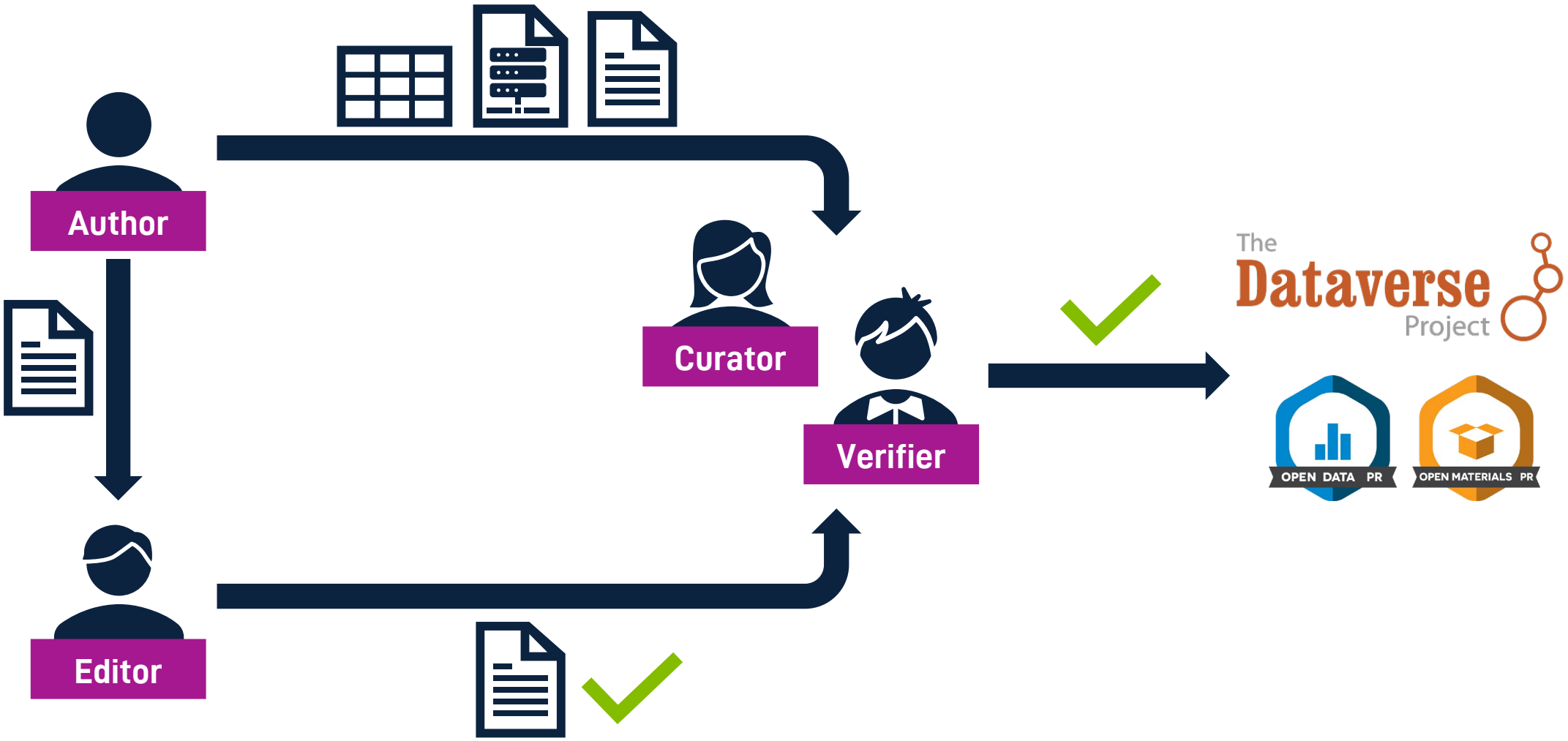
Supporting Computational Reproducibility:

Updates on CoRe2 Development and AJPS Verification Policy Implementation

Mandy Gooch, Research Data Archivist

UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

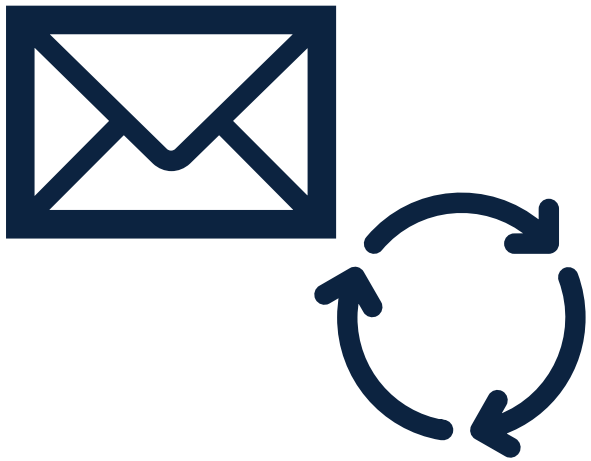
Publication & Data Curation + Verification



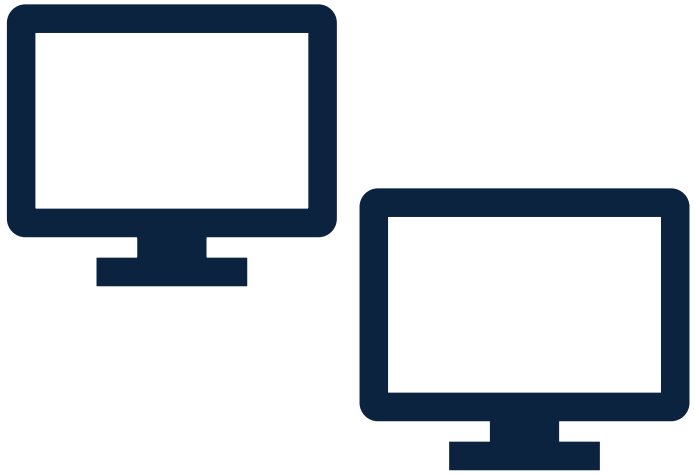
UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

