Charleston Library Conference

Building Capacity in Your Library for Research Data Management Support (Or What We Learned From Offering to Review DMPs)

William M. Cross
NCSU Libraries, wmcross@ncsu.edu

Hilary M. Davis
NCSU Libraries, hmdavis4@ncsu.edu

Building Capacity in Your Library for Research Data Management Support (Or What We Learned From Offering to Review DMPs)

William M. Cross, Director, Copyright and Digital Scholarship, NCSU Libraries
Hilary M. Davis, Interim Head, Collection Management & Director of Research Data Services,
NCSU Libraries

Abstract

In our evolving effort to build infrastructure and support around research data management needs, we found traction in launching a data management plan review service. In doing so, we have been able to achieve multiple goals: 1) support the research process; 2) create active learning situations for subject liaisons to engage in and learn how to support data management planning; 3) find resonance with campus-sponsored research officers; 4) collaborate with other campus research support groups including campus IT, the institutional review board, and statistical consulting; 5) and participate in the national dialogue about the tensions of data management.

Introduction

Many models of supporting research data management needs exist. Some libraries invest in spaces to showcase and deliver services for researchers who need support managing data. Examples include the new Research HUB at the University of North Carolina-Chapel Hill, which includes consultation space, workshop space, hands-on data wrangling, deposit, and visualization support. Some libraries have a web presence typically defined by a web guide or libguide that may or may not have librarian support associated with the guide. This approach, while potentially more hands-off, is also low-cost. And a few libraries have taken the approach of hiring a dedicated team of data curation specialists (with deep expertise in data curation and/or data wrangling), with or without dedicated space. This approach centralizes the support in a few library staff experts. In its efforts to provide research data management support at North Carolina State University (NCSU), the NCSU Libraries took a hybrid approach. We started with a web guide, then built up support across our subject liaison librarians, guided by a core team composed of subject liaisons, technology-focused librarians, and curation/preservation librarians. We have iteratively rolled out workshops, presentations, and a data management plan (DMP) review service as part of our growing portfolio. This paper will describe the components of our data management plan review service as a means for building capacity for supporting research data management at our campus. We will explain how different stakeholders participate in the process, and how subject liaisons learn to engage in research data management to lend insights for leveraging subject liaisons in research engagement. We end with a view to the future and explore ways to tap into a broader networks at your own institutions and nationally.

Components of Our Research Data Management Support Portfolio

The mainstays of the NCSU Libraries' support for research data management includes our digital repository services which focus on ETDs, scholarly publications, and technical reports, but have very limited support for datasets. We extend deep expertise in issues and guidance around copyright, intellectual property, open data, and the publishing landscape. Our consultation support provides help to researchers to find and gain access to datasets (with dedicated librarians for geospatial and social science data services). As partners for Dryad (biosciences data repository), we provide technical infrastructure and systems administration for researchers who leverage Dryad. Since 2011, we developed and have maintained a data management planning guide and adopted the DMPTool in 2012.

We deliver this support primarily through workshops and presentations to both broad groups as well as specific academic departments; via consultations to individuals and lab groups; and through referrals directed to our librarians from other units on campus or from the Libraries to other units on campus with specific expertise on issues such as statistical consulting, data security, and institutional review board protocols.

All of this activity is coordinated by the NCSU Libraries' Research Data Committee, which acts as a hub for the research data management support. The committee is composed of eight subject specialist and digital technologies librarians from across the Libraries.

In August 2013 we launched DMP Review service because we wanted to engage more fully in the active phase of planning a data management strategy. We heard from the Research Administration unit on campus that there was an imminent need to help researchers design their initial DMPs to make their data available and manage their data. No other unit provided this hands-on assistance to researchers and we were confident that we could fill that gap. We recognized that it could be a technical conversation (what is data and how to manage it) and that we could help researchers explore those questions.

Equally important, we wanted to know how to support the active phase of data management planning ourselves and we needed to provide practical experiences for our subject liaison librarians to gain skills to better support researchers' data management needs.

