College of Saint Benedict and Saint John's University

DigitalCommons@CSB/SJU

Library Staff Publications   **Library**

2015

Revolutionary or evolutionary? Making research data management manageable

Kristin Partlo

Diana Symons
College of Saint Benedict/Saint John's University, dsymons@csbsju.edu

Jonathan D. Carlson
College of Saint Benedict/Saint John's University, jcarlson@csbsju.edu

Follow this and additional works at: https://digitalcommons.csbsju.edu/library_pubs

Part of the Cataloging and Metadata Commons, and the Scholarly Publishing Commons

Revolutionary or Evolutionary?
Making Research Data Management Manageable

Kristin Partlo
Diana Symons
Jonathan D. Carlson[1]

Introduction

This chapter investigates the role of academic librarians, particularly those at small liberal arts institutions, in providing research data management services. Research data management may not seem like an obvious fit for curricular libraries whose primary mission is supporting teaching rather than faculty research, nor is data curation an obvious need for schools without a data repository or staff who specialize in the preservation and dissemination of data. Yet numerous reports cite data management and data services as critical services for the future of academic libraries (ACRL Planning and Review Committee, 2013; Johnson, 2014; Cox, 2013; Tenopir, 2012). The question raised, then, is how and why are data management services important in the liberal arts context? What can librarians at these institutions do to develop expertise in this growing area of the profession? What services are college and university libraries beginning to provide, and how successfully can existing models be adapted to other institutions? Does the addition of data services transform the mission of liberal arts libraries, and if so, is that transition revolutionary or evolutionary? Liberal arts librarians, as they have with numerous other shifts and trends in librarianship, can turn to models in the literature from research universities, develop communities of practice amongst themselves, and also innovate

---

[1] Kristin Partlo, Reference & Instruction Librarian for Social Sciences & Data, can be reached at kpartlo@carleton.edu. Diana Symons, Social Sciences Librarian, can be reached at dsymons@csbsju.edu. Jonathan D. Carlson, Science Librarian, can be reached at jcarlson@csbsju.edu.

from within their own unique contexts. The authors argue that such collaboration and innovation reflect an evolutionary process as librarians build on existing skills, strategies, workflows, and knowledge. The following pages of this chapter survey the current environment, offer case studies from two small liberal arts institutions, the College of Saint Benedict/Saint John's University and Carleton College, and provide readers with recommended action steps to develop a path of gradual, manageable, shared, and sustainable work in research data management.

Background: Librarianship and Research Data Management Services

Research data are critical components of scholarly communication. Good data management practices and services are needed to store, preserve, and share research data effectively. Using definitions from the ACTI Data Management Glossary, **data management** is "the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets." Data management is something that must happen throughout the lifecycle of data, not just at the end. Proper planning and documentation must take place from the grant-writing stages through collection and analysis to the preparation for dissemination and storage if the data are to be useful to other researchers. This process has not been uniformly taught in Ph.D. programs, and while faculty are aware of the value of proper data management they often struggle to find the time for it or prioritize it above other research concerns (Jahnke, 2012). **Data management services**, then, are "services provided to researchers and faculty to assist them with the creation of data management plans, metadata, storage planning, and curation requirements" (ACTI Data Management Glossary, 2013). **Data curation** goes beyond assisting researchers with data management by providing additional services to preserve and add value to the data over time,

extending from the lifecycle of the research project to its potential reuse. Traditionally handled by data archives, "data curation activities enable data discovery and retrieval, maintain its quality, add value, and provide for reuse over time, and this new field includes authentication, archiving, management, preservation, retrieval, and representation" (University of Illinois, 2014). Finally, **research data services** is a broader concept, encompassing the suite of services provided by libraries (or archives, research institutes, computing centers, or academic departments) in support of collections of data, usually in the social sciences (Geraci, 2012, p. 2.1), such as selection and acquisition, cataloging and access services, reference or discovery services, instruction services, hardware and software support, and public relations and marketing of such services. In recent years, librarians have been broadening the scope of their research data services to address data needs in the sciences and humanities as well.

The research data management practices needed to support a growing culture of data sharing serve to protect the original researchers' work, benefit the larger research community, and add to the public good. Such practices protect researchers from data loss (a concern due to changing technologies and file formats, bit rot, or inconsistent individual data storage and management practices on the part of research assistants, who may only be involved with a project for a brief period). These practices also ensure that researchers will be able to understand and access their own data when they revisit it in the future and provide support if their research is questioned or challenged (Coates, 2014; Shorish, 2012, p. 266). Good data management practices benefit other researchers by allowing them to replicate findings, mine data, use data in teaching, or otherwise reuse and repurpose others' data in new ways (McLure, 2014, p. 141).

The push for more deliberate and comprehensive research data management is the result of many interrelated trends in scholarly communication, including calls for greater

accountability, transparency, and accessibility, as well as changes in publishing technologies, practices, and available platforms. Most large research universities and some small colleges have had data specialists or research data programs for several years. However, recent data mandates from several funding agencies, like the National Institute of Health's (NIH) 2003 data sharing requirement[2] and the National Science Foundation's (NSF) decision in 2011 to require a data management plan (DMP) in all grant applications, served as a catalyst for a much broader swath of American librarians and information professionals to seriously consider offering research data management services.

The NIH, NSF, and other recent data mandates illuminated gaps in campus services that librarians are uniquely positioned to address, particularly in the areas of data management, the creation of local data collections, data literacy and education, and additional data access, preservation, and cataloging work. These are areas described by Haendel as "major bottlenecks in the scholarly communication cycle" (Haendel, 2012, p. 1). Much of the professional library and information science (LIS) literature emphasizes how research data management can be seen as a natural extension of librarians' traditional work of organizing information and making it accessible, the distinguishing difference being that librarians have historically focused on managing the published finished product, rather than the raw data that goes into such products (Nielsen, 2014, p. 221). Librarians have a range of professional interests and strengths (including, according to the NSF, "reliability, expertise in resource sharing, policy development, annotation and selection, and institutional commitment to sustaining access over long periods of time") that would make them assets in research data management projects (as quoted in Gold, 2010, p. 11). Most librarians have strong connections to faculty and the ability to work with

---

[2] http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

other campus partners like research and grants offices and information and academic technology services. An increasing number of academic librarians are also developing or supporting institutional and discipline-specific repositories and are committed to the open access movement to make information more discoverable and accessible. Other professional strengths include librarians' organizational skills, experience developing and applying best practices and preservation standards, metadata expertise, and familiarity with user needs assessment and education (McEwen, 2014; Antell, 2014; Gold, 2010).

Embarking on a path to offer research data management services can be exciting and professionally rewarding. Librarians may discover opportunities to meaningfully expand their roles and develop new expertise, to work closely with faculty at all stages of the research process, and to form new partnerships on campus and with individuals at other institutions (Antell, 2014, p. 558). Research data management is an opportunity for librarians to be involved in the entire knowledge and research lifecycle, both as custodians of published information (traditionally books and journal articles) and now through involvement during the development, research, and publication process (Gold, 2010, p. 2). These new roles will keep individuals engaged and help the library profession demonstrate its relevance and usefulness in a changing academic landscape.

