

## ***CASE 5***

# Developing a Recordkeeping Framework for Social Scientists Conducting Data-Intensive Research

**AUTHOR:** **ERIN O'MEARA**  
Electronic Records Archivist, University of Oregon  
([erino@uoregon.edu](mailto:erino@uoregon.edu))

**PAPER DATE:** April 2008

**CASE STUDY DATE:** 2006–ongoing

**ISSUE:** Identifying research data associated records on campus and developing a set of tools for faculty to better manage it and potentially deposit into an appropriate repository.

**KEYWORDS:** Data management issues, Faculty research data, Institutional repository

*Copyright by Erin O'Meara.*

## Institutional Context

Four years after it was established by an act of the state legislature, the [University of Oregon](#) opened in Eugene in 1876. In 2007, the total enrollment was more than 20,000 undergraduate and graduate students. The university is comprised of nine schools and colleges with more than 1,600 faculty members. Sponsored program awards have grown to more than \$86 million annually. The university is the state's only member in the Association of American Universities.

The [UO Libraries](#) contains over 2.6 million volumes and 18,000 periodicals and is the second largest library in the Pacific Northwest. There are four branch libraries within the library system. [Special Collections and University Archives](#) are housed in the main library and report to the Associate University Librarian for Collections and Access. Special Collections was formally established in the library in 1947. Many of the early collections were from the history department's collecting efforts, which began in 1875. University Archives was established in 1948 in the library. University Archives split from the library and rejoined the libraries in 1998. In 2007, the records management and electronic records component of University Archives was moved out of Special Collections and University Archives. The unit was renamed the [Electronic Records Program](#) and reports directly to the Associate University Library for Collections and Access.

## Background

Over the past two years, throughout the University of Oregon (UO) campus community (in computer, faculty and library meetings), "research computing needs" became a hot topic and discussions began to occur about infrastructure and support. Like most mid-sized universities, computing on the UO campus is semi-decentralized. There is a campus computing center ([Information Services](#)), separate departmental computing staff and computer labs managed by various units—library, schools and colleges, as well as the [Social Science Instructional Lab](#) (SSIL). In addition, there is a group called Departmental Computing that tries to link all staff that deal with computing issues on campus, regardless of departmental affiliation (primarily through an active listserv). There are computation-intensive resources available to anyone who has an email account with the university. Specifically, within Information Services, there are six dual processors that can handle computing-intensive research. In addition, there are a few consultants on the Information Services staff that provide assistance regarding statistical analysis and scripting. Other than IT support from Information Services and the SSIL Lab, researchers and graduate students receive no advice or assistance when conducting data-intensive research.

On campus, issues concerning the handling and management of research data traditionally have been viewed as the responsibility of the researcher. The university's sponsored projects department, [Office of Research Services and Administration](#) (ORSA),

deals primarily with grant management and does not manage or store research data. The [Office for Responsible Conduct of Research](#) (ORCR, which includes the Office for Protection of Human Subjects) “ensures that all participants in the conduct of research at the University have a basic understanding of research compliance and integrity issues.”<sup>1</sup> This office focuses on such compliance issues as human subjects review and conflicts of interest. The library hopes to play a part in providing guidance on the proper management and sharing of research data that would complement these other issues and bring research at the university to a more consistent level of compliance.

## **Gaining Momentum and Attention through a Research Project**

Based on the needs expressed in numerous venues on campus, I decided to apply for a [National Historical Publications and Records Commission](#) (NHPRC) Electronic Records Fellowship in order to develop a recordkeeping framework for social scientists who conduct data-intensive research. By developing an intellectual framework of guidelines, templates, and best practices to utilize throughout the research process, the university could begin to build greater support for faculty research. In addition to creating this type of information, another goal of the fellowship project is to pilot a technical framework for ingesting research data and records associated with the research process.

I first undertook background research and a basic literature review in order to familiarize myself with the topic. I conducted three interviews with faculty members who are social scientists involved in creating and/or managing social science datasets. I also interviewed a graduate student who works at SSIL and the Data Services Lab. My questions sought to elicit several pieces of information:

- 1) What types of records, and in what formats, are being generated by the research process?
- 2) How are social science faculty members currently managing these digital objects?
- 3) What are the perceived needs of the faculty from campus for these activities?

