2014

Research Data Management on a Shoestring Budget

Margaret Henderson
Virginia Commonwealth University, mehenderson@vcu.edu

Regina Raboin
Tufts University

Yasmeen Shorish
James Madison University

Steve Van Tuyl
Carnegie Mellon University

Research Data Management on a Shoestring Budget

by Margaret Henderson, Regina Raboin, Yasmeen Shorish and Steve Van Tuyl

Research Data Access & Preservation

EDITOR'S SUMMARY

In response to mandates from the federal government, many academic librarians face a new challenge to become service providers for research data management. A panel at RDAP14 convened representatives from Virginia Commonwealth University, James Madison University and a collaboration of seven New England libraries to discuss their strategies in response to the regulations when faced with limited resources. A common theme was to take advantage of work already done by others, including faculty surveys and existing data management resources such as DMPTool and Data Curation and Profiles Toolkit. Spreading the word on data issues to spark collaboration and gaining the support of library administration are key. Positive response from stakeholders can provide momentum to request additional resources and provide instruction on data management planning.

KEYWORDS

academic libraries	funding
data curation	collaboration
research data sets	strategic planning

Margaret Henderson is director of research data management at the Virginia Commonwealth University Libraries. She can be reached at mehenderson<at>vcu.edu.

Regina Raboin is the data management services group coordinator and science librarian at Tufts University. She can be reached at regina.raboin<at>tufts.edu.

Yasmeen Shorish is physical & life sciences librarian at James Madison University. She can be reached at shorisyl<at>jmu.edu.

Steve Van Tuyl is data services librarian at Carnegie Mellon University. He can be reached at svantuyl<at>andrew.cmu.edu.

Changes to the regulatory environment around research data management over the past few years, specifically the mandates to manage research data from the National Institutes of Health (NIH) and the National Science Foundation (NSF), have resulted in a great deal of research and discussion about roles, responsibilities and infrastructures for research data curation. Results include a large body of tools such as DMPTool, Data Curation and Profiles Toolkit, training programs (at various MLIS granting institutions) and reports. One major result, however, is that academic libraries, research administration offices and IT units at colleges and universities in the United States are stepping into the role of service providers for research data management.

In February 2013, moreover, the White House Office of Science and Technology Policy (OSTP) changed the landscape of the discussion around research data management by hinting strongly at more widespread requirements [1]. The consequence, in the communities of academic librarians and affiliates who are trying to assess how to approach this problem, is a broad uptick in activity around research data management at institutions that were previously not engaged on the topic.

We convened this panel on shoestring budgets to discuss the experiences of a variety of institutions standing up research data management services on a relatively short timeline and with limited resources. Panelists represented data services providers at an assortment of institutions (a research university, a master's comprehensive university and a multi-university collaborative) and discussed a range of issues related to creating functional research data management programs with few resources.

The Research University

Virginia Commonwealth University (VCU) is a large, urban research university with over 30,000 students on two campuses, with a library on each campus. Less than a year ago, the libraries hired a director of research data management (DRDM) – a new position to both the libraries and the university.

Given the size of the research data management problem, the DRDM and an associate university librarian created a plan that relied on the second mover advantage – the idea that VCU could learn from the successes and mistakes of other institutions in order to build its program. They decided to forego one element of data collection that is very common at many institutions initiating research data management programs – the faculty survey. There were already too many surveys running at the university, including an Educause survey by the IT group, and the Ithaka S+R Faculty Survey, which was going to be sent out by VCU libraries in early 2014. By reading articles and other resources, such as the Purdue Data Curation Profiles [2], they were able to build a picture of what was going on in our institution.

VCU's final plan consisted of these elements:

1. *Create a web presence/communication plan.* Set up research guides with lots of links and tools and work with the library public relations (PR) department to create handouts. The PR department also promoted the creation of the DRDM position online and in some newsletters.