Liaison and subject librarians, in particular, have strong working relationships with faculty members in their liaison departments, are familiar with discipline-specific publication and research practices, and can foster campus-wide communication between disparate departments and offices (Shorish, 2014). The professional literature surrounding data management services repeatedly emphasizes librarians' roles as facilitators and communicators (McLure, 2014; Wang, 2013; Carlson, 2012; Partlo, forthcoming). The academic library's

mission to the whole college helps it break down departmental silos and situates it well to address the convergence of data management needs coming from the social sciences, sciences, and humanities.

This new call to facilitate research data management, however, may cause feelings more akin to trepidation or anxiety than excitement. Despite librarians' complementary suite of skills, expertise, and professional interests, there remain concerns (both external to and within the profession) that librarians do not possess adequate technical or domain-specific knowledge to provide effective research data management services. How can librarians, particularly those without existing expertise or who might be working alone at smaller institutions, learn enough to navigate a wide variety of file formats, available technologies, funding requirements, and metadata schemas and standards, let alone find the time and resources to move beyond this initial information-gathering and expertise-building stage?

Librarians providing research data management services will need to develop "a set of combined competencies" that have traditionally been spread across several domains (McLure, 2014, p. 143). Fortunately, service models, best practices, and tools are increasingly being developed and shared across institutions, and opportunities for professional development are expanding, which makes the prospect of developing desired competencies less daunting. Readers will find a list of recommended readings, resources, and other tools to help get started in "Appendix A: Professional Development Resources."

Literature Review: Data Services Models

It is beyond the scope of this chapter to summarize the long history of research data services that have developed at research universities or the more recent addition of data

management services to that environment. Interested readers can turn to the IASSIST website[3]

and its journal, the *IASSIST Quarterly*, or Charles Bailey's "Research Data Curation

Bibliography" to find abundant literature documenting this work. Of particular interest to

librarians in small institutions are the ways in which models from research universities can be

adapted or scaled to schools that prioritize teaching over research.

Some readers at smaller schools may be interested in adding to existing data services, but

others will be developing their data management support alongside data services in general.

When building research services from the ground up, superb guidance is provided by Geraci,

Humphrey, and Jacobs (2012) on establishing suitable levels of collection, reference, and

computing services for a given institution and developing strategies to promote and integrate

those services on campus. Geraci, Humphrey, and Jacobs' model has been taught for more than

ten years in an Interuniversity Consortium for Public and Social Research (ICPSR) workshop on

implementing data services and has been adapted to many institutions. One example is Minglu

Wang's (2013) case study of the implementation at Rutgers University. Such case studies are

useful since they practically illustrate ways of adapting a model to an institution, even if such

case studies may reflect dramatically different staffing models, needs, and user populations than

are found at most small colleges. Scaling down a carefully thought-out and well-documented

process can be easier and more efficient than starting from scratch.

In addition to the conceptualization and structuring of service levels, research universities

provide examples of how working groups and departments can be created to support campus

research data management. A sampling of such groups include Cornell University's Research

---

[3] For a sense of how long this history is, see especially the 1982 special issue of Library Trends devoted to Data Libraries for the Social Sciences, edited by Kathleen Heim and the 1996 "Bibliography of Selected Works" that traces the history of collaboration between data archives and traditional archives at http://www.iassistdata.org/publications/bibliography.html.

Data Management Support Group, Johns Hopkins University's Data Management Group, Purdue University's Distributed Data Curation Center, University of Massachusetts' Data Working Group, MIT's Data Initiatives Group, and University of Minnesota's E-Science and Data Services Collaborative.[4] Teaching-centric libraries may not be able to devote whole departments to data management or data curation, but working groups, task forces, and interest groups are still achievable and can be an easy fit at small institutions where staff are accustomed to serving in many different capacities and collaborating across departmental divisions. For the newly-minted data librarian at a non-research library, isolation can be an issue, as can be feelings of insufficient skills or specialization. Working groups not only provide new services and drive new policy, but also serve as local support networks, bringing together people with complementary skills around a common goal. Groups might consist of subject and liaison librarians, metadata librarians, digital archivists, academic technologists, GIS specialists, collections librarians, and people engaged in digital scholarship. These types of groups have been essential for all three authors of this chapter.

University libraries' approach to user studies may need to be translated to a smaller environment. While surveys, focus groups, in-depth interviews, and other mixed-method approaches have been used to assess faculty research data practices at universities, liberal arts schools do not always develop services through such formal processes. Library staff at smaller institutions who have multiple areas of responsibility may not be supported with time release or the resources needed to employ such methods. If one's institutional culture encourages taking an organic rather than a scientific approach to creating services, more valued or realistic strategies may include one-on-one conversation, consensus building, and recruiting faculty champions. In

---

[4] For a more complete literature review of case studies written about university services for research data management, see Corrall, Kennan, and Afzal 2013, pp. 645-648.

small communities, informal approaches and word of mouth go a long way. A few strategic conversations and interactions with faculty can be highly effective. In fact, some steps for developing services might be easier or more straightforward at a smaller campus: getting to know the needs of 250 faculty members, for example, is a very different endeavor than assessing the needs of a faculty body of 2,500 or more.

Even if they cannot be implemented directly, resources developed for university environments can be adapted for use in smaller institutions. Gold suggests that those new to data management services "draw on a growing set of resources created by research library leadership (e.g. ARL [Association of Research Libraries]) and leaders in the campus technology community" in order to offer realistic and sustainable research data services (Gold, 2010, p. 10-11). For example, librarians at Colorado State University adapted the Purdue University Libraries' Data Curation Profiles Toolkit to study their researchers' needs (McLure, 2014). Librarians at Carleton have also adapted the interview instruments of the Toolkit to create a handout and a summarized framework to use in conversations with faculty about their data management practices in their research and teaching (see Appendix B). Additionally, Carleton librarians have used this summary as a tool for talking with and training other librarians and student digital humanities associates in the basics of data management. It might have been easy to dismiss the structured approach of the Toolkit as a poor fit for the smaller Carleton environment, but instead it has become a rich resource to which the librarians have returned repeatedly. The librarians at CSB/SJU will continue to adapt this resource for their own faculty needs assessment.

Librarians at liberal arts colleges or other small or teaching-centric institutions have unique considerations to address when first developing research data management services,

particularly in areas related to staffing, resources, researcher needs, student learning, and

potential campus or consortial partnerships. The professional literature is beginning to include

examples of models developed at liberal arts institutions (Dahl, 2014; Goldstein, 2011; Toups,

2013) or other institutions with a teaching emphasis (Scaramozzino, 2012; Shorish, 2012) that

address these considerations directly. As data services evolve and expertise grows at smaller

institutions, these institutions are contributing to and building their own community of practice.

