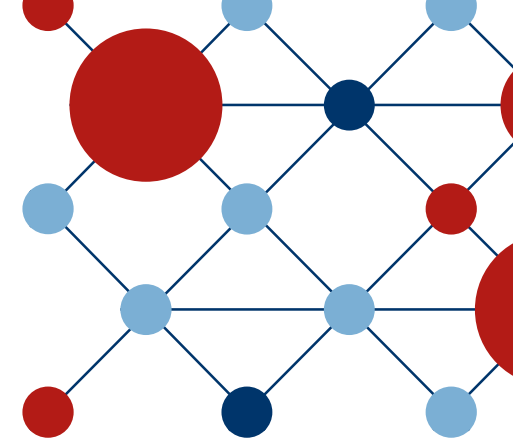


CURATING FOR REPRODUCIBILITY: WHY AND HOW TO REVIEW DATA & CODE

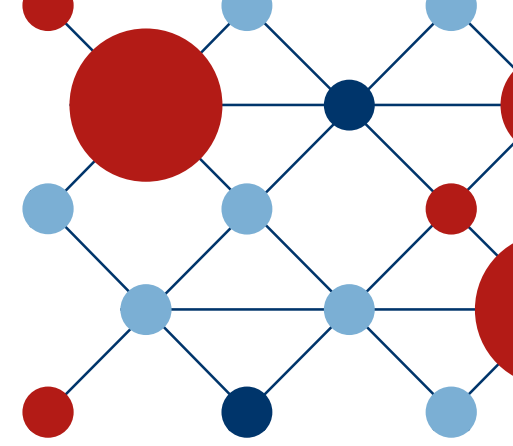




CURATING FOR REPRODUCIBILITY

THE CURE CONSORTIUM

- **Florio Arguillas**, Research Associate
Cornell Institute for Social and Economic Research (CISER)
Cornell University
- **Thu-Mai Christian**, Assistant Director for Archives
Odum Institute for Research in Social Science
University of North Carolina at Chapel Hill
- **Limor Peer**, Associate Director for Research
Institution for Social and Policy Studies (ISPS)
Yale University



CURATING FOR REPRODUCIBILITY

THE CURE CONSORTIUM

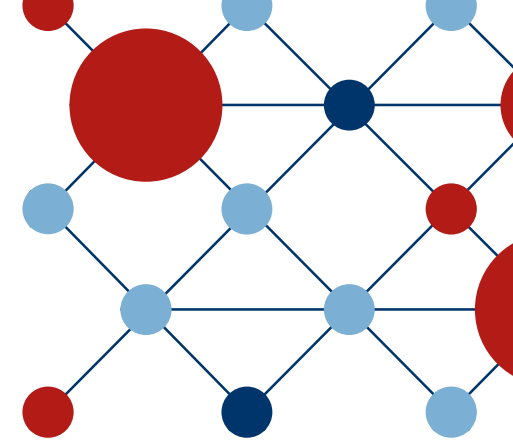
Establish Standards

Share Practices

Promote Data Quality Review



<https://cure.web.unc.edu/>



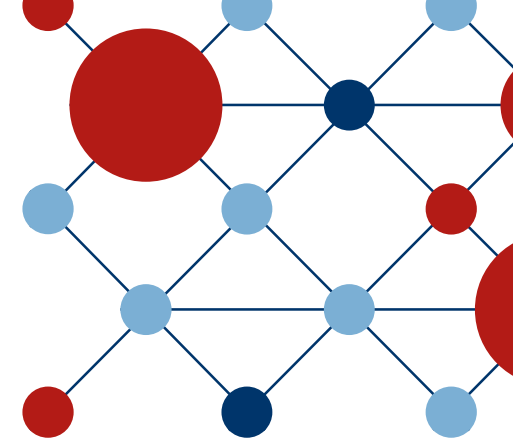
CURATING FOR REPRODUCIBILITY: WHY AND HOW TO REVIEW DATA & CODE

WHY

- What is curating for reproducibility?
- The impetus for curating for reproducibility
- Models of CURE practice

HOW

- Hands-on: Data & code review
- Demo: Data Curation⁺ Tool



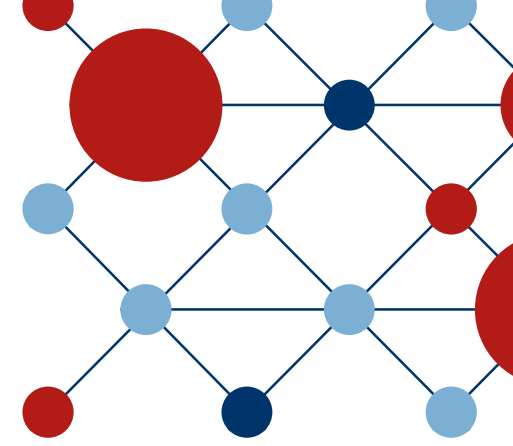
CURATING FOR REPRODUCIBILITY

DATA SHARING

- a** To reproduce or to verify research
- a** To make the results of publicly funded research available to the public
- a** To enable others to ask new questions of extant data
- a** To advance the state of research and innovation

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078. <http://doi.org/10.1002/asi.22634>

BUT...

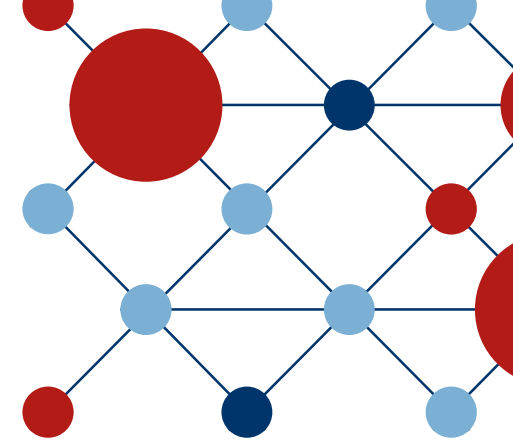


CURATING FOR REPRODUCIBILITY

DATA SHARING

Because there are more ways to share data, and because the scholarly landscape supports and encourages that, there is a proliferation of data files on many different types of systems that **do not meet the criterion of quality...**

Peer, L., Green, A., & Stephenson, E. (2014). Committing to data quality review. *International Journal of Digital Curation*, 9(1). <http://doi.org/10.2218/ijdc.v9i1.317>

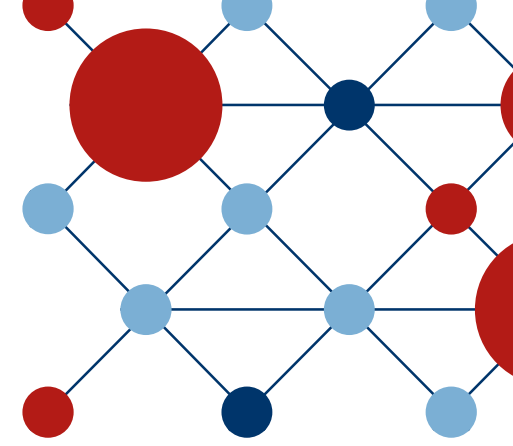


CURATING FOR REPRODUCIBILITY

DATA QUALITY

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

King, G. (1995). Replication, replication. *PS: Political Science & Politics*, 28(3), 444–452. <http://doi.org/10.2307/420301>

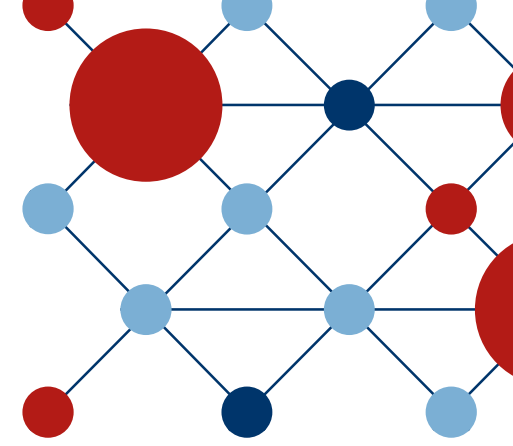


CURATING FOR REPRODUCIBILITY

DATA QUALITY

A set of measures that determine if data are **independently understandable for informed reuse**.

