipbstark](#) [#openscience](#)

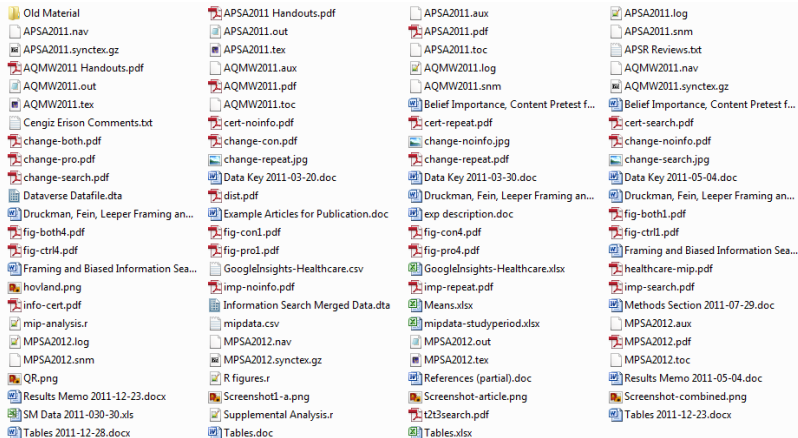


**What makes up the ideal
reproducible research product?**

**What makes up the ideal
reproducible research product?**

Nobody seems to agree!

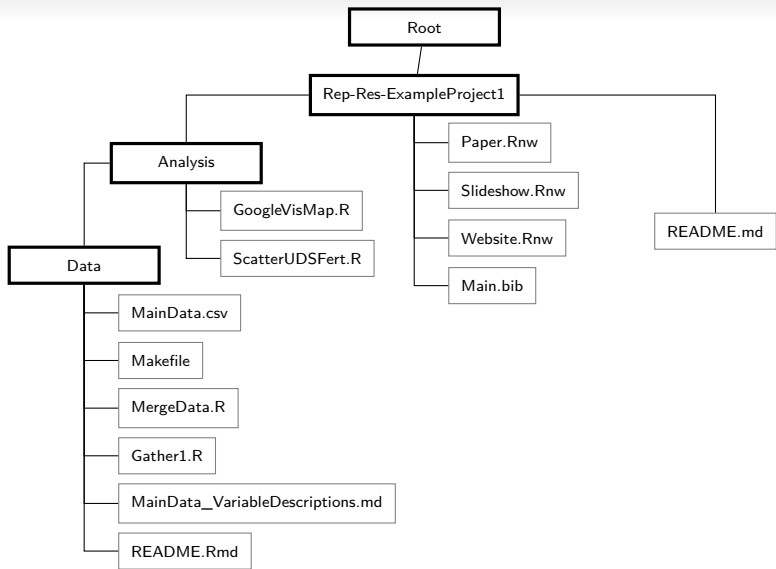
Old Material	APSA2011 Handouts.pdf	APSA2011.aux	APSA2011.log
APSA2011.nav	APSA2011.out	APSA2011.pdf	APSA2011.snm
APSA2011.synctex.gz	APSA2011.tex	APSA2011.toc	APSR Reviews.txt
AQMW2011 Handouts.pdf	AQMW2011.aux	AQMW2011.log	AQMW2011.nav
AQMW2011.out	AQMW2011.pdf	AQMW2011.snm	AQMW2011.synctex.gz
AQMW2011.tex	AQMW2011.toc	Belief Importance, Content Pretest f...	Belief Importance, Content Pretest f...
Cengiz Erison Comments.txt	cert-noinfo.pdf	cert-repeat.pdf	cert-search.pdf
change-both.pdf	change-con.pdf	change-noinfo.jpg	change-noinfo.pdf
change-pro.pdf	change-repeat.jpg	change-repeat.pdf	change-search.jpg
change-search.pdf	Data Key 2011-03-20.doc	Data Key 2011-03-30.doc	Data Key 2011-05-04.doc
Dataverse Datafile.dta	dist.pdf	Druckman, Fein, Leeper Framing an...	Druckman, Fein, Leeper Framing an...
Druckman, Fein, Leeper Framing an...	Example Articles for Publication.doc	exp description.doc	fig-both1.pdf
fig-both4.pdf	fig-con1.pdf	fig-con4.pdf	fig-ctrl1.pdf
fig-ctrl4.pdf	fig-pro1.pdf	fig-pro4.pdf	Framing and Biased Information Sea...
Framing and Biased Information Sea...	GoogleInsights-Healthcare.csv	GoogleInsights-Healthcare.xlsx	healthcare-mip.pdf
hovland.png	imp-noinfo.pdf	imp-repeat.pdf	imp-search.pdf
info-cert.pdf	Information Search Merged Data.dta	Means.xlsx	Methods Section 2011-07-29.doc
mip-analysis.r	mipdata.csv	mipdata-studyperiod.xlsx	MPSA2012.aux
MPSA2012.log	MPSA2012.nav	MPSA2012.out	MPSA2012.pdf
MPSA2012.snm	MPSA2012.synctex.gz	MPSA2012.tex	MPSA2012.toc
QR.png	R figures.r	References (partial).doc	Results Memo 2011-05-04.doc
Results Memo 2011-12-23.docx	Screenshot-1-a.png	Screenshot-article.png	Screenshot-combined.png
SM Data 2011-030-30.xls	Supplemental Analysis.r	t2t3search.pdf	Tables 2011-12-23.docx
Tables 2011-12-28.docx	Tables.doc	Tables.xlsx	



