Business Librarians Donning the Data Hat: Perspectives on This New Challenge

Journal: Public Services Quarterly

Manuscript ID: WPSQ-2014-0016

Manuscript Type: Column

Keywords: business librarianship, data librarianship, data management, data management plans

Abstract: This column is written by Justina M. Elmore and Charissa O. Jefferson. Justina M. Elmore (elmore@geneseo.edu) is the Business and Data Librarian at the State University of New York at Geneseo's Milne Library. She holds a B.A. in English from Colorado State University at Pueblo and M.L.S. from the University of Buffalo. Charissa O. Jefferson (charissa.jefferson@csun.edu) is the Business and Data Librarian at the Oviatt Library at California State University, Northridge. She holds a B.A. in Women's Studies from California State University, Long Beach, M.A. in Cultural Studies from Claremont Graduate University, and M.L.S. from University of North Texas.
Business Librarians Donning the Data Hat: Perspectives on This New Challenge

Justina M. Elmore
Milne Library, SUNY Geneseo, New York

Charissa O. Jefferson
Oviatt Library, California State University, Northridge, California

URL: http://mc.manuscriptcentral.com/psq  E-mail: elizabeth.blakesley@gmail.com
Business librarians donning the data hat: Perspectives on this new challenge

Data Services: An Introduction

The concept of ever-shifting job descriptions is not new to librarians. As technology has changed, so have libraries and the nature of the work performed within them. For business librarians, a quick look at ALA’s joblist shows a recent trend in the addition of data services to our job descriptions, offering a glimpse into the future of academic business librarianship. As far back as 1973, libraries were grappling with the data explosion (McAnally & Downs). Yet, we work in an environment today where the world’s knowledge is said to double every two years (Lohr, 2012). It isn’t any wonder that libraries have begun exploring the roles of data management with more urgency.

Much has recently been written about data services and how colleges and universities are positioning to meet new needs and requirements. For instance, Information Standards Quarterly (ISQ) dedicated its fall 2013 issue to data related issues (Callaghan, 2013), organizations have developed models for managing data (e.g. Digital Curation Centre’s Data Curation Lifecycle model & the Data Conservancy Stack Model for Data Management out of Johns Hopkins University), many are developing extensive digital repositories, and yet others have focused on teaching data management best practices and creating tools and processes to assist faculty with data management plans (DMPs) for grant proposals. Further, open data initiatives like the Denton Declaration and the White House Memo on Open Data are gathering momentum and more funders and publishers are requiring that authors make their underlying data available (Harvey, 2010).

Data management and resource sharing plans are now mandated for applicant’s submissions for certain federal research grants from both the National Science Foundation (NSF)

URL: http://mc.manuscriptcentral.com/psq  E-mail: elizabeth.blakesley@gmail.com
and National Institutes of Health (NIH). The NSF requires a Data Management Plan as a supplementary document to the grant application, which must indicate how researchers will store and make their data available. The NIH has a similar policy called a Data Resource Sharing Plan for grants of more than $500,000, which specifically encourages open access for analysis and to further opportunities for reuse in new scientific research. Under these recent federal policies and mandates, there is no single standard for the data itself. Even the definition of data among scholars is questionable because data will look different between the different disciplines. The development of data management services in such a data saturated environment is still new and unclear. The professional research community can only look to those who are developing best practices for a model.

**Pre-Existing Use and Environmental Scan**

With so much going on under the auspices of data management services, approaches to providing such services varies widely between organizations, especially between large universities with a greater graduate student population and smaller colleges with mostly undergraduate student populations. For instance, many larger research centers focus on data management instruction at the graduate level, whereas smaller colleges with fewer graduate students are incorporating data management best practices into instruction undergraduate level. The number of institutional repositories has increased dramatically over the past decade

(Triumph, 2012). For example, the University of Buffalo has developed a robust institutional repository for containing articles, data and other scholarly output along with services to provide assistance with preservation, curation, alternative publishing and data sharing. Alternatively, Syracuse University has plans to develop a data specific repository separate from the one that holds scholarly publications (Warren & Carrier, 2014). Further, Purdue University Research

URL: http://mc.manuscriptcentral.com/psq  E-mail: elizabeth.blakesley@gmail.com
Repository (PURR) serves as hub for faculty and students to store and share data in addition to offering extensive data services, with many of these projects supported through grant funding. Purdue has created data literacy standards and offers a data curation toolkit to assist researchers in developing data management plans. Purdue’s libraries also offer data management resources and assistance in data management planning by offering consults and data citation guidance. Additionally, University of California UC3 California Curation Center at California Digital Library offers Data Management Guidelines. This guideline outlines various funding agency requirements and provides tips for creating data sharing and management plans respectively.