Peer, L., Green, A., & Stephenson, E. (2014). Committing to data quality review. *International Journal of Digital Curation*, 9(1).
<http://doi.org/10.2218/ijdc.v9i1.317>

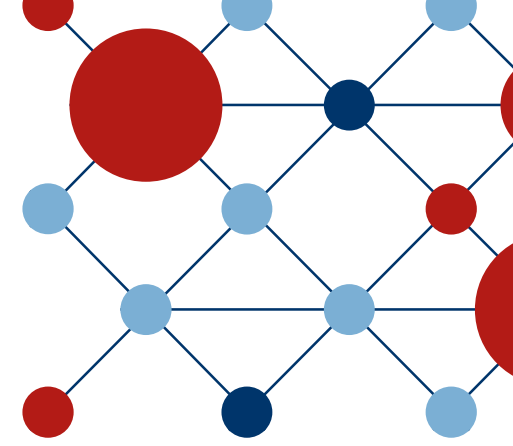


CURATING FOR REPRODUCIBILITY

DATA QUALITY

Could the published computational findings be **reproduced on an independent system by using the data and code** provided?

Stodden, V., McNutt, M., Bailey, D. H., Deelman, E., Gil, Y., Hanson, B., . . . Tauber, M. (2016). Enhancing reproducibility for computational methods. *Science*, *354*(6317), 1240–1241. <https://doi.org/10.1126/science.aah6168>



CURATING FOR REPRODUCIBILITY

DATA QUALITY REVIEW



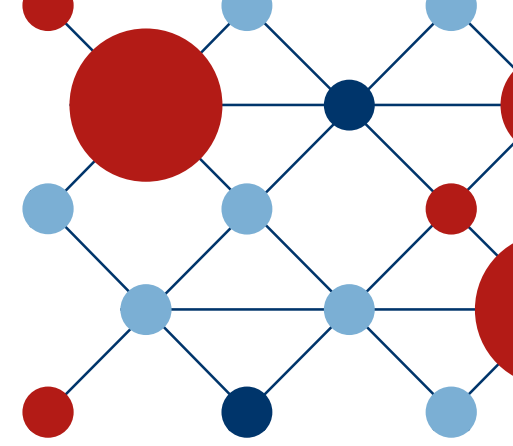
**FILE
REVIEW**



**DOC
REVIEW**



**DATA
REVIEW**



CURATING FOR REPRODUCIBILITY

DATA QUALITY REVIEW



**FILE
REVIEW**



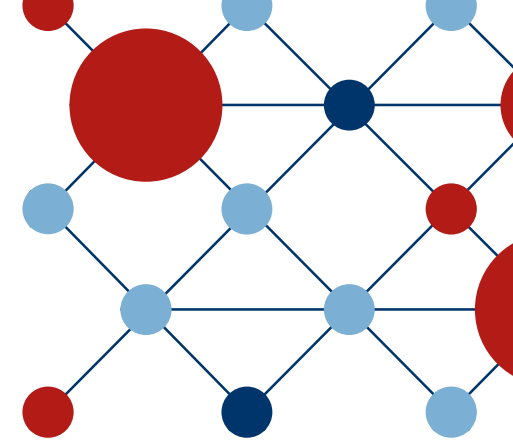
**DOC
REVIEW**



**DATA
REVIEW**



**CODE
REVIEW**

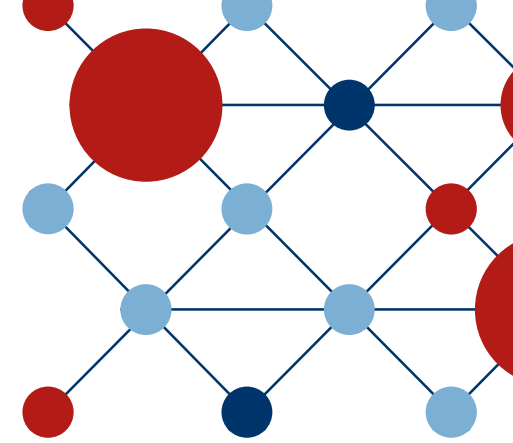


CURATING FOR REPRODUCIBILITY

DATA QUALITY REVIEW



- ✓ Assign persistent identifier
- ✓ Create study citation and study-level metadata record
- ✓ Record file size details
- ✓ Check for presence of all files
- ✓ Verify content of files matches expected format
- ✓ Create non-proprietary versions of files
- ✓ Implement migration strategy for file formats

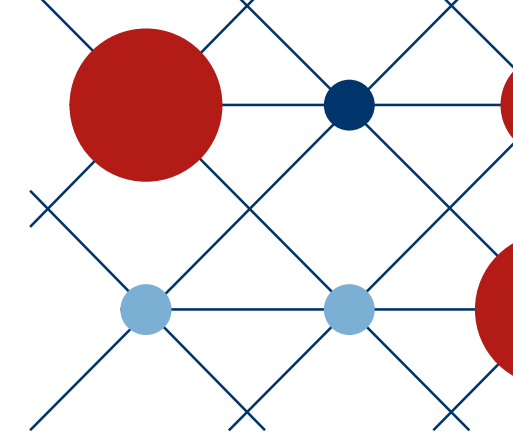


CURATING FOR REPRODUCIBILITY

DATA QUALITY REVIEW



- ✓ Confirm presence of comprehensive descriptive information necessary for informed reuse
 - Data definitions
 - Variable construction
 - Methodology
 - Sampling information
 - Original data source citation
 - Analysis software version
- ✓ Link to related research products

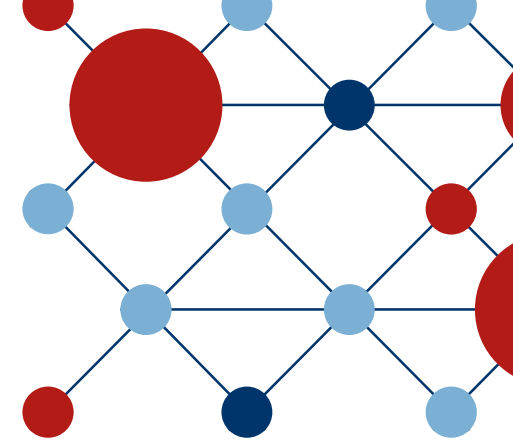


CURATING FOR REPRODUCIBILITY

DATA QUALITY REVIEW



- ✓ Check for undocumented variable and value information
- ✓ Examine data for inconsistencies and errors
 - Discrepancies in number of observations
 - Out-of-range or wild codes
 - Undefined null values
- ✓ Review data for confidentiality issues



