

Where Do We Go From Here: Choosing a Framework for Assessing Research Data Services and Training

William M. Cross

North Carolina State University Libraries, wmcross@ncsu.edu

Hilary M. Davis

North Carolina State University Libraries, hmdavis4@ncsu.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/charleston>



Part of the [Scholarly Communication Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at: <http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

William M. Cross and Hilary M. Davis, "Where Do We Go From Here: Choosing a Framework for Assessing Research Data Services and Training" (2015). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284316312>

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

Where Do We Go From Here: Choosing a Framework for Assessing Research Data Services and Training

William M. Cross, Director, Copyright and Digital Scholarship, North Carolina State University Libraries

Hilary M. Davis, Interim Head, Collection Management & Director of Research Data Services, North Carolina State University Libraries

Abstract

Research data management has become a critical issue for campus researchers, funding agencies, and libraries, who have made substantial investments of time, energy, and resources into support for managing and sharing data. As data management programs proliferate, however, assessment of research data services has become a notorious challenge for libraries. How can we know—and demonstrate—that our efforts are having an impact, and how can we learn to make them even more effective?

In this session, we will present a survey of several frameworks for assessing research data management services. We will lead a discussion about the application of different frameworks for assessing or auditing existing skill sets, external facing services, and capacity to support an array of research data services. This discussion will be grounded in a demonstration of how we applied one framework to audit the North Carolina State University (NCSU) Libraries’s “training ground” model, which serves the dual purpose of developing competencies within our librarians and supporting researchers in their needs to manage, preserve, and share research assets.

Through an active discussion of our efforts, and the efforts of libraries around the world, we can chart a course for effective research data management that can help guide libraries already deep into the process as well as those just getting their feet wet.

Note: This presentation and conference paper is derived in part from the following publication: Davis, H. M., & Cross, W. M. (2015). Using a data management plan review service as a training ground for librarians. *Journal of Librarianship and Scholarly Communication*, 3(2), eP1243.

Introduction

Research data management (RDM) has become a critical issue for campus researchers, funding agencies, and libraries, who have made substantial investments of time, energy, and resources into support for managing and sharing data. As data management programs proliferate, however, assessment of research data services has become a notorious challenge for libraries. Because the practice of data management itself is relatively new and is evolving quickly, nascent library programs that support data management have no established best practices. A program like the Research Data Committee (RDC) in the NCSU Libraries must develop a strategy that helps us understand and demonstrate the impact of our efforts.

These proceedings present an overview on the NCSU Libraries’s initial attempts to develop such a strategy. We begin with a survey of several frameworks for assessing research data management services reviewed by the NCSU Libraries. We identify three primary frameworks for assessment of an RDM program and discuss other potential frameworks that were considered. Next, we describe the results of the Libraries’s audit process based on one framework, the Joint Task Force on Librarians’ Competencies in Support of E-Research and Scholarly Communication (2014). Informed by this audit, we conclude by discussing next steps for the NCSU Libraries RDM program. We anticipate that our experience can be used by other libraries to inform their own programs.

The NCSU Libraries RDM Program and Frameworks for Addressing Program Needs

The NCSU Libraries Research Data Management program has been described in some detail in previous Charleston Conference Proceedings and a recent scholarly article (Davis & Cross, 2015). In brief, the program is more than two years old and uses a committee structure in order to be team-based, light, and nimble. Our “training ground model” (described below) serves to both meet the needs of researchers and prepare librarians in many different departments for greater engagement with research data.

Our audit of the program focused primarily on three related areas: our externally facing services, our existing skillsets, and our capacity to meet the needs of researchers. By focusing on these three areas, we hoped to identify services that are important to our organization as it works to fulfill its role in supporting research data management. Having identified those services, we also hoped that our audit would help to prioritize certain services over others. After a review of the literature, we identified three promising frameworks for assessing a research data management program.

