Building Data Services From the Ground Up: Strategies and Resources

Heather L. Coates
Indiana University - Purdue University Indianapolis

Let us know how access to this document benefits you.

Follow this and additional works at: https://escholarship.umassmed.edu/jeslib

🌀 Part of the Library and Information Science Commons

UMass Chan
MEDICAL SCHOOL

Journal of eScience Librarianship
putting the pieces together: theory and practice

Building Data Services From the Ground Up: Strategies and Resources

Heather L. Coates

Indiana University - Purdue University Indianapolis, Indianapolis, IN, USA

Abstract

There is a scarcity of practical guidance for developing data services in an academic library. Data services, like many areas of research, require the expertise and resources of teams spanning many disciplines. While library professionals are embedded into the teaching activities of our institutions, fewer of us are embedded in activities occurring across the full research life cycle. The significant challenges of managing, preserving, and sharing data for reuse demand that we take a more active role. Providing support for funder data management plans is just one option in the data services landscape. Awareness of the institutional and library culture in which we operate places an emphasis on the importance of relationships. Understanding the various cultures in which our researchers operate is crucial for delivering data services that are relevant and utilized. The goal of this article is to guide data specialists through this landscape by providing key resources and strategies for developing locally relevant services and by pointing to active communities of librarians and researchers tackling the challenges associated with digital research data.

Introduction

There is little practical guidance for developing data services in academic libraries (Reznik-Zellen, Adamick, and McGinty 2012). Much of the data services literature is composed of case studies conducted by teams at research-intensive institutions and high-level reports on data challenges facing scientific research (Committee on Science 2009, Jones, Pryor, and Whyte 2013, National Science Board 2005). I am a solo data specialist on a health sciences campus within a large multi-campus institution. Although my institution provides significant computing, software, and storage resources; busy students, staff, and faculty need additional training and support to effectively in-corporate these resources into their research processes. Many day-to-day issues that my colleagues and I struggle with most are not discussed in literature. These issues include strategies for building relationships with busy researchers, demonstrating the value of the library perspective in this arena, advocating for improved research practices, and creating strategies for engaging with high-level administrators about these issues. An invitation to present at the *Doing it Your Way: Approaches to Research Data Management for Libraries* Symposium served as the incubator for many of the ideas offered here. My goal is that this article will begin to fill the gap by providing practical strategies, key resources, and point to communities of data specialists.[1]

Correspondence to Heather Coates: hcoates@iupui.edu

Until recently, the journal article was the ultimate product of the research process and the primary means for disseminating scientific knowledge. Data is increasingly viewed as a valuable product, particularly as consumer access to powerful computing and user-friendly tools expands. However, many researchers have neither the time nor the training to manage their data in ways that facilitate the reproducibility, openness, and interoperability encouraged by funding agency policies. Despite the hype about the potential of data big and small, there are few incentives for researchers and institutions to reallocate already limited resources towards data management, sharing, and curation. The increasing competition for limited funding dollars coupled with the heavy research administrative burdens (Schneider et al. 2012) contributes to the time pressures many academic researchers experience. These factors encourage data hoarding rather than sharing. This shifting environment presents exciting opportunities for libraries to support the transition to greater sharing of research data.

Navigating the Data Services Landscape

Changing the cultural context for research data management and sharing will be neither quick nor simple. Promoting incremental changes in research practices is one viable approach to increasing the suitability and availability of data for re-use. This approach opens up many options for developing services, while enabling us to build on what we are already doing. The challenge then is to identify feasible options appropriate to institutional circumstances. While there is no recipe for developing data services, there are common activities in the process. These include identifying the available institutional resources (i.e., environmental scan), choosing an approach, eliciting the needs of researchers (i.e., requirements analysis), identifying partners and collaborators, conducting

pilots, and building a community. Once you have invested in services that address high-priority needs, ongoing maintenance involves evaluating services, outreach and promotion, and maintaining relationships. Although it seems daunting, we can support gradual change toward improved data practices by meeting researchers where they are in the process, and by being aware of the challenges they face.