CURATING FOR REPRODUCIBILITY

DATA QUALITY REVIEW

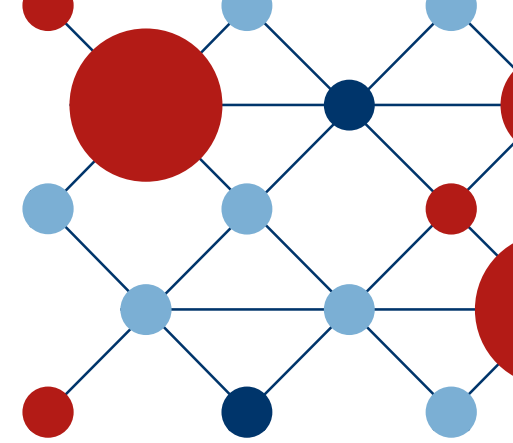


- ✓ Convert absolute file paths to relative file paths
- ✓ Check code for presence of non-executable comments that document analysis processes
- ✓ Identify packages required to execute code
- ✓ Execute code to ensure code is error-free
- ✓ Compare code output to findings presented in article

CURATING FOR REPRODUCIBILITY

MODELS OF PRACTICE

- 1. Institution for Social and Policy Studies (ISPS)**
Aligning Data Curation Workflows with Data Quality Review
- 2. Cornell Institute for Social and Economic Research**



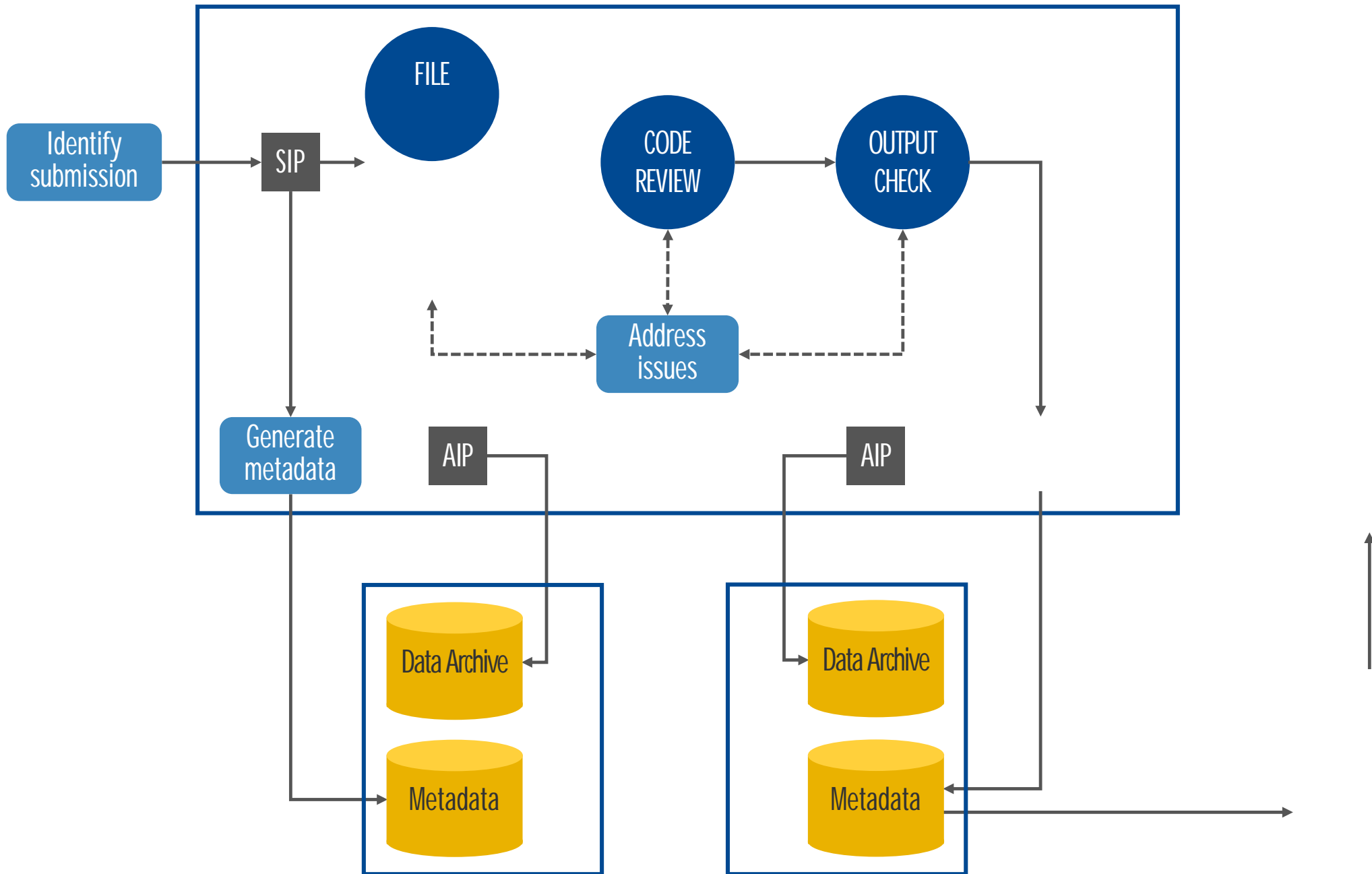
CURATING FOR REPRODUCIBILITY

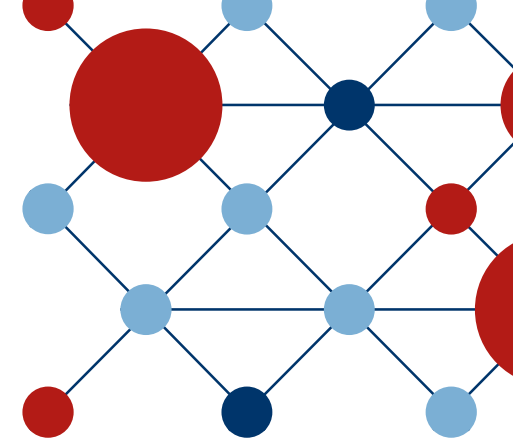
MODELS OF PRACTICE

Institution for Social and Policy Studies (ISPS)

Aligning Data Curation Workflows with Data Quality Review

- ISPS was founded in 1968 as an interdisciplinary center to support social science and public policy research at Yale University
- ISPS Data Archive captures and preserves intellectual output of ISPS-affiliated scholars
- ISPS data archivists developed a data curation workflow that implements the ideals of scientific reproducibility and transparency





