September 2013

Changing Library Operations-Data Curation

Allen McKiel

Western Oregon University, mckiel@wou.edu

Jim Dooley

University of California, Merced, jdooley@ucmerced.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

McKiel, Allen and Dooley, Jim (2013) "Changing Library Operations-Data Curation," *Against the Grain*: Vol. 25: Iss. 4, Article 44.
DOI: <https://doi.org/10.7771/2380-176X.6596>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Changing Library Operations — Data Curation

Column Editors: **Allen McKiel** (Dean of Library Services, Western Oregon University) <mckiel@wou.edu>

and **Jim Dooley** (Head of Collection Services, University of California, Merced) <jdooley@ucmerced.edu>

Column Editor's Note: As **Allen McKiel** explained in his note accompanying his first article for this column, he and I have been presenting together at the **Charleston Conference** on changing library operations for over five years. The column is an extension of the topics we have been discussing. This article on the importance of academic libraries providing data curation services is my first contribution to this effort. — **JD**

The phrase “transition from print to electronic” implies an orderly progression from one state to another. For most libraries, however, the reality has been the addition of electronic content and services on top of existing print-based content and services. If the transition has an end point, it is far in the future. For the present and the immediate future, libraries will have to support both print and electronic formats. At the same time, the types of content and services being developed in the electronic realm clearly originate in the print world, and the evolution of technical services has largely involved the adaptation of print-based systems and workflows to handle electronic resources.

Data curation services are different in that they require libraries to work in largely new areas and library staff to develop new skill sets. This is very different from being at an awkward place in a transition and represents new work permanently added to existing work. Development of these new services requires the library to work with campus IT departments in ways not previously required. Even if libraries didn't set out to become campus leaders in data curation, governmental and institutional requirements are forcing them to play an important role. While attention may be focused on research universities, at this point even liberal arts college libraries need to address these issues.

Aside from the question of library involvement, research data curation has become increasingly important simply because the amount of data produced has increased exponentially in a short time. This phenomenon is often referred to as “Big Data.” It is not uncommon for even relatively small institutions to have research centers or institutes that produce data that must be managed. Advances in computer science have allowed researchers to work with ever-larger data sets, both numeric and textual. Research data output is by no means limited to the hard sciences; social scientists are producing large quantities of survey data. Even in the humanities, data intensive digital humanities projects are becoming much more common. These developments have led administrators to focus on the increasing need to store, back up, and migrate massive quantities of data. Storing data on hard drives under researchers' desks isn't a viable solu-

tion any more. At the same time, researchers themselves are looking for ways to cite and share data for publication and to meet new requirements of funding bodies.

All this begs the question of why libraries should become involved in data curation at all. Why not leave these issues to campus IT departments and/or Offices of Research? The answers will vary depending on local conditions, but will likely involve some combination of capacity and opportunity. For many institutions, central IT offices or Offices of Research may simply not have the capacity to take on data curation activities. To do so may require significant expansion of their missions and corresponding increases in resources that may simply not be available. In such circumstances, partnership with the library or even the library assuming a leadership role may be the best way to address the issues.

For the library, involvement with data curation is an opportunity to provide new services to the campus community and at the same time broaden and redefine its mission. Academic libraries have always been about providing access to information resources. Traditionally these resources have been created by and acquired from others outside the university. Over the past few years, many academic libraries have come to recognize that faculty research output is also an important information resource that the library should be managing. The response to this realization has taken several forms: advocacy of open access publishing alternatives; institutional repositories; formal electronic publishing; management of university presses. Involvement with data curation can be seen as one more aspect of managing the intellectual capital of the university.

Research universities depend on external funding to support many of their activities and thus have a vested interest in ensuring that their faculty are competitive in the pursuit of research grants. New requirements for data management plans and open access to data, as well as cuts in federal government support, have raised the bar for what faculty need to do to remain competitive. Libraries have traditionally placed considerable emphasis on user services, outreach and instruction. Librarians are thus particularly well qualified to provide instruction and guidance to faculty in meeting these new requirements.

The **University of California, Merced (UC Merced)**, the tenth **University of California** campus, began operation in September 2005, expects to have 6,300 FTE in September

2013 and is planning for 10,000 FTE by 2020. It hopes to achieve a **Carnegie Classification** as a Research University (high research activity) by 2015. In addition to research conducted within its four schools, it houses the following interdisciplinary research institutes: the **Center for Excellence on Health Disparities**, the **Health Sciences Research Institute**, the **Sierra Nevada Research Institute**, the **UC Advanced Solar Technologies Institute**, and the **UC Merced Energy Research Institute**.

This relatively rapid growth in research activity has not been accompanied by a corresponding growth in support infrastructure. Under the circumstances, the library was best positioned to provide the level of support needed by faculty. Librarians were most knowledgeable about data curation needs and about resources available to meet these needs.

In 2012 the library hired a data curation librarian to help faculty manage their data and prepare data management plans required by external funding agencies. Also in 2012 the library participated in the second cohort of the **DuraSpace/Association of Research Libraries (ARL) / Digital Library Federation E-Science Institute** even though the library is not a member of **ARL**. The purpose of the institute is “to help academic and research libraries develop a strategic agenda for e-research support, with a particular focus on the sciences.” <http://duraspace.org/e-science-institute>. One outcome from the development of this strategic agenda was an increased awareness on the part of senior campus administration of the importance of adequate campus-wide support for research computing. The 2013 library strategic agenda declares that the library will “serve as the campus clearing house for data curation” and will “work strategically with faculty to meet their needs for data curation.” In the **UC Merced** context the phrase “campus clearing house” is significant in that it acknowledges that the library will not have to create data curation solutions from scratch. Rather the goal will be to connect faculty with the resources available through the **University of California (UC)** system.

