



Oregon State University Libraries & Press Strategic Agenda for Research Data Services

May 2013

Shan Sutton

Associate University Librarian for Research & Scholarly Communication
OSU Libraries & Press

David Barber

Senior Program Manager
Information Services

Amanda Whitmire

Assistant Professor, Data Management Specialist
OSU Libraries & Press



This work is licensed under a Creative Commons
Attribution 3.0 Unported License (CC-BY)

Suggested citation: Sutton, Shan; Barber, David; Whitmire, Amanda L. (2013): *Oregon State University Libraries and Press Strategic Agenda for Research Data Services*. Oregon State University Libraries.
<http://hdl.handle.net/1957/38794>

BACKGROUND STATEMENT

As a major research institution, Oregon State University supports scholarship and investigation in a broad range of disciplines. Its faculty and students increasingly require infrastructure and services to support research that involves generating, sharing, and preserving data in a variety of formats and sizes. The OSU Strategic Plan addresses the need to strengthen the university's commitment to these endeavors in Goal 3:

“Improve the physical and information infrastructure that supports the education, research and outreach and engagement missions of the university, including construction of the remaining facilities targeted for the public phase of the Campaign for OSU (e.g. the Linus Pauling Science Center, the Student Success Center, and the four Cultural Centers); continued upgrading of classrooms and research facilities; and constant enhancements to the backbone structure of information technology.”

Achieving this goal will require collaboration across units to leverage the skills and resources that are distributed through both centralized and decentralized structures at Oregon State. Within this context, OSU Libraries is uniquely positioned to provide leadership in coordinating support for data-driven research. The newly-developed OSU Libraries and Press Strategic Plan asserts the need to develop this capacity as it shifts from the traditional library model of “information gatekeeper” to being an active participant in the creation and curation of knowledge at OSU. In achieving this transformation, OSU Libraries “will curate the intellectual output of the OSU community as a trusted repository and work with other campus units to provide the tools and infrastructure necessary to support the entire life-cycle of that data; including the selection, maintenance, preservation, discovery, and long-term access.” (Excerpt from OSU Libraries and Press Strategic Plan Goal 3: Play a Fundamental Role in Knowledge Creation and Dissemination.)

In support of this goal, OSU Libraries is broadening its programs in support of data-driven research. In order to strengthen its emerging research data services, OSU Libraries participated in the E-Science Institute in 2012. The E-Science Institute was developed by the Association of Research Libraries and the Digital Library Federation to enable academic libraries to “boost institutional support of e-research and the management and preservation of our scientific and scholarly record.” This program involved a cohort of twenty-two, three-person teams from various universities that completed a three-month online training program before meeting for an in-person capstone event. The OSU team was made up of Shan Sutton and Amanda Whitmire of OSU Libraries, and David Barber of Information Services.

The primary outcome of participation in the E-Science Institute is the Strategic Agenda for Research Data Services. It will guide the development of these services at OSU Libraries from January 2013 to June 2014. The Strategic Agenda is informed by best practices in data management at peer universities such as Purdue. Major elements of the Strategic Agenda include an instructional program on data management, robust data management consultation and planning services, facilitation of OSU participation in open data initiatives, and the development of additional data preservation and sharing infrastructures. OSU researchers will inherently utilize different combinations of these services according to the nature of their data, their aspirations for preserving and sharing it, and the data services provided by their colleges and departments.

In building an integrated program to support data-driven research, OSU Libraries will serve as a catalyst in furthering the OSU research agenda across disciplinary boundaries. Research universities across the country are pursuing similar strategies to meet the multi-faceted challenges of preserving and sharing the data generated by their faculty members. Failure to do so will lead to a loss of research productivity and possibly federal research funding. Within the context of increasing requirements from federal agencies for open access to data from federally funded research, OSU faculty will be better positioned to compete for grant funding. Successful implementation of the Strategic Agenda will also facilitate broad public dissemination of OSU research outputs. This scenario not only benefits the scholarly community, it also supports OSU's land grant mission by helping to ensure the University's scholarship is appropriately preserved and made accessible to the citizens of Oregon and a global audience.

