

Robert Lundeen OSU Library Faculty Development Grant END OF FUNDING PERIOD REPORT

TITLE OF PROJECT: *Professional development in data curation*

Applicant(s): Amanda L. Whitmire

E-mail: Amanda.whitmire@oregonstate.edu

Telephone: 541-737-3133

Time Period of Grant: July 28 – August 3, 2013

Amount of Award Received: \$2,850

I. PROJECT SUMMARY

Goal 3 in the *OSU Libraries & Press Strategic Plan 2012-2017* describes a commitment to “curate the intellectual output of the Oregon State University community ... to support the entire life cycle of that data including selection, certification, maintenance, preservation, discovery, and long-term access.” In order to succeed in this aspect of the goal, OSUL&P needs to have faculty on board who have expertise in data curation. I proposed to use a Lundeen Award to facilitate professional development in data curation by attending a five-day workshop, *Curating and Managing Research Data for Re-Use*¹, at the Inter-university Consortium for Political and Social Research (ICPSR). The workshop was intended to, “assist individuals working with data to apply efficient curation practices to ensure the usability and safekeeping of data resources. Participants learned about best practices for managing research data, how to apply them to daily operations, and the types of tools that can assist in curation efforts. This workshop was for individuals actively engaged in the curation and management of research data, particularly data librarians, data archivists, and data stewards with responsibilities for data management”.

The Lundeen Award is intended to boost professional capacity in areas that are critical to advancing the Library mission and that benefit our patrons². While I have significant practical experience in data curation, I lacked formal training in this area. Attending this workshop enhanced my ability to knowledgeably curate research data in ScholarsArchive@OSU (SA@OSU) and to assist patrons with curating their own data in the event that SA@OSU is not a suitable option for them. This both addresses our Strategic Plan and benefits the faculty and students who will increasingly rely on OSUL&P to assist with the curation of the products of their research.

Aside from explicit training in data curation, there were two important indirect benefits of attending the workshop that have value which should not be underestimated: gaining hands-on experience with social science data (my background is in the natural sciences), and networking with the instructors and other participants. The workshop was designed to be interactive and hands-on, using a social science dataset as a case study in the process of curation through the ICPSR

¹ “Curating and Managing Research Data for Re-Use,” *ICPSR Summer Program in Quantitative Methods of Social Research*, accessed April 23, 2013, <http://www.icpsr.umich.edu/icpsrweb/sumprog/courses/0149>.

² “Lundeen Award | OSU Libraries | Oregon State University,” *OSU Libraries and Press*, accessed April 23, 2013, <http://osulibrary.oregonstate.edu/awards/lundeen-award>.

deposit process. While it is difficult to imagine that curating social science data is significantly different from curating data produced by scientific research, the workshop broadened my experience and increased my data literacy. This workshop provided an opportunity to gain familiarity with a social science dataset in a scenario that didn't involve taking risks with a 'real' dataset. The ICPSR workshop was taught by experts in data curation and a colleague from a peer institution who is leading the way in developing academic data services (see Appendix II for an instructor list). Attendees of this workshop were peers at universities who are also developing library-based data curation services. Data curation is relatively new to the scope of library support services, and a significant amount of practical information is being exchanged peer-to-peer. The value of these interactions can't be assessed or quantified, but they are essential to faculty professional development and are impossible to achieve via email. There is no substitute for being present.

In summary, I attended the ICPSR workshop *Curating and Managing Research Data for Re-Use* in order to build professional capacity within OSUL&P that would empower us to achieve our goal of supporting the whole lifecycle of data (Goal 3 of the Strategic Plan) and provide more knowledgeable support for patrons in the curation of their data. Participation also facilitated productive interactions with peers within a burgeoning community of practice.