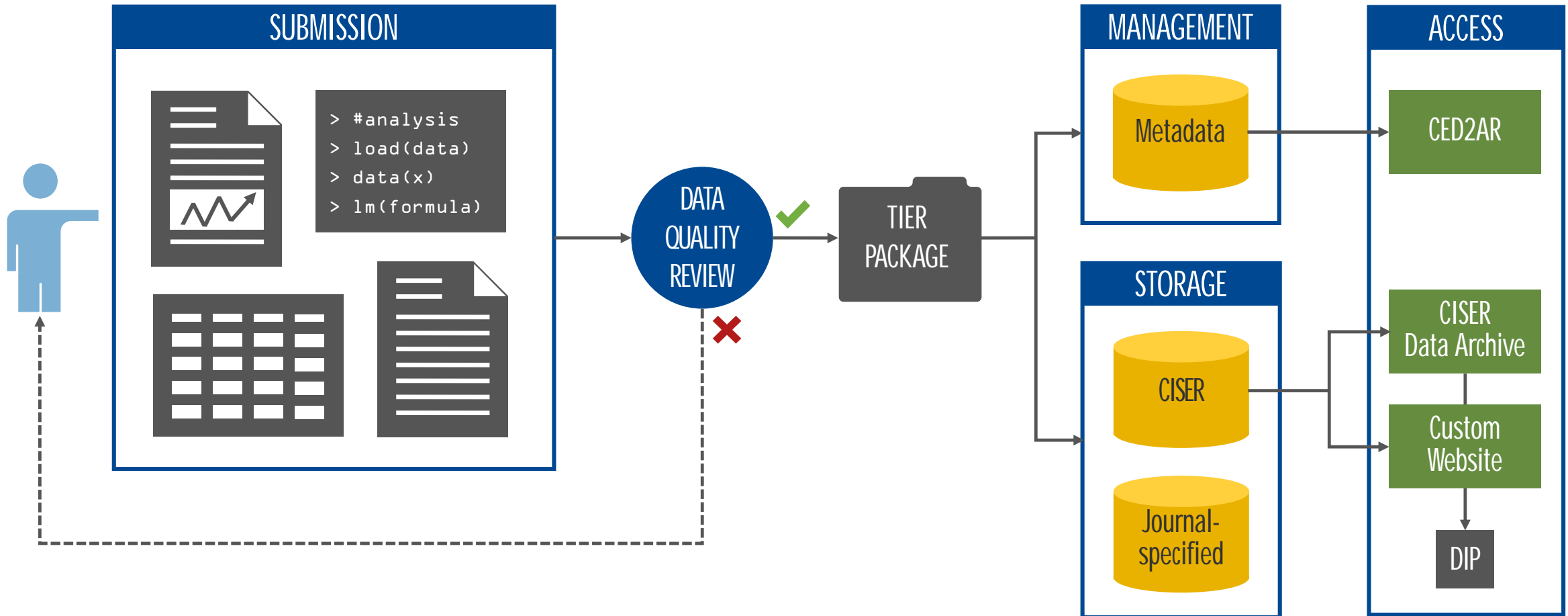
CURATING FOR REPRODUCIBILITY

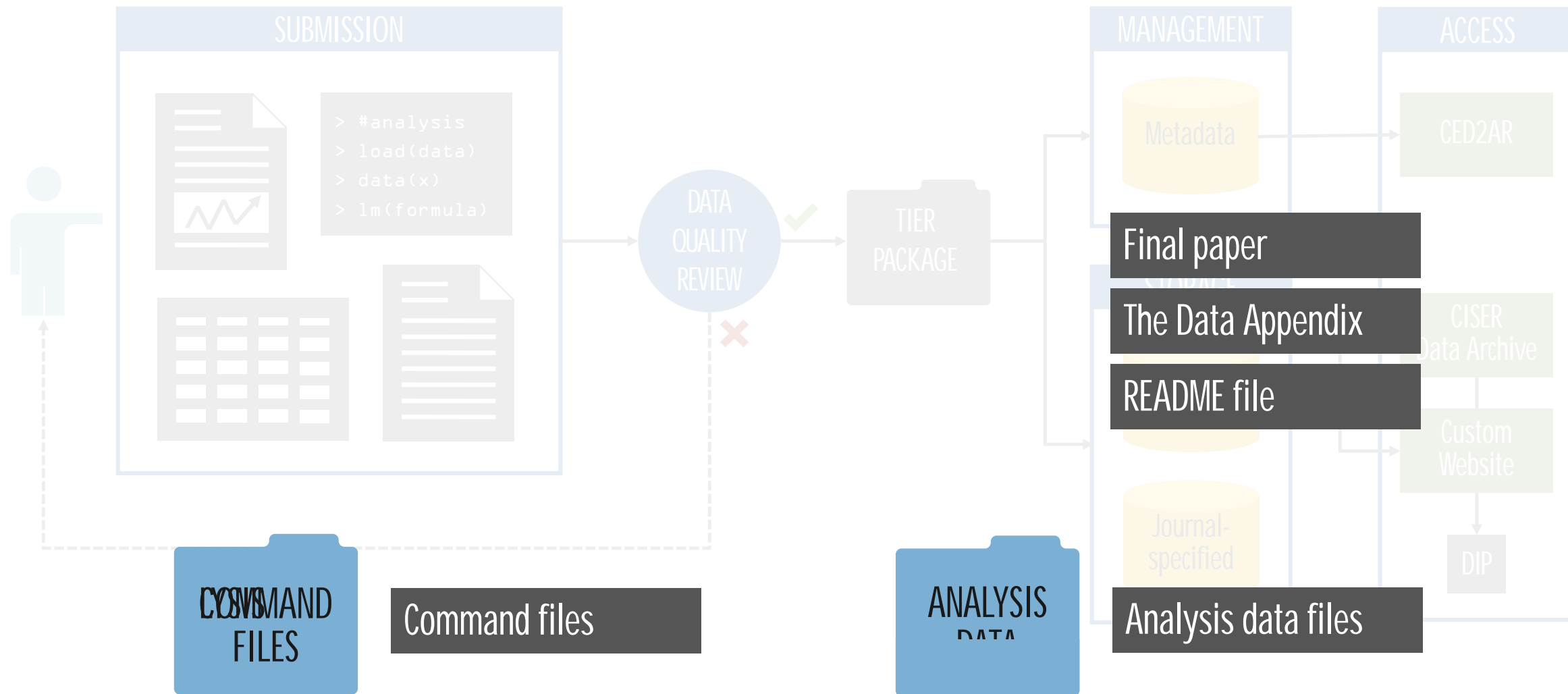
MODELS OF PRACTICE

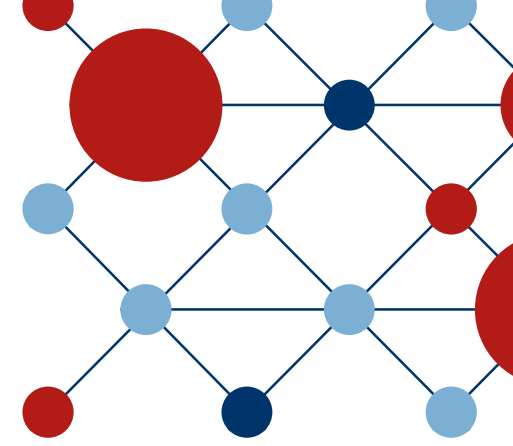
Cornell Institute for Social and Economic Research

Providing Data Curation and Reproduction of Results (R²) Services

- CISER was founded in 1981 to support the evolving computational and data needs of social scientists and economists throughout the entire research lifecycle
- The CISER Data Archive provides access to approximately 27,000 social and economic dataset files
- CISER staff offers appraisal, curation, and replication services to researchers preparing for manuscript submission to scholarly journals