Two reports provide a detailed description of the data services landscape from leaders in the field. The Association of College and Research Libraries (ACRL) white paper (Tenopir, Birch, and Allard 2012) reports on early progress in research data services prior to the NSF data management plan requirement, while the ARL Spec Kit (Fearon et al. 2013) characterizes data services after implementation. Core services identified (Tenopir, Birch, and Allard 2012, Fearon et al. 2013) include: data reference and guides; institutional repositories (IR); GIS and spatial analysis support; purchasing, acquisition, and licensing data; data management plan consultation and education; metadata and standards support; and data services training. Both reports are useful for identifying possible data services that might be appropriate for your institution.

Attending national conferences on research data management and curation topics provides opportunities to discuss pragmatic issues and develop relationships with other data specialists. Well-known events include the Research Data Access and Preservation Summit (RDAP), International Digital Curation Conference (IDCC), and the annual meeting of the International Association for Social Science Information Services and Technology (IASSIST). There are also data-related sessions at annual meetings of the American Library Association (ALA), Medical Library Association (MLA), and ACRL. If travel is not a viable option, many data spe-

---

[1] Given that many of the professionals providing data services in academic libraries and research institutions have graduate degrees and backgrounds outside of library and information science, I refer to this broad group of practitioners as data specialists.

cialists are active on Twitter and various mailing lists like the New England e-Science portal, ACRL Digital Curation Interest Group, the ALA Digital Preservation list, DataONE, Research Data Alliance, and the Digital Preservation Network. These activities are more than professional development; they are an important venue for engaging with the data specialist community and building your professional network.

Scanning the Environment

In developing data services, scanning the environment is crucial for understanding how your work could fit into existing resources and services related to research data. The process is also useful for identifying gaps in services and key members of the institutional research network. The ability to provide relevant and valuable services depends on the ability listen to stakeholders – research teams, administrators, and IT teams – to gather the needs they have identified, and to help them to identify unrecognized needs.

Institutional policies do not always facilitate rapid surveys, so consider a variety of options for gathering requirements at individual, team, and departmental levels. Two commonly used tools are the Data Curation Profiles (DCP; Witt et al. 2009) and the Data Asset Framework (DAF; Jones, Ball, and Ekmekcioglu 2008). The DCP is a semi-structured interview designed to convey detailed information about the data and processes related to a particular project in the voice of the interviewee (i.e., investigator, collaborator, research assistant, etc. Although it is time-intensive (5-10+ hours per profile) and difficult to use for broad-scale information gathering, it is useful for examining the intricacies of a particular project. This can be a good first step in exploring embedded data services. A growing collection of completed profiles is available online (Carlson and Brandt 2014). In contrast, the DAF was designed primarily as a high-level institutional survey of data assets and needs (Jones, Ross, and Ruusalepp 2009), al-

though it can be applied to research centers or departments. The survey process incorporates input from multiple stakeholders and units and identifies gaps, risks, and strategies for addressing such issues. The DAF has been adapted and used at other institutions like Georgia Tech (Rolando et al. 2013) and the University of Houston (Peters & Riley Dryden 2011). Another tool, developed by the University of Virginia, is DMVitals. It offers a rating system for information gathered during a structured interview with individual researchers (Sallans & Lake 2012). Both scope and process for these tools differ widely. The choice of tools should be shaped by how you choose to develop services. Selecting a tool can be daunting; the nuances of a tool are difficult to determine without actually using it first. One timesaving approach is to consult with data specialists who have already used the tools under consideration.

A logical next step after the environmental scan is a needs assessment. The Digital Curation Centre (DCC) has developed two relevant resources. Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO) is designed to provide benchmarks and identify gaps between current and required support capabilities. Recently, they released a how-to guide for eliciting requirements (Whyte & Allard, 2014). The guide focuses on three key areas that can influence requirements: ideas, artefacts, and facilities; research stakeholders; and institutional rules and research norms. A standardized approach simply does not work for delivering support to such a diverse range of professionals. The guide also discusses the costs and benefits of a variety of approaches, from case studies to focus groups and usage scenarios. Although a formal approach may not be necessary to gather needs, it can be helpful to structure and focus the conversation. While data specialists may use the research life cycle to frame how research data services can support a project, researchers typically do not think about their research processes in this way. Reaching a shared understanding

about needs is an ongoing negotiation.