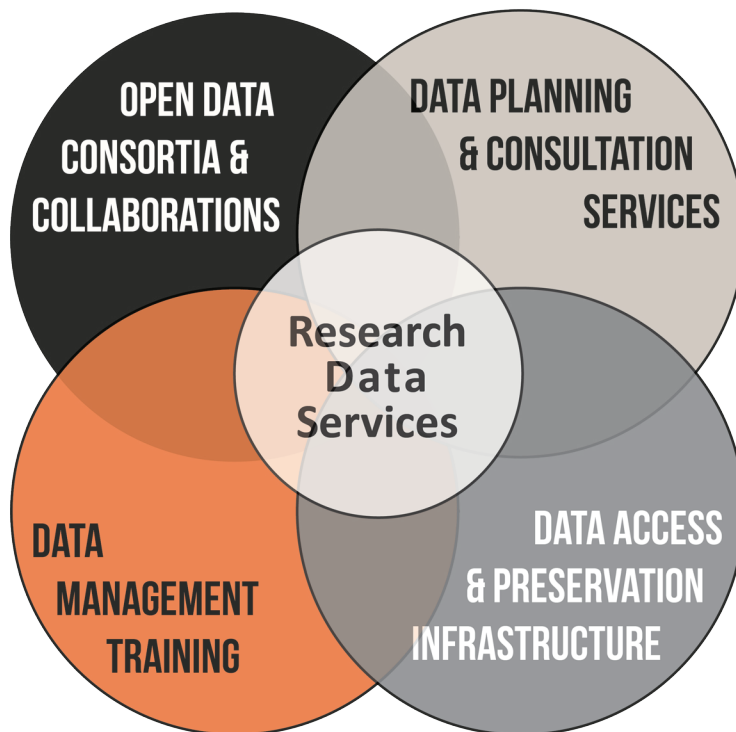


Figure 1. The OSU Libraries Research Data Services model. Our services and support will overlap in these four areas, with the goal of providing comprehensive, integrated support over the entire data lifecycle.

GOALS AND ACTIVITIES

PHASE I (January-December 2013)

1. Develop a university-wide perspective on the nature of OSU research data and data management practices

1a. Activity: Gather information from a broad range of OSU faculty on their research data and data management practices through a needs assessment survey.

The Library will conduct research to more fully determine which colleges and units need support services for data curation by conducting a campus-wide data management needs assessment. The primary tool for conducting the needs assessment will be an online survey. Topics covered in the survey will include¹:

- data formats and volume;
- data stewardship roles and responsibilities;
- potential data management services;
- data management funding;
- data management standards and policies (professional community, agency, and university);
- data sharing, dissemination, discoverability and collaboration;
- metadata practices;
- protecting and preserving data

The survey will be distributed to faculty across tenure/non-tenure track levels in order to reach faculty research assistants who may be managing a lab and handling the bulk of the research data management responsibilities. We are also interested in understanding the data landscape in the humanities and social sciences. Thus, all campus colleges will be included in the survey and needs assessment process. Follow-up activities with specific units, potentially including more targeted surveys or focus groups, may subsequently occur. The survey results will be combined with other sources of information to develop a report on data management practices and needs at OSU. The survey results will also likely inform research by Library faculty that will lead to articles and conference presentations.

Lead: Amanda Whitmire

Partners: Shan Sutton, Michael Boock (OSU Libraries and Press), Maura Valentino (OSU Libraries and Press), David Barber, Rich Holdren (Research Office)

Timeline: Design and administer the initial survey by June 2013

¹ Based on: Gary Marchionini, *Research Data Stewardship at UNC: Recommendations for Scholarly Practice and Leadership* (Chapel Hill: University of North Carolina, February 15, 2012), http://sils.unc.edu/sites/default/files/general/research/UNC_Research_Data_Stewardship_Report.pdf.

1b. Activity: To broaden understanding of university guidelines that are relevant to data management planning, the Library will collaborate with the Research Office to clarify and disseminate OSU data ownership, stewardship and retention policies.

University-wide policies related to data management are not currently well defined or communicated. The Library will work closely with the Research Office to identify, clarify where needed, and disseminate OSU data ownership, stewardship and retention policies. Making these resources accessible through a centralized web site will encourage uniform adoption of such policies by researchers and foster a campus culture of data management best practices.