I chose subjects that provided an accurate representation of the variety of disciplines, methods, and data types that are being created and analyzed at the UO. One of the subjects that participated was an archaeologist who creates GIS, digital photographs, archaeological site data, moving images, as well as other data types during and after fieldwork. Another subject was a sociologist who creates and manipulates quantitative data in highly organized data sets with collaborators at other institutions. The third subject was a public policy faculty member who uses sensitive data regarding public health. The last subject was a geography graduate student who also works in the university’s lab for social science instruction and helps manage the license for access to

---

<sup>1</sup> From the ORCR website, [http://research.uoregon.edu/research\\_orcr.html](http://research.uoregon.edu/research_orcr.html).

the Inter-University Consortium for Political and Social Research (ICPSR) data to campus users.

I conducted semi-structured interviews that used a two-part questionnaire. Interviews lasted approximately 1–2 hours. Data was transcribed and then thematic memos were written for each interview. A larger thematic memo was created to distill large themes that permeated the interviews. The interviews provided a sampling of what types of data were being created, how they were being managed and shared, and what needs faculty expressed in terms of recordkeeping and IT support for their data-intensive research. These interviews informed what type of documentation to provide and the format for the website that I was creating for data management and recordkeeping throughout the research process. It also informed the direction that I pursued in order to develop a process for the actual management of the research data and associated records. Were faculty members going to buy into a “do-it-yourself” deposit model or would they need a mitigated process? Would domain-specific archives be more suitable than an institutional repository?

## **Findings**

Overall, I found that the interviewees’ experiences and data management practices varied widely. There was a general sense of frustration, and apathy from faculty about recordkeeping. This arose mainly from the issue of limited resources (e.g. time, finances, constraints from tenure process, etc.). Faculty members never really saw how recordkeeping could realistically be a priority that they could address. On a positive note, the majority of interviewees saw benefits from receiving assistance and/or guidance from the library or other units on campus.

Below are a few quotes from the transcribed interviews. It is interesting to see the wide variety of situations that some of the researchers are faced with.

One interviewee discusses the disconnect between infrastructure at institutions and granting agencies’ requirements: “There’s no departmental system for archiving data or anything like that. Now it’s becoming more and more common to have...I think you told me this and I used it in my proposal for the NSF. You know, digital archiving and all this stuff. I actually even mentioned this...”

Instead of systematic deposit, some faculty members are using their disciplinary networks to manage their papers and research data: “One of my primary [PhD] advisors, I’m in the will, I’m supposed to deal with one set of her papers.”

This response touches on the generational issues that some of the early-career faculty witness: “Some of my older collaborators...it’s not always true. I have an older collaborator on this project and he seems to be pretty good. But this is what I’m saying about this could be a dark age for information and imagery like visual imagery because these people don’t know what they’re doing. And there really isn’t a lot of support for

faculty.” Another interviewee brings up the pedagogical issue that accompanies this digital divide: “My [thesis] advisor barely knows how to use her computer. I’m unlikely to get instruction on data management from her.”

One interviewee who would definitely be considered a “power user” or model records creator, has explicit and linked documentation to data as well as a server that is backed up on a very regular basis. One pet peeve involved version control, especially because of work that was being done between different institutions: “If you make changes to a file you must be sure that there’s a version number or I will personally fly out there and wring your neck.” This interviewee is also very concerned with replication of results, something that many social scientists do not see as very realistic based on their methods and levels of funding: “So that one [project] was early early early and I had no idea what I was doing and I made tons of mistakes. There I basically just had a folder and I kept all of the files in the folder and about half-way through for a bunch of technological reasons I decided that I had to switch computer programs and not use SPSS anymore and switch to something else. And that, of course, just created an absolute crisis and that was also the process of doing that paper that I realized I had created models and graphs and figures and so on that I couldn’t reproduce, I couldn’t figure out how the hell I had gotten them.”

The interviewee changed the way the files were organized, based on a colleague’s use of Stata (a statistical analysis program), and began using something they called Do files: “There are the files that tell the computer what to do with the data. So my own procedure is to always have a master file that calls all of the other files in order so that at any point you can go back to that analysis, press go, and it will take the original pieces, reassemble them and spit out the tables that you printed in the paper. In my way of thinking, if you can’t do that than your project isn’t replicable. Almost nobody can do that; even I can’t do it on a bad day.”

Many faculty members view their research as sufficiently small to ignore both recordkeeping requirements and calls to submit data in a repository for data sharing. “Basically you publish and then someone asks you for it, then depending on the situation, you can share it but it’s not like I’m doing national studies where it costs millions of dollars and the funders are like ‘oh, you need to make this publicly available.’”

## Outcomes

Currently, the website that will contain the recordkeeping information is still a work in progress. Web pages have all been created, and reviewed by relevant subject specialist librarians, but will also be revised based on usability testing done with a different set of campus researchers (see below).<sup>2</sup> The organization of the websites was influenced by the *ICPSR Guide to Social Science Data Preparation and Archiving*.<sup>3</sup> This manual had been

---

<sup>2</sup> Homepage is located at, <http://libweb.uoregon.edu/records/datasets/index.html>.