Choosing an Approach to Service Development

Choosing a strategy for developing data services is strongly informed by two key environmental factors: 1) Have high-level administrators identified data management and sharing as an institutional priority? 2) Who will be responsible for providing data services? If a team approach is chosen, the composition of the team and division of responsibilities will have a strong impact on the types of data services offered. If a data specialist is solely responsible, it is vital to manage expectations regarding the scalability and sustainability of services. For the purposes of this discussion, I narrowed the spectrum of options down to two approaches. These include the traditional top-down approach, represented by the DCC guide "How to Develop Research Data Management Services" (Jones, Pryor, & Whyte 2013), and the bottom-up approach, represented by Brian Mathews' white paper "Think Like a Startup" (2012).

Both Australia and the U.K. have organized a well-coordinated, top-down approach facilitated by national data services. In Australia, the National Collaborative Research Infrastructure Strategy led to the development of the Australia National Data Service (ANDS). In the U.K., the Research Councils (RCUK) issued common principles on data policy. Both organizations produce resources that can be adapted for outreach and education (2014b, 2014a, Jones, Pryor, and Whyte 2013). The DCC guide to developing data services assumes consistent agency policies and strong institutional leadership in research data stewardship (Jones, Pryor, and Whyte 2013). It outlines responsibilities, strategies, and processes for senior administration and various stakeholders. In contrast, the research data policy environment in the United States is far more fragmented. We are still awaiting the release of agency policies resulting from the OSTP memorandum on access to the results of publicly funded research (Holdren 2013). The lack of coordinated policies is reflected in the disengagement of many administrators from research data stewardship issues. The reality is that many libraries interested in establishing data services do not have high-level institutional support, so we must build support from the bottom-up.

Brian Mathews, an advocate for further innovation in academic libraries, offers several suggestions relevant to data services. Academic libraries often rely on a traditional, front-loaded planning process (Learn-Build-Measure; Mathews 2012) that strives for perfection in a new product or service before launching it. Mathews' (2012) article makes a compelling case for a more experimental approach. This "Build-Measure-Learn" (Lean Startup Method, Eric Ries 2014) approach comes from the lean startup movement and emphasizes continuous innovation. The idea is that small, early failures produce a better product in the end. Adopting this iterative and flexible approach can help data services remain relevant in a rapidly changing research environment.

As the policy landscape changes, infrastructure advances, and interdisciplinary standards emerge, data specialists must stay informed in order to identify new opportunities and recognize when services are no longer relevant. Most data services will need to adapt quickly in order to remain relevant. The exception to this are institutional repositories (IR), which require the robust infrastructure, deep expertise, and long-term commitment best facilitated by a top-down approach. For example, the impetus for many data services programs was the 2011 National Science Foundation data management plan requirement. As faculty and departments become more adept at creating these plans, the demand for library support in creating these will wane. In some cases, this has already happened.

While new federal funding agency policies

will soon be released in response to the 2013 Office of Science and Technology Policy memorandum (Holdren 2013), there are other opportunities for data specialists to provide useful services in a rapid, responsive way. Providing education, consultation, and resources for researchers to improve their data management practices could have a great impact on the quality and accessibility of research data produced. Key areas for improvement include file organization, storage and backup procedures, documentation, and data publishing and citation. Other common entry points to data services include providing a mediated deposit into domain or subject repositories and support for locating and citing existing data. These services require expertise and time, but are not necessarily dependent on administrative approval or funding. Even though many librarians do not have the expertise to facilitate the use of data, we can refer to campus units who do. Students, staff, and faculty alike are often unaware of the research support services available to them. Similarly, many datasets can be shared appropriately using existing repositories; not all institutions need to host an IR. Serving as a navigator to various research support units is a valuable service (Peters and Riley 2011), as is providing targeted training to fill the gaps in education. These roles position the library as a central access point for research support.

