

**Society of American Archivists  
Council Conference Call  
March 17, 2021  
Virtual Meeting**

**SAA Dataverse Collection Development Policy  
(Prepared by the SAA Committee on Research, Data, and Assessment)**

**BACKGROUND**

In May 2020 the SAA Council approved three recommendations from the Committee on Research, Data, and Assessment (CORDA) to establish the “SAA Dataverse”:

THAT the SAA Council approve collaborating with the Odum Institute, University of North Carolina at Chapel Hill, to establish and support a pilot instance of Dataverse for a duration of five to six years, followed by an explicit decision by the Council to extend the service; and

THAT the SAA Executive Director be authorized to negotiate a Memorandum of Understanding with the Odum Institute to outline the rights and responsibilities of SAA and the Odum Institute for deposited data, liabilities for data reuse, and technical support; and

THAT the Council affirm the responsibilities of CORDA to populate the SAA Dataverse with datasets and establish policies governing contributions and data use.

Read the full Committee report [here](#).

Council members reviewed the Odum-SAA agreement online in December and it was executed in early January.

Since May 2020 [CORDA](#) has conducted a series of online open houses to inform and educate members about the SAA Dataverse and to gather their questions and comments. On February 12, 2021, CORDA held an [online open house](#) to gather feedback on its draft Collection Development Policy. More than 120 people registered, 60 individuals attended and provided feedback, and two people submitted written feedback via the “Suggestion Box” form provided to all registrants.

**DISCUSSION**

As noted when news of the Dataverse was released: “The SAA Dataverse will support active data reuse and reevaluation on behalf of the entire archival community and will welcome data deposits from our designated communities.” Further, from CORDA’s May 2020 discussion document:

The three truly distinguishing features of a robust data repository are 1) full and complete documentation on the datasets [i.e., codebooks and reports]; 2) cross-searching across the metadata

for multiple datasets [i.e., aggregation]; and 3) the possibility of joining data from the repository with easy-to-use analytical tools and the guidance needed to analyze the data.

A data repository is fundamentally about fostering data reuse for purposes of deepening our understanding of SAA’s activities and contributions over time, the status and value of archival organizations, and the evolving professional status of archivists. The types of datasets envisioned for an SAA-managed repository are national-level surveys of archivists and archival organizations (e.g., A\*CENSUS I, A\*CENSUS II, past surveys of archivists or archival organizations), data gathered by SAA sections and other component groups, data gathered initially by large archival organizations that likely have broader uses, and research data gathered by archivists as part of their work or studies.

To serve these purposes well, it was important that CORDA establish a collection development policy as soon as possible to guide those who may contribute data. The policy was written to reference aspects of the signed Memorandum of Understanding with the Odum Institute, including the Odum Terms of Use and Odum Data Deposit form. Drafts of the policy were reviewed by CORDA members and others, including the Executive Director and CORDA Council liaison, and the following policy reflects feedback from various stakeholders as well as the 60 attendees at the February 12 open house.

## **RECOMMENDATION**

**THAT the SAA Dataverse Collection Development Policy developed by the Committee on Research, Data, and Assessment be approved.**

### **SAA Dataverse Collection Development Policy**

#### **Statement of Purpose**

The Society of American Archivists (SAA) Dataverse is an SAA data service that was established to support the needs and interests of SAA’s members and the broader archives community. The SAA Dataverse supports the reuse of datasets for purposes of fostering knowledge, insights, and a deeper understanding of archival organizations, the status of archivists, and the impact of archives and archival work on the broader society. Deposited datasets should be “actionable” in that they should support direct analysis and interpretation. The SAA Dataverse welcomes deposits of collections of quantitative or qualitative data and associated documentation.

#### **Guiding Principles and Collecting Focus**

- The SAA Dataverse supports active data reuse and reevaluation on behalf of the archives community.
- The SAA Dataverse will adhere to [FAIR Guiding Principles for Scientific Data Management and Stewardship](#) (Findable, Accessible, Interoperable, and Reusable).
- The SAA Dataverse systems, policies, and procedures align with standards for trustworthy digital repositories, providing long-term data archiving, bit-level preservation, and data sharing that foster analysis, reuse, and new interpretations.
- The scope of the data and associated documentation must have clear potential value to one or more of four designated communities:
  1. Archives and records professions in North America

2. Society of American Archivists
3. Archival repositories
4. Individual professional archivists and archival scholars

The SAA Dataverse welcomes data created by the SAA community and allied organizations and individuals.