ADMINISTRATION

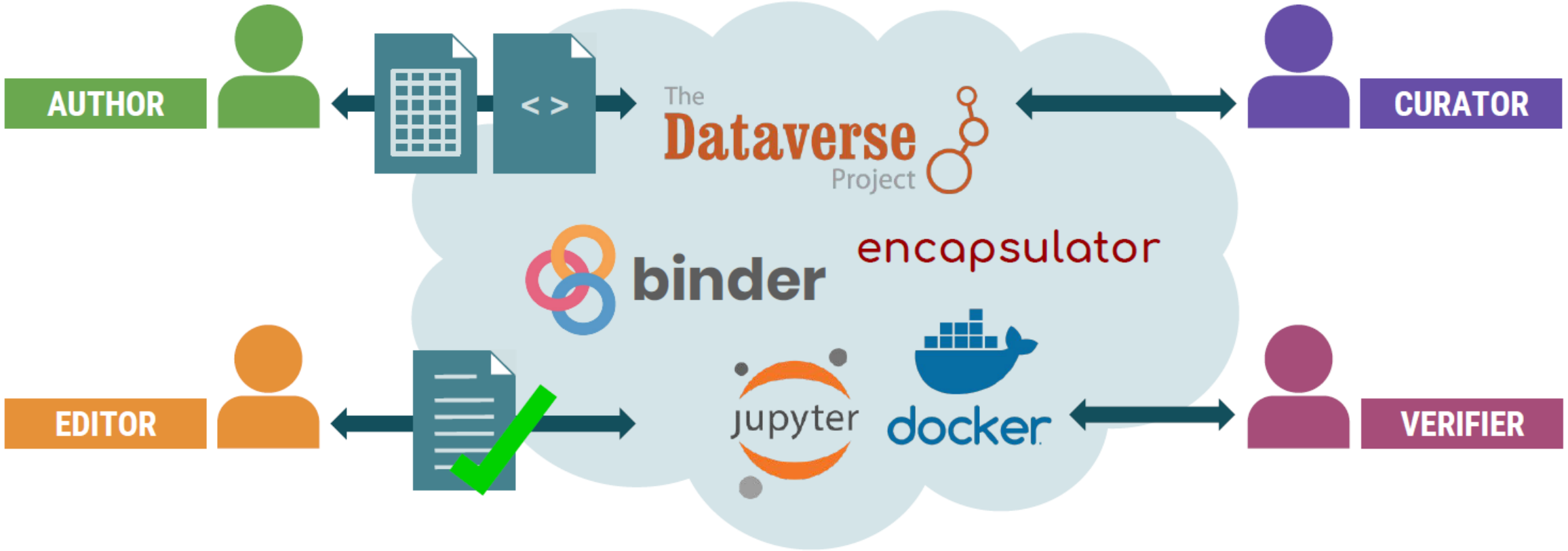
COORDINATION



COMPUTATION



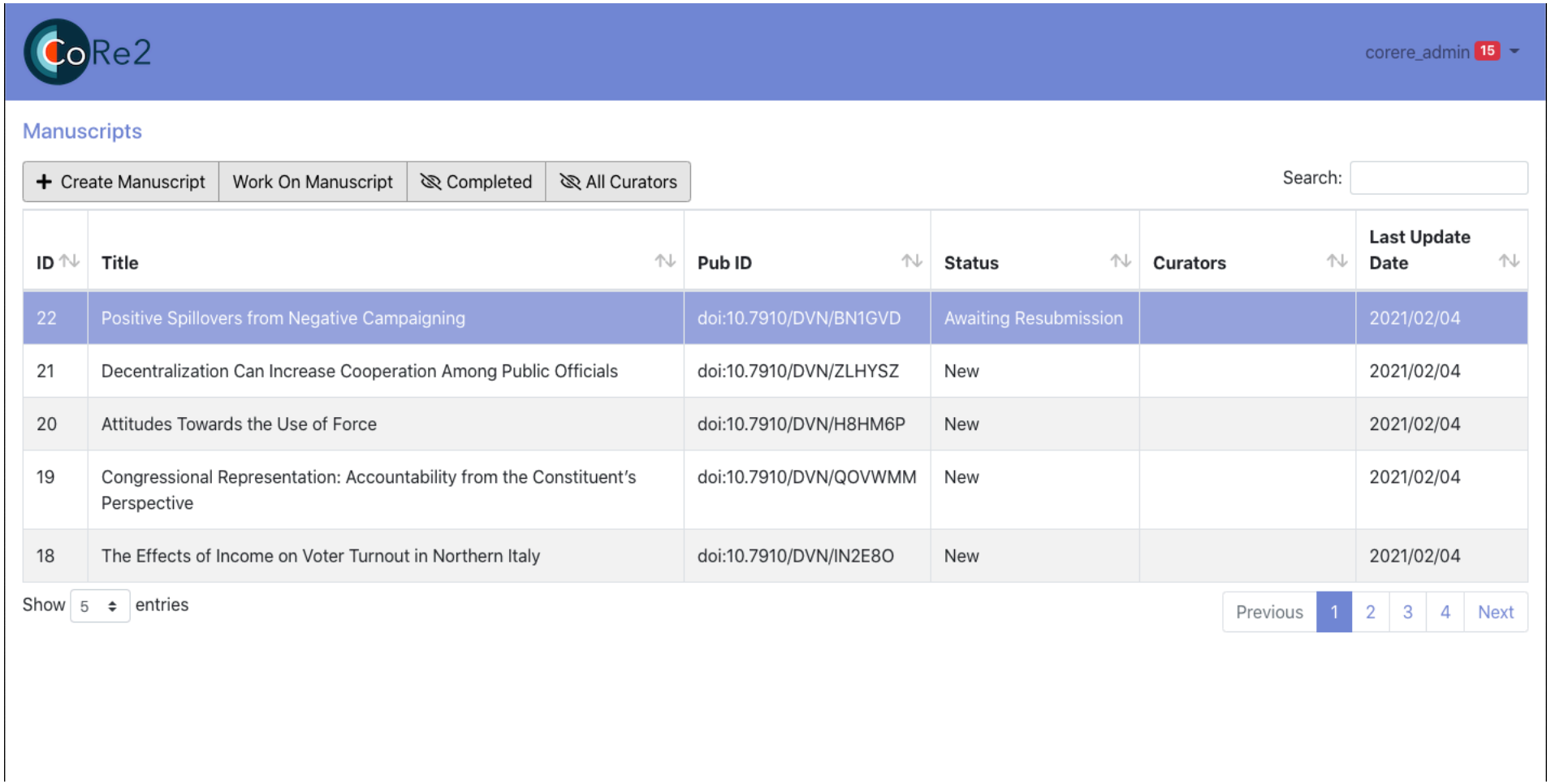
Confirmable Reproducible Research (CoRe2) Environment