Building a Community

Given the complexity of modern academic research, no one person can know everything necessary to carry out a successful research project. Like research, data management is a team sport. Potential partners within the institution include grants coordinators, sponsored projects, college research deans, institutional review boards, IT system administrators, research support offices, statistical consulting, data security offices, copyright/legal offices, and commercialization offices (Fearon et al. 2013, Hofelich Mohr and Lindsay 2014). Both the processes of scanning the environment and gathering needs are excellent opportunities to begin

developing relationships. Libraries are already members of the institutional research support community. As part of this community, our relationships and social interactions with researchers are as important as our services. Integrating the library into the research life cycle at our institutions means that we as individuals become part of our researchers' social networks.

Since colleagues and classmates are a common trusted source of information for faculty and students respectively, our personal relationships may be the most important tool we have. These relationships are also critical for building trust and social capital within the institution. Significant value arises from the social capital generated through interactions taking place in or facilitated by libraries (Johnson 2012). Building enduring, collaborative relationships with researchers is an important skillset that is poorly documented. The literature about building these relationships is sparse. There is some discussion of the social capital generated by public libraries (Vårheim, Steinmo, and Ide 2008, Johnson 2012), while academic libraries seem to be re-examining personalized services and one-on-one consultations (Nolin 2013). Practical strategies for establishing relationships with researchers, other than participating in departmental meetings and functions, are highly dependent on institutional culture. Documenting and sharing strategies for cultivating relationships with patrons would benefit all library professionals.

Conclusion

Librarians embarking on data services face an exciting but uncertain journey. The path for developing data services at your institution will likely differ from mine. Despite the scarcity of practical guidance in the literature, there is an abundance of knowledge embedded within communities of data specialists. Some of these communities are associated with library organizations, while others are tied to particular research topics or methods. Resources and support provided

by library and institutional administration will largely determine your approach for service development. Limited resources can be maximized with support from your professional network. As a data specialist, having a broad, diverse professional network is tremendously helpful. Engage with professionals outside of libraries and include people with diverse backgrounds, areas of expertise, and perspectives on research data. The earlier you can engage others in the process, the better. Regardless of your available resources, key elements of the service development process include environmental scanning, needs assessment, building a community of collaborators and users, and continuous feedback and improvement of services. Possibly the most important thing to do in developing any new service is to accept failure as part of the process, to view it as an opportunity to learn and improve.

Data specialists have important roles in advocating for and training researchers to use effective research data practices. Given the library's longstanding role in scholarly communication, data specialists are in a strong position to examine the changes in data practices as well as gaps in the sociotechnical system of academic research. Although preserving valuable research data is a truly important endeavor, I believe success of the data specialist in the next decade will more accurately be reflected by the strength of our relationships with researchers and other campus units than the number of datasets deposited in our IR. Shifting the research practices on our campuses towards more efficient workflows and sustainable infrastructure is a long-term goal requiring deep knowledge of both institutional and disciplinary practices. The contribution of data services towards this goal will not be easy to measure, but that makes them no less important.

References

Australia National Data Service. 2014a. "ANDS Guides and Other Resources." Accessed August 3, 2014. http://www.ands.org.au/guides/index.html

Australia National Data Service. 2014b. "Data Management: Overview." Accessed August 3, 2014. http://www.ands.org.au/datamanagement/index.html

Carlson, Jake, and D. Scott Brandt. 2014. Data Curation Profiles Directory. West Lafayette, IN: Purdue University.

Committee on Science, Engineering, and Public Policy (U.S.). Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age. 2009. *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*: The National Academies Press. http://www.nap.edu/catalog.php?record_id=12615

Fearon, David, Betsy Gunia, Barbara Pralle, Sherry Lake, and Andrew L. Sallans. 2013. *Research data management services*. Edited by L. A. George, *SPEC Kits*. Washington, D.C.: Association of Research Libraries.

Hofelich Mohr, Alicia, and Thomas Lindsay. 2014. "It takes a village: Strengthening data management through collaboration with diverse institutional offices." International Association for Social Science Information Services & Technology, Toronto, CA, June 4, 2014.

Holdren, John P. 2013. Increasing access to the results of federally funded scientific research. edited by Executive Office of the President Office of Science and Technology Policy. Washington, D.C.: White House.