II. ACCOMPLISHMENTS and EVALUATION

1. Describe what was achieved during the time period of the grant.

Attending the workshop facilitated my professional development in the area of data curation, which has improved the capabilities of OSUL&P as a whole to act as a steward of the research outputs of the university. The specific learning outcomes of the workshop were achieved:

- Gain understanding of and ability to apply the best practices for managing research data, including management of confidential data and effective documentation practices;
 - Coming from the natural sciences, it was very beneficial to spend focused time interacting with a social science dataset, and learn about considerations for privacy, confidentiality, de-identification, etc.
- Increased ability to both conduct and teach data review and preparation for deposit into a repository (SA@OSU or otherwise);
 - I haven't yet curated a social science dataset in SA@OSU, but half of my students in GRAD521 were from the social sciences. My experience at the ICPSR workshop was paramount in giving me the ability to teach these students knowledgeably.
- Increased capability to create, comply with, and evaluate required data management plans;
 - As with the previous point, my interaction with a social science dataset and its documentation (metadata) substantially increased my ability to review social science data management plans in a meaningful way.
- Understanding of data repository requirements and assessment;

- During the workshop we did an activity centered on comparing and assessing data repositories. I liked the activity so much that I replicated it for my GRAD521 students.
- Ability to identify and implement the tools that can assist in curation efforts
 - We also did an activity using various metadata creation tools, and learned about software for running checksums and “bagging” dataset files together for preservation. Again, I found this experience so useful that I replicated it for my students.

2. What aspects were completed as proposed? If your study could not be completed as proposed, explain how your plans were altered.

The project was completed as planned.

3. Did the project accomplish what it intended? Did it make a difference?

Without a doubt. I think my comments in Question I.1. summarize things pretty well. While I still haven't had the opportunity to curate a social science dataset in SA@OSU, the workshop did confer the anticipated skills and knowledge. I am much more familiar with the formats and types of social science data and metadata, and the tools used to create them. As I mentioned, this experience has already been significant in improving my ability to teach students in the social sciences and be able to review social science data management plans. These were both unexpected benefits of attending the workshop, beyond being able to curate data for our repository. The experience of learning about the ICPSR's process for curating the datasets in their repository was also very informative to improving our process here.

a. *Include any relevant quantitative data, if applicable.*

N/A

b. *Include any anecdotes, if applicable.*

None. But, if you want to see how datasets are being curated in SA@OSU, feel free to browse around our Collection: <http://ir.library.oregonstate.edu/xmlui/handle/1957/40397>

4. What would you do differently next time, if anything?

I wouldn't change anything.

III. IS YOUR PROJECT COMPLETED?

B. Yes No

IV. FINANCIAL STATEMENT

Explain how the funds were spent. Attach your original budget and indicate how well your estimates matched with actual expenditures. Receipts or documentation of expenses are desirable but not essential.

JoLynn has all of the receipts on file. The flight was more expensive because I made a mistake and did not ask ahead of time when the workshop ended on the last day. It ended at noon rather than 5 pm, and I was lucky that OSULP was gracious enough to let me reschedule my flight to be able to get home a day earlier. It's a mistake that I won't make again. Everything else is in the ballpark of the proposed budget.

| Budget category | Proposed Budget | Actual Cost |
|------------------------|------------------------|--------------------|
| Flight | \$550.00 | \$834.60 |
| Agent booking fee | \$25.00 | \$50.00 |
| Ground transportation | \$109.00 | \$107.00 |
| Hotel | \$666.00 | \$521.70 |
| Registration | \$1500.00 | \$1500.00 |
| TOTAL COST | \$2850.00 | \$3013.30 |

V. SHARING YOUR PRODUCT/RESULTS

As a professional development exercise, most of the impact of this project is disseminated indirectly through my increased skills and knowledge. However, as I mentioned previously, I've already passed along some of the workshop experiences to my students, and have been able to apply the experience toward reviewing DMPs for faculty. Rather than rehash the workshop content in my LFA Seminar, I gave a presentation that updated the campus community on our progress and actions in data services over the last 18 months. I briefly described how the ICPSR workshop provided me with professional development across our various activities.

VI. NOTE

Information included in this report may be reprinted or posted on the web for dissemination.

Adapted from Librarians Association of the University of California Research Grants Program
<http://www.ucop.edu/lauc/committees/rpd/endoffundingreport.doc>