Framework 1: “Joint Task Force Framework”

The first framework we identified was the the Joint Task Force on Librarians’ Competencies in Support of E-Research and Scholarly Communication (Joint Task Force, 2014). This framework was developed by a collection of the major players in academic libraries including the Association of Research Libraries (ARL), the Canadian Association of Research Libraries (CARL), the Confederation of Open Access Repositories (COAR), and the Association of European Research Libraries (LIBER). Designed to “outline the competencies needed by librarians in [an] evolving environment,” the Joint Task Force framework is a natural fit for evaluation of a library program in RDM.

The Joint Task Force is grounded in three broad areas that are analyzed through specific competencies: providing access to data, advocacy

and support for managing data, and managing data collections. As discussed below, this detailed framework which aims “to identify skill gaps in their institution” and “carry out self-assessments” proved to be the cornerstone of our own audit in the NCSU Libraries.

Framework 2: “20 Needed Competencies for Science & Technology Librarians”

While the Joint Task Force framework was the centerpiece of our audit, another framework also provides valuable guidance for assessment of library RDM programs. A group of scholars at the University of Massachusetts Medical School library worked together to identify and describe competencies “needed in order to cultivate [RDM] skills, establish [RDM] roles, and engage successfully in eScience” (Creamer, Morales, Crespo, Kafel, & Martin, 2012).

This framework is grounded in the twenty “needed competencies” identified in the survey. The framework divides these into two distinct areas. The first, “data literacy,” includes data sharing, open access, data services, “data interview,” data life cycle, IP and copyright, compliance with mandates, data literacy, data management plans, access or locate datasets, and data security. In addition, this framework identifies a number of “technical competencies” such as web 2.0 technologies, data archiving, managing and curating data, metadata standards, digital databases, IR, data mining and visualization, and digital lab notebook applications.

This framework is well organized and provides a rich and detailed set of competencies. It also reflects the health science context that generated it. As such, the framework was useful for establishing context but not a perfect fit for an academic library RDM program.

Framework 3: “Data Management as A Research Tool (DART)”

The third framework we investigated is, in many ways, the most promising. Funded by an IMLS grant, a group of scholars led by Amanda Whitmire has reviewed a large group of data management plans (DMPs) in order to expose the

common practices and hurdles faced by scholars in order to provide insight into the detailed individual data management habits of scholars. They have used this review to create a framework branded as “Data Management as a Research Tool” (DART, 2015).

This work builds on research done by Mischo, Schlembach, and O’Donnell (2014) and Parham and Doty (2012) and has generated an analytic rubric “to standardize the review of data management plans as a means to inform targeted expansion or development of research data services at academic libraries” (DART Project, 2015). Early discussion of the framework has been promising and presentations at conferences in the fall of 2015 suggest significant value for academic libraries looking to audit their RDM programs. Unfortunately, no formal documentation was available when we conducted our audit, and only preliminary results were available at the time of our presentation at the Charleston Conference. The DART framework shows great promise and can be expected to be worth watching for future assessment, but simply was not available to inform our audit.

Bringing the Frameworks Together

In addition to these three frameworks, we reviewed several other promising frameworks that informed our efforts. Work done by Raboin, Reznik-Zellen, and Salo (2012) describe shared competencies across three surveyed libraries that include “garnering institutional support, managing the integration of services with new or existing staff structures, and continuing to meet researchers’ needs as they evolve.” We also incorporated research from peer institutions and model RDM programs such as those documented in Zilinski, Chan-Park, Dasler, and Nicholls’ (2013) discussion of commonalities across RDM services at Purdue, Baylor, Maryland, and Michigan.

As discussed above, the Joint Task Force framework ultimately proved to be the most appropriate for our audit. However, we were careful to look to each of the frameworks to add context and nuance to our analysis. Each added something of value to our audit and each provided information that has helped us understand and

improve our program. Our discussion below is presented in terms of the Joint Task Force, but subsequent audits are likely to incorporate information from the nascent DART framework as well as more recent scholarship that has been—and will be—produced after our initial audit.