Lead: Amanda Whitmire

Partner: Rich Holdren

Timeline: Completed (<http://cdss.library.oregonstate.edu/research-data-services>)

2. Empower the next generation of scholars to effectively manage their research data

2a. Activity: Develop and implement a graduate-level research data management course.

As a sun-, space-, sea- and land-grant institution that is committed to teaching and research excellence, one of OSU's core missions is to prepare our students with rigorous coursework that puts them at a competitive advantage when they graduate. A well-designed course in research data management (RDM) ties directly into this mission by providing graduate students with a solid foundation upon which to build their professional careers.

The Library will lead a collaboration with the Graduate School and Research Office to develop an outstanding education program in RDM for graduate students. The course will cover all aspects of research data management best practices. Course content includes²: an overview of research data management, definitions and best practices; types, formats and stages of research data; metadata; data storage, backup and security; legal and ethical considerations of research data; data sharing and reuse; and, archiving and preservation. The final project for the course will be composition of a detailed data management plan for each student's thesis or dissertation research project.

Lead: Amanda Whitmire

Partners: Shan Sutton, Michael Boock, Anita Azarenko (Graduate School), Barbara Bond (Graduate School), Rich Holdren

Timeline: Teach the course in Winter 2014.

² Based on: Mary Piorun et al., "Teaching Research Data Management: An Undergraduate/Graduate Curriculum," *Journal of eScience Librarianship* 1, no. 1 (February 14, 2012): 46–50, doi:10.7191/jeslib.2012.1003.

3. Enhance OSU data sharing and discoverability

3a. Activity: Investigate integrating OSU research data into consortial and disciplinary data networks, such as DataONE and the Northwest Knowledge Network.

One of the most important roles that the Library can play is to be a facilitator in making OSU research data more discoverable through participation in open data initiatives. Examples include DataONE (dataone.org), a newly established international network providing access to earth and environmental science data from multiple providers such as the Partnership for Interdisciplinary Studies of Coastal Oceans (PISCO) and the Long Term Ecological Research Network (LTER); and the Northwest Knowledge Network, a data management system developed at the University of Idaho with other northwest universities to preserve and distribute cross-disciplinary data. The Library will take a lead role in working with selected OSU faculty and units such as the Institute for Natural Resources to determine if these and other open data initiatives should be pursued to make OSU research data more accessible and widely-used. This activity is likely to require additional investments in equipment, staffing, and training as involvement in data networks increases over time.

Lead: Amanda Whitmire

Partners: Shan Sutton, Michael Boock, various Associate Deans for Research and other faculty, Institute for Natural Resources

Timeline: Ongoing, but investigation of at least three consortial/disciplinary networks will be conducted by September 2013

4. Establish OSU Libraries as a central, essential resource for OSU research data services

4a. Activity: OSU Libraries will promote its role in research data services.

The Library will better define the parameters of its data management capabilities (abilities and weaknesses; e.g. ScholarsArchive@OSU; metadata practices), and will communicate those capabilities to stakeholders. Results from the needs assessment will be used to identify areas where current Library resources are being underutilized, and we will conduct targeted outreach to better advertise them to stakeholders. The Library will develop and pursue a coordinated marketing plan to inform faculty of research data services.

Lead: Amanda Whitmire

Timeline: Ongoing, but initial marketing plan will be designed and implemented by December 2013

4b. Activity: OSU Libraries will provide comprehensive information and support for research data management best practices and the development of data management plans.

In order to demonstrate leadership in research data support on campus, it is critical that the Library provide faculty and graduate students with a local service point when developing data management plans or working to improve their research data management practices. The Library's Data Management Specialist will coordinate the

development of a robust point-of-service support resource to meet this goal. Support will include:

- A comprehensive web site with:
 - Information about federal agency requirements for data management plans and tools/resources available to create them (e.g. DMP Tool, boilerplate language for use of ScholarsArchive@OSU)
 - Data management best practices including: file formats and naming conventions, data storage and backup, human subjects and IRB, intellectual property (IP) and copyright, metadata, archiving and preservation, citing and publishing data, data repositories, version control, and how to keep a laboratory notebook
 - Links to useful information not housed on our page, e.g. video tutorials, agency web sites, and online tools
 - List of campus resources and contacts for data management issues
 - Summary of OSU policies on data ownership and retention, IRB, IP, and responsible conduct or research, with links to OSU web sites with full information
- Provide expedient assistance in answering questions from faculty and graduate students, and directing them to the appropriate on-campus or virtual resource(s).
- Seminars and/or workshops on creating data management plans for faculty