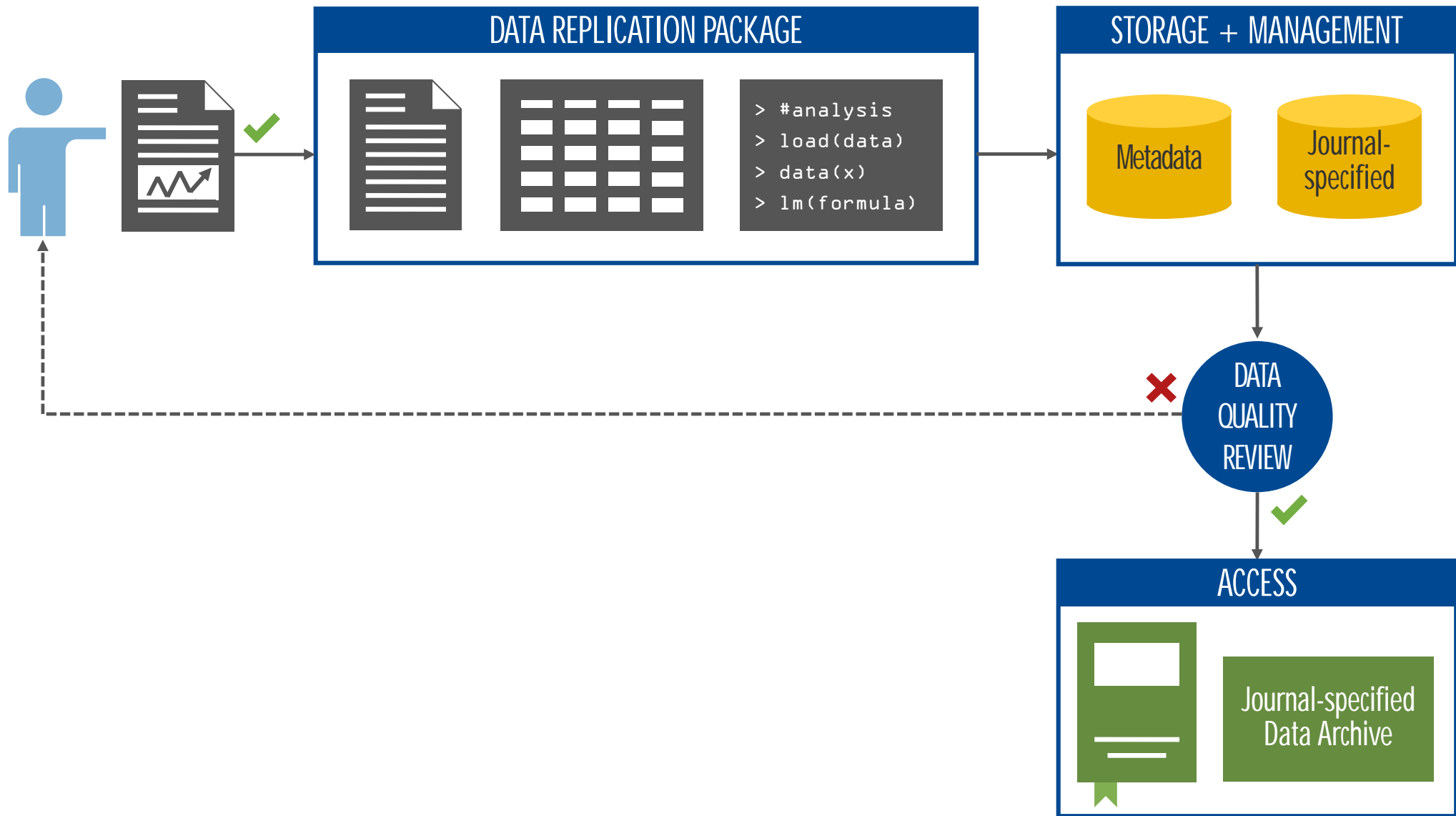


CURATING FOR REPRODUCIBILITY

MODELS OF PRACTICE

Odum Institute for Research in Social Science Enforcing Journal Data Replication Policies

- Founded in 1924, the Odum Institute is considered the oldest university-based interdisciplinary social science institute
- The Odum Institute hosts the open access UNC Dataverse
- Odum Institute data archivists and statisticians work together to offer data and code review services that support enforcement of robust journal data replication policies





AJPS AMERICAN JOURNAL
of POLITICAL SCIENCE

*AJPS, South Kedzie Hall, 368 Farm Lane, East Lansing, MI 48824
ajps@msu.edu, (517) 884-7836*

GUIDELINES FOR PREPARING REPLICATION FILES

Version 2.1, May 19, 2016

**William G. Jacoby
Robert N. Lupton**

Michigan State University

The American Journal of Political Science requires the authors of all empirical or quantitative articles to provide replication files when the article enters the production stage of the publication process. The replication files must be made available as a Dataset (i.e., a collection of files) on the Harvard Dataverse Network. Instructions for getting started are available in the "Quick Reference for Uploading Replication Files," located in the AJPS Dataverse on the Harvard Dataverse website. The replication files for each article must be located in the AJPS Dataverse on the Harvard Dataverse website. The replication files for each article must be located in the AJPS Dataverse on the Harvard Dataverse website.



http://www.ciser.cornell.edu/ASPs/search_athena.asp?IDTITLE=2782

The screenshot shows the Cornell University Cornell Institute for Social and Economic Research (CISER) website. The header includes the Cornell University logo and navigation links: About Us, Computing, Data, Training, and Support. The main content area is titled 'CISER Data Archive: Online Catalog'. On the left, there are links for 'About the Archive' (About Us, Location and Hours, News and Announcements, Policies) and 'Finding and Using Archive Data' (Search Archive Holdings, Browse Holdings by Subject, How to Locate Our Data). The main article is titled 'Eating Heavily: Men Eat More in the Company of Women' and includes bibliographic information: 'Kevin M. Kniffin, Ozge Sigirci, and Brian Wansink 2016. Evolutionary Psychological Science (2016) 2:38-46 [producer]. Springer International Publishing 2015 [distributor]. Codebook: R2E-KNIFFIN-2016. This study includes files created by Cornell researchers and/or staff.' There is a 'View Abstract' link and a 'User note' stating: 'The de-identified data (eliminating height, weight and age variables) can be found at: <https://doi.org/10.6077/J5CISER2783>'.

“Statistical heartburn: An attempt to digest four pizza publications from the Cornell Food and Brand Lab”

van der Zee, T., Anaya, J., & Brown, N. J. L. (2017). Statistical heartburn: An attempt to digest four pizza publications from the Cornell Food and Brand Lab. *PeerJ Preprints*, 5:e2748v1.

<https://doi.org/10.7287/peerj.preprints.2748v1>

partners, women eating with men tended to estimate themselves to have eaten more and reported feeling like they were rushed and overate. In addition to expanding upon previous research concerning women's eating behaviors, our findings concerning male overconsumption in the presence of women appear to present an example of self-handicap behavior.