NCSU Libraries’ Training Ground Model

Like many libraries, we are in a position of providing support at the point of need, while also training subject liaison librarians in the practical issues and realities of managing research data and research processes. We deployed a Data Management Plan (DMP) Review service to the NC State campus to provide service at the point of need, and to gain access to DMP drafts of our researchers for training purposes. The DMP Review service is managed by a committee of librarians with diverse experience in data management and domain expertise (the Research Data Committee, or RDC). The RDC serves three primary roles: (1) to develop and implement training for librarians and researchers; (2) promote and market those services across campus and to targeted groups; and (3) provide services at the moment of need to support RDM for our university. This group also invites subject liaison librarians to participate in the DMP Review process. Committee membership has strengths in digital scholarship, digital preservation, metadata, geospatial and numeric data, copyright and intellectual property rights, and the grant review process, and includes liaisons with subject expertise. Members rotate annually, allowing new members to get exposure to RDM and inform the future of the committee. To scale out awareness and knowledge within the NCSU Libraries, RDC members have offered workshops and seminars for other library staff on institutional review board protocols, directives for public access to federally funded research, strategies for reviewing DMPs, data rights and ownership, and data management for students. Leveraging what we learned through deploying RDM support for campus, these workshops have facilitated rich, engaging training opportunities grounded in actual issues faced by researchers.

The combination of the DMP Review service and the internal training coordinated by our

Table 1. Core competencies for providing access to data and relevant experience gained through the training ground model.

Core Competencies	Relevant Experience Gained Through Training Ground model
Identifying datasets, discovery tools	Re3data.org and Data Citation Index to identify datasets and repositories for data deposit
Data centers, repositories, and collections	Mediate with data repositories for eligibility and criteria for data deposit; how data is organized in repositories; emerging standards for data citation
Data organization and structure within these collections	
Data citation/referencing	
Data licensing and intellectual property policies and principles	Licensing language and terms for external datasets; knowledge of licensing options to support sharing and reuse of data
Data manipulation/analysis techniques and tools	Expertise limited to a few librarians with specific geospatial and data analysis skills

committee constitutes our training ground model. This model aims to develop needed competencies for our librarians and provide support to researchers through relevant services and partnerships. The heart of the training ground model is the use of actual DMPs to give librarians hands-on experience and opportunities for peer-to-peer training while reviewing DMPs. We all benefit from seeing how we each find problems and propose solutions, and we each play the role of trainer/trainee on any given DMP. The model presents a reinforcing cycle of support and training: skills gained from the DMP review process provide hands-on training for librarians, sustained marketing lets researchers know that we are a source of help, and, at the point of need, the Committee’s shared expertise facilitates immediate support for researchers on a deadline.

Audit of Training Ground Model

In order to identify gaps in our program, we conducted an audit of the DMP Review service in terms of its capacity to develop and enhance

competencies as identified by the Joint Task Force on Librarians’ Competencies in Support of E-Research and Scholarly Communication. The lack of established benchmarks for what counts as quality service and impact limited our ability to empirically measure the effectiveness or impact of our services; therefore, an audit enabled us to take a first step toward an assessment of our model. The Joint Task Force framework presents competencies in three areas: (1) Providing access to data; (2) advocacy and support for managing data; and (3) managing data collections.

Providing Access to Data

Providing access to data includes competencies for finding datasets, leveraging data centers and how data are organized and structured in those collections, methods for citing data, licensing to support sharing and reuse, and tools and techniques for data manipulation. The DMP Review service builds competencies in finding and identifying datasets, but this is limited since most of the researchers who submit DMPs are using

Table 2. Core competencies for advocacy and support for managing data and relevant experience gained through the training ground model.