ALFRED P. SLOAN FOUNDATION

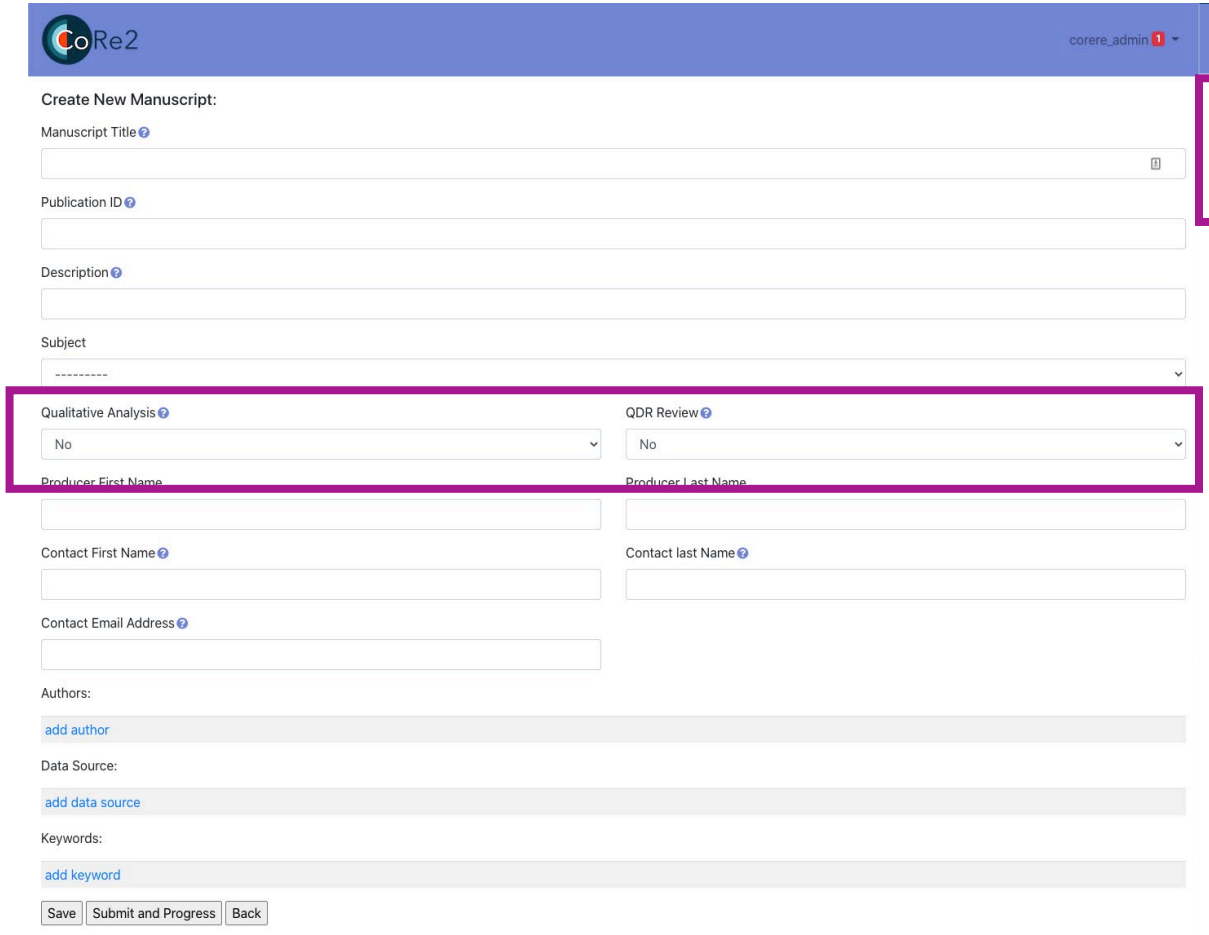
UNC ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

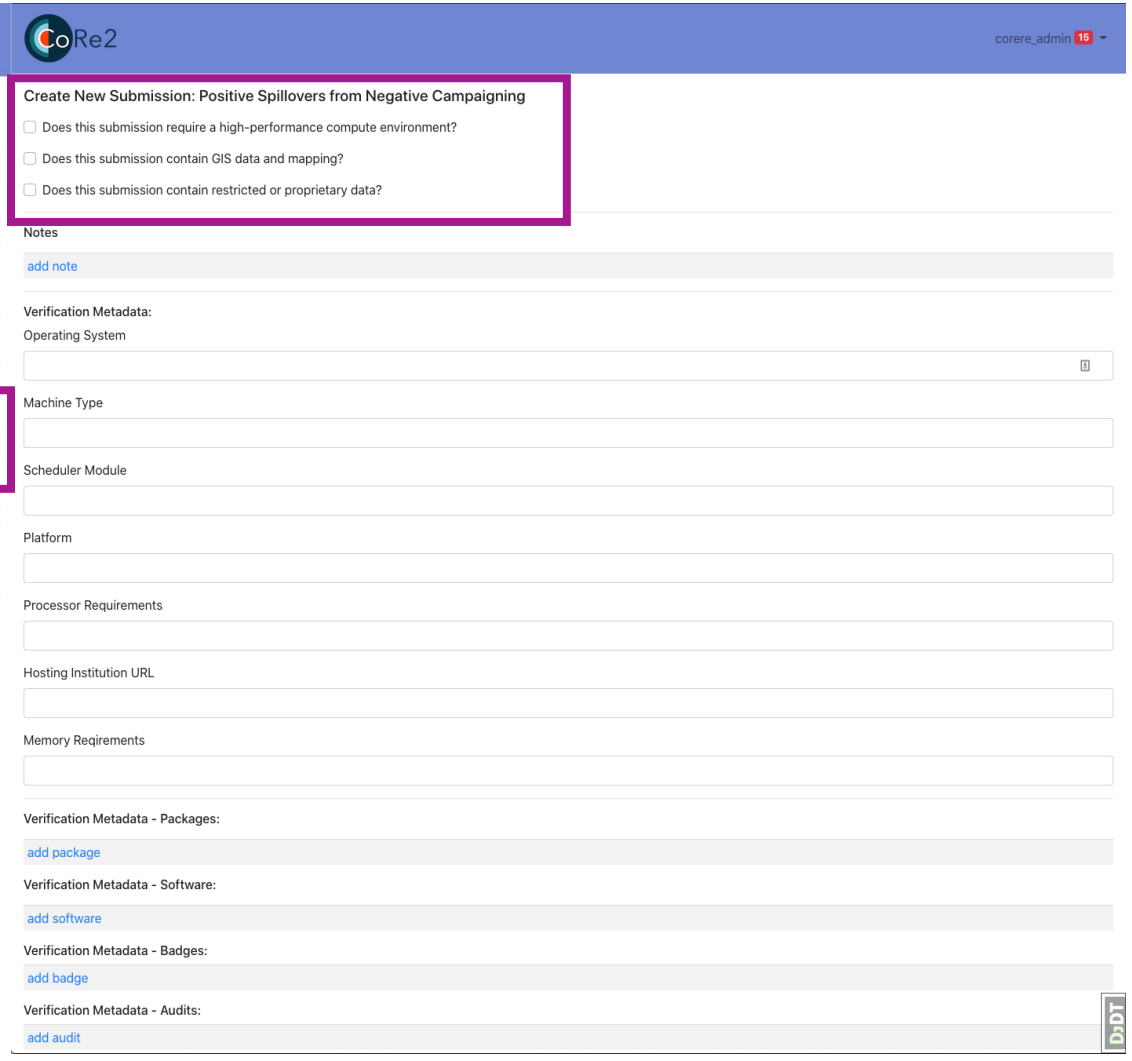
Confirmable Reproducible Research (CoRe2) Environment



UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

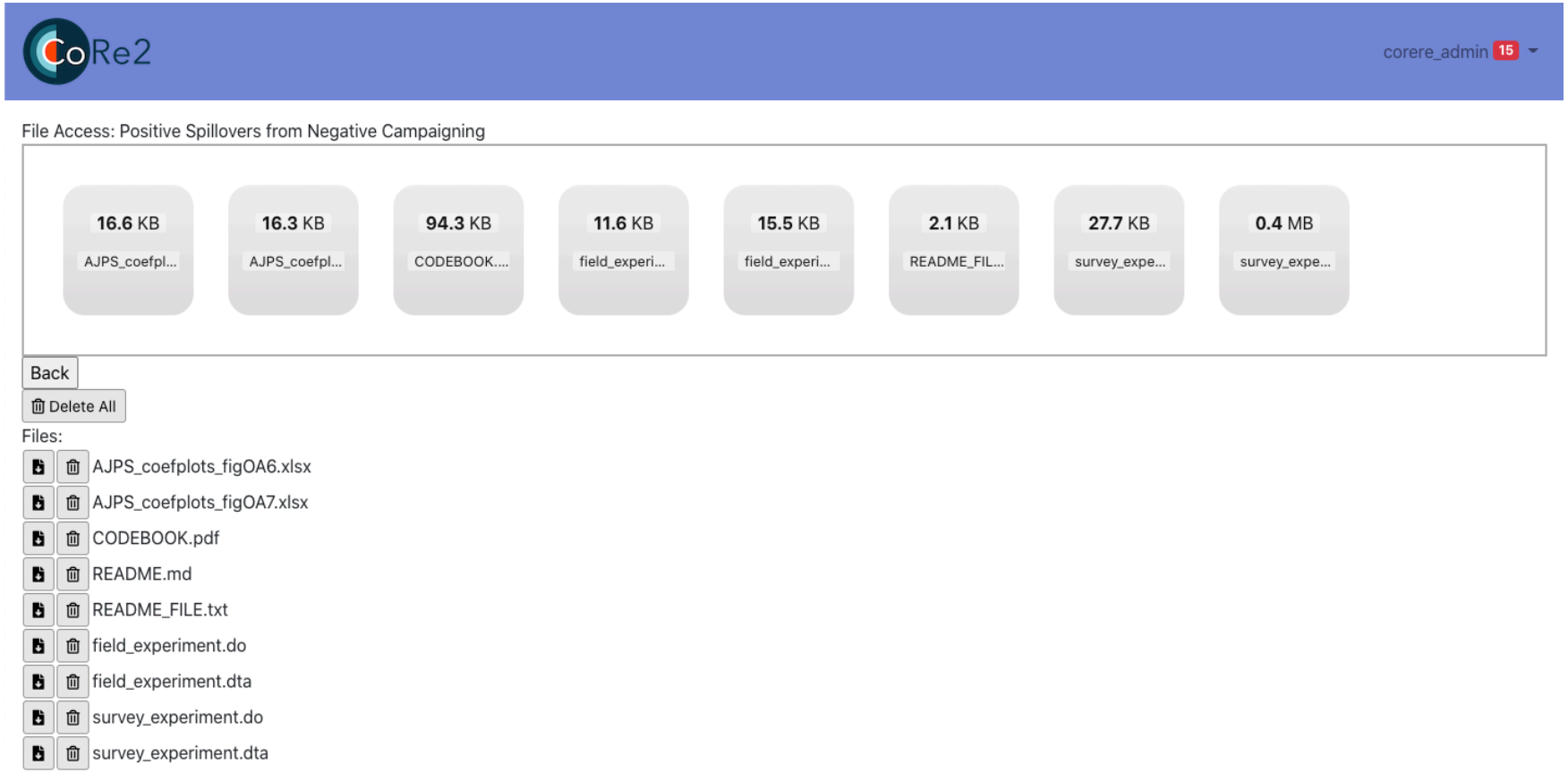
Confirmable Reproducible Research (CoRe2) Environment





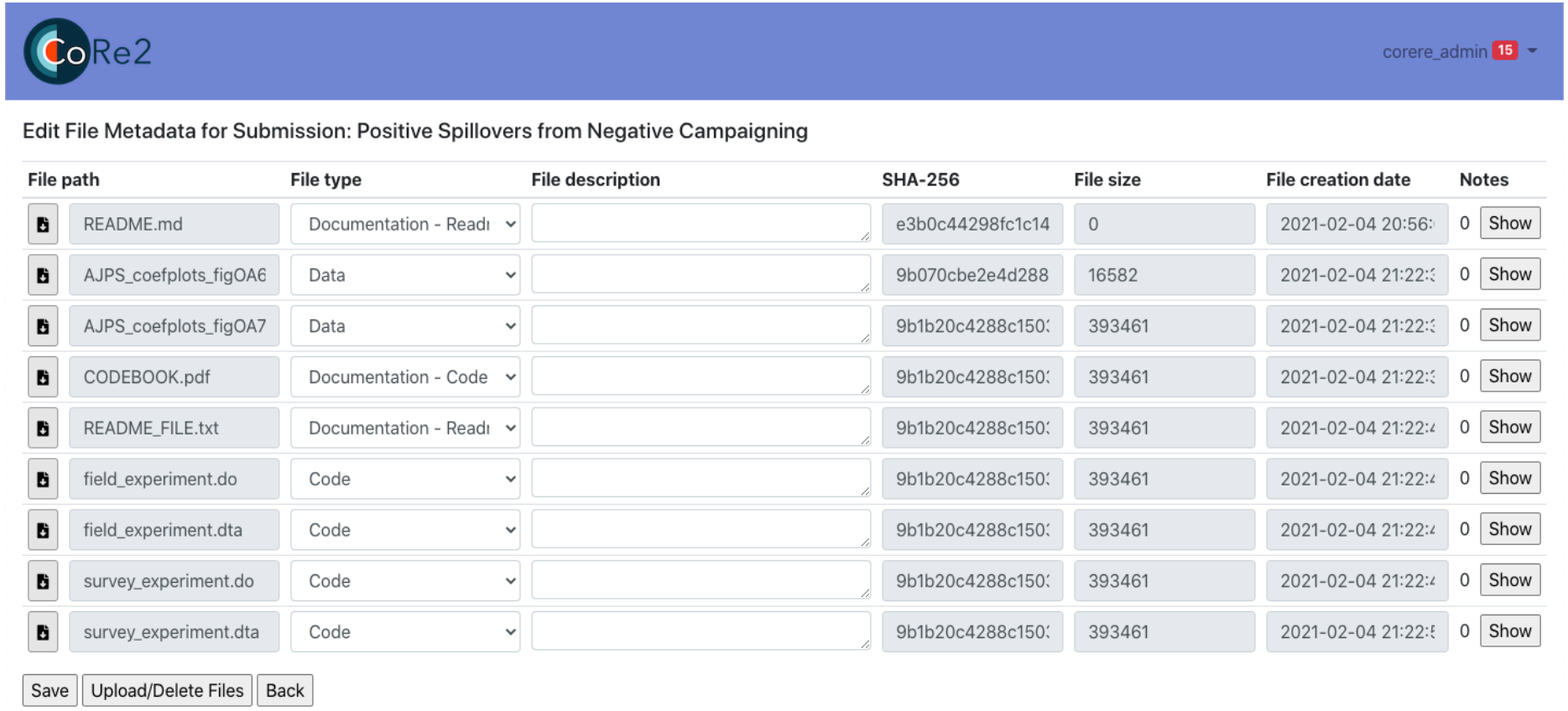
UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

