

THE ODUM INSTITUTE

FOR RESEARCH IN SOCIAL SCIENCE

Odum Institute Data Archive COLLECTION DEVELOPMENT POLICY

Introduction

Beginning in 1965 with the acquisition of the Louis Harris Data Center polls, the Odum Institute has been acquiring, preserving, and disseminating data of interest to the social science research community. Collections include national- and state-level polling data, census data, and survey data focused on the Southern United States. In recent years due to the increasingly interdisciplinary nature of scholarly research, the Odum Institute Data Archive catalog has expanded its scope to include data produced in domains other than the social sciences.

While the collecting mission of the Odum Institute Data Archive remains focused on social science data, it invites contributions from other disciplinary domains engaged in research projects that include a social component. These often include, but are not limited to, data from the behavioral, health, and environmental sciences.

Designated Community

Data in the Odum Institute Data Archive collections are accessed and used primarily by members of the academic community engaged in social science research or research that includes a social component. These members include research faculty, students, and other individuals that participate in such academic research. Because the Odum Institute Data Archive provides free and open access to its collections, data are also accessed by journalists, policymakers, citizen scientists, and others interested in the collections.

Selection

The Odum Institute Data Archive identifies and solicits social science data considered by the Designated Community to be of significance to the study and understanding of society and social relationships. The Data Archive accepts donations of data that complement the scope of the Odum Institute Data Archive collection, particularly those datasets that focus on topics related to the Southern region of the United States and state-level public opinion polls. The Data Archive also prioritizes data considered to be at risk of being lost.

Appraisal

The Odum Institute Data Archive performs an appraisal of all data submissions based on the collecting scope of the Odum Institute Data Archive and established standards of professional archival practice. Data submissions are appraised according to primary and secondary criteria.

Primary criteria:

- The data have substantive value to social science research
- The data have influence on the body of social science knowledge
- The data have enduring value to the Designated Community

- The data are unique (i.e., the data are not available in another repository)
- The data cover a significant or useful timeframe or date span for study

Secondary criteria:

- The data support or expand upon subject area concentrations
- The data address substantive gaps in existing holdings
- The data are in sound physical condition
- The data meet quality standards for accuracy and interpretability
- The data are accompanied by complete and readable documentation

Accepted File Formats

The Odum Institute Data Archive accepts data in a variety of formats, but prefers fully documented SPSS (.sav), R (.RData), Stata (.dta), or Microsoft Excel (.xlsx) files containing variable and value labels with complete and accurate documentation. For all other file types, the Data Archive prefers formats that are:

- Widely adopted by the designated community
- Able to be converted or transferred to formats widely adopted by the designated community
- Non-proprietary or open source
- Free of external software dependencies
- Well-documented

Only in rare cases is the Odum Institute Data Archive willing to process and preserve data stored in file formats that are obsolete, have atypical software dependencies, are in rare use by the broader research community, or for which the Data Archive does not have the expertise or tools to provide full data curation. For these exceptional file formats, the Odum Institute Data Archive is able only to provide original bit-level preservation and access, which does not guarantee the future usability of the data.

Levels of Curation

The Odum Institute Data Archive employs three primary levels of curation for its collections: minimal, routine, and intensive. These curation levels are assigned to data submissions based on the specific processing requirements of the data as well as the value of the data to the Designated Community as determined during the appraisal process. For data submissions that require curation beyond the intensive level or otherwise require specialized processing beyond the capabilities of the Odum Institute Data Archive, the Odum Institute Data Archive will seek partner archives to provide the necessary level of curation for and stewardship of these data.

Minimal Curation

Minimal curation is assigned to data submissions for which a significant amount of the required processing has already been completed by the depositor. The following are minimal curation tasks that are completed prior to data archiving and distribution:

- Files are reviewed for completeness
- Data are reviewed to detect the presence of direct identifiers
- Common file formats are normalized to preferred file formats
- Discovery metadata are generated

Routine Curation

Routine curation is assigned to data submissions that appear complete and for which the cost of intensive processing may not be justified as determined during the appraisal process. In addition to the tasks associated with minimal curation, the following routine curation tasks are completed prior to data archiving and distribution:

- Data are reviewed for accuracy and interpretability
- Provisions are put in place to resolve confidentiality concerns
- Documentation is reviewed for the inclusion of complete data definitions
- Data are converted programmatically to preferred file formats
- Additional descriptive metadata are generated to facilitate discovery and reuse

Intensive Curation

Data submissions requiring intensive curation include data that are considered of great potential value to the Designated Community based on popularity, quality, methodological rigor, rarity, and/or relevance to current public policy or research foci. Intensive curation is also assigned to important datasets stored in endangered formats and are at risk of loss. In addition to the tasks associated with minimal and routine curation, these data undergo various intensive curation tasks according to the specific needs of the data. Intensive curation may include the following tasks:

- Data containing personally identifiable information or protected health information are de-identified to produce a public-use version of the data
- Missing descriptive information is recovered from available resources and assembled into a more complete document set
- Replication datasets that underlie published articles are verified to ensure that author-provided analysis code and data reproduce the findings presented in the article

The substantial cost of labor and expertise required to perform intensive curation may require that the Odum Institute Data Archive recover these costs through service contracts or grant funding.

Self-Archiving

The Odum Institute Data Archive also offers to its user community the option to self-archive their data. Using the UNC Dataverse archival platform hosted by the Odum Institute Data Archive, individuals may deposit their data in an open-access Dataverse that is administered by the individual themselves. The Odum Institute Data Archive does not perform curation on these data; however, the data are periodically audited for appropriateness, quality, and policy compliance.

Self-depositors should consult [UNC Dataverse Support](#) to prepare materials for submission. Individuals who opt for the self-archive option are required to review, agree to, and abide by the [UNC Dataverse Terms of Use](#) and all other Odum Institute Data Archive policies and guidelines.

Policy Review

The Odum Institute Data Archive Collection Development Policy is subject to three-year review. The current policy was approved and issued on May 1, 2017.