Confession: This is my PhD dissertation.

What makes up the ideal reproducible research product?

- Gandrud's template
- rOpenSci's "Research Compendium"
- Project TIER



```
project
|- DESCRIPTION      # project metadata and dependencies
|- README.md       # top-level description of content
|
|- data/           # raw data, not changed once created
| +- my_data.csv   # data files in open formats
|
|- analysis/       # any programmatic code
| +- my_scripts.R  # R code used to analyse data
```

TIER Protocol Documentation

Original Data

Original data files
Importable data files (if necessary)

Metadata

The Metadata Guide
Supplementary metadata documents
(if necessary)

Analysis Data

Analysis data files

Documents

The final paper
The Data Appendix
The Read Me file

Command Files

Command files

What makes up the ideal reproducible research product?

- Gandrud's template
- rOpenSci's "Research Compendium"
- Project TIER

What makes up the ideal reproducible research product?

- Gandrud's template
- rOpenSci's "Research Compendium"
- Project TIER
- Docker container?

What makes up the ideal reproducible research product?

- Gandrud's template
- rOpenSci's "Research Compendium"
- Project TIER
- Docker container?
- Virtual machine?

What makes up the ideal reproducible research product?

- Gandrud's template
- rOpenSci's "Research Compendium"
- Project TIER
- Docker container?
- Virtual machine?
- ???

What makes up the ideal reproducible research product?

- Gandrud's template
- rOpenSci's "Research Compendium"
- Project TIER
- Docker container?
- Virtual machine?
- ???

More is probably better for *reproducibility*, but declining marginal returns for the researcher.

What makes up the ideal reproducible research product?

- Big disagreements
 - What exactly is being reproduced?
 - What is being assumed about software, hardware, data formats, etc.?
 - What tools are best? packages? make? docker?

What makes up the ideal reproducible research product?

- Big disagreements
 - What exactly is being reproduced?
 - What is being assumed about software, hardware, data formats, etc.?
 - What tools are best? packages? make? docker?
- Consensus is not possible!

HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.



YEAH!

SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.

What makes up the ideal reproducible research product?

- Big disagreements
 - What exactly is being reproduced?
 - What is being assumed about software, hardware, data formats, etc.?
 - What tools are best? packages? make? docker?
- Consensus is not possible!

What makes up the ideal reproducible research product?