Many services deployed by the ten **University of California** campuses are developed and maintained by the **California Digital Library (CDL)** which is often referred to as the eleventh UC library. The **CDL** has established the **University of California Curation Center (UC3)** to help “researchers and the UC libraries manage, preserve, and provide access to their

continued on page 77

Changing Library Operations from page 76

important digital assets.” Current services are the Merritt repository; EZID (easy-eye-dee), a service to create and manage persistent identifiers; DataUp, a tool to manage data in Excel spreadsheets; and the DMP (Data Management Plan) Tool to help researchers create and manage data management plans. Detailed information about these services can be found at <http://www.cdlib.org/services/uc3>.

The Merritt repository, named for Lake Merritt in downtown Oakland where the CDL has its offices, may be thought of as the UC institutional repository. It began as a service that provided permanent storage for digital objects. A front end is now available to allow viewing of publicly-available material including publications and electronic theses and dissertations (ETDs). Planning is currently underway for a new interface including the potential to expose content for search engine indexing. In the beginning Merritt only accepted textual files; it can now accept audio and video files and, most recently, data sets with persistent URLs. In addition to Merritt, UC researchers also have access to data management services provided by the **San Diego Supercomputer Center**.

As was noted above, one of the principal drivers for library involvement with data curation is grantor insistence on the creation of data management plans for funded research. The **National Science Foundation** began requiring a formal data management plan as part of the proposal process in January 2011. Many other funding agencies have followed. Currently data management plans are also required by the **National Institutes of Health**, the **Institute for Museum and Library Services**, the **National Endowment for the Humanities**, and the **National Oceanic and Atmospheric Administration**, among others. The memorandum from the federal **Office of Science and Technology Policy** of February 22, 2013 entitled *Increasing Access to the Results of Federally Funded Scientific Research* signals an increasing federal government interest in openness and access to federally funded research. Data management and preservation requirements are an integral part of this effort. To assist researchers in meeting these requirements, UC3 and external partners have created the DMP Tool. This tool helps users create ready-to-use data management plans for specific funding agencies, provides guidance on how to manage data, and provides information on resources available to help meet granting agency data management requirements. As of June 2013 the DMP Tool was being used by over 5,600 researchers at over 700 institutions.

Another service of the **UC Curation Center** is EZID (easy-eye-dee) which allows researchers to create persistent identifiers, including ARKs and Data Cite DOIs. These facilitate both the citing of data sets in publications and the sharing of data sets among researchers. As of June 2013 the service was being used by forty-nine academic institutions and other organizations.

These three components: an institutional repository providing preservation and access, a tool to help researchers create data management plans and a tool for the creation of persistent identifiers for data sets are basic to any data curation service. In addition to these, UC3 also supports DataUp, a tool to facilitate the management and archiving of tabular data on spreadsheets such as Excel. This tool checks for possible formatting issues, creates metadata and describes the attributes of the data set in standardized format, obtains a unique identifier for the data set, generates a citation, and posts the data and metadata to a repository.

One of the clearest explanations of why libraries should embrace new roles in data curation is found in the **ARL** report *New Roles for New Times: Digital Curation for Preservation* by **Tyler Walters** and **Katherine Skinner** http://www.arl.org/storage/documents/publications/nrnt_digital_curation17mar11.pdf.

“We suggest how research libraries need to be repositioned as vibrant knowledge branches that reach throughout their campuses to provide curatorial guidance and expertise for digital content, wherever it may be created or maintained. We argue that libraries can no longer expect that researchers and scholars will come to them for advice and assistance; libraries must instead find new ways to reach them wherever they may be. Research and learning activities are increasingly intra- and inter-institutional, collaborative, interdisciplinary, international, and virtual. We show how the library must adjust its service offerings to

Basch is in business to help you save time & money!

We offer multiple platforms designed for libraries & corporate procurement.

The Basch Advantage

Call us today to hear about our complete subscription management solutions!

- E-journal activation & linkage to paid and open access content
- E-journal URL maintenance
- Custom & branded e-portal
- Site license management
- Publisher license requirements
- Check-in software for print titles
- Overdue issue notification
- Publication schedules
- Publisher payment history
- Express payments to publishers
- Responsive & empowered staff
- Consolidation services
- Online renewals
- Real time reports

10 Ferry Street, Concord, NH 03301
(P) 603-229-0662 (F) 603-226-9443
www.basch.com

this new landscape in order to remain viable.” Increasingly, this vision is being adopted by all types and sizes of academic libraries, not just those considered traditional research libraries.

All academic libraries face budget and staffing constraints. In many cases, involvement in new work such as data curation will of necessity involve giving up some traditional work and repurposing existing staff. This can be difficult, maybe even painful, but it is necessary if academic libraries are to adapt to the new realities of faculty research support. 🌱

Rumors from page 75

And speaking of medical content, did y’all notice the mini-interview with **Tom Richardson** of the **New England Journal of Medicine** about the newly-established **NEJM Group**?

<http://www.against-the-grain.com/2013/08/a-mini-interview-with-tom-richardson-director-institutional-sales-and-service-at-the-new-england-journal-of-medicine/>

And — also — did you notice that we are doing weekly **Rumors** online? <http://www.against-the-grain.com/category/atg-originals/rumors/>

It has been **raining non-stop in Charleston** all summer and fall. Here’s hoping it’s finished in time for the **Charleston Conference!** Have you registered yet? www.katina.info/conference

Meanwhile — see y’all in November or online at www.against-the-grain.com/.

Love from Yr. Ed. 🌱