The Stata code (Eating_Heavily_Script.do) and data (PizzaStudy.txt) associated with this study reproduced: a) Tables in Comment_Eating_Heavily.pdf that did not involve age, weight, and height variables, which were removed to de-identify the dataset; and b) output log appended at the bottom of the Comment_Eating_Heavily.pdf

Search Website

Search

SAMPLE STUDY

- The study was questioned for inconsistencies
- Authors could not locate their analysis code to reproduce the study
- To refute the criticism, authors had to hire:
 - A statistician to reproduce the study
 - An outside reviewer to review the text, tables, and Stata outputs
 - **CISER to reproduce the output produced by the statistician**
- Re-analysis re-affirmed signature findings of the study, although numbers were not replicated

```

*****
***** Eating Heavily *****
*****

clear
log using "<path>\Eating_Heavily.smcl", replace text
import delimited "<path>\PizzaStudy.txt"

//Labeling the variables
label variable treatment "The manipulation group"
label define treatment1 1 "$4" 2 "$8"
label value treatment treatment1
label variable pieces "How many pieces of pizza did you eat
today?"
label variable gender "Gender"
label define gender1 1 "Male" 2 "Female"

***** Table 1 - Descriptive statistics of the sample
tab mmff
ttest age if mmff ==1 | mmff == 2, by(mmff) unequal
ttest age if mmff ==3 | mmff == 4, by(mmff) unequal

```

COMMAND FILE

- ✓ Curate prior to processing analytical code
- ✓ Label all variables and values
- ✓ Comment code to describe processes and map to paper sections
- ✓ Order code outputs in the same order as they appear in paper
- ✓ Anonymize file paths

```
name: <unnamed>
log: <path>\Eating_Heavily.smcl
log type: text
opened on: 27 Mar 2017, 13:00:06
. import delimited "<path>\PizzaStudy.txt"
(30 vars, 139 obs)
.
. //Labeling the variables
. label variable treatment "The manipulation group"
. label define treatment1 1 "$4" 2 "$8"
. label value treatment treatment1
. label variable pieces "How many pieces of pizza did you
eat today?"
. label variable gender "Gender"
. label define gender1 1 "Male" 2 "Female"
.
. // Anova results in the text
. anova pieces mmff if mmff == 1 | mmff == 2 // pizza
consumption - males eating with males or females
      Number of obs = 65 R-squared = 0.1574
      Root MSE = 1.62753 Adj R-squared = 0.1441
```

COMPARISON OUTPUT FILE

- ✓ Produce comparison output file (i.e., log file) to document results of code review
- ✓ Share comparison output file to enable re-users to compare it to their output and be confident that they have processed the materials for reproduction correctly

PACKAGING THE MATERIALS

Eating Heavily: Men Eat More in the Company of Women



Bibliographic Information:

Kevin M. Kniffin, Ozge Sigirci, and Brian Wansink 2016. Evolutionary Psychological Science (2016) 2:38-46 [producer]. Springer International Publishing 2015 [distributor]. Codebook: R2E-KNIFFIN-2016. This study includes files created by Cornell researchers and/or staff.

[View Abstract](#)

User note: The de-identified data (eliminating height, weight and age variables) can be found at: <https://doi.org/10.6077/J5CISER2783>

File Information:

Type of File	Directory \ File Name	Size / Size Zipped
 Documentation	V:\r2e\KNIFFIN-2016\Comment_Eating_Heavily.pdf	363 KB / 331 KB
 Stata Program	V:\r2e\KNIFFIN-2016\Eating_Heavily_Script.do	7 KB / 2 KB

ANALYSIS DATASET

COMPARISON OUTPUT

COMMAND FILE

Abstract: Sexual selection has been commonly considered by evolutionary psychologists interested in eating disorders among women;

men. We present the results of a field study through which we find that men eat more food when sharing a meal with women than with men. Notably, men appear to eat larger quantities of both unhealthy (pizza) and healthy (salad) food when in the company of women. More specifically, men eating with women ate 93% more pizza (1.44 more slices) and 86% more salad. Additionally, while women do not eat significantly differently as a function of the sex of their dining partners, women eating with men tended to estimate themselves to have eaten more and reported feeling like they were rushed and overate. In addition to expanding upon previous research concerning women's eating behaviors, our findings concerning male overconsumption in the presence of women appear to present an example of self-handicap behavior.

however, comparable attention has not been paid to problematic eating by men. The Stata code (Eating_Heavily_Script.do) and data (PizzaStudy.txt) associated with this study reproduced: a) Tables in the Appendix showing the variables, which were removed to de-identify the dataset; and b) output.log appended at the bottom of the Comment_Eating_Heavily.pdf

associated with this study reproduced: a) Tables in the Appendix showing the variables, which were removed to de-identify the dataset; and b) output.log appended at the bottom of the Comment_Eating_Heavily.pdf

The Stata code (Eating_Heavily_Script.do) and data (PizzaStudy.txt) associated with this study reproduced: a) Tables in the Appendix showing the variables, which were removed to de-identify the dataset; and b) output.log appended at the bottom of the Comment_Eating_Heavily.pdf

http://www.ciser.cornell.edu/ASPs/search_athena.asp?IDTITLE=2782

The screenshot shows the Cornell University CISER Data Archive website. The header includes the Cornell University logo and navigation links for About Us, Computing, Data, Training, and Support. The main content area displays the title 'Eating Heavily: Men Eat More in the Company of Women' and provides bibliographic information: Kevin M. Kniffin, Ozge Sigirci, and Brian Wansink (2016). It also includes a 'View Abstract' link and a 'User note' regarding de-identified data. A table lists files for download, including 'Documentation' (363 KB / 331 KB zipped) and 'Stata Program' (7 KB / 2 KB zipped).

Type of File	Directory \ File Name	Size / Size Zipped
Documentation	V:\r2e\KNIFFIN-2016\Comment_Eating_Heavily.pdf	363 KB / 331 KB
Stata Program	V:\r2e\KNIFFIN-2016\Eating_Heavily_Script.do	7 KB / 2 KB

ADVANTAGES OF SHARING DATA & CODE

- Research transparency
- Accelerate advancement of science
- No one asking you for access to data and code

It can be stressful when someone requests your data and code and you are not confident about their quality—or if you can't find them. **Your reputation could suffer!**

Abstract: Sexual selection has been commonly considered however, comparable attention has not been paid to problem and that men eat more food when sharing a meal with women (pizza) and healthy (salad) food when in the company of women (pizza) and 86% more salad. Additionally, while women do not eat more food with male partners, women eating with men tended to estimate themselves as eating more food in addition to expanding upon previous research concerning the presence of women appear to present an example of self-harm.

The Stata code (Eating_Heavily_Script.do) and data (PizzaSlices_2010-2011.dta) are available for download. The data are in a format that does not include the names of the participants. The data also include the output log appended at the bottom of the Comment_Eating_Heavily.pdf that did not involve age, weight, and height variables, which were removed to de-identify the dataset; and

http://www.ciser.cornell.edu/ASPs/search_athena.asp?IDTITLE=2782

The screenshot displays the Cornell University CISER Data Archive website. The header includes the Cornell University logo and navigation links for About Us, Computing, Data, Training, and Support. The main content area is titled "CISER Data Archive: Online Catalog" and features a sidebar with navigation options like "About the Archive" and "Finding and Using Archive Data". The main section displays search results for the study "Eating Heavily: Men Eat More in the Company of Women" by Kevin M. Kniffin, Ozge Sigirci, and Brian Wansink (2016). It includes bibliographic information, a table of file information, and a snippet of the abstract.

Eating Heavily: Men Eat More in the Company of Women

Bibliographic Information:
Kevin M. Kniffin, Ozge Sigirci, and Brian Wansink 2016. Evolutionary Psychological Science (2016) 2:38-46 [producer]. Springer International Publishing 2015 [distributor]. Codebook: R2E-KNIFFIN-2016. This study includes files created by Cornell researchers and/or staff.