Lead: Amanda Whitmire

Partners: Maura Valentino

Timeline: Completed (<http://cdss.library.oregonstate.edu/data-management-plans>)

PHASE II (July 2013-June 2014)

5. Provide professional development for library faculty members in research data management

5a. Activity: The Library will develop activities and training to increase library faculty members' familiarity with research processes and data management (RDM) practices.

As RDM becomes a more central element of the Library's services, more Library faculty will become involved in its promotion and implementation. Articulation of the roles and levels of expertise expected of Library faculty in different units and positions, including altering position descriptions when warranted, will ensure a shared understanding of expectations in this area. This is especially important for liaison librarians who engage with faculty in various units at OSU. As part of documenting what we expect Library faculty to do and know with regard to RDM, there is a need to develop training to raise familiarity with research processes and RDM practices. One possibility is an RDM Boot Camp for Library faculty that employs models developed by similar programs such as the GWLA Science Boot Camp for Librarians. A RDM Boot Camp could be developed in partnership with library faculty at the University of Oregon and Portland State University.

Lead: Shan Sutton

Partners: Amanda Whitmire, Michael Boock, Margaret Mellinger (OSU Libraries and Press), OSU Libraries and Press Administration, Management, and Planning Group (LAMP), colleagues at the University of Oregon Libraries and Portland State University Library

Timeline: Develop initial expectations in this area, and host an RDM boot camp by June 2014

6. Preserve and disseminate graduate student research data

6a. Activity: In collaboration with the Graduate School, conduct a survey of graduate students to document the nature of data they are producing and compiling in the course of their thesis/dissertation research. The results of this survey will inform the advisability and viability of requiring the deposit of thesis/dissertation datasets into ScholarsArchive@OSU.

OSU graduate students are currently required to deposit an electronic copy of their thesis/dissertation in ScholarsArchive@OSU. The Library will collaborate with the Graduate School to investigate the advisability and viability of establishing a related requirement that graduate students deposit data related to their thesis/dissertation in ScholarsArchive@OSU, another data repository at OSU, or an Open Access repository outside of OSU. This would ensure the data generated by OSU graduate students are preserved and made accessible to other scholars, and will increase the impact of the research they conducted at OSU. If the datasets were stored in another Open Access repository, students would be required to deposit dataset-level metadata in ScholarsArchive@OSU with a description of the dataset(s) and link to its/their location.

A few factors must be considered before implementing such a requirement:

- What is the nature of the data being produced and compiled by graduate students at OSU in terms of formats, sizes, etc.?
- Is ScholarsArchive@OSU able to accept, store and share the multitude of datasets that it would receive as a result of this requirement (e.g. large files, hundreds/thousands of files for a given project, etc.)?
- Are graduate students adequately trained to reasonably meet such a requirement? (The grad-level research data management course will address this need, but will not be in place until 2014.)

The survey will be a first step in determining the nature of the data, and will inform subsequent steps in this process. The survey results will also likely inform research by Library faculty that will lead to articles and conference presentations.

Lead: Amanda Whitmire

Partners: Shan Sutton, Michael Boock, Anita Azarenko, Barb Bond

Timeline: Design and administer the survey by June 2014

7. Explore options for enhanced data storage and management at OSU

7a. Activity: The Library will work with Information Services on assessing options for developing data storage and sharing systems to meet the short- and long-term needs of OSU faculty.

While much of the research data at OSU is managed by faculty in individual units, the needs assessment will likely reveal the need for more centralized infrastructure to enhance the preservation and accessibility of unit-based systems, as well as provide options for faculty in units that do not have strong data management infrastructure. As Information Services continues to develop infrastructure for storing and preserving data, it is essential for the Library to be involved in this process to ensure its RDM services align with these systems to optimize the preservation and discovery/sharing of OSU research data. Progress in this area is of increasing importance with the expansion of requirements from federal agencies for the accessibility of data associated with federally funded research. This activity will require additional investments in equipment, and possibly staffing, to build out data storage and sharing systems.