Nuts and Bolts of the Data Management Plan (DMP) Review Service

Recognizing the need for a DMP review service and the components of our portfolio, we designed our service to be team-based, light, and nimble. In order to review the DMP's we knew we would need expertise from several fields and that we would not have a full-time research data librarian, so we gathered experts from across the Libraries who could work together and share knowledge on the Research Data Committee (RDC). This gave us

a team that, working together, could cover most of the substantive issues around data management and sharing—from filetypes to funder mandates. It also meant that the RDC would have established relationships with many departments, creating a broad network for outreach across campus.

This team-based approach also served as a training ground where experts could share their knowledge with the rest of the RDC, so all members could learn from one another. In the past year, we have also begun to rotate new members onto the RDC, sending trained librarians back "into the field" and bringing in new members for this "on-the-job training" in data management and sharing. Rotating members in and out of the RDC has also continued to expand our network with stakeholders across campus and created new relationships for the RDC to build upon.

The process for reviewing a DMP begins with a submission of a draft DMP by the researcher to a library web form that is forwarded to an internal RDC email distribution list. Members of the RDC monitor the distribution list and the first person who has the ability to reply back to the researcher reaches out to gather the relevant bits of information needed to review the DMP. The first responder also coordinates with relevant subject specialists and creates a simple Google document of the draft DMP that all RDC members and subject specialists can contribute to. Each member reviews the DMP, adding comments that would improve the DMP or, where appropriate, raising questions that could be best answered by another RDC member with relevant expertise.

Once all comments have been made, the team leader compiles and synthesizes them into a manageable format. We are conscious of the danger of overwhelming a researcher with too many comments or comments that are daunting or overly complex. Our aim is to offer comments that are detailed enough to be put into practice and offer actionable suggestions for the researcher. Generally, these comments do not exceed a single page. To support both DMP review and drafting feedback, we have compiled several documents describing established practice based

on our own past experience as well as materials from other scholars such as Dorothea Salo.

Having offered the service for more than a year, we have also begun some preliminary assessment. Measuring success can be complex, and we want to be particularly sensitive to the time of the researchers who have reached out to us. Since our earliest users have been some of our most active researchers, a time-consuming or detailed survey might be an unwelcome addition to their already busy schedule. Our DMP review service also deals with content that requires confidentiality as well as impacting funding and promotion/tenure decisions, so we want to be mindful of avoiding any questions that might be uncomfortable for our researchers to answer. Because our initial group of researchers would offer a small sample size, we also rejected seeking any quantitative outcomes measurements.

These complications, along with fundamental questions about what constitutes a "successful" program, have led us to rely on a basic set of questions about how our researchers experienced the review. We asked them simply to let us know if the service was helpful for them. Early results have been very positive, with several researchers noting that they appreciated the service in strong terms. "The service was GREAT," one wrote "It really helped me craft a strong data management plan for the project!" Other researchers make it a point to indicate specific suggestions they found helpful, such as data security guidance and help locating repositories for data storage. We will continue to consider strategies for and conduct assessment, but at this stage the project seems to be valued highly by researchers and is certainly valuable for us in the Libraries as we engage with these important issues.

What We Learned From Our DMP Review Service

Supporting the Research Process

Through the DMP Review service and other consultations we have conducted, we have a better picture of where we can have impact on the research process. We have found a lot of momentum in helping to train graduate students

in data information literacy and have been invited to deliver presentations to academic departments about the fundamentals of data management planning. As part of more robust and in-depth effort to increase data literacy for researchers and students in NCSU College of Agriculture and Life Sciences, several of our subject liaison librarians are actively developing and delivering customized hands-on training for those stakeholders.

By reviewing and providing feedback about DMPs, we have found synergy in establishing connections with other experts across campus, including statistical consultation support, institutional review board (IRB) compliance officers, data security experts, and the technology transfer unit.

In many ways, we act as advocates for researchers with publishers (e.g., to help researchers maintaining the rights they need to comply with funding public and open science mandates), and with grant funding agencies (e.g., working as mediators between our NIH-funded researchers and the NIH compliance group to work out trouble spots in the NIH submission system). Most close to home, we help to shape the campus-wide dialogue and drive action toward creating a system that supports activities like assigning author identifiers, DOI minting, providing short- and long-term data storage, and establishing policies or expectations for data sharing, data integrity, and data security.