View Abstract
User note: The de-identified data (eliminating height, weight and age variables) can be found at: <https://doi.org/10.6077/J5CISER2783>

Type of File	Directory \ File Name	Size / Size Zipped
Documentation	V:\r2e\KNIFFIN-2016\Comment_Eating_Heavily.pdf	363 KB / 331 KB
Stata Program	V:\r2e\KNIFFIN-2016\Eating_Heavily_Script.do	7 KB / 2 KB
		Opinion Research

Abstract: Sexual selection has been commonly considered however, comparable attention has not been paid to problem and that men eat more food when sharing a meal with women (pizza) and healthy (salad) food when in the company of women (pizza) and 86% more salad. Additionally, while women do not eat more food with their partners, women eating with men tended to estimate themselves as eating more. In addition to expanding upon previous research concerning the presence of women appear to present an example of self-harm.

The Stata code (Eating_Heavily_Script.do) and data (PizzaS...
Comment_Eating_Heavily.pdf that did not involve age, weight, and height variables, which were removed to de-identify the dataset; and
b) output log appended at the bottom of the Comment_Eating_Heavily.pdf

ADVANTAGES OF SHARING DATA & CODE

- a** Research transparency
- a** Accelerate advancement of science
- a** No one asking you for access to data and code

It can be stressful when someone requests your data and code and you are not confident about their quality—or if you can't find them. Your reputation could suffer!

R² SERVICE REQUIREMENTS: ARTICLE

- ➔ Highlight all sections (e.g., paragraphs, sentences, tables, charts) that reference output derived from your data.

(C→B→M vs. C→B). According to life table estimates treating separation as a competing risk, the share of cohabiting parents who married after having a child dropped from 59% in the earlier period to 48% in the later period; of those marrying, the average duration to marriage increased

ut here > < Table 1 abo

cohabitation between the 1995 and 2006-
years prior to interview, the share
decline was almost completely offset by a
n 17% to 36%. Table 1 also shows
en by marital status at birth. Births to
ne college educated. Half of these married
od (compared to 28% of married mothers
p the educational ranks as well, but the
Of those cohabiting at birth, there has
some college (with the some college
accounted for 5% or less of all
ences have been much less significant, to

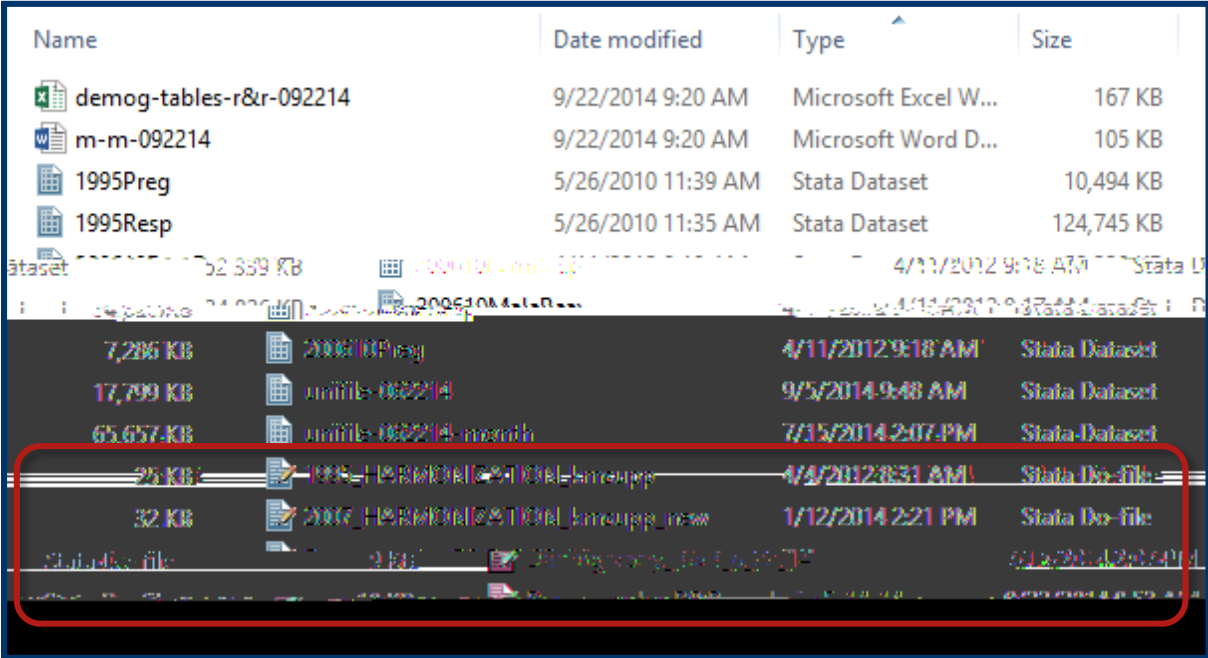
Table 1 shows a striking shift from marriage to
2010 surveys. Among unions bearing children in the 10
married at union start dropped from half to 30%. This d
doubling of the share who were cohabiting at birth: from
substantial shifts in the education distributions of wom
married couples are increasingly concentrated among th
mothers were college graduates in the more recent peri
in the 1995 period). Cohabiting mothers have moved up
progression stops short of college in both time periods:
been a shift from mothers with a high school degree to
group increasing from 17% to 29%). College graduates
cohabiting births in both periods.
Changes in prior union and childbearing experie

R² SERVICE REQUIREMENTS: CODE

- ➔ Specify the sequence of execution if it consists of multiple files. Prefix the filename with Step #.
- ➔ Add comments that map sections of code to results in paper. Make sure every command that generates results is preceded by a comment that indicates which result the command generates. For example:

*The following command generates column 1 of Table 1

*The following command generates the mean age mentioned on page 3, paragraph 3



Name	Date modified	Type	Size
demog-tables-r&r-092214	9/22/2014 9:20 AM	Microsoft Excel W...	167 KB
m-m-092214	9/22/2014 9:20 AM	Microsoft Word D...	105 KB
1995Preg	5/26/2010 11:39 AM	Stata Dataset	10,494 KB
1995Resp	5/26/2010 11:35 AM	Stata Dataset	124,745 KB
2005Preg	4/11/2012 9:18 AM	Stata Dataset	7,286 KB
2005Resp	4/11/2012 9:18 AM	Stata Dataset	17,799 KB
2007_HARMONIZATION_kmeupp	4/1/2012 8:31 AM	Stata Do-file	25 KB
2007_HARMONIZATION_kmeupp_new	1/12/2014 2:21 PM	Stata Do-file	32 KB

R² SERVICE REQUIREMENTS: DATA

- ➔ Free of errors and inconsistencies
- ➔ All variables and values labeled
- ➔ Data are anonymized (if needed)

Variables Manager

Filter variables here

Drag a column header here to group by that column.