Student data literacy and learning is one area, for example, that has received particular attention

in discussions about liberal arts and teaching institutions' data services.

Current undergraduate and graduate students are increasingly involved in research

projects and will be tomorrow's professional researchers, so it is fitting for liberal arts

institutions to consider student data literacy and education needs when planning data

management services. As Dahl explains,

> At liberal arts colleges, the big opportunities around data services center on students.
> Teaching students how to find, analyze, and present data supports advanced forms of
> inquiry in the liberal arts and builds critical competencies that students will need beyond
> college.[...]Training students to implement data management practices as apprentices in
> the research process is an opportunity to build data literacy among future scientists and
> scholars. (Dahl, 2014, n.p.)

Students could benefit from introductory workshops on statistical software and data-

naming and storage best practices. Goals to promote data literacy, teach and train with data, and

help undergraduate researchers find, interpret, use, and preserve data all dovetail with the

teaching mission at liberal arts institutions. According to Gabridge,

> Librarian subject liaisons already teach students how to be self-sufficient, independent
> information consumers. This role can be easily extended to include instruction on data
> management and planning. Liaisons can offer seminars and other support mechanisms
> (Web page, tutorials) to help student researchers understand what to do with their data
> and increase their awareness of library resources. (Gabridge, 2009, p. 17)

While students' needs for help with quantitative literacy competencies may be relatively clear, librarians must still work closely with faculty when developing related services. There are increasing numbers of data management workshops and curricula shared online, but librarians at smaller colleges should not expect faculty to immediately rely upon them for such expertise. Instead, librarians can position familiar services such as information literacy and course-integrated instruction as part of the network of support for faculty teaching around data. Much of the LIS literature on research data management services revolves around librarians' evolving role in supporting research, but at smaller institutions librarians' roles may focus more on supporting the teaching of data literacy and the development of future researchers' data management habits. Such change, which grows out of librarians' recognized strengths, represents less of a dramatic revolution of services and more of an evolution. The case studies in the next section illustrate this transition at the authors' institutions.

Case Studies

The following two case studies from the chapter authors' institutions show what might be possible at both the outset and then several years into providing research data management services. The CSB/SJU Libraries began providing research data management services only a couple of years ago, while the efforts of librarians and other staff at Carleton College reflect possibilities for a more mature program of services.

Case Study 1: College of Saint Benedict/Saint John's University

The College of Saint Benedict and Saint John's University (CSB/SJU) are two Catholic, Benedictine, residential liberal arts institutions in central Minnesota with a shared enrollment of

approximately 4,000 undergraduate students. As separate women's and men's schools with a

unique coordinate relationship, they offer a shared curriculum and the CSB/SJU Libraries

operate two physical libraries under one joint library service. In 2012, after a faculty member

inquired about assistance with their grant proposal's DMP, the CSB/SJU Libraries created an ad-

hoc Data Management Task Force (consisting of the science and social sciences liaison

librarians) to investigate possible services. The task force members assisted with this initial DMP

and, as they continued identifying useful information, best practices, and available tools, set up

an online guide to recommend existing data management resources to researchers while also

highlighting the Libraries' nascent data management services

(http://libguides.csbsju.edu/data_mgmt). This guide was shared with the colleges' external grants

office, which now refers researchers to the Libraries as needed.

The Data Management Task Force has since assisted other faculty members with their

DMPs, facilitated conversations between a faculty member and Information Technology

Services in order to transform a massive spreadsheet into a secure web-searchable database, and

investigated the capabilities of the existing institutional repository[5] as a data repository.

CSB/SJU uses Digital Commons as its institutional repository platform, and bepress has begun

promoting their product as an option for data storage, but there remain concerns about the

findability of data files and data sets in Digital Commons. Digital Commons is crawled by

Google Scholar, but neither Digital Commons nor Google Scholar have a specific notation or

limiter to designate something as a data file; the only way to find a data file in Digital Commons

is through serendipity or while looking at an article that is associated with that data. Further

---

[5] DigitalCommons@CSB/SJU, which can be accessed at http://digitalcommons.csbsju.edu/.

considerations for comparing the strengths and limitations of institutional data repositories and disciplinary repositories are provided near the end of this chapter.

After working on these initial projects, the librarians at CSB/SJU have realized that starting small (and while juggling several other library projects and initiatives) can still lead to progress and positive results. A faculty member recently asked the Data Management Task Force to weigh in on several shared storage options for documents, including SharePoint, Google Drive, and Dropbox. Even a basic familiarity with these types of tools and best practices for personal data curation (which many librarians may erroneously assume is common knowledge) can help improve students' and researchers' data management habits, and these types of questions may also serve to raise the library's profile and spark further discussions about research data management services.

The CSB/SJU Libraries do not currently have a data, metadata, or digital projects librarian, so one of the Data Management Task Force members' top priorities has been to build expertise and develop a community of practice with librarians at other small colleges in Minnesota. To better equip themselves for further involvement in research data management services, the chapter authors from CSB/SJU have partnered with mentors and collaborators, including coauthor Kristin Partlo; presented at regional conferences; and contributed to broader discussions on research data management, facilitated in large part through the Liberal Arts Research Data Support (LAReDaS) listserv for liberal arts colleges. They also plan to leverage existing on-campus partnerships with individual faculty researchers and academic departments (developed in their roles as liaison librarians) and with campus offices (developed as representatives from the institutional repository committee) to offer a more robust suite of services in the future.

Case Study 2: Carleton College

Carleton College is a highly selective residential liberal arts college in south central Minnesota with an enrollment of 2000 undergraduates. The library and Information Technology Services (ITS) have been offering data services in formal partnership for about ten years and informally for around ten years before that. In 2004, a librarian was hired in a new combined social science liaison and data librarian position with the charge to collaborate with the social science academic technologist from ITS to develop and provide joint data services. This hire coincided with the beginning of a vigorous faculty initiative around quantitative reasoning (QR) across the curriculum. These staff participated in the faculty initiative, emphasizing their services and ways that staff support can increase the potential of what faculty can accomplish in teaching quantitative reasoning. Through building the library's statistical and datasets collections, outreach through the QR initiative, and helping students find data as part of the information literacy offerings, the library built its campus reputation as a major part of a distributed model of data services that also included the GIS specialist and the Writing Center in addition to ITS. For about two years, the various providers of data services on campus formed a group, Research Data Services and Support Group (RDSSG), to assess needs and look for gaps in current services. That group is not currently meeting, but the relationships formed through the group remain strong and comprise the backbone of the distributed campus model.