<i>Core Competencies</i>	<i>Relevant Experience Gained Through Training Ground model</i>
Funders' policies and requirements	Wide range of funding agency requirements for DMPs
Data management plans	Understanding main elements of DMPs; exposure to successful and unsuccessful examples across multiple disciplines
Research practices and workflows	Awareness of researchers' practices and literacy regarding best practices for managing data
Data centers, repositories, and collections for deposit	Identify eligibility for data deposit; investigate how data is organized in repositories; exposure to emerging standards for data citation
Articulate benefits of data sharing and reuse	Ways to help researchers effectively share data and publications while maximizing intellectual property rights
Data sharing options, open access, IPR, licenses	Practices spanning use of commercial grade cloud storage to storage on local media; benefits that established repositories provide (e.g., persistent identifiers, protection of sensitive data, citation tracking, preservation services)
Disciplinary norms and standards for data management	
Data structures, types, and formats	Ways to standardize file-naming, migrate to non-proprietary file formats, document and describe data structure through metadata
Best practices for managing data, standards, metadata, and vocabularies	Variation in metadata standards; in cases where standards do not exist, offer to assist in creating standard practice for research projects
Data publication requirements of specific journals	Data sharing expectations of journals that require data publication; terms in publishers' author agreements that may limit researchers' ability to meet sharing funders' public access mandates
Data audit (i.e., identify range of datasets on campus) and assessment tools (e.g., Data Curation Profiles)	Limited experience in identifying datasets on campus and in employing assessment tools

their own data or have already identified external data sources. The DMP Review service builds competencies in data citation as well as licensing data, for example, when we help researchers find appropriate licenses for data and software. One gap area for our model is that data manipulation and analysis is limited to a few people and is a skill set that needs to be developed more broadly. Table 1 provides an overview of core competencies as outlined by the Joint Task Force

focused on the benefits of sharing data and reuse of data, open access, intellectual property rights, licensing data collections, and disciplinary standards for data management, as well as data types and formats. The DMP Review service supports many of these competencies. For example, we have learned about the drawbacks of commercial-grade cloud storage (e.g., Dropbox and Google Drive) and the benefits of data repositories. Benefits of data repositories

and the relevant experience gained through our model.

Advocacy and Support for Managing Data

This category presents competencies in knowledge, expertise, and awareness of funders' policies, DMPs, research practices and workflows, and data repositories to support RDM. By reviewing DMP drafts, we have been exposed to a wide range of practices and are becoming attuned to variable components of DMPs across and within funding agencies (e.g., NSF, USDA, NIH, DOE). The DMP Review service has provided a practical context for us to learn about data repositories across multiple disciplines; in some cases we have contacted data repositories on behalf of researchers to identify eligibility requirements and criteria needed to contribute data.

This category also includes competencies

Table 3. Core competencies for managing data collections and relevant experience gained through the training ground model.

<i>Core Competencies</i>	<i>Relevant Experience Gained Through Training Ground Model</i>
Metadata standards and schemas, data formats, domain ontologies, data citation, data licensing, and identifiers	Variation in metadata standards; offer to create metadata schema; standards for data citation; licensing options to support sharing and reuse of data; document and author identifiers
Selection and appraisal tech for datasets	Limited experience due to lack of campus data repository
Discovery tools	Some experience in these areas when providing support for researchers in identifying disciplinary or generic data repositories for data deposit; expertise in a few librarians who manage locally curated special collections and university archives
Database design types and structures	
Data linking and data integration techniques	
Data storage infrastructures	
Digital preservation metadata	
Forensic procedures in digital curation	

represent disciplinary norms and standards including persistent identifiers to aid in citation and sharing; support with sensitive data; and tracking citations to databases to monitor the impact of data in the research community.

Finally, this category includes competencies in metadata standards, complying with data publication requirements (and funders' requirements), and a local audit of the range of datasets on campus and use of assessment tools such as the Data Curation Profiles. The audit showed us that we have room for improvement in learning how to develop metadata for research data where standards are lacking. In terms of data publication requirements, as publishers and funders firm up their plans, this will be a constant area for growth for us. When our campus conducts a data inventory, we will be in a position to inform that process. Table 2 provides an overview of core competencies as outlined by the

Joint Task Force and the relevant experience ssgained through our model.

Managing Data Collections

Managing data collections competencies include the application of metadata standards and best practices including ontologies and identifiers, selection and appraisal of datasets, discovery tools, database design, data integration and linking, storage infrastructures, preservation metadata, and digital forensics for curation. The biggest gap area for us is the selection and appraisal of data for local storage since we do not currently have a local data repository. Similarly, our model does not fully support competency development in the areas of data integration in the context of data storage infrastructure, digital preservation metadata, and curation forensics. Expertise is concentrated in a few and needs to be broadened to more library staff. Table 3 provides an overview of core competencies as outlined by

the Joint Task Force and the relevant experience gained through our model.

