

## **Data Management Plan**

People: Brooks Hefner, Edward Timke, Kevin Hegg, and James Madison University and University of California, Berkeley student research assistants (to be determined).

Kevin Hegg, with Libraries and Educational Technologies (LET) at James Madison University, will be responsible for the training of students for access and data entry.

Yasmeen Shorish (Data Services Librarian) at James Madison University will advise on data management protocols and procedures.

Brooks Hefner will be ultimately responsible for the stewardship of the data.

The Center for Open Science's Open Science Framework (OSF) will be responsible for dissemination. James Madison University's IT will maintain a dark archive copy as well.

## **Expected data**

The project will generate tabular data in spreadsheet format (CSV), digitized from print copies of reports submitted to the Audit Bureau of Circulations, and held at the Library of Congress. The data will be transcribed into spreadsheets via the GoogleSheets interface. Spreadsheet data will be transformed to a SQL database on a LET server via Python scripts, undergo quality assurance (QA), and then be outputted to JSON. JSON data will underlie the visualization tools hosted on the project website.

LET servers undergo weekly back-up to tape. QA and JSON will be stored on the OSF. Data stored on the OSF is backed by a \$250,000 preservation fund that will provide for persistence of your data, even if the Center for Open Science runs out of funding.

## **Period of data retention**

Upon conclusion of the grant period, all data detailed above (CSV, SQL, JSON, associated scripts and metadata) will be fully available and accessible via the Open Science Framework for an indefinite period of time.

## **Data formats and dissemination**

Data will exist in CSV, SQL, Python, and JSON formats. We will utilize MySQLdump and ReproZip to make SQL data available in a shareable format. Metadata information will be captured in ReadMe files, and a data dictionary will accompany the spreadsheet data. This information will define variable names and units, and will follow spreadsheet best practices (e.g., atomized unit-level information per column).

Data - including scripts, ReadMe files, and procedural information - will be available via the Open Science Framework without embargo or restriction at the conclusion of the grant period. A DOI will be generated for the project space.

**Data storage and preservation of access**

Data will be available at the Open Science Framework project space. A backup copy will be housed on JMU central servers, where the project data will be stored, backed up, preserved (replicated to one offsite location), and made accessible for no less than ten years. Data stored on the OSF is backed by a \$250,000 preservation fund that will provide for persistence of your data, even if the Center for Open Science runs out of funding.