Confirmable Reproducible Research (CoRe2) Environment



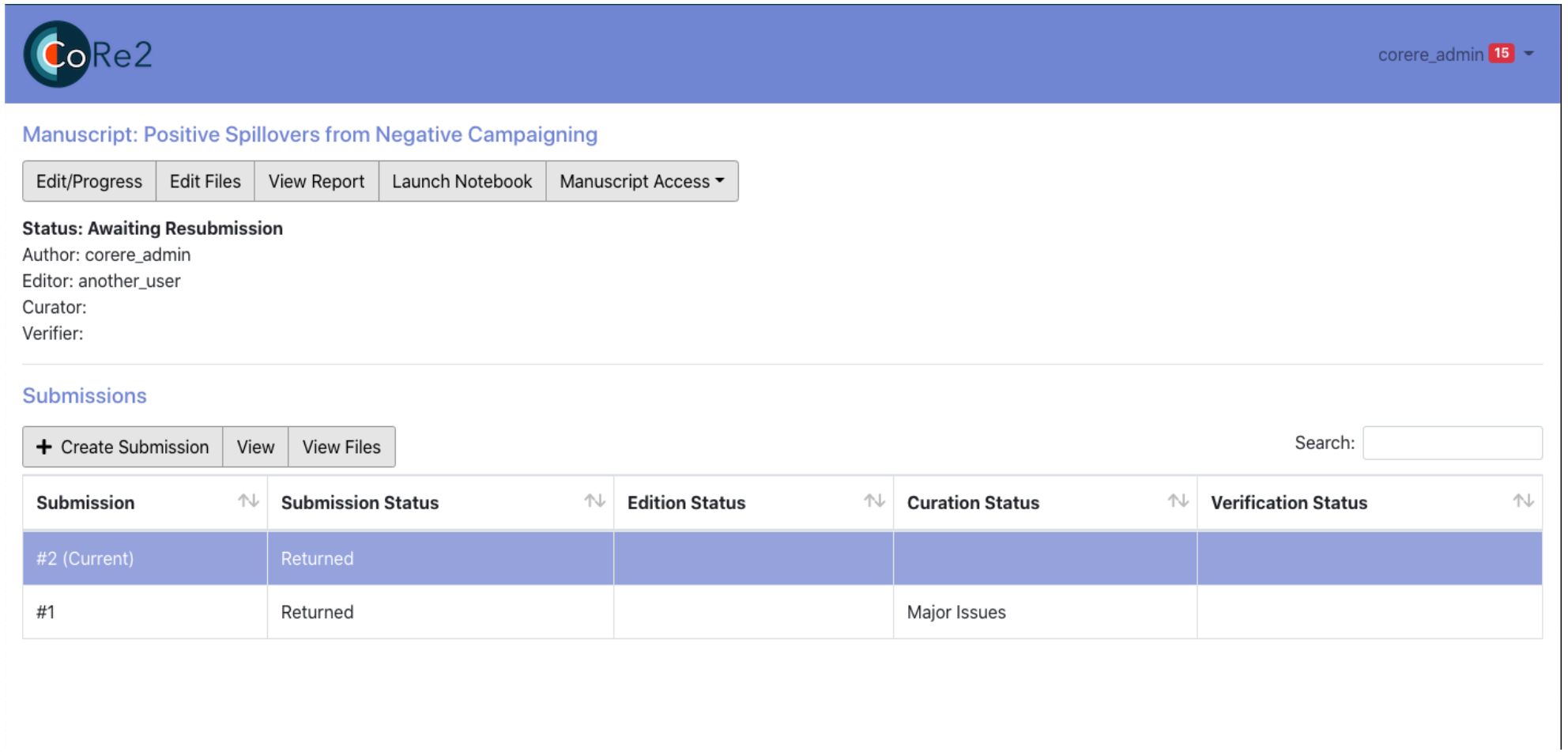
UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

Confirmable Reproducible Research (CoRe2) Environment



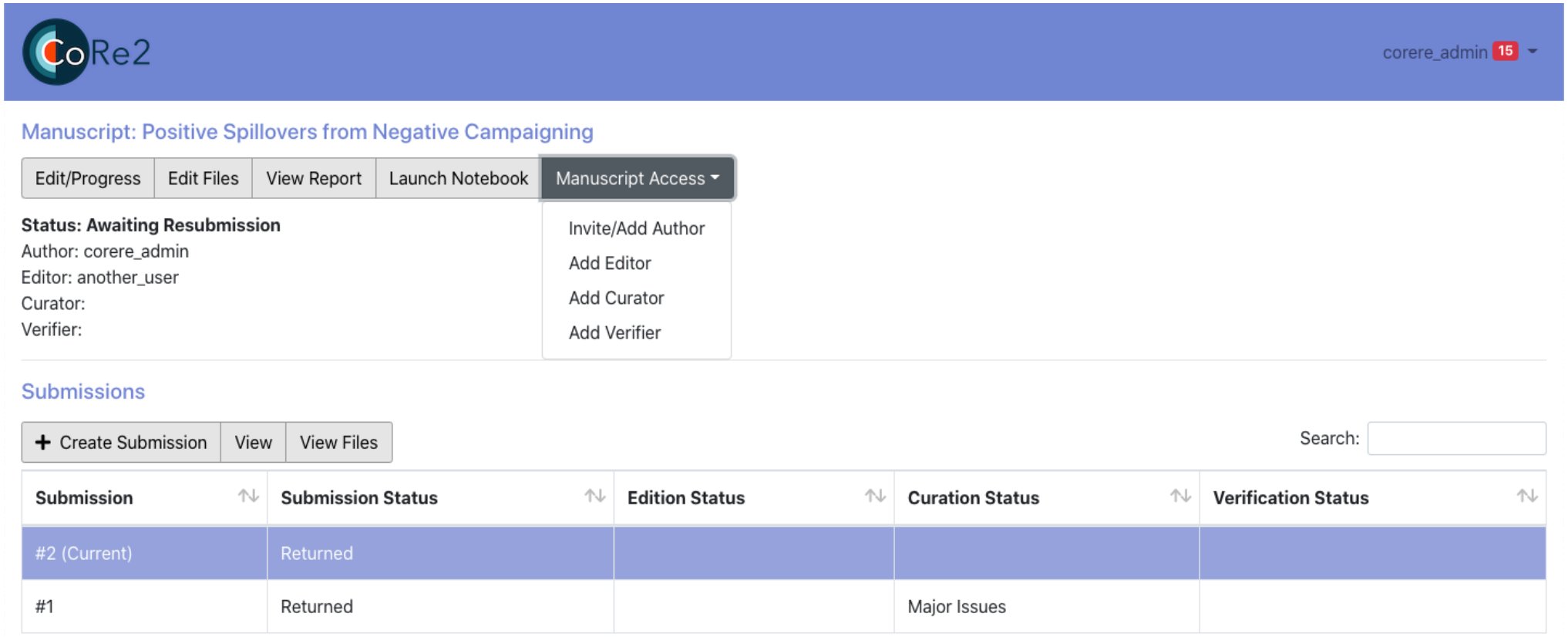
UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