Lead: Amanda Whitmire

Partners: David Barber, Michael Boock, Evviva Weinraub (OSU Libraries and Press)

Timeline: Ongoing, Library representatives will participate in relevant IS planning discussions

7b. Activity: Assess the Library's data management capacity through ScholarsArchive@OSU (using the DSpace platform) and the viability of alternative models.

As the Library considers expanding its data management infrastructure, it will need to document the capacity and limitations of the current DSpace platform while also exploring alternatives such as Hydra/Fedora. There is a related need to determine if the Library should pursue a bifurcated model (such as the one in place at Purdue) in which publications and similar items are stored in one repository, with research data residing in another repository, and metadata/links directing users between them. The current collaboration with the University of Oregon Library (UO) on Hydra development will inform this process, as will the above-mentioned partnership with Information Services involving data storage and preservation systems. As with the related Activity 7a above, progress in this area is of increasing importance with the expansion of requirements from federal agencies for the accessibility of data associated with federally funded research. A transition from DSpace to a new platform will likely require significant investments in equipment, staffing, and training.

Lead: Michael Boock

Partners: Amanda Whitmire, Tom Johnson (OSU Libraries and Press), Karen Estlund (UO Library), Brian Westra (UO Library)

Timeline: Ongoing, begin analysis of the viability of the Hydra/Fedora platform as a replacement for DSpace by June 2014

SELECTED MEASURES OF SUCCESS

All of the activities described above will involve some form of assessment. Selected measures of success will highlight progress in specific activities.

Activity 2a: Develop and implement a graduate-level research data management course.

Indicators of Success:

- Number of students enrolled and number of academic units represented.
- Results of a follow up survey documenting impact on their data management behavior.

Outcome:

- OSU Libraries will improve the data management capacity of graduate students through a research data management course.

Activity 3a: Investigate integrating OSU research data into consortial and disciplinary data networks, such as DataONE and the Northwest Knowledge Network.

Indicators of Success:

- OSU research data is integrated into at least one consortial/disciplinary data network.
- Use statistics from a network in which OSU participates indicates that OSU data is being consistently accessed by researchers.

Outcome:

- OSU research data is more widely accessible to the scholarly community and the public via consortial/disciplinary data networks.

Activity 4b: OSU Libraries will provide comprehensive information and support for research data management best practices and the development of data management plans.

Indicators of Success:

- Annually, an increasing percentage of OSU faculty's federal grant submissions include data management plans generated by the DMPTool.

Outcome:

- OSU faculty grant proposals are more competitive by having robust data management plans that are facilitated by OSU Libraries training, consulting services, and tools.

Activity 5a: The Library will develop activities and training to increase library faculty members' familiarity with research processes and data management (RDM) practices.

Indicators of Success:

- A pre- and post-test given to library faculty who participate in an RDM bootcamp demonstrates all attendees increased their knowledge of research data issues.

Outcome:

- Members of the OSU library faculty are familiar with research processes and data management practices to an appropriate level for their roles in support of the Library’s research data services.