Just as the faculty QR initiative emphasized QR across the curriculum, Carleton's model of data literacy from the library has always emphasized a strong partnership between the data librarian and the rest of the liaison librarians. Whenever possible, data instruction is done by the subject liaison in consultation with the data librarian. Students seeking help are directed first to

their subject librarian and then to the data librarian when necessary. This has been an intentional choice from the beginning to integrate data as a format and information type into the rest of the collections and information literacy programming, in order to avoid letting data services function in isolation. Even though data instruction from librarians began with assisting economics senior capstone projects, once word got out that librarians could help students find and evaluate data, services were soon in demand in political science and sociology and now in many other fields. In order to meet all of that need, the involvement of the other instruction librarians has been essential.

By developing the library's data services (and by extension its role in campus data services) as an extension of the liaison model, librarians were able to incorporate a whole new service by building on existing expertise in the library, partnerships with ITS, relationships with faculty, and expectations of the librarians' role in information literacy on campus. More important, and not entirely expected, these strengths were increased in the process. Because a portion of the social science liaison's time is dedicated to developing data services, and the library provided extra professional development to improve her expertise, she pushed to include data instruction as an aspect of her regular library instruction. This relatively minor shift in instruction – and the marketing of that instruction – along with visible participation in the faculty initiative resulted in an expanded notion on the part of faculty of how librarians could help students. Demand for liaison support in the social sciences grew so rapidly that a second liaison librarian was hired in 2008. Meanwhile, the whole team of liaisons developed their ability to integrate data sources and data literacy into their instruction when appropriate, allowing for much broader reach of data instruction across the disciplines than would have been possible otherwise. The partnership between librarians and academic technologists around data support

became a model for cross-departmental collaboration of support staff. In this small community, focusing on slow change through professional relationships and word of mouth allowed change to happen just about as quickly as staff could manage to keep up with it.

Now, as more attention turns toward adding data management services to existing data collection development and data reference and instruction services, library and ITS staff at Carleton are adjusting the model accordingly. The social science academic technologist has been working more with researchers on their DMPs and consulting with them about data storage, documentation, and preservation. The data librarian has been working with the liaison librarians to increase their comfort in discussing data management with faculty and students through informal training and raising awareness appropriate to the institutional culture. In particular, the data librarian has been working closely with the digital humanities librarian to identify ways that data management practices and language developed in the social sciences and sciences can be adapted to the humanities. While the academic technologist addresses more of the technical issues of data management, the librarians are focusing energies on finding new ways in their information literacy instruction to help make sure students know not just how to find, assemble, collect, interpret, analyze, and describe quantitative information, but also to be aware of the complete lifecycle of data and to know how to document, cite, and share data for further use. When students learn to think of their PDF readings and associated libraries of bibliographic information as a kind of data that they manage throughout the lifecycle of their research paper, they are gaining foundational understandings that can be built upon later. The library is a major player in campus data services efforts, but the library's role has more to do with education and classroom support than with faculty research, so librarians have focused their data management support on education rather than data management consultation, though in 2015 the data librarian

and academic technologist will begin offering data management plan consultation drop-in hours. Without an institutional repository, the library does not have the staff or workflows in place to systematically reach out to faculty about their publications, so this work has been done more by the academic technologist in partnership with the college's grants coordinator. The data librarian and the academic technologist meet approximately every two weeks to share information and coordinate efforts.

Finally, the issue of a local data repository remains an open question. For several years, the campus has used a pilot instance of Dataverse (an open source data repository solution for sharing, citing, and exploring data) as a proof of concept. The collection consists of some local replication datasets about the campus, some student project datasets, and a dataset used by a research methods class, none of which would have been good fits for a disciplinary repository. While networked storage space may have been sufficient in several instances to solve the storage need, a data repository simultaneously addresses issues of storage and dissemination, versioning, and user-inputted metadata. Additionally, it provides an opportunity to give students experience depositing data and going through the steps and thought processes necessary to transform their data from something organized to suit their own workflow to something that can be shared, understood, and reused by someone else. The project is currently on hold as the campus considers implementing an institutional repository and until the new version of Dataverse is released. What the academic technologist and librarians have learned so far from the project is that there are potential uses to be explored for a small-scale liberal arts college data repository that have more to do with supporting classroom teaching (e.g., a mechanism for managing and sharing teaching datasets with students) and pedagogical goals (e.g., data management education for students).

Next Steps at CSB/SJU and Carleton

The next steps in providing data management services at these institutions will be on several fronts. At CSB/SJU, librarians plan to conduct a more comprehensive environmental scan and needs assessment, along with targeted outreach, in order to better determine where there might be gaps in data education, access, and management. While librarians on the Data Management Task Force have learned to respond to needs as they arise, this approach is a somewhat reactive one to assisting with data management. Now librarians are ready to develop more deliberate and scalable services. The following areas are the current priorities:

- **Needs assessment:** Conduct a faculty-wide survey with follow-up interviews or focus groups.

- **Campus partners:** Meet with additional campus offices in order to build awareness of the Libraries' existing data management services, to foster deep inter-office collaboration, and to begin larger conversations about data management on campus. These offices might include the external grants office, Information Technology Services, the Institutional Review Board (IRB), and the Office of Experiential Learning & Community Engagement (ELCE), which supports undergraduate research.

- **Outreach to researchers:** Send out periodic emails reminding faculty and undergraduate researchers that help is available for writing DMP plans or assisting with other data management needs. The Data Management Task Force is also interested in investigating partnerships with Learning Enhancement Service (LES) to explore pedagogical and instruction-based concerns relating to data literacy, and perhaps working with faculty

researchers who are also "library champions" to develop an on-campus Forum presentation to encourage further faculty discussion about research data management.

- **Student data literacy:** Identify departments to collaborate with on instruction. In the past, for example, the MapCores (Mathematics, Physics, Computer Science Research Scholars) program has required students to locate and use public data sets in their assignments; librarians need to identify other programs and courses that include similar data requirements. Explore working with ELCE, Media Services, or IT Services to develop workshops on best practices for data management for summer undergraduate student researchers.

- **Institutional repository:** Continue to explore the possibilities and limitations of the CSB/SJU institutional repository as a data repository.

- **Digital humanities:** Collaborate with the CSB/SJU Libraries' new Digital Scholarship Task Force to expand data services appropriately to CSB/SJU's humanities and social science divisions.

At Carleton, the focus will be on exploring the question of a data repository, education and outreach, building campus relationships, and connecting with colleagues at other institutions. Regarding the storage and dissemination of data at Carleton, the librarian and ITS academic technologist will have conversations with faculty about potential use cases for the Dataverse pilot. There are people on campus who may be interested in it for organizing, documenting, and sharing their teaching data with others on campus; for example, multiple sections of a methods course that all use similar data could save and document these datasets in a central location. There are also faculty who teach their students about data management who might want to give students the opportunity to experience submitting their data to a repository as an assignment.