Confirmable Reproducible Research (CoRe2) Environment



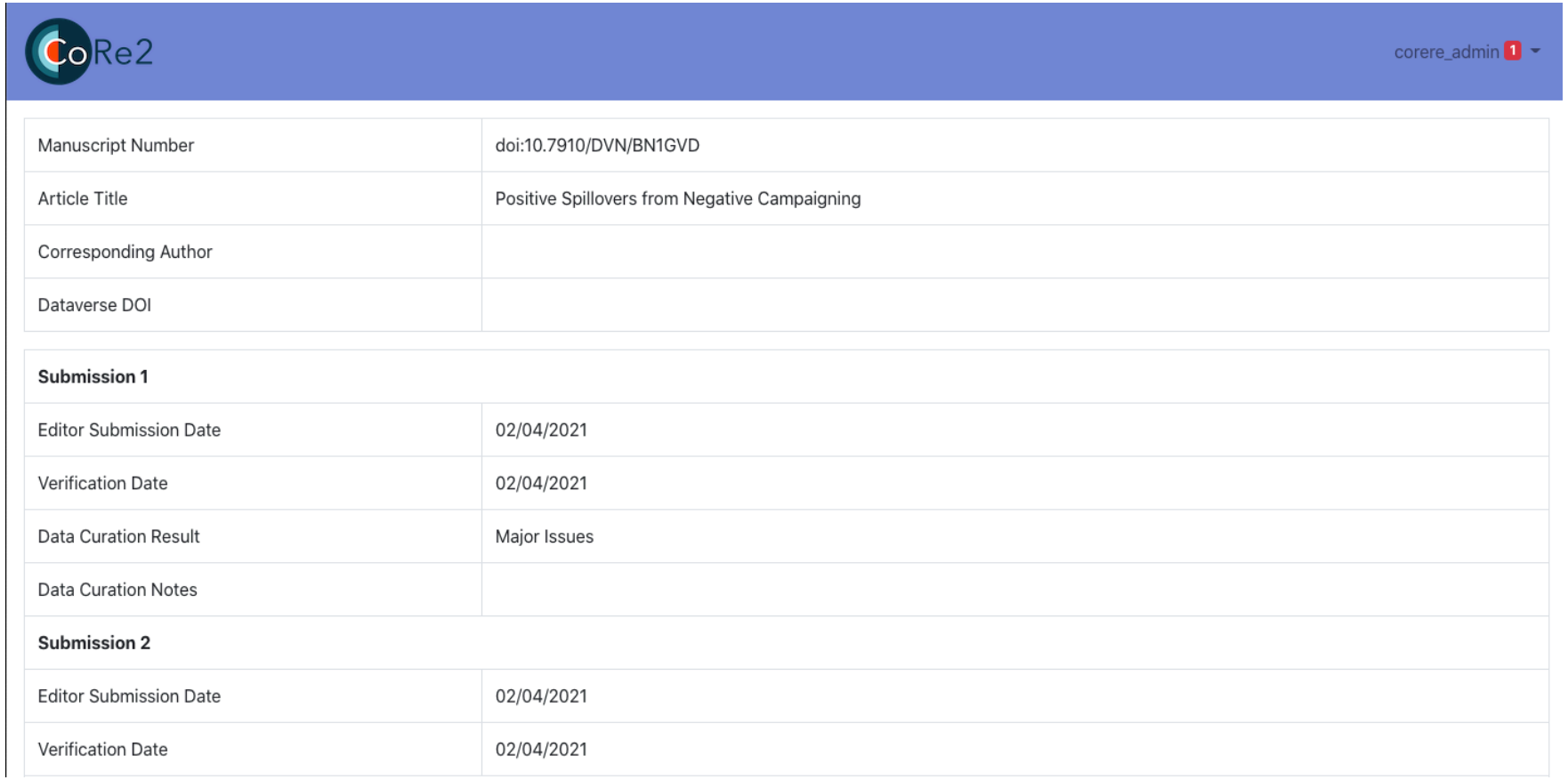
UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

Confirmable Reproducible Research (CoRe2) Environment



UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

Confirmable Reproducible Research (CoRe2) Environment



Manuscript Number	doi:10.7910/DVN/BN1GVD
Article Title	Positive Spillovers from Negative Campaigning
Corresponding Author	
Dataverse DOI	

Submission 1	
Editor Submission Date	02/04/2021
Verification Date	02/04/2021
Data Curation Result	Major Issues
Data Curation Notes	
Submission 2	
Editor Submission Date	02/04/2021
Verification Date	02/04/2021

UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

Confirmable Reproducible Research (CoRe2) Environment

TERMINOLOGY

NOTIFICATIONS & TOOL TIPS

METADATA

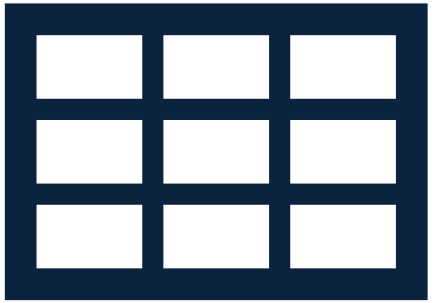
METRICS & REPORTS

Common Issues

README

CODEBOOK

ANALYSIS DATA



README

UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

ODUM INSTITUTE
DATA ARCHIVE

AJPS README Guide

Essential Information:

1. Title and Author(s) of article
2. Compute environment dependencies:
   a. operating system
   b. statistical software version
   c. packages used in analysis
3. List of all files within your verification package in AJPS Dataverse and descriptions of each file
4. Must be saved as either a .TXT or .PDF
5. Full data citations and access instructions for all original data sources in the README or Codebook.