Conclusions and Next Steps

The primary objectives of the DMP Review service is to help our researchers be more competitive in the research proposal process and to make research assets more widely available. By equipping librarians with the experience to develop competencies and confidence to support researchers, we can deliver help at the point of need. Conducting our audit has shown us where we have been effective at helping librarians develop skills to support researchers at the grant proposal stage. The DMP Review service has given us the credibility and experience to advocate on behalf of researchers. We have noted the lack of a local data repository as a major liability for research competitiveness and have shared concerns of researchers with IT leaders, leading to the development of a proposal for a local storage solution in the coming months.

The DMP Review service has created opportunities to learn valuable skills while providing time-sensitive service to researchers.

However, the gaps identified by the audit point out areas for improvement. To address gaps, we are engaging in the development of a community of practice for data science and data visualization. We are continuing to build out a program of formal, informal training, mentoring, and a support network. We piloted a data science short course for our librarians focusing on topics like data manipulation, data wrangling, first step data analysis, data visualization, and storage infrastructure. We intend to double back and cover the bases to ensure that all of our librarians engaged in research data management support have opportunities to handle raw data, create metadata and discovery-enabled files, get under hood with data repositories, and develop a foundation in digital forensics. We continue to seek partnerships with other units across campus that offer unique skills and support such as data security, options for campus-provided data storage, institutional review board compliance support, technology transfer, and grant proposal development. We are also seeking opportunities to go beyond DMPs to become research partners. In select cases, we are engaged in grant proposals that embed librarians as grant funded senior personnel.

References

- Creamer, A., Morales, M. E., Crespo, J., Kafel, D., & Martin, E. R. (2012). An assessment of needed competencies to promote the data curation and management librarianship of health sciences and science and technology librarians in New England. *Journal of eScience Librarianship*, 1(1), Article 4. <http://dx.doi.org/10.7191/jeslib.2012.1006>
- Cross, W., & Davis, H. (2014). Building capacity for research data management support. *Charleston Conference*, November 5–8, 2014. <http://dx.doi.org/10.5703/1288284315632>
- DART Project: Data Management Plans as a Research Tool. (2015). Project overview. Retrieved from <http://dmpresearch.library.oregonstate.edu/project-overview>
- Davis, H. M., & Cross, W. M. (2015). Using a data management plan review service as a training ground for librarians. *Journal of Librarianship and Scholarly Communication*, 3(2), eP1243. <http://dx.doi.org/10.7710/2162-3309.1243>
- Joint Task Force on Librarians' Competencies in Support of E-Research and Scholarly Communication. (2014). Librarians' competencies profile for research data management. Retrieved from <https://docs.google.com/document/d/15Kqg6w6AFsdbbaxcRCx7DZ3OERnQk6E0UhQ4UeRu0Cc/>

- Mischo, W. H., Schlembach, M. C., & O'Donnell, M. N. (2014). An analysis of data management plans in University of Illinois National Science Foundation grant proposals. *Journal of eScience Librarianship*, 3(1): Article 3. <http://dx.doi.org/10.7191/jeslib.2014.1060>
- Parham, S. W., & Doty, C. (2012). NSF DMP content analysis: What are researchers saying? *Bulletin of the American Society for Information Science and Technology*, 39(1), 37–38. <http://dx.doi.org/10.1002/bult.2012.1720390113>
- Raboin, R., Reznik-Zellen, R. C., & Salo, D. (2012). Forging new service paths: Institutional approaches to providing research data management services. *Journal of eScience Librarianship*, 1(3), Article 2. <http://dx.doi.org/10.7191/jeslib.2012.1021>
- Zilinski, L. D., Chan-Park, C., Dasler, R., & Nicholls, N. (2013). Carpe data: Data curation services at four different institutions. Presented at the 2013 DLF Forum, Austin, TX. http://docs.lib.purdue.edu/lib_fspress/39/