2. *Conduct an environmental scan for data and data management resources.* Comb through websites, call people to see what they can help with and compile a resource list for research data management. Through this process, they found that some individuals felt territorial, and it was necessary for the library to ensure them that the program was there to fill in the service gaps, not take over.

3. *Pick the low hanging fruit*. Many data services have started by helping with NSF data management plans (DMP), just like many relationships have started with NIH Public Access Policy discussions, and more will start with the OSTP memo.

4. *Talk to researchers, students and others involved in data.* Identify faculty, students and administrators who are interested in data issues and people who are looking for help. Through these interactions the library has formed collaborative relationships where possible. One key set of relationships is with technology services and a couple of divisions of the University's Office of Research.

5. *Educate everyone you can.* Seek to engage colleagues, administration, faculty, staff and students on the topic of research data management. The staff spoke to and taught classes for as many groups as possible.

Lastly, the backing of library administration was very important. Starting any sort of new service is hard, and knowing the administration is supporting these efforts makes things much easier.

The Master's Comprehensive University

James Madison University (JMU) is a predominantly undergraduate institution of nearly 20,000 students in the Shenandoah Valley in Virginia. This master's comprehensive university used existing resources to create data management services, built on that success to advocate for additional resources and laid the groundwork for future data curation services. Prior to the fall of 2011, JMU lacked any formal infrastructure to assist researchers in meeting the National Science Foundation's (NSF) data management plan (DMP) requirement. With the hiring of a new science librarian, the library began to offer support to researchers via DMP consulting, albeit in an ad hoc manner given the lack of additional resources to offer a more comprehensive, systematic approach. Despite the lack of additional funding or release time, the library began investigating cross-campus collaborations and outside resources to create a more robust support system. The library initiated coordination with a variety of campus partners to determine researcher needs on campus and the limits of the support available to respond to those needs.

While the NSF requirement was the impetus for starting these conversations, it became clear that there were other areas where offering data management support would be appreciated. While all parties on campus recognized that there was a need that required a systematic response, there were no additional funds or personnel available to create a product to meet that need. Utilizing existing tools, such as the DMPTool,

allowed JMU to quickly address researcher concerns, while investigating what measures could be taken to build additional support.

The library also used information learned from hosting the ACRL Scholarly Communication Roadshow in 2012 to perform an environmental scan of the campus. A taskforce of library faculty met with every college and leadership organization on campus to discuss scholarly communication needs, including data management and curation. Undeniably, constituents wanted the library to take a leadership role in providing support for created digital works, including research data. The positive response by faculty to these actions positioned the library to ask for additional resources to expand services. JMU is currently rolling out an institutional repository and has recently hired a metadata librarian and digital collections librarian to help support any future data curation efforts.

These positions meet current needs and have the potential to be involved in building infrastructure for additional, more curatorial, efforts like RDF and linked data. While it is possible that the library will grow data services in the future, there is still a need to improve the existing infrastructure and to continue to process digital project backlogs. Could the increase in service stimulate enough growth to warrant additional positions, such as a scholarly communications librarian? Or additional staff to aid in digital collections? These considerations are currently unknown, although it may be advantageous to have a single, discipline-agnostic champion in the library who can speak to nontraditional scholarly products such as data.

The Multi-Institution Collaborative

A new service and role for librarians is teaching research data management (RDM) best practices to undergraduate and graduate students and researchers. Funded by the National Library of Medicine (NLM) with the principal investigator and project coordinators from the University of Massachusetts Medical School Lamar Soutter Library, an open, case-based, modular course for librarians to use to teach RDM was developed by seven libraries in the New England region. The New England Collaborative Data Management Curriculum (NECDMC) [3] is an open curriculum, with a CC-BY license. NECDMC is an instructional tool for teaching data management best practices to undergraduates, graduate students and researchers in the health sciences, sciences and engineering disciplines. Each of the curriculum's seven online instructional modules aligns with the National Science Foundation's data management plan recommendations and addresses universal data management challenges. Each module provides a lesson plan, PowerPoint activities and resources that can be used out-of-the-box or customized to fit an institution's or audience's needs. In addition to the workshop materials, the NECDMC site provides additional information and training content [4].