README Template

```
[Article Title]
[Author(s)] – [Affiliation] – [E-mail] – [Twitter (optional)]

Compute Environment:
[Operating system]
[Statistical software versions]
[Estimated run time for all analyses]
[Packages used in analysis]
```

R PRO-TIP

As part of the base package the command "R.Version" lists the R version and the computer platform.

After a script has been run, the command "(.packages())" will list all of the packages that have been loaded during the session.

STATA PRO-TIP

The 'version' command will tell you which version you are currently using. The command 'dis ""c(machine_type)"" will also tell you if you are using a 64 or 32 bit machine type. Please also be sure to include the type of Stata being used (IC, SE, or MP).

README

UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

STATA PRO-TIP

You can open your .do file on a clean install of Stata that does not have any user packages (such as your University virtual lab, etc.) and run your script to see if there are any errors that pop up due to missing packages.

Alternatively, you can also use the "-ado-" command which will list all user programs that have been installed on your Stata version, however these are not necessarily packages from the list that are actually used in your script.

IMPORTANT

If your analyses were run on a high-performance compute (HPC) environment, please provide the following information:

The script(s) used to schedule the resources (i.e. SLURM or PBS scripts, etc.). These scripts include important information such as the memory needed, the time needed and the number of processes to run (nodes, tasks, etc.).

List of files:

(Note: if there is a folder hierarchy you can list the files in that structure here

DATA FILES

- Datafile.dta – main analysis dataset constructed from original source data; this data is used to generate the results within the manuscript

- Appendixdatafile.dta – data used to generate results within appendix

- Originalsourcedata.dta – original source data used to compile main analysis data set. Source: AuthorLastName, AuthorFirstName. Title of Dataset, Year. Name of distributor [distributor], date of distribution as YYYY-MM-DD. DOI or persistent URL

- Originalsourcedata1.dta – original source data used to compile main analysis data set. Source: AuthorLastName, AuthorFirstName. Title of Dataset, Year. Name of distributor [distributor], date of distribution as YYYY-MM-DD. DOI or persistent URL

- Originalsourcedata2.dta – original source data used to compile main analysis data set. Source: AuthorLastName, AuthorFirstName. Title of Dataset, Year. Name of distributor [distributor], date of distribution as YYYY-MM-DD. DOI or persistent URL

IMPORTANT

If original source data cannot be shared within Dataverse due to access restrictions or proprietary nature of the data, please provide in-depth instructions for requesting access to data and the exact file names you requested. Please also make sure to inform the editors that you cannot share specific data due to copyright and access restrictions.

README

CODE

— Mainanalysis.do - uses datafile.dta to construct Figure 1, Figure 2, and Table 3 from the manuscript

— Appendixanalysis.do - uses Appendixdatafile.dta to construct Appendix Table 1, Table 2, and Figure 1

— Constructanalysisdata.do - uses originalsourcedata.dta, or originalsourcedata1.dta, and original sourcedata2.dta to construct datafile.dta, which is used to create manuscript results.

PRO-TIP

If using programs that do not have a coding environment (such as Excel, ArcGIS, QGIS, GRASS, etc.), you must instead provide detailed, step-by-step instructions for creating results using these programs.

DOCUMENTATION

— README.txt - This file, which lists all files and compute environment dependencies necessary to verify the results from this manuscript

— Codebook.PDF - codebook for datafile.dta, the analysis dataset used to generate results from this manuscript.
(Note: If citations for original source data are not included in your README, they must be included in your codebook. See the AJPS Codebook Resource for more information and examples).

Examples of Complete and Successful READMEs

1. Slough, Tara; Fariss, Christopher, 2020, "Replication Data for: Misgovernance and Human Rights: The Case of Illegal Detention without Intent", https://doi.org/10.7910/DVN/Q5PV4U, Harvard Dataverse, V1 README

2. Becher, Michael; Brouard, Sylvain, 2020, "Replication Data for: Executive Accountability Beyond Outcomes: Experimental Evidence on Public Evaluations of Powerful Prime Ministers", https://doi.org/10.7910/DVN/B9JAV0, Harvard Dataverse, V1, UNF:6:g64TwQKsliSH+90FpwQjsQ== [fileUNF] README

3. Bisgaard, Martin; Rune Slothuus, 2020, "Replication Data for: How Political Parties Shape Public Opinion in the Real World", https://doi.org/10.7910/DVN/Z5BTCQ, Harvard Dataverse, V1, UNF:6:FQVZkZHntORBI5mWkh3MdA== [fileUNF] README

UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

CODEBOOK

UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

ODUM INSTITUTE
DATA ARCHIVE

AJPS Codebook Guide

Essential Information:

1. Title and Author(s) of article
2. If the codebook covers multiple datasets, please divide the codebook into sections where each section corresponds to one of the datasets.
3. All variables used in the analysis dataset(s) must be fully defined within the codebook. If there are variables in the analysis data that are not used to produce results within the manuscript or appendices, they should be removed from the dataset. There must be a 1:1 match between variables in codebook and variables in analysis data.
4. All values (including missing values) must be included for each variable with complete value labels. This includes any categorical, dichotomous, binary, and/or ordinal variables.
5. Must be saved as a PDF.
6. If full data citations for original data sources have not been included in the README, they must be included in the codebook. We recommend placing data citations in a References section at the bottom of the codebook.

Codebook Template

[Article Title]

[Authors] – [Affiliation] – [E-mail] – [Twitter (optional)]

A. Codebook for datafile.dta

Variable	Variable Definition	Value & Value Label (s)	Source
Gender	Respondent is male or female	1 = Male 2 = Female	
Age	Age of respondent	1 = 18 – 25 2 = 26 – 30 3 = 31 – 35 4 = 36 – 40 5 = 40 & older . = Missing	
Female_labor	Percentage of women on the labor market		The World Bank (2017)
ChRef (binary)	Whether the sponsor is Chair of the committees to which the bill is referred	1 = Sponsor is Chair of one of the committees to which the bill is referred 0 = Otherwise	

B. References

The World Bank. 2017. "Gender Data Portal."
http://datatopics.worldbank.org/gender/about (August, 2017).

R PRO-TIP

Both the 'code book' and 'dataMaid' library packages can be used to generate codebooks. Again, please review the output to ensure that all variables, variable definitions, values and value labels are present

CODEBOOK

UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

STATA PRO-TIP

The 'codebook' or 'label book' commands can be used to generate some of the codebook. These commands depend on the robustness of your data within Stata so you will need to review the output to see if any variables, variable definitions, values or value labels are missing. You will also need to add original data source citations to the final product (if they have not been included in the README)

What should I include in a data citation?