**Needs Assessment and Implementation**

The expectations and data needs at California State University at Northridge (CSUN) are still evasive because the majority of the campus community is not yet aware of the importance of data management. The need for a web resource that librarians at CSUN’s Oviatt Library can utilize when consulting with faculty and graduate students who need to manage their data has been identified. This web resource will serve as a guide to best practices and tools in efforts to manage data and result in better metadata for access and retrieval. The institution also needs a cohesive standard on data literacy competencies. This will improve issues surrounding management, access and preservation that could impede the ability to fully utilize the capabilities of ScholarWorks, the institutional repository (IR).

Many institutions are unsure of data needs and budget constraints do not allow for new data librarian positions, which is why it is often tacked onto an already existing title, in this case business. Part of the expectations for the data librarian is to educate the campus community on the importance of data management and sharing by providing resources for researchers to implement in their own respective fields. These expectations extend to ScholarWorks, to which

**URL:** http://mc.manuscriptcentral.com/psq  **E-mail:** elizabeth.blakesley@gmail.com
all NSF and some NIH grant applicants point to for their data management plan. ScholarWorks is maintained on campus by a digital initiatives librarian as well as the library project manager for the IR.

As the Business and Data Librarian, there is an expectation to expose researchers to ScholarWorks, especially if they included ScholarWorks on the DMP for their grant proposal. The need to streamline a workflow for follow-up for researchers who have stated in their grant applications that they will utilize ScholarWorks for data management to ensure that they’re proceeding according to plan is crucial. The Business and Data Librarian has the challenge of reaching out to those researchers to ensure that they are conscientious of the terms of their grants.

Until very recently there has not been a lot of enforcement by universities nor funding agencies that require DMPs. However, funding agencies have begun to hold researchers much more accountable, even withholding funding from those who have not complied with DMP requirements (Noorden, 2014). This presents an opportunity for partnership with CSUN’s Office of Sponsored Projects and Research. This office works directly with researchers in the grant application process, but there is potential for more thorough follow-up during the course of the research project. Since the Projects Office directly benefits from researchers who have received funding on behalf of the university, it would be in their best interest to ensure researchers are doing what they promised in their application. The Business and Data Librarian can partner with this office to educate grant writing consultants on how to implement the expected outcome and refer the researcher to the appropriate contact for ScholarWorks.

In juxtaposition to larger institutions which usually have a number of business librarians or even dedicated data librarians across multiple campus libraries, SUNY Geneseo’s Milne
library has one business librarian who serves a small liberal arts college with a student population composed largely of undergraduate students. As with any small academic library, Milne’s subject librarians are required to wear a number of hats and serve in a variety of roles—and the business librarian is no exception. Prior to the Fall 2012 semester, the business librarian was responsible for several liaison roles for the school of business and several other departments in the social sciences, which include responsibilities for collection development, library instruction, scholarly communication, web development, marketing library services and the like. When the position became vacant and the job posting was developed during the summer of 2012, the position title was changed to Business and Data Librarian with data services added as a new component.

Since library data services initiatives are relatively new to Milne, librarians have endeavored to discover the true needs of our students and faculty by employing an environmental scan. Data-related services are being fleshed out based on discoveries made through a series of faculty discussions, webinars, workshops and research. Milne librarians have made a concerted effort to reach out to interested faculty and staff across campus, which led to a series of interdisciplinary discussions across campus in a summer of data exploration. During these discussions, several areas where the library could serve were identified:

- Provide assistance locating discipline specific repositories for finding secondary research data and publishing primary research data.
- Provide embedded library instruction or workshops for undergraduate students to improve skills in managing both laboratory and their own data.
- Provide assistance in developing data management plans for funding applications.
- Identify faculty work or research projects that could/should be digitally curated.

URL: http://mc.manuscriptcentral.com/psq E-mail: elizabeth.blakesley@gmail.com
Milne Library encourages data sharing by offering a suite of scholarship and publishing services, which includes suggestions and assistance with finding, managing and storing data, data citation (citing secondary data or generating recommended citations for their own primary research data), and locating discipline-specific repositories and their required metadata standards for contributing content. To meet faculty needs for finding secondary data or for making their own data openly available, Milne has developed a robust list of links for finding data and has also linked to resources such as DataBib, a searchable catalog for finding data repositories for multiple disciplines. These resources are available on a library guide for students and faculty in search of open data sets or a list of repositories to which they might contribute their data.