<sup>3</sup> Located at, <http://www.icpsr.umich.edu/ICPSR/access/dataprep.pdf>.

cited by a few of the interviewees but none of them had finished reading it due to its length and level of detail. This part of the findings was disheartening, but caused me to simplify the type of information that is so useful within the ICPSR document and other similar manuals like *Guidance on Data Management* from the UK Data Archives.<sup>4</sup>


#### Core pages for the four main areas of research.

The screenshot shows the University of Oregon Libraries website. The header includes the University of Oregon logo and name, followed by 'UO Libraries'. Navigation tabs include Catalog, Articles, Journal Titles, Course Reserves, and Library Site. A search bar is present with a 'Search' button. Below the header, a navigation bar lists: Home, Find Resources, Research Assistance, Library Services, About the Libraries, Ask a Librarian, and My Account. The main content area is titled 'Project planning/application for funding' and contains three sections:

- Review of existing datasets**: Are there existing datasets available that are relevant to your research questions? (Accompanied by an image of glasses on a document).
- Review challenges for management and preservation**: Are there issues if privacy, consent or format with the data that you will create and/or use? (Accompanied by an image of a glass and a white object).
- Determine cost of preserving research materials**: Will the data that you create be very costly to manage and preserve through time? (Accompanied by an image of US dollar bills).

At the bottom of the page, it says 'Maintained by: Erin O'Meara, [erino@uoregon.edu](mailto:erino@uoregon.edu) Last Modified: 07/25/2007'. The footer contains the University of Oregon address and contact information, along with links for 'Contact Us', 'Make a Gift', and 'Site Index'.

<sup>4</sup> Located at, <http://www.data-archive.ac.uk/news/publications/reluaug2006.pdf>.


UNIVERSITY OF OREGON

# UO Libraries

[Catalog](#)
[Articles](#)
[Journal Titles](#)
[Course Reserves](#)
[Library Site](#)

Look for books, magazines, videos, music, and more.


Search for:

[Summit](#)
[WorldCat](#)

[Home](#)
[Find Resources](#)
[Research Assistance](#)
[Library Services](#)
[About the Libraries](#)
[Ask a Librarian](#)
[My Account](#)


[Datasets Home](#)
[Project Planning and Funding](#)
[Project Preparation](#)
[Conducting Research](#)
[Closing The Project](#)

## Project preparation




[Create recordkeeping plan](#)

Develop a plan that lays out how the data and records will be created and managed throughout the research process.



[Develop records and data workflows](#)

Map how information will flow and be used during the research process.




[Conduct pilot test](#)

Run through workflows with test data and refine as needed.

Maintained by: Erin O'Meara, [erino@uoregon.edu](mailto:erino@uoregon.edu)  
Last Modified: 07/25/2007

University of Oregon | 1501 Kincaid Street | Eugene, OR 97403-1299 | T: (541) 346-3053 | F: (541) 346-3485

[Contact Us](#)
[Make a Gift](#)
[Site Index](#)


UNIVERSITY OF OREGON

# UO Libraries

[Catalog](#)
[Articles](#)
[Journal Titles](#)
[Course Reserves](#)
[Library Site](#)

Look for books, magazines, videos, music, and more.


Search for:

[Summit](#)
[WorldCat](#)

[Home](#)
[Find Resources](#)
[Research Assistance](#)
[Library Services](#)
[About the Libraries](#)
[Ask a Librarian](#)
[My Account](#)


[Datasets Home](#)
[Project Planning and Funding](#)
[Project Preparation](#)
[Conducting Research](#)
[Closing The Project](#)

## Conduct Research




[Refine filing and naming conventions while gathering data](#)

Ensure adequate and consistent documentation while gathering data.



[Manage data and records](#)

Review file structure and monitor its utilization.