- SAA requires that SAA-sponsored data be submitted for deposit. This includes data gathered by the Council and component groups, work supported by SAA Foundation funding, and selected materials related to the SAA Research Forum.
- Archival repositories that gather data of interest and usefulness beyond the originating organization.
- Individual professional archivists and archival scholars who gather data of broad interest and value.
- Authors publishing research in *American Archivist*, a peer-reviewed journal published by the Society of American Archivists.

SAA will prioritize data identified by and solicited for deposit by CORDA in its role as developer and manager of the SAA Dataverse repository.

### **Submission Requirements and Data Formats**

Datasets and associated documentation must conform to the SAA Dataverse Terms of Use and SAA Dataverse Data Deposit Form. [Link expected to be available in Spring 2021.] Submissions will be reviewed by a designated review committee of SAA's Committee on Research, Data, and Analysis (CORDA), and ingest will depend on adherence to the Guiding Principles and Collecting Focus. Review does not include assessment of methodology used.

Data should provide information about:

- Origins of the dataset;
- Methodologies for gathering, preparing, and analyzing the data;
- Any analyses or reports that are derived from the data, either published or unpublished.

Associated documentation can take the form of:

- Tabular data (including Excel, CSV, R, Stata, SPSS, tab-delimited files);
- Image files (.jpg, .png, .tif);
- Geospatial files; and/or
- File types that are widely accepted, that can be converted, and that are non-proprietary/open source (PDF, etc).

All contextual information necessary to understand and use the data (such as readme files, checklists, etc) is a mandatory component of the dataset. The submitter must remove all personally identifiable information (PII). The submitter must supply metadata to the fullest extent possible at the time of ingestion.

### **Preservation:**

The SAA Dataverse is hosted by the Odum Institute for Research in Social Science at the University of North Carolina at Chapel Hill, which is a trusted repository for research data, datasets, and associated research whose mission is to provide long-term data archiving and data

sharing. The Odum Institute's systems, policies, and procedures have been developed in alignment with standards for trustworthy digital repositories as outlined in [ISO 14721 Reference Model for an Open Archival Information System \(OAIS\)](#) and ISO 16363 Audit and Certification of Trustworthy Digital Repositories. Data is preserved at the bit-level. Deposited items will be retained for the lifetime of the SAA Dataverse.

**Copyright and Intellectual Property Rights:**

All datasets and supporting documentation must be granted for distribution in the SAA Dataverse. Uploaded materials must be made freely available via a Creative Commons license. Datasets and associated documentation may be deposited in repositories outside of the SAA Dataverse if permission and copyright allow.

**Deaccessioning:**

Deaccessioning records from the SAA Dataverse should occur only if there is a legal or otherwise valid reason for the dataset to no longer be accessible to the public. If a record must be deaccessioned, a version of the dataset or the entire dataset may be deaccessioned. If a record was published for any amount of time, a "tombstone page" will remain for records or data that have been deaccessioned to indicate its removal. Submitters must work with CODA to deaccession data or records.

**Support Statement:** The Collection Development Policy, which references the Odum Institute's Terms of Use and Data Deposit form, will serve as an important guide for those individuals and institutions that wish to contribute data to the SAA Dataverse.

**Strategic Priorities:** All aspects of the SAA Dataverse support SAA's Goal 3: Advancing the Field, and particularly 3.2. (Foster and disseminate research in and about the field) and 3.3. (Participate actively in relevant partnerships and collaborations to enhance professional knowledge).

**Fiscal Impact:** No direct expenses. Development and fielding of the policy required a significant time commitment by several CODA volunteers; implementation of the policy will require a nominal amount of staff time and significant support from CODA volunteers to respond to queries from potential depositors.