- Big disagreements
 - What exactly is being reproduced?
 - What is being assumed about software, hardware, data formats, etc.?
 - What tools are best? packages? make? docker?
- Consensus is not possible! *Nudge* instead.

What makes up the ideal reproducible research product?

- Big disagreements
 - What exactly is being reproduced?
 - What is being assumed about software, hardware, data formats, etc.?
 - What tools are best? packages? make? docker?
- Consensus is not possible! *Nudge* instead.
 - Provide templates to use when *starting* a project

What makes up the ideal reproducible research product?

- Big disagreements
 - What exactly is being reproduced?
 - What is being assumed about software, hardware, data formats, etc.?
 - What tools are best? packages? make? docker?
- Consensus is not possible! *Nudge* instead.
 - Provide templates to use when *starting* a project
 - Provide exemplars to show how to conceptualize the organization of a project



This repository Search

Pull requests Issues Marketplace Gist



IQSS / dataverse

Watch 44

Unstar 185

Fork 104

Code

Issues 811

Pull requests 14

Projects 1

Insights

Dataset: File Hierarchy + Directory Structure #2249

New issue

Open

eaquigley opened this issue on Jun 11, 2015 · 19 comments



eaquigley commented on Jun 11, 2015

Member



User community request to be able to organize files in a dataset in a hierarchical manner so when a user exports files from Google Drive, Dropbox, or OSF (or any other location), the file structure is maintained without causing the user extra work on the Dataverse side.



eaquigley added **Component: UX & Upgrade** **Type: Suggestion** **Status: Design** **Component: File Upload & Handling** labels on Jun 11, 2015



eaquigley self-assigned this on Jun 11, 2015

Assignees

No one assigned

Labels

Component: File Upload & Han...

Component: UX & UI

Status: Triaged

Type: Suggestion

Projects

i Use the check boxes next to the file name to download multiple files. Data files will be downloaded in their default format. You can also download all the files in a category by checking the box next to the category name. You will be prompted to save a single archive file. Study files that have restricted access will not be downloaded.

 Select all files

Total Number of Files: 10

Total Downloads: 48

Downloads of Files in This Version: 44

 Codebook

 Study 1 (Lab) Key.docx

application/octet-stream - 21 KB - 7 downloads



Download

Study 1 data key

 Study 2 (Exit Poll) Key.docx

application/octet-stream - 16 KB - 3 downloads



Download

Study 2 data key

 Data

 Study 1 (Lab) Data.tab

Tab Delimited - 54 KB - 18 downloads + analyses



Download as...

Study 1 Data

TABULAR DATA

647 Cases

25 Variables



Access Analysis + Subsetting

View Data Citation [+]

 Study 2 (Exit Poll) Data.tab

Tab Delimited - 42 KB - 4 downloads + analyses



Download as...

Study 2 Data

TABULAR DATA

765 Cases

22 Variables



Access Analysis + Subsetting

View Data Citation [+]