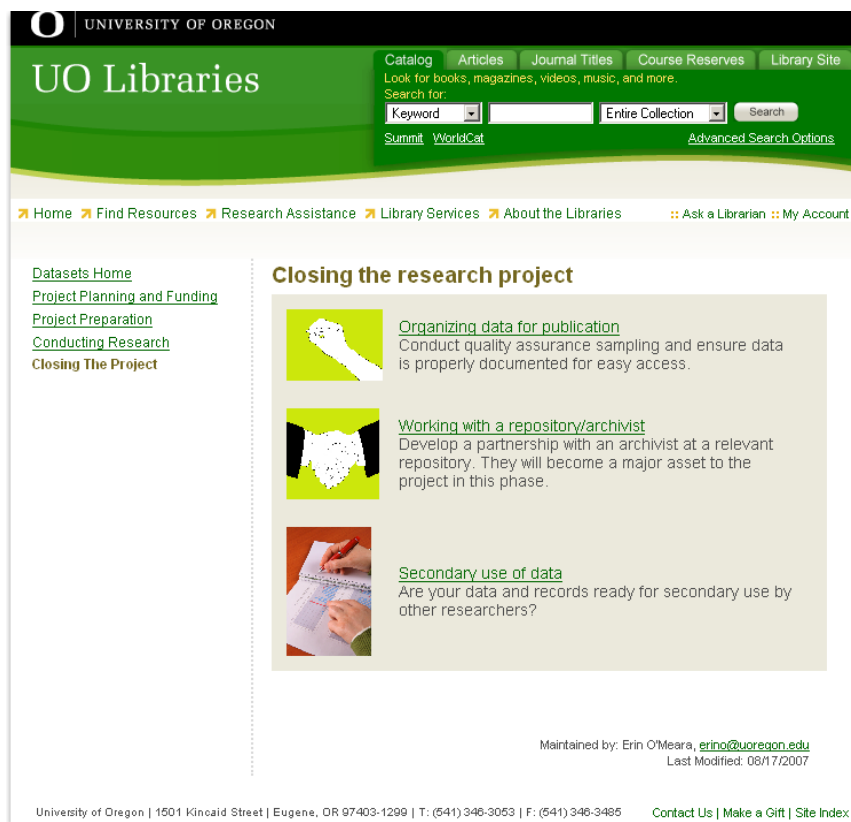
[Protecting research materials](#)

Ensure research data and records are being backed up regularly and that there are proper access controls in place.

Maintained by: Erin O'Meara, [erino@uoregon.edu](mailto:erino@uoregon.edu)  
Last Modified: 07/25/2007

University of Oregon | 1501 Kincaid Street | Eugene, OR 97403-1299 | T: (541) 346-3053 | F: (541) 346-3485

[Contact Us](#)
[Make a Gift](#)
[Site Index](#)



The second outcome, which involved creating a technical prototype that will deal with ingest and access to research data and associated records at the end of this project, proved a bit ambitious and constituted an oversimplified solution to a larger problem. Once I began talking to researchers and reading literature about institutional repositories and research data,<sup>5</sup> I realized that doing this as a stand-alone service would probably not receive long-term buy-in from the faculty. Rather, collaboration with research units on campus and domain-specific archives would create a stronger environment that could cultivate faculty participation. Perhaps this might encourage faculty to spend some of their time preparing their records for deposit. As this is a case study still in progress, I hope to pursue a collaborative approach to this type of information management situation.<sup>6</sup> I will be proposing a shared model where the library acts as a conduit for ingest and consults with the faculty member on appropriate placement of materials (whether in the IR, in a domain-specific repository or somewhere else).<sup>7</sup>

<sup>5</sup> Such as, Ann Green and Myron P. Gutmann. "Building Partnerships among Social Science Researchers, Institution-Based Repositories and Domain Specific Data Archives." *OCLC Systems and Services*, vol. 23, no. 1 (2007).

<sup>6</sup> As of winter 2008, two positions have been created in the library that will add expertise to this area. The two positions are Head, Scholarly Communications and Data Services and Informatics Librarian.

<sup>7</sup> One of the key stakeholders, the Director of Academic Technology position in campus IS, was vacant through the research period. Now that the position has been filled, collaboration and dialogue can take place with everyone at the table.



## Analysis

I conclude that consistently applied data management techniques for faculty is not a realistic goal in a decentralized environment such as a medium-sized university. Appropriate training tools and visible resources will hopefully provide a reasonable level of support for faculty dealing with large volumes of data. Developing the tools and programs that support computing-intensive research involves working with stakeholders from various units on campus. It is challenging to address effective collaboration across major campus units. Such units often have a somewhat shared mission regarding information management, but they often might interpret its execution quite differently. Our next step is to develop infrastructure that includes the most appropriate campus units but also provides the best user access and support. Our goal is to create a system that is visible, campus-wide, and utilized and appreciated by the faculty.

Does your university archives have born-digital records?

Share how you are effectively managing these digital records by submitting a case study to *Campus Case Studies*.

Visit [www.archivists.org/publications/epubs/CampusCaseStudies/](http://www.archivists.org/publications/epubs/CampusCaseStudies/).