After launching the online course, the researchers conducted a webinar to introduce librarians to RDM and writing data management plans (DMPs). Following the webinar, the researchers held an in-person train-the-trainer workshop [5] for librarians who had attended the webinar from the United States and Canada. The workshop introduced participants to the course materials and provided instruction on teaching the curriculum. Post-workshop, the NECDMC training team identified a sample of librarians interested in piloting the course in their settings. After completing the pilot course at their libraries, the participants conducted a post-course self-evaluation, describing their participants, setting and methods used. They also conducted post-course evaluations of the students' perceptions of the course's content and of the methods and conduct of the course. The researchers analyzed these results and then conducted follow-up qualitative interviews with the pilot instructors. The website also has Community Bulletin Board and Join the Collaboration sections that provide information on how participating libraries are using NECDMC and links to resources used and developed by the partners and pilot institutions.

Eighteen institutions have enlisted to use NECDMC, and of these, nine are in the process of piloting the curriculum. Examples of these pilots are a semester-long, for-credit course; librarian professional development; workshops for graduate and undergraduate students, professional organizations and conferences; library-school for-credit course; and one module per week for researchers. Recently, Module 1 was translated into French and presented at the University of Montreal [6].

Conclusions

The institutions represented on this panel are just a few among many that have initiated new services around research data management with little financial or personnel support. This lack of support is not insurmountable, and the members of this panel have outlined a number of service points that can be provided under these conditions. Broadly categorized they include the following:

1. *Understand the research data landscape.* The VCU and JMU examples, while different in execution, highlight the importance of surveying the landscape of research data support needs when planning research data services. This environmental scan can be internal or external, based on a formal survey or on meetings with stakeholders, can include formal or informal reporting. Regardless of how the scan is conducted, this step is crucial to targeting services to the right audience and at the right level.

2. *Take advantage of existing resources and tools.* Each of the panelists in this session identified existing resources and tools provided by external entities (e.g., DMPTool, the Data Curation Profiles Toolkit, etc.) or units within their organizations (e.g., public relations offices, local administration). While lack of resourcing may be a long-term issue, there are many tools at one's disposal to springboard research data services at one's institution.

3. *Collaborate to create new tools.* The NECDMC case is a great example of how a group of collaborating institutions has identified an internal need (the need for a data management curriculum) and has, in turn, provided tools to help others with that need. The success of NECDMC is a testament to the value of collaborating to create tools and services. Some of the aforementioned resources and tools, such as DMP Tool, were created in this same spirit of collaboration.

While the discussion on this panel was helpful for framing the challenge of providing research data services in a resource-scarce environment, it was clear from the questions during the session that these problems were common among panel attendees, regardless of institution size and type. It is crucial that communities in academia tasked with providing research data management services continue to share tools, ideas and inspiration to the benefit of all. ∎

Resources Mentioned in the Article

[1] U.S. Office of Science and Technology Policy. (February 22, 2012). Expanding public access to the results of federally funded research [Web log post]. Retrieved from www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

[2] *Data Curation Profiles Toolkit*: www.datacurationprofiles.org

[3] New England Collaborative Data Management Curriculum: http://library.umassmed.edu/necdmc/

[4] NECDMC Slides from Workshops: http://library.umassmed.edu/necdmc/slides_from_workshops

[5] NECDMC Train-the-Trainer: http://library.umassmed.edu/necdmc/necdmc_train_trainer_emartin.pptx

[6] Clairoux, N. (May 22, 2014). Introduction à la gestion des données de recherché, www.slideshare.net/nclairoux/gestion-donnes-recherche-plan-de-gestion-des-donnes-archivage-prservation