 Experimental Materials

 Con Pretreatment Articles.doc

MS Word - 28 KB - 1 download



Download

Con Pretreatment Articles

 Control Pretreatment Articles.doc

MS Word - 23 KB - 1 download



Download

Control Pretreatment Articles

 Pro Pretreatment Articles.doc

MS Word - 28 KB - 2 downloads



Download

Pro Pretreatment Articles

 Questionnaire.doc

MS Word - 97 KB - 4 downloads



Download

Questionnaire

 Replication code

 Study 1 (Lab).do

Stata Syntax - 5 KB - 3 downloads



Download

Study 1 Replication code










 Study 2 (Exit Poll).do

Stata Syntax - 2 KB - 1 download



Download

Study 2 Replication code

<input type="checkbox"/>	 <p>Control Pretreatment Articles.doc MS Word - 23.5 KB - Jul 26, 2013 - 4 Downloads Control Pretreatment Articles Experimental Materials</p>	Download
<input type="checkbox"/>	 <p>Pro Pretreatment Articles.doc MS Word - 28.5 KB - Jul 26, 2013 - 5 Downloads Pro Pretreatment Articles Experimental Materials</p>	Download
<input type="checkbox"/>	 <p>Questionnaire.doc MS Word - 97.0 KB - Jul 26, 2013 - 11 Downloads Questionnaire Experimental Materials</p>	Download
<input type="checkbox"/>	 <p>Study 1 (Lab) Data.tab Tabular Data - 54.7 KB - Dec 9, 2011 - 32 Downloads 25 Variables, 647 Observations - UNF:5:gD3DdWpenPRpthWUWvA== Study 1 Data Data</p>	Explore Download
<input type="checkbox"/>	 <p>Study 1 (Lab) Key.docx MS Word (docx) - 21.4 KB - Dec 9, 2011 - 10 Downloads Study 1 data key Codebook</p>	Download
<input type="checkbox"/>	 <p>Study 1 (Lab).do Stata Syntax - 5.0 KB - Jul 26, 2013 - 6 Downloads Study 1 Replication code Replication code</p>	Download
<input type="checkbox"/>	 <p>Study 2 (Exit Poll) Data.tab Tabular Data - 42.9 KB - Dec 9, 2011 - 21 Downloads 22 Variables, 765 Observations - UNF:5:1q2402xO+rIXWfvAeJCG== Study 2 Data Data</p>	Explore Download
<input type="checkbox"/>	 <p>Study 2 (Exit Poll) Key.docx MS Word (docx) - 16.1 KB - Dec 9, 2011 - 6 Downloads Study 2 data key Codebook</p>	Download
<input type="checkbox"/>	 <p>Study 2 (Exit Poll).do Stata Syntax - 2.6 KB - Jul 26, 2013 - 3 Downloads Study 2 Replication code Replication code</p>	Download

The Advice to Researchers

- 1 Reproducibility isn't just one more burden
- 2 It's about helping your (future) *yourself* first
- 3 Be reproducible for science second

The Advice to Researchers

- 1 Reproducibility isn't just one more burden
- 2 It's about helping your (future) *yourself* first
- 3 Be reproducible for science second

```
#DEAR FUTURE SELF,  
#  
# YOU'RE LOOKING AT THIS FILE BECAUSE  
# THE PARSE FUNCTION FINALLY BROKE.  
#  
# IT'S NOT FIXABLE. YOU HAVE TO REWRITE IT.  
# SINCERELY, PAST SELF
```

DEAR PAST SELF, IT'S KINDA
CREEPY HOW YOU DO THAT.

```
#ALSO, IT'S PROBABLY AT LEAST  
# 2013. DID YOU EVER TAKE  
# THAT TRIP TO ICELAND?
```

STOP JUDGING ME!



Irreproducibility

Irreproducibility

- Fabrication

Irreproducibility

- Fabrication
- Human error

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Proprietary data and file formats

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Proprietary data and file formats
- Unavailable data

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Proprietary data and file formats
- Unavailable data
- Analysis uses proprietary software/hardware

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Proprietary data and file formats
- Unavailable data
- Analysis uses proprietary software/hardware
- Analysis unavailable

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Proprietary data and file formats
- Unavailable data
- Analysis uses proprietary software/hardware
- Analysis unavailable
- “Available from the author”

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Proprietary data and file formats
- Unavailable data
- Analysis uses proprietary software/hardware
- Analysis unavailable
- “Available from the author”

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Proprietary data and file formats
- Unavailable data
- Analysis uses proprietary software/hardware
- Analysis unavailable
- “Available from the author (now deceased)”



Kaitlin Thaney

@kaythaney



Following

"'Reproducible research' is a redundant term. 'Irreproducible research' just used to be known as 'bullshit'." - [@fperez_org](#)
::slow clap::



RETWEETS

122

FAVORITES

61



6:11 PM - 8 May 2014