Creating Learning Opportunities for Librarians

Our goal is for subject liaisons to learn to engage in research data management in a collaborative, supportive environment. Our approach has emphasized the importance and value of leveraging subject liaisons in research engagement. Since we started integrating subject liaisons into our RDM support strategies (July 2013), we have held training sessions/workshops on institutional review board (IRB) protocols, directives for public access to federally funded research, reviewing example DMPs (repeated multiple times), data rights and ownership, data management for students, and statistical consulting services across campus. We are

planning a session with our campus IT to have a discussion with us about supporting data security.

The main component lacking in all of these sessions was real-life examples to build skills and drive home the value of being able to engage with researchers in RDM. The DMP Review service offered exactly that in a safe environment where subject liaisons could engage in the process on their own terms, learn from one another, and gain an understanding of the various components of support across campus.

Early reactions from subject liaison librarian were mixed: some were more comfortable observing the process, while others adopted the new role more readily. As subject liaisons had opportunities to get involved directly with reviewing DMPs, presenting workshops on research data management fundamentals, and consult with researchers, they see the value in being able to handle those conversations and have grown into their new roles in an organic, iterative way.

List of Most Common Mistakes (aka Learning Opportunities for Librarians)

The process of reviewing a DMP might be a little daunting to the uninitiated, but the learning curve is not as difficult to overcome as some might think. When we review DMPs, liaisons can observe and learn from the group process and get involved as much as they are comfortable. Once they have a foot in the door, they can learn how to handle some of the common mistakes. That generates confidence in their ability to take part in the conversation about RDM, and gives them a foundation to build on.

Some of the more common mistakes are listed as follows with opportunities for librarians to offer a value-added level of support:

- Forgetting to review the list of expected components—funding agencies expectations vary, even within different divisions of the same funding agency; the value that librarians bring is to review the specific requirements of the funding source that a researcher is applying for and let the researcher know which are missing. We have found it extremely

useful to review the funding agency guidelines for every DMP that has been submitted to us for review and feedback.

- Neglecting to be more specific about types of data being produced by the research—DMP requirements dictate more specificity beyond just "observational" or "experimental" data; the value that librarians can bring is to help elucidate the data types by asking the researcher questions that get at what types of data will be produced.

- Proprietary vs. nonproprietary formats—it has happened to every one of us: we try to open a file that is in a format for which we don't have the right software or the right version on our computers. Researchers don't often think about converting final data files into non-proprietary formats, but librarians do because we help our patrons overcome this problem all of the time. We help the researcher think ahead and figure out which formats will achieve their goals of future compatibility and ease of sharing.

- Data? I don't have data—several of the DMPs we have reviewed start with "no data will be generated from this project" and then they go on to describe the software code that will be created and the analyses that will be conducted to test models. The value that librarians bring is to remind the researcher about what is considered data and what isn't. We make sure the researchers are aware of this and help them think through their plan for active and ongoing management and sharing of their software code or other data that they don't realize is considered a research asset to funders.

- Describing data—most of the DMPs we have reviewed do not say much more than "we will provide documentation for the data produced." The value that librarians bring is that we recognize that this is too vague to pass muster and we can make easy suggestions about

standard schema to use and even offer consultation to help set up a schema.

- Providing data on request or promising to post it on a researcher's website—this is sometimes considered inadequate, depending on the community of interest. Some PIs are unaware that disciplinary data repositories exist. We have helped connect researchers with disciplinary repositories and often do the legwork to find out logistics, cost, and policies up front.

- Sharing the publications—after reviewing some DMPs, we have suggested that researchers leverage our institutional repositories (text or data or both); if your repository is really only set up for text objects as opposed to data, you can still encourage the researcher to deposit their version of their article into your repository to enhance discovery and uptake. This will also open the door to some useful conversations about author's rights.

- Expectations for re-use by others—we have seen researchers that do not address how they might expect others to re-use their data and we have seen cases where researchers try to write a contract for re-use in the DMP. Librarians can guide researchers to some easily digestible options and take this opportunity to address the balance that can be struck between sharing and intellectual property rights while helping to reinforce data citation practices so that researchers get credit for their data too.

Once we have subject liaisons go through a couple of actual DMP review experiences, they have a better sense of what it takes to have conversations about researchers and their data. We have experts that can help them go the extra mile once the conversation reaches a more advanced level (e.g., metadata generation) and we can leverage other experts on campus (e.g., IRB group).