Johnson, Catherine A. 2012. "How do public libraries create social capital? An analysis of interactions between library staff and patrons." *Library & Information Science Research* 34, no. 1:52-62, http://dx.doi.org/10.1016/j.lisr.2011.07.009

Jones, S., S. Ross, and R. Ruusalepp. 2009. Data Audit Framework Methodology, v1.8. HATII, University of Glasgow: Glasgow, UK.

Jones, Sarah, Alexander Ball, and Çuna Ekmekcioglu. 2008. "The Data Audit Framework: A First Step in the Data Management Challenge." *International Journal of Digital Curation* 3, no. 2:112-120, http://dx.doi.org/10.2218/ijdc.v3i2.62

Jones, Sarah, Graham Pryor, and Angus Whyte. 2013. How to Develop Research Data Management Services - a guide for HEIs. In *DCC How-to Guides*. Edinburgh: Digital Curation Centre.

Mathews, Brian. 2012. Think Like a Start-up: A white paper to inspire library entrepreneurialism. Virginia Tech University. http://hdl.handle.net/10919/18649

National Science Board. 2005. *Long-lived digital data collections : enabling research and education in the 21st century*. [Arlington, VA]: National Science Foundation.

Nolin, Jan Michael. 2013. "The special librarian and personalized meta-services." *Library Review* 62, no. 8/9:508-524, http://dx.doi.org/10.1108/lr-02-2013-0015

Peters, Christie, and Anita Riley. 2011. "Starting from scratch: Assessing the library's role in data management on campus." Special Libraries Association, Philadelphia, PA, June 15, 2011.

Peters, Christie, and Anita Riley Dryden. 2011. "Assessing the Academic Library's Role in Campus-Wide Research Data Management: A First Step at the University of Houson." *Science & Technology Libraries*: 30, no. 4:387-403, http://dx.doi.org/10.1080/0194262X.2011.626340

Reznik-Zellen, Rebecca, Jessica Adamick, and Stephen McGinty. 2012. "Tiers of Research Data Support Services." *Journal of*

eScience Librarianship 1, no. 1:27-35, http://dx.doi.org/10.7191/jeslib.2012.1002

Ries, Eric. 2014. The Lean Startup Methodology. Accessed October 22, 2014. http://theleanstartup.com/principles

Rolando, Elizabeth, Chris Doty, Wendy Hagenmaier, Alison Valk, Susan Wells Parnham. 2013. *Institutional Readiness for Data Stewardship: Findings and Recommendations from the Research Data Assessment*. Accessed October 22, 2014. http://hdl.handle.net/1853/48188

Sallans, Andrew, and Sherry Lake. 2012. "DMVitals: A Data Management Assessment Recommendations Tool." Paper presented at the Annual Meeting of the International Association for Social Science Information Services & Technology, Washington, D.C, June 2012. http://www.slideshare.net/ASallans/dmvitals-a-data-management-assessment-recommendations-tool-iassist-2012

Schneider, Sandra L., Kirsten K. Ness, Sara Rockwell, Kelly Shaver, and Randy Brutkiewicz. 2012. 2012 Faculty Workload Survey: Research Report.

Tenopir, Carol , Ben Birch, and Suzie Allard. 2012. "Academic libraries and research data services current practices and plans for the future; an ACRL white paper." Accessed 2012. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Vårheim, Andreas, Sven Steinmo, and Eisaku Ide. 2008. "Do libraries matter? Public libraries and the creation of social capital." *Journal of Documentation* 64, no. 6:877-892, http://dx.doi.org/10.1108/00220410810912433

Whyte, A and Allard, S. (Eds) 2014. "How to Discover Research Data Management Service Requirements." DCC How-to Guides.

Edinburgh: Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides

Witt, Michael, Jacob Carlson, D. Scott Brandt, and Melissa H. Cragin. 2009. "Constructing Data Curation Profiles." *International Journal of Digital Curation* 4, no. 3:93-103, http://dx.doi.org/10.2218/ijdc.v4i3.117

Disclosure: The author reports no conflicts of interest.