

The SURA/ASERL Research Data Management Collaboration

In 1999 the US Office of Management and Budget (OMB) Circular A-110 was amended to require Federal awarding agencies to ensure that all data produced under an award will be made available to the public through the procedures established under the Freedom of Information Act (FOIA). The federal government strengthened its position on research data management in January 2011 when the National Science Foundation (NSF) instituted a requirement that all proposals submitted to that agency must include a supplementary document of no more than two pages labeled “Data Management Plan” (DMP). The DMP should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. Specifically, the plan should address five points: expected data, the period of data retention, data formats and dissemination, roles and responsibilities and data storage and preservation of access. Other research funding organizations and even some journal publishers are instituting similar requirements.

Recognizing that advances in institutional research data management would greatly benefit from collaborative efforts between institutional research officers, technical library staff and central IT organizations, the Southeastern Universities Research Association (SURA) teamed with the Association of Southeastern Research Libraries (ASERL) to sponsor a meeting, in the Spring of 2012, that brought together University Research Officers, Chief Information Officers and library technology leaders from SURA member institutions to explore issues and possible solutions to the rapidly increasing demands for improving the management of institutional research data. This effort was motivated by multiple factors including:

- New requirements from funding agencies;
- Need to effectively manage the ever-increasing size of research data sets;
- Ensuring research integrity, replication and the accuracy and reliability of data and records;
- Enhancing data security and minimizing the risk of data loss;
- Preventing duplication of effort by enabling others to use data; and
- Complying with practices conducted in industry and commerce.

Building on outcomes from that initial meeting, SURA established the Research Data Management Working Group (RDM), a joint project between SURA and ASERL to explore multi-institutional projects targeted at improving the way our members manage institutional research data. This group now consists of over 55 individuals (IT and Library professionals) from more than 25 SURA and ASERL member institutions. The group conducts its work through bi-weekly calls and occasional face-to-face meetings as needed. The group’s work is focused on prioritizing and developing research data management topics identified through a SURA member community survey, developing a set of projects and goals to improve the institutional management of research data and identifying individuals willing to provide leadership for specific projects.

Over the past year the SURA RDM group focused its efforts on the development of an institutional “Step-By-Step Guide to Data Management,” which is being used to identify gaps in existing RDM processes and guide the future efforts of the group. The group also built a discipline specific metadata scheme directory to assist researchers in finding existing metadata models for their research data and co-sponsored a research data management policy paper, “Model Language for Research Data Management Policy”. These products and other materials associated with the work of this group are available on the SURA/ ASERL RDM WiKi: <http://www.lib.ua.edu/wiki/sura/>.

A copy of the SURA/ASERL RDM Group “Step-By-Step Guide to Data Management” can also be found at: <http://sura.org/news/docs/RDMStepGuide101512.pdf>

Current efforts of the SURA/ASERL Research Data Management (RDM) group are focused on a multiple institution pilot implementation of the Dataverse Network (<http://thedata.org/>). The Dataverse Network is an open source application that facilitates the ability to publish, share, reference, extract and analyze research data. It helps to make research data available to others, and allows the replication of others information. Researchers, data authors, publishers, data distributors, and affiliated institutions all receive appropriate credit. A Dataverse Network hosts multiple dataverses. Each dataverse contains studies or collections of studies, and each study contains cataloging information that describes the data plus the actual data (or pointers to the actual data) and complementary files. Additionally, each dataverse can support the harvesting of cataloging information to then serve as a federated search that links to other repositories hosting the actual data files. The University of North Carolina’s Odum Institute has been involved in the implementation and development of the Dataverse Project and has agreed to work with the SURA RDM Working Group on this Dataverse Pilot Project.

A Call for Participation was released to the SURA/ASERL RDM community on early January, 2013 seeking participants in a multi-institutional Dataverse Network pilot project. Through this Call for Participation we are seeking 5-6 schools to participate in this pilot project by identifying a single point of contact responsible for coordinating the efforts of their institution. Working with SURA and the Odum Institute each pilot project participant will be expected to:

- Identify a limited number of research domains and a sample set of research data from researchers at their institution to serve as test data to contribute to the pilot;
- Coordinate training and support for the Dataverse environment pilot at their institution;
- Assist with the customization and institutional branding of the Dataverse environment for their institution;
- Document their institution’s experience with the Dataverse environment;
- Meet by phone and web conference on a regular basis to manage the progress of this pilot.

The expectation is that this will be a 6 month pilot project with a starting date in mid-Feb. 2013 with a rough timeline as follows:

- 1 month - training and initial configuration of institutional Dataverse implementations
- 3 months – loading of research data into institutional Dataverse implementations
- 2 months – evaluation of usefulness of Dataverse tools
- Report to community by late summer 2013

A key goal for this pilot project will be to select a diverse set of research data types and domains to enable a robust evaluation of the tools provided by the Dataverse Network. The evaluation results will then inform next steps in terms of direction, enhancements and gaps regarding a larger deployment of the Dataverse Network for the SURA and ASERL community.

The Southeastern Universities Research Association (SURA) is a consortium of over 60 leading research institutions in the southern United States and the District of Columbia established in 1980 as a non-stock, nonprofit corporation. SURA serves as an entity through which colleges, universities, and other organizations cooperate with one another, and with government and industry in acquiring, developing, and using laboratories and other research facilities and in furthering knowledge and the application of that knowledge in the physical, biological, and other natural sciences and engineering. SURA operates the Thomas Jefferson National Accelerator Facility (Jefferson Lab) for the U.S. Department of Energy through Jefferson Science Associates, LLC, a SURA/Computer Sciences Corporation joint venture. SURA sponsors other collaborative research activities with its member institutions, government agencies, and other researchers working to explore and advance the transformative nature of information technology on the regional, national and international fronts, and facilitating a better understanding of coastal, ocean and environmental phenomena that plays a prominent role in our lives. For more information about SURA visit: www.sura.org

For more than 50 years, the Association of Southeastern Research Libraries has brought together leaders from research, federal, and state libraries in the region to foster a high standard of library excellence through inter-institutional resource sharing and other collaborative efforts. Today, ASERL is the largest regional research library consortium in the United States. By working together, ASERL members continue to provide and maintain top quality resources and services for the students, faculty, and citizens of their respective communities. For more information about ASERL visit: <http://www.aserl.org/>

The Odum Institute provides services for researchers in managing, archiving and preserving social science data. The archive staff provides comprehensive data management consultation and support through all phases of the research lifecycle. The Institute houses one of the oldest and largest catalogs of machine readable data in the United States. The Odum Archive encourages researchers to deposit their data on any social science or health topic. For more information about the Odum Institute visit: <http://www.irss.unc.edu/odum/home2.jsp>