SUNY Geneseo does not currently have an institutional policy for data management and data sharing, nor is there any SUNY-wide policy (nor recommendations) on this issue. However, incremental steps are being taken towards better stewardship of the research data created at SUNY Geneseo. Librarians coordinate with the Office of Sponsored Research to assist faculty in developing data management plans. The library provides a link to the DMPTool, developed by the University of California Curation Center and the California Digital Library, to assist faculty in creating their DMPs or quickly determining the requirements for specific funding agencies. Despite the fact that faculty do not apply for large grants where funders require DMPs, librarians have begun having conversations with faculty about data management and open data initiatives. One useful strategy has been to work data management best practices into library instruction sessions (some embedded and some one-shot) and then use that as a springboard for conversations with the faculty member about their own scholarship. Milne librarians have also made themselves available to assist faculty in finding relevant services and resources within the library and across campus for long term archiving and preservation. Rather than create a small

URL: http://mc.manuscriptcentral.com/psq E-mail: elizabeth.blakesley@gmail.com
institutional repository at SUNY Geneseo, the library, campus IT, and the Office of Institutional Research will collaborate with one of the larger university research centers within the SUNY System to store/preserve large or restricted data sets that cannot be contributed to an open data repository.

Also new for the Spring 2014 semester, Milne Library now offers a one-shot data management workshop for undergraduates each semester. However, since a one-shot workshop may not appropriately address discipline specific considerations, instruction librarians have also begun working with faculty to incorporate data management education into embedded library instruction sessions where applicable. To further that aim, the two librarians involved in the initial environmental scan are working with instruction librarians to begin mapping current library instruction to the twelve data literacy core competencies developed by Purdue (Carlson, Fosmier, Miller, & Nelson, 2011). In addition, progress has been made in discovering discipline specific metadata standards so liaison librarians can make recommendations to or better assist faculty with developing and/or reviewing their data management plans.

The survey of SUNY Geneseo faculty interested in data issues conducted this past summer revealed that none of the faculty who responded had published/shared data nor tracked citations of their data for tenure or promotion. Only two faculty members (social science & statistics) were interested in working with the library to acquire and reuse secondary datasets for their own scholarship. Thus, later discussions among liaison librarians included conversations involving staff from institutional research on ways that our institution could encourage data sharing and reuse among faculty. As a direct result, Milne included data services among the suite of scholarship and publishing services mentioned above. When approaching faculty about openly sharing data, it is helpful to be armed with evidence that researchers who provide open access to
their data report higher citation rates for their publications (Piwowar, Day & Fridsma, 2007) and that reuse continues to increase among datasets that are shared (Piwowar & Vision, 2013), especially given that most faculty see publishing their data as more work with not a lot more value added for the effort. However, to fully legitimize data publishing, it must be better recognized for purposes of granting promotion and tenure.

Currently, there is no great demand at SUNY Geneseo for data curation with respect to faculty research. However, librarians there have begun developing skill sets in this area in anticipation of future needs. Such initiatives are, as yet, an untapped opportunity. In the meantime, librarians will continue to monitor faculty publications and approach them about curation opportunities, creating data profiles based on Purdue’s Curation Profiles Toolkit.

Conclusion

While excited for the opportunities for growth in this area for both academia as a whole and more organically within the institution, it is difficult not to become overwhelmed with what UKOLN calls “the burgeoning data deluge” (Lyon, 2007). How do we know if/when we’ve done enough to prepare for needs that may arise in the near and far future? In addition to the need for better recognition of data publishing within the tenure and promotion process, there are a number of other foreseeable issues to overcome as libraries work to build data management and curation initiatives. Assuring that data is made available in re-usable formats is a factor. Data is now being created on a larger scale than ever before and current techniques used to assure effective backup (e.g. storage capacity, avoiding disk failure or migration errors, and the time and frequency needed to verify data integrity) can break down (Rosenthal, 2010). Content creators from the various disciplines have differing objectives and have little incentive to go beyond short term data storage/preservation to meet their own needs for publication, even with

URL: http://mc.manuscriptcentral.com/psq  E-mail: elizabeth.blakesley@gmail.com
pressure from funders and publishers to make data openly available (Harvey, 2010). Just getting all of the stakeholders within the organization on the same page and using the same terminology will be a feat (e.g. Is archiving a subset of preservation or vise versa? Are data management and data curation interchangeable terms? Or, is data management a subset of data curation or vice versa?). Additionally, most libraries are just developing data services and having new or little infrastructure creates the potential for problems. Too, institutional repositories are relatively new and it is unknown if they'll stand the test of time (Lyon, 2007).

