CORNELL CENTER for SOCIAL SCIENCES

CISER CORNELL INSTITUTE for Social and Economic Research

Now part of CCSS

Accelerates, enhances, and amplifies social science research at Cornell.

Preparing a Data Archive or Repository for Changing Research Data and Materials Retention Policies

Jonathan Bohan
Data Archive Specialist

Archival requirements for researchers are changing rapidly Journal requirements

"In general, all computer code central to the findings being reported should be available to readers to ensure reproducibility." – *Science* Journals

- "AGU encourages authors to identify and archive their data in approved data centers." "Data include .
- .. New code/computer software used to generate results or analyses reported in the paper." <u>American Geophysical Union</u>

Authors "must provide materials that are sufficient to enable interested researchers to verify all of the analytic results that are reported in the text and supporting materials." – <u>American Journal of Political Science</u>

More examples at https://libraries.mit.edu/data-management/share/journal-requirements/

University Requirements

"Reproducibility, which is the ability to verify research findings by other members of the scientific community or by using other methods, is essential to the advancement of science. This ability requires access to relevant research data, materials, documents, protocols, methods, and procedures" – <u>Cornell University Interim Research Data Retention Policy</u>

"In practice, scientific data include both intangible data (statistics, findings, conclusions, etc.) and tangible data . . . records should include sufficient detail to permit examination for the purpose of replicating the research . . . " - <u>University of Pittsburgh Guidelines on Research Data Management</u>

"Research data include, but are not limited to, laboratory notebooks, as well as any other records that are necessary for the reconstruction and evaluation of reported results of research and the events and processes leading to those results, regardless of the form or the media on which they are recorded" – <u>University of Mississippi Medical Center Policy on Research Data Retention</u>

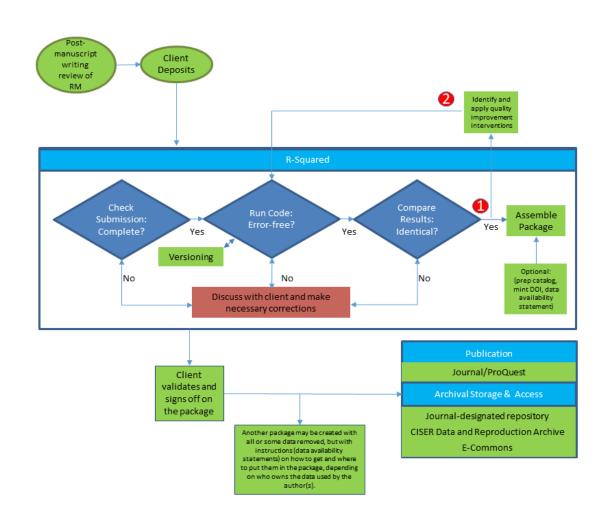
Considerations for Data Librarians and Archivists

- Code is not data; requires extensive documentation explaining processes and the data itself to be useful.
- Storing code with data can be wasteful of valuable storage space; data can be used for multiple publications. Separate catalog record for the data?
- Most data archives have not been designed with code archiving in mind.

 Possibility of dozens to hundreds of files which need to be run in a precise order using specific software.
- Materials are tied to a specific publication and can be viewed as an extension of that publication requires special citation of that publication, separate from any later citations of the archival material.

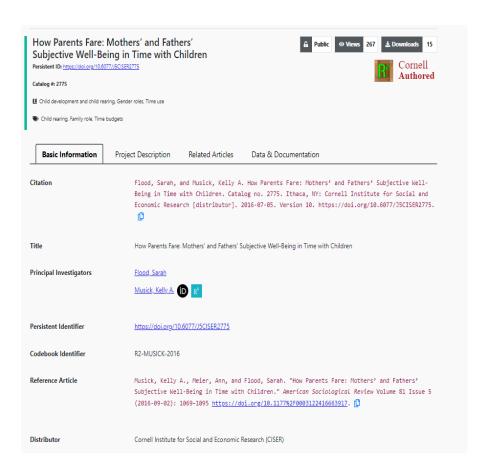
Results Reproduction Service

Computationally reproduces the results of research to ensure Reproducibility and Transparency – "think of it as enhanced proofreading for your Data and Code."



Results Reproduced Catalog Record

- Suggested citation
- Persistent identifier
- Links to the researchers'
 ORCID and/or ResearchGate
 profiles
- Special citation for the "reference article."



Thank You!

Jonathan Bohan

Cornell Institute for Social and Economic Research, Cornell Center for Social Sciences

https://ciser.cornell.edu/

https://socialsciences.cornell.edu/

jb2287@cornell.edu