These data might not be kept permanently, but the feasibility of allowing faculty to use the nascent data repository as a teaching tool has not been fully fleshed out. Since the library may never have an institutional repository or permanent data repository, it is important to assess what the needs are on campus and think creatively about how the library can help to meet them. In the absence of a local repository, additional outreach and education and resource guides can be designed to help faculty find appropriate repositories for their data when writing DMPs.

In the area of data literacy education, the data librarian, digital humanities librarian, and a librarian at a neighboring school are developing data management teaching modules as a resource for other instruction librarians. Teaching librarians need to be ready to field surprise data management questions that pop up unexpectedly in class or in consultation, especially since they are not doing more focused outreach, like stand-alone workshops, about data management at Carleton. Reinforcing the message that all liaison librarians can provide some basic level of data management support by virtue of their disciplinary expertise remains a priority. Beyond that, if librarians have at the ready sets of illustrative examples, discussion prompts, and concise lists of tips, they can be even more effective with serendipitous instruction. Additionally, the same group will provide a training session at both schools on data management for all library staff, focusing on how librarians can use data management principles in their day-to-day work. This will both be a service to the staff and also an outreach strategy to make all staff feel less anxious about this topic and more prepared to discuss data management.

More work can be done to build connections on campus, such as working with the grants coordinator, the people who manage the institutional administrative data and campus records management, the IRB, the archives, and faculty managing data outside the sciences and social sciences. In the winter of 2014-2015, the data librarian and academic technologist will be

offering limited data management drop-in hours, which will be listed on the grants coordinator's checklist for faculty. A librarian could easily fill her calendar with discussions with potential partners, but with other work to be done, these efforts must be approached strategically as well.

CSB/SJU and Carleton are at different places in their implementation of data management services, and each have unique campus cultures that will affect how they provide such services going forward. However, their shared interests and goals inspire them to look for more ways to reach out to librarians at other small colleges. The scholarly communication landscape is in flux, the work of data management often comes in fits and starts, and librarians have other responsibilities to balance, so it can be challenging to ever feel adequately prepared for these new roles. Academic librarians need to pool expertise, experience, and ideas in order to learn from one another about how to support research data services, particularly in smaller liberal arts environments.

Making Data Management Manageable

The following suggestions are possible action steps, pulled from the LIS literature and from the authors' experiences at CSB/SJU and Carleton, for librarians interested in developing research data management services at their own institutions. These steps are quite broad, and will need to be applied in ways that fit each unique campus environment, librarians' and collaborators' skill sets, and faculty research culture. While these steps are here arranged in a somewhat linear fashion, in implementation some steps will be more important than others at particular institutions, some steps are best addressed concurrently, and, just as in research, the process of developing services is an iterative one.

Information Gathering

Read widely to learn more about the research data and scholarly communication landscapes. Learn from the successes and challenges experienced in other academic libraries. Begin to weigh how much domain-specific expertise you might need to develop in order to feel confident in your work with research data management services. Try not to get "stuck" at this stage, though. Dorothea Salo, who teaches graduate courses on digital curation and research data management, boldly states that "no one is an expert in data management."[6] Do not tell yourself that you have to know everything before moving forward and starting a needs assessment.

Needs Assessment

As you gather information, it is equally important to assess the research data management needs at your institution. Planning discussions with faculty researchers will help you better understand your institution's unique research environment and help you determine the scope and focus of the data services your library should work to provide. You may want to conduct this needs assessment through an online faculty survey, one-on-one researcher interviews, or a focus group.

Identifying Collaborators

Research data management is not something that can be done in isolation. Identifying collaborators and developing successful partnerships both on and off campus will be crucial to any coordinated, sustained program of services. Look for experts in your existing professional network and consider your local context when identifying others to work with, since partnership

---

[6] Syllabi for Dorothea Salo's courses can be found at http://dsalo.info/teaching/; quote via personal communication.

opportunities will vary according to one's unique campus environment. Consider creating a data

management task force, interest group, or support team if one does not already exist. This group

may want to begin creating an online guide or other resources that educate readers about research

data management and, eventually, the services you provide. While some schools are creating

data centers, others have distributed services and each model comes with its advantages and

disadvantages. Even without a physical location or branded identity, much can be accomplished

through regular communications with academic technologists, grants offices, offices of research,

IRBs, GIS specialists, math and statistics centers, digital humanities centers, and writing centers.

As student peer support grows, opportunities for cross training can be especially rich and has the

added benefit of modeling the ways experts rely on each other. As awareness grows, the library

can be involved in the development of a campus data management policy. Erway (2013)

provides helpful guidance on working with partners to advocate for a data management policy on

your campus.

Develop Initial Services

Librarians at small liberal arts institutions should not be intimidated by the admittedly

daunting task of getting started with research data services. Instead, start small and be strategic

when building a suite of services with collaborators. McLure offers good recommendations for

librarians at this stage:

> [W]hile librarians undertake to develop new skills and knowledge appropriate to new
> roles in data curation, they may also leverage their existing skills and expertise to pursue
> education and outreach efforts. Such efforts may immediately and positively begin to
> support researchers' connection and communication with other campus researchers;
> promote more widespread awareness and use of the libraries' data-management plan
> templates and repository; enhance understanding of the data life cycle and data-
> management considerations associated with each stage of the life cycle; aid identification
> of existing local, regional, national, or international repositories for data sharing;

facilitate identification and use of appropriate metadata standards; and encourage preplanning for sharing data files to accompany journal manuscripts. These goals are likely to be relevant, practical starting points for other libraries, too. (McLure, 2014, p. 158)

Work to identify initial data services you can realistically offer. Assess your ability to provide services related to data literacy and education, data discovery and retrieval, data management, or data curation, and then begin developing your model of data services, useful templates, and specific workflow scenarios.

Repository Considerations

If your institution already has an institutional or data repository, it may be a good idea at this point to compare the merits of these local repositories with disciplinary repositories. To aid you in this comparison, most large disciplinary repositories are described in the Registry of Research Data Repositories (re3data.org) and are indexed by Thomson Reuters Data Citation Index. Consider what you and your researchers need a data repository to do. If you are adding researchers' data to share it and make it more widely available to the public, then an institutional data repository may not be the best (or at least only) option to explore, since data stored in institutional repositories are relatively hard to find. If a researcher is looking for data to (re)use, the most likely place to look is a disciplinary data repository. In addition, most disciplinary repositories have metadata standards in place for ingest, have checksums set up to detect bit rot, and have a migration plan in place as file formats become obsolete. Consider adding data associated with student papers to your institutional repository, as the data then lives with the paper that describes and analyzes it, and since this work may otherwise not be captured. You will want to assess your institutional repository to see what file sizes and file types are allowed, and if

comparable standards and preservation policies are in place, before deciding what type of

repository to recommend to your faculty and student researchers.

Outreach to Faculty

As discussed earlier in the literature review section, there are advantages and

disadvantages to small institutions when it comes to faculty outreach. Consider which faculty

members might be potential champions of your services and spend extra time with them,

learning about their research needs, classroom assignments, and ways the library can work with

them. Use some of the literature about faculty surveys to craft informal talking points and hold

conversations with a carefully selected few who will spread the word amongst their colleagues.