In spite of these hurdles, libraries are making great strides in the development of data services. A number of open source tools are available online for those getting started with data services and digital initiatives. The URL’s for some of these resources can be found in the suggested reading section of this article. A great place to start is Digital Curation 101: How to Manage Research Data and the tools and services provided on the Digital Curation Centre (DCC) website. Tufts University at UMASS created the New England Collaborative Data Management Curriculum (NECDEMC), which is an instructional tool for teaching data management best practices. It provides modules, activities and lesson plans which can be easily adopted by librarians, undergraduates, graduates, researchers. Other resources for those looking to beef up their skills in data management include the University of Edinburgh’s Do-It-Yourself Research Data Management Training Kit for Librarians, ARL’s resources for e-research, the University of Minnesota’s Data Management Course and the e-Science Portal for New England Librarians.

Furthermore, many librarians who work with data at their research institution offer libguides and websites to guide patrons through the data management resources. These libguides are also helpful for other librarians looking for guidance on creating similar guides for their

URL: http://mc.manuscriptcentral.com/psq E-mail: elizabeth.blakesley@gmail.com
institutions. Some academic libguides on data management worth visiting are available from Massachusetts Institute of Technology, University of Minnesota, North Carolina State University, University of Nebraska-Lincoln, University of Colorado Boulder, Purdue University and Carnegie Mellon (URLs are provided in the suggested websites section).

There are various professional development opportunities available in the form of workshops, webinars and summits. The Special Libraries Association (SLA) offers such webinars as Data 101: A Gentle Introduction taught by Kim Silk, Data Librarian at University of Toronto and Suddenly... I'm Consulting on Data Management Plans taught by Kiyomi D. Deards from the University of Nebraska - Lincoln. The Association for Information Science & Technology (ASIS&T) annually offers a three day Research Data Access and Preservation (RDAP) Summit. The annual Digital Library Federation (DLF) Forum usually has some great sessions on data services. The University of Massachusetts e-Science Initiatives provides an annual symposium and their website has links to documents from previous symposiums. ICPSR will offer a five day workshop called Curating and Managing Research Data for Re-Use on July 28 - August 1, 2014. Syracuse University iSchool offers a MOOC called Introduction to Data Science Open Online Course. Additionally, the e-Science Portal for New England Librarians provides list of other professional development opportunities and a recommended list of blogs that cover data services. You might also consider signing up for rss or email alerts from the Journal of eScience Librarianship. Or, network with others on the Data Management for Librarians Mendeley Group and the New Business Librarians group through ALA Connect.

URL: http://mc.manuscriptcentral.com/psq E-mail: elizabeth.blakesley@gmail.com
References


URL: http://mc.manuscriptcentral.com/psq E-mail: elizabeth.blakesley@gmail.com


Triumph, T. F. (2012). Starting, strengthening, and managing institutional repositories by


DOI:10.7710/2162-3309.1024


**Suggested Reading and Websites**


California State University at Northridge. CSUN Institutional Repository - ScholarWorks. Retrieved from http://scholarworks.csun.edu/


Data Conservancy. (n.d.). Data Stack Model. Retrieved from

URL: http://mc.manuscriptcentral.com/psq E-mail: elizabeth.blakesley@gmail.com
http://dataconservancy.org/community/


URL: http://mc.manuscriptcentral.com/psq E-mail: elizabeth.blakesley@gmail.com


DOI:10.2218/ijdc.v4i2.105


Purdue University. (n.d.). Research and Data Services LibGuide. Retrieved from https://www.lib.purdue.edu/research


URL: http://mc.manuscriptcentral.com/psq  E-mail: elizabeth.blakesley@gmail.com


University of Massachusetts and New England Area Librarian eScience Symposium Retrieved from http://escholarship.umassmed.edu/escience_symposium/2014/


University of Minnesota. (n.d.). Data Management Course. Retrieved from

URL: http://mc.manuscriptcentral.com/psq E-mail: elizabeth.blakesley@gmail.com
https://sites.google.com/a/umn.edu/data-management-course_structures/home-1


URL: http://mc.manuscriptcentral.com/psq E-mail: elizabeth.blakesley@gmail.com