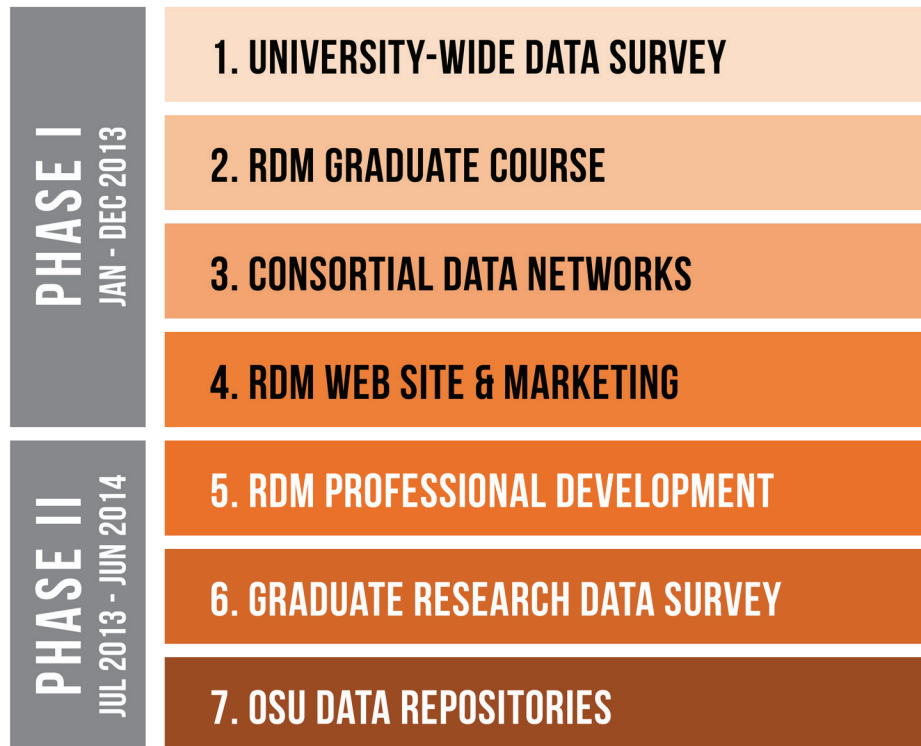


Figure 2. Summary and timeline of activities in the Strategic Agenda for Research Data Services

ORGANIZATIONAL IMPLICATIONS

Budget

Most of the activities outlined in this strategic agenda can be initially supported with existing staff and resources. However, as the research data services model gains traction there will be a corresponding need to secure additional funding, including grants, to support the expansion of these endeavors, potentially in partnership with other units such as Information Services or the Research Office. The outcomes of three activities in particular will likely require additional investment of funds and resources: 3a. Integrating OSU data into consortial and disciplinary data networks; 7a. Developing data storage and sharing systems in conjunction with Information Services; and 7b. Assessing alternative models for the Library data management capacity.

Staff capacity

The Library currently has several positions within the Center for Digital Scholarship and Services that are dedicated to varying degrees to Research Data Services. As the service model evolves and expands, there will be a corresponding need to have Library faculty and staff in other units assume some duties in this area. Two possibilities that merit specific attention are adding duties to liaison librarians' responsibilities as part of a larger shift in the roles of liaison librarians at OSU, and also exploring the potential for assigning data services-related duties to Library faculty and staff as the demands for their time in more traditional areas such as collection development and cataloging are reduced over time.

Tools and resources

OSU has significant existing cyber infrastructure resources. The Information Services (IS) Community Network provides a wide range of services including workstation support, secure file storage, support for remote access to servers and workstations, and software licensing. The Oregon Explorer (OE), a collaborative effort between OSU Libraries and the Institute for Natural Resources, is a web-based resource that uses advanced information technology to access and integrate data from state and federal agencies, local governments, university scientists, citizens, and K-12 educators and students. While OE is currently restricted in scope and resources, it will likely expand in these areas, and provides a model for how to implement a collaborative data storage and discovery tool. The Library needs to partner with IS to determine the extent to which it can provide massive digital storage for research data and provide access to discovery tools.

Institutional partners

The OSU Libraries values partnerships in our efforts to provide comprehensive research data services for the OSU community. We have a close working relationship with Information Services, who are a natural partner in all matters related to centralized computing infrastructure. The Research Office can play an important role in the development of centralized research data services by acting as a liaison between the Library/IS and campus stakeholders. The Research Office can effectively bring the appropriate campus stakeholders into the planning and implementation processes. The Library acknowledges the necessity for, and thus encourages, a faculty-generated and owned vision of data management models and policies on campus. We are committed to transparency in our efforts to develop and expand research data services at OSU, and we value input from and collaboration with all of our stakeholders. While we have good relationships with staff in the Research Office, IS, and within the faculty, these groups traditionally view the Library as a service organization as opposed to a partner in the research process. We strive to engage in active partnerships with these groups in developing a campus strategy for building capacity in research data services.

External partners

The principle of preserving research data and making it more accessible is driving universities and other entities to develop partnerships around these issues. As the Library undertakes this strategic agenda, it will proactively look to collaborate with other institutions within both existing frameworks such as DataONE and the Northwest Knowledge Network, and in developing new initiatives and seeking grant funding. Strong potential partners include the University of Oregon, as well as members of the Greater Western Library Alliance.