If workshops on data management would not draw an interested crowd, look instead for ways to

inject data management principles into other existing venues. For example, at Carleton, staff

have looked for ways to be included as presenter-participants in faculty-led faculty development

workshops and learning and teaching center presentations. Sometimes just being present is

powerful outreach if you can attend events or professional development for faculty on related

topics like digital privacy, writing with data, or the ethics of using quantitative information.

Integrating data management into existing conversations has the added benefits of preventing

research data services from becoming an isolated service and finding an audience for an often

unpopular topic. As Lynch (2013, p. 397) points out, for the near term it will only be a minority

of faculty who are committed to the ideals of sharing research data or preserving it for the long

term, while most will see it merely as a hoop to jump through for the funding agencies. Liberal

arts librarians can take advantage of their relationships with faculty and work with them directly,

appealing to their more immediate concerns and helping them identify data management

solutions that might achieve more than one goal at once (e.g., working together to incorporate consistent data documentation protocols into training for student research assistants). Finally, as you expand data literacy and data management discussions beyond individual faculty members, you can help shape the campus discussion or institutional culture regarding issues as varied as data management and curation, data literacy, statistical software, scholarly communication, digital scholarship, and open access.

Consider Student Data Needs

Small liberal arts institutions, in particular, may have a stronger focus on student outcomes and teaching rather than research. Make sure to consider student data literacy and data management needs as you develop your services. For initiatives on teaching with or about data, it may be useful to adapt existing teaching tools from other institutions. ICPSR has a rich collection of resources available to the public that includes data-driven learning guides, exercise sets, and a crosstab assignment builder. Resources listed in Appendix A, such as "Using Data in the Classroom" from the Science Education Resource Center (SERC), TeachingWithData.org, the New England Collaborative Data Management Curriculum, the University of Minnesota's materials for a graduate workshop on data management, and the ICPSR Data Management and Curation resources, are full of curricular resources such as lesson plans, sample data, and assignments. The University of Minnesota's workshop focuses on overarching concepts, like "How to Inventory, Store, and Backup Your Data" and "How to Share Your Data and Ethically Reuse Data Created by Others," that could easily be adapted to the needs of undergraduate researchers at smaller institutions (Johnston, 2014, p. 431-434), and Mooney et al.'s recent

(2014) article may inspire you to develop faculty/librarian collaborations for undergraduate research training.

Long-term Scalability

Once your initial suite of services has been developed and you have begun to clearly demarcate who on campus can offer what services, you and your collaborators will need to assess your services, define areas for growth, and consider scalability and long-term planning for future research data management initiatives. Shorish emphasizes that "long-term planning must occur in order to ensure that future directions are sustainable. Possible considerations include additional library staff, an institutional repository, a distributed repository in cooperation with other peer institutions, or a researcher targeted service" (Shorish, 2012, p. 269). On a small campus, services like data management often grow slowly, and in large part through word of mouth and personal relationships. There is an inherent challenge in determining how many resources to devote to a set of needs that may not be evident until faculty suddenly come seeking assistance. Data management questions may then fall on one's doorstep with all the complexity of a faculty member's research or teaching, and librarians must decide if they will draw a hard line about the limits of their services or view unanticipated situations like these as an opportunity to form partnerships and develop additional services in order to be ready for the next challenge.

Conclusion: Evolutionary or Revolutionary?

Working to provide research data management services and to foster data literacy can meaningfully expand academic librarians' roles. Although local contexts may vary considerably, this is no less true at liberal arts institutions than at larger research universities. Data

management services, which tend to get classified as a research service, might at first seem out of scope for librarians in curricular libraries. Insofar as curricular libraries and liberal arts institutions are supporting the development of future researchers, however, then librarian support of student education in this area also constitutes a type of data management service. Additionally, the boundaries between faculty researchers and student learners are blurring as students increasingly participate in faculty research. If students are engaged in research, either with faculty or independently as part of their education, then supporting that research is arguably within the scope of a teaching library.

As the role of academic libraries shifts to being involved in the production of research more broadly, the boundaries of the curricular library become less clear, and librarians can participate in the curation and dissemination of original content in local digital repositories. In this way, libraries and research data management services are transforming together, evolving to meet the needs of scholars, teachers, and students in the shifting environment of scholarly communication. While research data services including data management and curation are dramatically changing academic libraries, they are doing so in an evolutionary rather than revolutionary fashion. Research data services can be integrated into existing library services, workflows, and goals, and the foundational skills necessary to provide basic support in data management draw on existing librarian skills in and commitments to organization, metadata creation, the sharing and preservation of information, and bridging the disparate standards and cultures of academic disciplines to meet needs across the curriculum. Helping faculty researchers identify disciplinary data repositories or developing consortial solutions with local college and university partners to tougher challenges like long term data preservation solutions are scalable and strategic uses of librarians' time and expertise. This type of work, if done in stages and in a

manageable way, need not entail a dramatic shift of mission, simply efforts to thoughtfully re-skill, re-position, and re-structure library staff and services. Academic libraries, as they have throughout their history, will evolve to meet the needs of their communities.

References

ACTI Data Management Working Group. (2013). "ACTI Data Management Glossary." https://net.educause.edu/ir/library/pdf/ACTI1304.pdf.

ACRL Planning and Review Committee. (2013). Environmental Scan 2013. Chicago, IL: Association of College and Research Libraries. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/EnvironmentalScan13.pdf

Antell, Karen, Bales Foote, Jody, Turner, Jaymie, and Brian Shults. (2014). "Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions." *College & Research Libraries* 75(4), 557-574. doi:10.5860/crl.75.4.557

Bailey, Charles W. (2014). "Research Data Curation Bibliography" on the Digital Scholarship web site. Accessed on 12/30/2014. http://digital-scholarship.org/rdcb/rdcb.htm.

Carlson, Jake. (2012). "Demystifying the Data Interview: Developing a Foundation for Reference Librarians to Talk with Researchers about Their Data." *Reference Services Review* 40 (1): 7–23.

Coates, Heather. (2014). "Ensuring Research Integrity: The Role of Data Management in Current Crises." *College & Research Libraries News* 5:598-601.