Finding Resonance With Campus-Sponsored Research Officers

In conjunction with our work with subject specialists, we have developed an important working relationship with our Sponsored Programs and Regulatory Compliance Services (SPARCS). In particular, the College Research Officers (CROs) have been critical contacts since CROs are the people who are on the ground supporting PIs in both the pre-award and post-award phases. Although CROs are not typically involved in writing the DMPs themselves, they are valuable for spreading the word about our support. As research moves from design to grant application and through the process it has been important to establish clear lines of responsibility and coordination, so research can be "handed off" to the appropriate body through recommendations and word of mouth.

This relationship is natural, of course, since we have the shared objective of supporting researchers. Indeed, our creation of our Data Management Planning Guide arose from a SPARCS request. As our project has continued to develop, we hope to continue to find opportunities for collaboration. SPARCS and the CROs have been particularly useful because they have an established infrastructure for information sharing that we have been able to tap in to. From access to listservs and mentions in their presentation to invitations to offer full workshops to their researchers, we have been able to share information about our service through this channel. Although we have not seen substantial adoption of our services directly by CROs, the real value has been putting our name into the ears and onto the tip of the tongues of the people who work directly with and advise researchers.

Collaboration With Other Campus Research Support Groups

We have had similar success with other partners across campus, particularly our campus institutional review board (IRB) and IT teams. As with SPARCS and CROs, we have worked to build lines of communication with these stakeholders so we can share information and tap into a broader collaborative network. In particular, we

have served as the canary in the coal mine for these stakeholders, taking the pain points from the DMPs and using them to make a pitch to campus for improved support and services. Where appropriate, we are able to use the examples from our service to create case studies and document examples where current services leave our researchers unprepared or in need of greater support.

As with our partnership with CROs, we have also offered workshops on IRB issues and Statistical Consulting Services, and held ongoing conversations about what we can learn from one another's experiences supporting research. We have not always been able to persuade these stakeholders to address our concerns— researchers currently face a major gap on campus around networked storage—but by raising those issues and keeping lines of communication open we believe we are continuing to build a case based on evidence we gather about the needs of researchers in this area.

What Is the Next Step/Future for Library Support RDM?

In terms of our next steps for local engagement, we are continuously exploring ways to enhance support for research data management and find the right fit and investment of our already overstretched subject liaison librarians. Dedicated space configured and promoted specifically for digital scholarship is an area of current investment to support needs such as consultation and collaboration space, hands-on data wrangling, and visualization support.

As we continue to push efforts and integrate into campus dialogue regarding enterprise-wide data storage and sharing infrastructure, we also are exploring ways to more formally become part of the grant application workflow (e.g., the DMPTool has a feature to enable researchers to request review of their DMPs) as well as the Responsible Conduct of Research (RCR) program at NC State.

We also recognize the importance of engaging at the regional and national level. So far, our role has focused on serving as a bridge, connecting

national projects with local practice on campus. We work to remain abreast of developing best practices and new tools and bring those to campus. We also regularly share updates about open data mandates and requirements, offering workshops, updates, and consultation about the ongoing response to the White House Office of Science and Technology Policy's Directive on open access. These have covered the development of practitioner-driven solutions such as SHARE and CHORUS as well as the new mandates from executive agencies such as the DOE.

We also work to share local concerns with national stakeholders. By maintaining relationships and lines of communication with researchers and librarians on campus, we are able to draft well-informed responses to calls for comment, requests for information, amicus briefs, and similar federal outreach opportunities. We also use information gathered through this service to inform conference presentations, scholarly writing, and active participation in discussions about open data and data sharing.

Along with this national engagement, we have focused on developing regional networks in our area. The North Carolina Research Triangle has a strong core of academic libraries, each with data management and data sharing programs, and we have worked to leverage those relationships into regular meetings and, where appropriate, partnerships.

These collaborations at the regional and national level will be critical as libraries across the profession work to develop infrastructure needed to support open access to research and research data envisioned by the OSTP Directive, other open data mandates, and subsequent policies and legal requirements. Projects like SHARE, in particular, are predicated on a robust, interoperable network of libraries, repositories, and practices. Without those networks of services and relationship, open access and open data would be much more difficult to manage. By developing them at our institution, we hope to empower our researchers, engage our library services, and facilitate more complete, transformative data sharing.