

UCSB Library Data Curation Pilot Projects

Status report, July 2015

The pilot projects are finally underway! Teams for each of the three pilot projects have been staffed, have held initial meetings with each other and with the associated faculty members, and have created project plans and timelines. Additionally, the advisory group for the pilot project activity as a whole has been assembled and has held an initial meeting.

Briefly,

- The Bren project (curation of masters student group project data and other artifacts) is focusing on student education first, and will attend to the historical group project data later. The timing of the pilot project is fortuitous in that we were able to just catch the tail end of the outgoing group projects and the startup of next year's projects. The team is currently surveying the data managers of the outgoing projects to understand the data types and curation issues the students typically encounter.
- The Porter project (curation of small-scale imagery generated by a single researcher) held an initial meeting to understand the faculty member's workflow and data management practices. The next step is to investigate possible repositories for her imagery.
- The Ford project (curation of a large, mostly geospatial data collection) is focusing on development and documentation of standard procedures that can be uniformly adopted by the faculty member's resident data manager and rotating group of student workers.

Challenges

A common theme across the projects is the definition of the Library's level of support for faculty research data, a level which is to be found on the spectrum between doing nothing (i.e., not within the Library's scope) and doing everything (acquiring the data as a Library collection). This mid level of support can be characterized as consultation, but therein lie a couple challenges.

The first challenge is the generic problem of providing advice. For advice to be useful, it must be received, understood, remembered, and ultimately acted upon. Thus to provide useful advice we will need to gauge its receptivity and effectiveness, and we will need to tailor the advice depending on the audience. Best practice type documents and metadata requirements will likely work well with Bren students, but faculty, by contrast, will likely require contextualization and interpretation of best practices, and customization for their particular data and situation.

The second challenge is that, since the advice addresses *how* faculty work with their data, we run the risk of the advice delving into the realm of personal work practices and habits. This is seen most acutely on the Porter project, where both the information flows and the management practices from the faculty member's hand-written field notebooks to the digital imagery that appears in her publications are fairly seamless and integrated. The Library will need to define a line demarcating what is within scope of data curation and what is not.

Opportunities

The role that subject librarians might play in curation is extremely promising. While subject librarians lack experience with data curation, they enjoy great rapport with faculty and speak the same language. Even if subject librarians never gain sufficient expertise to answer data curation questions directly themselves, but can only refer questions to other Library curators, nevertheless their role here and their ability to outreach to faculty is vital. But their success in playing that role is contingent on being able to recognize (potential) data curation issues and in being aware of available curation resources, both within the Library and without.

Another promising opportunity that has emerged is the role that the Library can play in the Bren school. Masters students there participate in group thesis projects, and the school has a well-established curriculum and well-honed infrastructure for these projects that includes the nomination of a "data manager" for each project and education of data managers by system support staff in the use of disk storage and other compute resources. But missing from this education are any topics related to data management and curation (including organization, identification, metadata, and provenance). The Bren school would welcome education in these areas; the infrastructure to provide that education is already in place; and the Library is in a prime position to provide it.

Next tasks

The central task of the pilot project leadership, in consultation with the advisory committee, is to define the actual services that the Library will be supporting and performing. While the pilot projects are intended to be instructive, they themselves, being multi-month in duration and staffed by multiple librarians, are not exact models. Future faculty interactions will need to be more transactional and efficient. We expect efficiencies to come as a result of librarian education in data curation tools and techniques; in the development of standard materials (e.g., best practice handouts) and other resources; and in improved understanding of common problems and solutions. Regarding the latter, as a reminder, the pilot projects were chosen precisely because of their ability to serve as models:

- the Bren project is a model for student education, development of best practices and repository requirements, and institutional relations;
- the Porter project, a model for single-researcher workflows and repository self-deposit; and
- the Ford project, a model for standardizing group practices.

The definition of standard models of interaction and the creation of pre-prepared resources will be key to the sustainability of the Library's services in this area.

Finally, as a first step towards defining the services, we will be starting to create a web presence for the Library's curation activities.