Corrall, Sheila, Kennan, Mary Anne, and Waseem Afzal. (2013). "Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research." *Library Trends* 61(3), 636-674. doi:10.1353/lib.2013.0005

Cox, Andrew M., and Corrall, Sheila. (2013). "Evolving Academic Library Specialties." *Journal of the American Society for Information Science and Technology*. Early View. doi:10.1002/asi.22847

Dahl, Mark. (2014). "Data-Driven Liberal Arts: the Library Role." The National Institute for Technology in Liberal Education's (NITLE) *Academic Commons*. http://www.academiccommons.org/2014/07/24/data-driven-liberal-arts-the-library-role/

Delserone, Leslie M. (2008). "At the Watershed: Preparing for Research Data Management and Stewardship at the University of Minnesota Libraries." *Library Trends* 57(2), 202-210. http://mwbdvjh.muse.jhu.edu/journals/library_trends/v057/57.2.delserone.pdf

Erway, Ricky. (2013). *Starting the Conversation: University-Wide Research Data Management Policy*. Dublin, OH: OCLC Research. http://www.oclc.org/research/publications/library/2013/2013-08.pdf

Federer, Lisa. (2013). "The Librarian as Research Informationist: A Case Study." *Journal of the Medical Library Association (JMLA)* 101(4):298-302. doi:10.3163/1536-5050.101.4.011

Gabridge, Tracy. (2009). "The Last Mile: Liaison Roles in Curating Science and Engineering Research Data." In *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC* (no. 265), 15-21. http://old.arl.org/bm~doc/rli-265-gabridge.pdf

Geraci, Diane, Humphrey, Chuck, and Jim Jacobs. (2012). *Data Basics: An Introductory Text*. Interuniversity Consortium for Political and Social Research, Ann Arbor, MI.

Gold, Anna. (2010). "Data Curation and Libraries: Short-term Developments, Long-term Prospects." *Library Publications and Administrative Reports*. Paper 27. 1-33. http://digitalcommons.calpoly.edu/lib_dean/27

Gold, Anna. (2007). "Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians." *D-Lib Magazine* 13(9/10), n.p. http://www.dlib.org/dlib/september07/gold/09gold-pt1.html

Goldstein, Sarah, and Sarah K. Oelker. (2011). "Planning for Data Curation in the Small Liberal Arts College Environment. *Sci-Tech News* 65(3), 5-11. http://jdc.jefferson.edu/scitechnews/vol65/iss3/4/

Haendel, Melissa A., Vasilevsky, Nicole A., and Jacqueline A. Wirz. (2012). "Dealing with Data: A Case Study on Information and Data Management Literacy." *PLoS Biology* 10(5), 1-4. doi:10.1371/journal.pbio.1001339

Jahnke, Lori M., and Asher, Andrew. (2012). "The Problem of Data: Data Management and Curation Practices Among University Researchers." In *The Problem of Data*. CLIR Publication no. 154. Washington, D.C.: Council on Library and Information Resources, 3-19. http://www.clir.org/pubs/reports/pub154

Johnson, Larry, Adams Becker, Samantha, Estrada, Victoria, and Freeman, Alex. (2014). *NMC Horizon Report: 2014 Library Edition*. Austin, Texas: The New Media Consortium. http://www.nmc.org/publications/2014-horizon-report-library

Johnston, Lisa, and Jon Jeffryes. (2014). "Steal this Idea: A Library Instructors' Guide to Educating Students in Data Management Skills." *College & Research Libraries News* 75(8), 431-434.

Lynch, Clifford. (2013). "The Next Generation of Challenges in the Curation of Scholarly Data." In *Research Data Management: Practical Strategies for Information Professionals*, by Joyce M. Ray. West Lafayette, IN: Purdue University Press.

McEwen, Leah, and Ye Li. (2014). "Academic Librarians at Play in the Field of Cheminformatics: Building the Case for Chemistry Research Data Management." *Journal of Computer-Aided Molecular Design* 28(10), 975-988. doi:10.1007/s10822-014-9777-4

McLure, Merinda, Level, Allison V., Cranston, Catherine L., Oehlerts, Beth, and Mike Culbertson. (2014). "Data Curation: A Study of Researcher Practices and Needs." *portal: Libraries & the Academy* 14(2), 139-164.

Mooney, Hailey, Collie, W. Aaron, Nicholson, Shawn W., and Marya R. Sosulski. (2014). "Collaborative Approaches to Undergraduate Research Training: Information Literacy and Data Management." *Advances in Social Work* 15(2), 1-22.

Nielsen, Hans Jørn, and Birger Hjørland. (2014). "Curating Research Data: The Potential Roles of Libraries and Information Professionals." *Journal of Documentation* 70(2), 221-240. doi:10.1108/JD-03-2013-0034

Partlo, Kristin. (forthcoming). "From Data to Creation of Meaning Part II: Data Librarian as Translator." *IASSIST Quarterly*.

Reisner, Barbara A., Vaughan, K.T.L, and Yasmeen L. Shorish. (2014). "Making Data Management Accessible in the Undergraduate Chemistry Curriculum." *Journal of Chemical Education,* Article ASAP. doi:10.1021/ed500099h

Scaramozzino, Jeanine Marie, Ramírez, Marisa L., and Karen J. McGaughey. (2012). "A Study of Faculty Data Curation Behaviors and Attitudes at a Teaching-Centered University." *College & Research Libraries* 73(4), 349-365. doi:10.5860/crl-255

Shorish, Yasmeen. (2012). "Data Curation Is for Everyone! The Case for Master's and Baccalaureate Institutional Engagement with Data Curation." *Journal of Web Librarianship* 6(4), 263-273.

Tenopir, Carol, Birch, Ben, and Allard, Suzie. (2012). *Academic Libraries and Research Data Services: Current Practices and Plans for the Future*. Chicago, IL: Association of College and Research Libraries. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Toups, Megan, and Michael Hughes. (2013). "When Data Curation Isn't: A Redefinition for Liberal Arts Universities." *Journal of Library Administration* 53(4), 223-233. doi:10.1080/01930826.2013.865386

University of Illinois at Urbana-Champaign, Graduate School of Library and Information Science. (2014). "Specialization in Data Curation." http://www.lis.illinois.edu/academics/degrees/specializations/data_curation

Van den Eynden, V., Corti, L., Woollard, M., Bishop, L., & Horton, L. (2011). Managing and sharing data: Best practice for researchers. Colchester, UK: UK Data Archive. http://www.data-archive.ac.uk/media/2894/managingsharing.pdf

Wang, Minglu. (2013). "Supporting the Research Process through Expanded Library Data Services." *Program: Electronic Library & Information Systems* 47(3), 282-303. doi:10.1108/PROG-04-2012-0010

Appendix A: Professional Development Resources

"There's no need to recreate a small universe of resources and tools related to data management. Many things your faculty may find helpful are already online, easily accessible, and ready to be shared." ~Sarah Goldstein & Sarah K. Oelker, Mount Holyoke College

General Guides for Getting Started

- Digital Curation Centre's *How to Develop Research Data Management Services:*

  http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services

- Digital Curation Centre's *Digital Curation Manual*:

  http://www.dcc.ac.uk/resources/curation-reference-manual

- DataONE Index of Resources: https://www.dataone.org/resources (includes a Librarian

  Outreach Kit at https://www.dataone.org/for-librarians)

- Digital Formats website (includes file format information):

  http://www.digitalpreservation.gov/formats/

- Department of Cultural Resources' Best Practices for File-Naming:

  http://www.ncdcr.gov/Portals/26/PDF/guidelines/filenaming.pdf

- EDUCAUSE's Data Management Glossary:

  http://www.educause.edu/library/resources/acti-data-management-glossary

- ICPSR resources for data management and curation:

  http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/index.html

Publications

- *IASSIST Quarterly*

Listservs & Organizations

- DCIG (ACRL Digital Curation Interest Group) listserv subscription:

  http://lists.ala.org/wws/info/acr-igdc-l

- LAReDaS (Liberal Arts Research Data Support: Librarians and Information

  Technologists Supporting the Use and Management of Research Data on their Campuses)

  listserv subscription: https://lists.carleton.edu/info/laredas

- IASSIST (International Association for Social Science Information Services &

  Technology) http://www.iassistdata.org/ has a very active listserv

- RDAP (ASIS&T Research Data and Preservation SIG) listserv subscription:

  http://mail.asis.org/mailman/listinfo/rdap

- DataONE (Data Observation Network for Earth): https://www.dataone.org/

- Data Conservancy: http://dataconservancy.org/

- Digital Curation Centre: http://www.dcc.ac.uk/

Training Opportunities & Conferences

- ICPSR's Summer Institute Programs, including "Providing Social Science Data Services:

  Strategies for Design and Operation"

  (http://www.icpsr.umich.edu/icpsrweb/sumprog/courses/0041) and "Curating and

  Managing Research Data for Reuse"

  (http://www.icpsr.umich.edu/icpsrweb/sumprog/courses/0149). See also their Official

  Representative annual meetings and instructional materials

  (http://www.icpsr.umich.edu/icpsrweb/instructors/index.jsp .

- Mantra Research Data Management Training

A free online course from Edina (the Jisc-designated national data center at the University of Edinburgh) intended for anyone who manages digital data as part of a research project. Of special interest is the "DIY Training Kit for Librarians."
http://datalib.edina.ac.uk/mantra/

- eScience Portal for New England Librarians. See especially the Professional Development section. http://esciencelibrary.umassmed.edu/

- Library schools are increasingly offering data science courses. Look for continuing education offerings in data curation and data management.

- IASSIST Annual Conference: http://www.iassistdata.org/conferences. Note that conference workshops offer training and a network of colleagues in social science data fields.

- Research Data Access and Preservation Summits: http://www.asis.org/rdap/

- International Conference on Open Repositories: conference website changes annually

Tools for Data Management

- Purdue University's Data Curation Profiles Toolkit: http://datacurationprofiles.org/

- DMPTool: https://dmp.cdlib.org/

- Databib: http://databib.org. Annotated bibliography of discipline-specific data repositories.

- Registry of Research Data Repositories: http://www.re3data.org/. Note that in 2015, Databib and Re3data.org will merge and retain the name Registry of Research Data Repositories.

- SHERPA/Juliet: http://www.sherpa.ac.uk/juliet/. Global listing of data archiving policies for funding agencies.

Resources for Teaching

- University of Minnesota's Data Management Workshop Series materials:

  http://z.umn.edu/teachdatamgmt

- New England Collaborative Data Management Curriculum:

  http://library.umassmed.edu/necdmc

- "Using Data in the Classroom" from the Science Education Resource Center (SERC):

  http://serc.carleton.edu/usingdata/index.html

- TeachingWithData.org: http://www.teachingwithdata.org/

- ICPSR Data Management and Curation resources:

  http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/index.html

Example Data Management Services Guides

- MIT Libraries Data Management Guide: http://libraries.mit.edu/data-management/

- University of Wisconsin-Madison Research Data Services: http://researchdata.wisc.edu/

- DM Vitals: University of Virginia Library Data Management Consulting Group:

  http://dmconsult.library.virginia.edu/

Appendix B: The DCP Toolkit -- Summary of Interview Elements for Discussion

The following is a summary by Kristin Partlo of the "DCP Interviewer's Manual" of the Data Curation Profiles Toolkit, developed by Scott Brandt and Jake Carlson of Purdue University Libraries, the Institute of Museum and Library Services, and the Distributed Data Curation Center (http://datacurationprofiles.org/). It has been useful in several contexts: talking to other reference and instruction librarians about data management and curation, talking with digital humanities practitioners about overlap and what can be learned from the field of social science data, and training digital humanities student associates on data management and curation.

1. The data set (brief description)

2. The lifecycle of the dataset

    a. Describe multiple stages of the data. For each, what are the number, size, and format of files?

3. Sharing

    a. For each stage, would you share with: no one, your collaborators, within your institution, within your field, beyond your field, anyone? Timing of sharing?

    b. What do you imagine the value to be to others? How might it be used?

    c. Would you want to place any conditions on the use?

4. Access

    a. Willingness to submit data to a repository? At what stages in the lifecycle? Embargo?

    b. Prioritizing needs such as: ability to cite the data, requiring users to cite the data, mirror sites and restricted access.

5. Transfer of data / ingest into a repository

   a. What preparatory actions would need to take place to make your data ready for ingest?

   b. Prioritizing needs such as: ability to self-deposit, to automate submission via some trigger such as date or event, and the ability to batch upload.

6. Organization and Description of Data

   a. Explain how the dataset is organized and how it has been described in documentation.

   b. Is the organization and description sufficient for someone else to use it?

   c. Prioritize need for ability to: make data available in multiple formats, apply standardized metadata from your field to the dataset.

7. Discovery

   a. Prioritize ability for researchers within your own field to find the dataset, outside your field, general public, using Internet search engines.

8. Intellectual property

   a. Who is the owner of the data? What are you funding sources?

   b. Are there requirements from funders to: draft a data management plan, share your data with others, preserve your data beyond the life of funding? Any privacy or confidentiality concerns

9. Tools

   a. Software or hardware tools used in generating the data, utilizing the data.

   b. Prioritize need for ability to connect data to visualization tools, and for others to comment on or annotate the dataset.

10. Linking/Interoperability

   a. In the journals where you publish, are data or supplemental materials accepted for publication?

   b. Prioritize needs to connect data with publications, support the use of APIs, connect or merge your data with other datasets.

11. Measuring impact

   a. Prioritize needs for usage statistics, gather information about the people who use the datasets.

   b. Any other measurements or analytics you want to apply to your data?

12. Data management

   a. What are the primary ways you currently manage your data, including storage media and tools?

   b. Do you make backup copies? If so, how often? Do you take security measures to protect your data? If so, what? Prioritize need to enable version control.

13. Data preservation

   a. What are the most important parts to maintain or preserve over time?

   b. How long would your data have value for others if preserved?

   c. Prioritize needs for: ability to audit for structural integrity of files over time, migrate to new formats over time, secondary storage site for the data, secondary storage at a different geographic location, documentation of any and all changes made to the dataset over time.