#	Name	Label	Type	Format	Value label	Notes
	am_hungry	I am hungry now	byte	%8.0g		
	feel_guilty	I feel guilty about how much I...	byte	%8.0g		
	portion_control	I use portion-control methods	byte	%8.0g		
	overate	I overate	byte	%8.0g		
	ate_more_general	I ate more than I should have	byte	%8.0g		
	felt_rushed	I felt rushed	byte	%8.0g		
	mmff	The type of groups	byte	%27.0g	mmff1	
	salad	Mark the amount of salad you...	float	%9.0g		
	calories	The amount of calories that p...	int	%8.0g		
	mixedgroup	The type of group	byte	%8.0g	yes_no	
	male_1	The number of males in groups	byte	%8.0g		
	group	Number of people in the group	byte	%8.0g		
	id	The ID of participants for res...	int	%8.0g		
	height_cm	Height in cm	float	%9.0g		
	weight_kg		float	%9.0g		
	bmi	BMI	float	%9.0g		
	male_c	With whom male participants ate	float	%38.0g	male_c1	

Variable properties

Name: treatment

Label: The amount of calories that p...

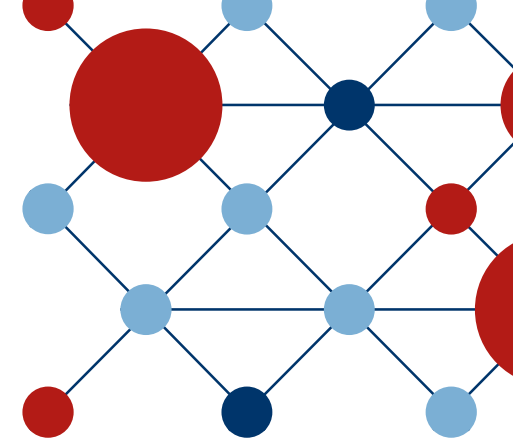
Type: byte

Format: %8.0g Create...

Value label: treatment1 Manage...

Notes: No notes Manage...

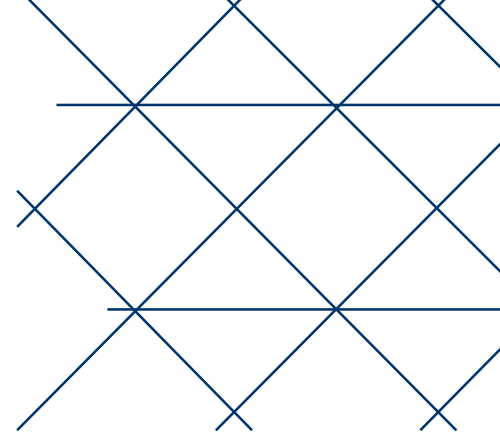
Vars: 40 CAP NUM



HANDS-ON DATA AND CODE REVIEW

PART 1: 20 MINS

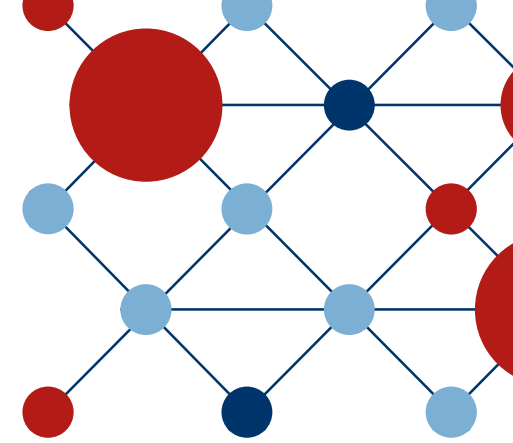
- ➔ Get a hard-copy of the 1st two pages of from the workshop instructors
- ➔ Open [this file](#) and go to page 4. The section that begins with START HERE marks the beginning of the output produced by the code
- ➔ Compare the output produced by the code to that of the paper. The comments on the command file will tell you which section of the paper the output refers to. On the paper, the table displays the old and new values. Compare the output to the new values, which are the below figures.
- ➔ Note the problems, issues, and inefficiencies encountered while comparing the output.



HANDS-ON DATA AND CODE REVIEW

PART 2: 15 MINS

- ➔ Discuss the problems, issues, and inefficiencies encountered while comparing the output
 - Table 1
 - Results in the text
 - Table 2
 - Table 3

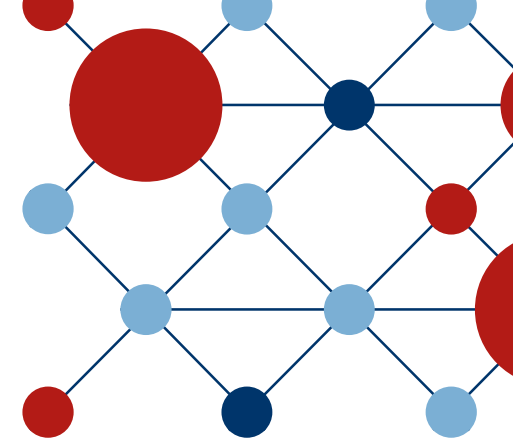


HANDS-ON DATA AND CODE REVIEW

PART 3: 5 MINS

- ➔ Show final code that addressed the issues
- ➔ Get a hard-copy of the 1st five pages of workshop instructors from the
- ➔ Open [this link](#) and go to page 9 and review the contents of the log file.
 - The variables now have variable and value labels
 - As soon as variables are created, they are labeled
 - Well commented code, you know what the code is doing
 - Code produced output that followed the order of the paper
 - Comparing table output is now easier in the eyes, more efficient, and not confusing





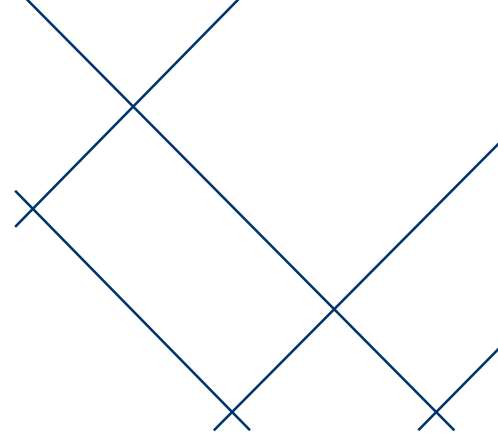
CURATION TOOL: YARD

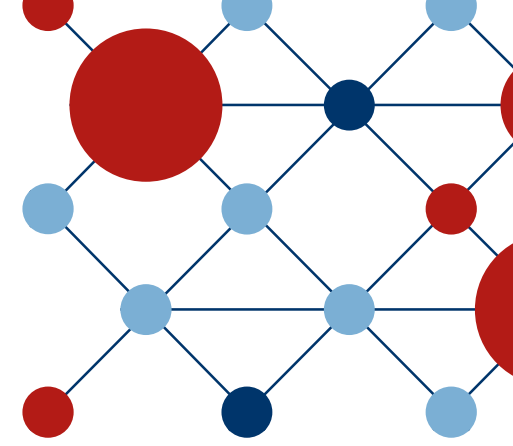
YALE APPLICATION FOR RESEARCH DATA

- Conceptualized by the Yale University Institution for Social and Policy Studies (ISPS) and Innovations for Poverty Action (IPA)
- Developed by Colectica
- Development begins 2014; Production and code release in 2017

CURATION TOOL: YARD

YALE APPLICATION FOR RESEARCH DATA





CURATION TOOL: DEMONSTRATION

YALE APPLICATION FOR RESEARCH DATA (YARD)

Log in

Yale
ISPS

Log in to the ISPS Data Curation Tool with your username and password.

Don't have a ISPS Data Curation Tool account?
[Create an account.](#)

Email

Password

Remember me

[Forgot your password?](#)

Log in

Joshua Dull, Research Data Support Specialist
Center for Science and Social Science Information
Yale University