A data citation should lead a secondary user directly back to the original source data. We recommend using ICPSR's Citing Data as guidance: https://www.icpsr.umich.edu/web/pages/datamanagement/citations.html

Here is an example of a complete data citation:

> US Census Bureau. 1993-2011. Public Elementary-Secondary Education Finance Data. https://www.census.gov/programs-surveys/school-finances/data/tables.html

PRO-TIP

The citation above tells you the data producer or author (US Census Bureau), the date of the data collection (1993 - 2011), the title of the dataset (Public Elementary-Secondary Education Finance Data) and provides a direct link to the tables used.

The user should be able to follow the citation and immediately gain access to the data. In some cases, this may not be possible. For example, with online databases, you may not be able to directly link a user back to the original data; however, you can provide the data citation alongside instructions with the key terms or query used to extract the original data.

If the data cannot be directly accessed due to access restrictions, licensing agreements, or copyright, please provide a data citation to the source of the original data as well as detailed instructions for requesting access to the exact data file(s) you used.

If you have additional questions or concerns about data access restrictions, please contact the AJPS Editors.

Examples of complete and successful codebooks:

1. Grumbach, Jacob M.; Frymer, Paul, 2020, "Replication Data for: Labor Unions and White Racial Politics", https://doi.org/10.7910/DVN/VJUOOV, Harvard Dataverse, V2, UNF:6:TBRE2ynz6RXYOx1Xb+EzBw== [fileUNF] CODEBOOK

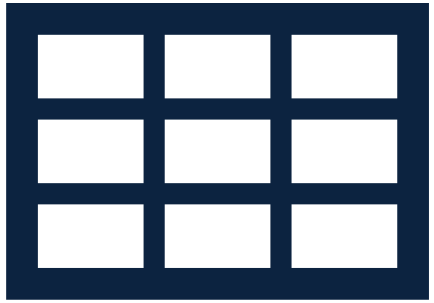
2. Zhang, Nan; Lee, Melissa M., 2020, "Replication Data for: Literacy and State-Society Interactions in 19th Century France", https://doi.org/10.7910/DVN/YQSBOA, Harvard Dataverse, V1, UNF:6:8QC42Rz+5l07o276blDJag== [fileUNF] CODEBOOK

3. Slough, Tara; Fariss, Christopher, 2020, "Replication Data for: Misgovernance and Human Rights: The Case of Illegal Detention without Intent", https://doi.org/10.7910/DVN/Q5PV4U, Harvard Dataverse, V1 CODEBOOK

UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

AJPS | AMERICAN JOURNAL of POLITICAL SCIENCE

ANALYSIS DATA



UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

ODUM INSTITUTE
DATA ARCHIVE

Analysis Data Construction Guide

Guidance for Original Data Sources:

In many cases analysis data are constructed from multiple original data sources, these original data can come from a variety of sources: other researchers, online databases, government agencies, and even hand-copied from tables on webpages. Authors are required to provide code and/or instructions for obtaining and constructing their analysis data as part of the data verification policy. If your particular use case is not described below, please contact the editors with questions.

For analysis data generated by code:

If you have code files that generate the analysis data files, please provide the code and (if possible) the original data files used by the code. This allows verifiers to run the code and ensure that it does, in fact, create the analysis data files used to generate the manuscript results.

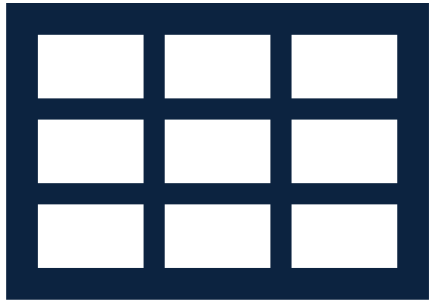
> **Example:**
> [Citation] Malis, Matt; Smith, Alastair, 2020, "Replication Data for: State Visits and Leader Survival", https://doi.org/10.7910/DVN/9QUMY3, Harvard Dataverse, V1, UNF:6:olMtwppSf2uU65yWocvahw==[fileUNF]

For original data downloaded via APIs:

If you used an API to acquire the original data used in your analysis, a full data citation of the original data source must be included and must have a stable URL or DOI leading directly to the data source. Complete instructions for using the API to acquire the exact data used must be included in the documentation (either the README or Codebook). If you are able, please provide the script used that calls the API and collects the data.

> **Example:**
> [Citation] Larson, Jennifer; Nagler, Jonathan; Ronen, Jonathan; Tucker, Joshua, 2019, "Replication Data for: Social Networks and Protest Participation: Evidence from 130 Million Twitter Users", https://doi.org/10.7910/DVN/RLLL1V, Harvard Dataverse, V1, UNF:6:yBovQ2lL0+OEnmzy7MxqBw==[fileUNF]

ANALYSIS DATA



For original data pulled from online databases:

If you have used data from online databases to construct your analysis data, please provide full data citations to the online database. In addition to the citation, detailed instructions containing key terms and queries used to navigate the database are required. This information can be provided in the README or codebook, whichever makes the most sense. The more information provided, the easier it will be for secondary users to acquire the original data used.

Example:
[Citation] Osgood, Iain; Cory, Jared; Lerner, Michael, 2020, "Replication Data for: Supply Chain Linkages and the Extended Carbon Coalition", https://doi.org/10.7910/DVN/W08NIR, Harvard Dataverse, V1, UNF:6:lRYb3/xoV6RX+cAeFkbfJQ== [fileUNF]

For original data hand copied from websites:

There are instances where you may have hand copied data from a site. In order to provide the most information on how these data were obtained, you must include a full data citation for the original data source. Direct links to the page(s) where the data were obtained must be provided with instructions for the specific variable(s) copied into your analysis data.

Example:
[Citation] Jansa, Joshua, 2019, "Replication Data for: Chasing Disparity: Economic Development Incentives and Income Inequality in the U.S. States", https://doi.org/10.15139/S3/4OX6SV, UNC Dataverse, V1, UNF:6:N2mzh6HQtQpZSBNSbk7+rW==[fileUNF]

For original data that cannot be shared due to licensing, access restrictions, or ethical concerns:

In some cases the main hurdle may be sharing the original data files. Please review the terms of use and possible restrictions for all original data sources. If there are no restrictions, you must share these data within the journal Dataverse and provide a full data citation for the original data source. Or, if possible, have the code call the data from the original source site.

If there are restrictions, authors will need to contact the original data producer to see if it is possible to share the original data files with editors and Odum Institute for verification purposes only. The data will be deleted upon successful verification.

If the data still cannot be shared due to restrictions, inform the editors of the restrictions and provide detailed instructions within the documentation on requesting the specific data files used to construct their analysis data. In some cases, Odum may be able to request access to the data and get permissions; however, if there are costs involved, the final decision will be made by the editors as to how access to original data sources are to be handled.

Thank you

CoRe2 Project
(http://core2project.org/)

CoRe2 Github
(https://github.com/OdumInstitute/dataverse-corere)

Mandy Gooch | agooch@unc.edu

UNC | ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE