

Joining in the Enterprise of Response in the Wake of the NSF Data Management Planning Requirement

**Patricia Hswe, Digital Collections Curator,
Pennsylvania State University Libraries**

and

**Ann Holt, Graduate Assistant, Office of Scholarly Communication,
Pennsylvania State University Libraries**

The data management plan requirement activated in January 2011 by the National Science Foundation (NSF) has mobilized many research libraries to develop and offer resources and services more specifically dedicated to guiding faculty and students to meet this new condition. At libraries, library associations, and data service organizations alike, a spate of new or revised web pages, as well as webinars, workshops, templates, and tutorials, has emerged in the months since the NSF's May 2010 press release.¹ A sense of urgency no doubt infuses this enterprise of response—and rightfully so. Funder requirements cannot be ignored. Such enterprise may hint at novel, even groundbreaking, roles for librarians and libraries, particularly as subject specialists, data curators, researchers, information technologists, and university administrators come together, perhaps for the first time, to address the requirement in actionable ways. Yet, it is also necessary and affirming to take momentary stock of the situation. What are libraries already doing in this space that would be valuable to apply and expand on? Who are the data specialists in our libraries whose expertise could be leveraged for purposes of both “inreach” (educating librarian colleagues in data management concepts and practice) and outreach (getting the word out to faculty researchers that the library is ready to help)?

This article affords an overview of the new, leading roles libraries can

adopt in the provision of data services, thus blending appraisal with advocacy. How are libraries currently giving assistance in data management planning? What recommendations can libraries make that draw from, and build on, these efforts? The article also reports on new communities of practice forming around the challenges of digital data issues, bringing together much needed knowledge and expertise not only from libraries but also from various other sectors of a university, including IT divisions, grant administration offices, and research institutes.

The NSF requirement may appear to cast libraries into uncharted territory, but there is arguably much territory already charted here—to the extent that some of it may need only to be remapped toward either specific or generalizable uses. An example is the challenge of developing a template broadly applicable to management of research data in a range of disciplines, yet sufficiently detailed and targeted both to meet the NSF requirement and to suit the particular community of interest (in the absence of more specific guidelines provided by an NSF directorate or division). Another example is found in practices familiar to subject specialist librarians with public service experience: just as the reference interview constitutes an important structured approach for determining the information need of users, so is the “data interview” a critical, deliberate process for helping researchers think through their data management needs.²

Similarly, information literacy practices can also be consulted. An effective understanding of data management planning involves reaching a level of literacy about data—i.e., what are the issues regarding description and documentation of data as well as their access, sharing, storage, and security? Tutorials and workshops in “data literacy” can be integrated in research methodology courses and certificate programs in research integrity, which many junior faculty and even graduate students conducting original research often are required to take. Given the increasing emphasis on the ability to understand and work with data, as well as to manage it, it becomes incumbent on librarians and faculty to work together to educate students early in their university and college careers about research data and, perhaps more crucial, to impart consistent advice on how to “do” data planning.

Subject specialists who have liaison librarian responsibilities have a prominent role to play in this realm, too, as suggested by Tracy Gabridge of MIT in “The Last Mile: Liaison Roles in Curating Science and Engineering

Research Data,” which appeared in the August 2009 issue of *Research Library Issues*.³ The “last mile,” borrowed from telecommunications jargon, refers to the final stage of a “research data cyberinfrastructure—the part of the network that will provide connections between the systems and the researchers and, ultimately, to new users of the data.”⁴ Well before the NSF announcement, Gabridge proposed data management planning as an area in which liaison librarians in science and engineering could expand on their already collaborative efforts, laying the groundwork for depositing data by conferring with faculty researchers from the moment data is created. “Librarians can put researchers in touch with standards applicable to their need, create a plan for managing the lifecycle of data in compliance with their grants, and create organizing strategies for documentation, files, backups, and more.”⁵ Resources like MIT’s Data Planning Checklist⁶ and digital curation guidelines provided by the Inter-University Consortium for Political and Social Research⁷ have been available in the last two to three years and are often the first steps for both librarians and researchers who are wrangling data plans.

More importantly, inasmuch as the NSF requirement marks a chance for libraries and campus entities such as research institutes and laboratories to join forces anew, or cohere substantially around a common challenge, it also occasions, if not necessitates, opportunities for cross-departmental collaborations within a library itself. The call from NSF means that librarians will need to have more than a satisfactory understanding of one another’s work—for reasons of efficiency, accurate referrals, and identification of gaps in services and specializations. Implicit in Gabridge’s foregoing appeal is the depth and range of librarian expertise that cuts across boundaries of practice and skill sets: subject specialists, metadata librarians, institutional repository coordinators, data curators, systems/IT librarians, copyright specialists, collection managers, and acquisition librarians (for advisement on data sharing and collection policies). The expertise may have to be mined differently than before, and a new framework, or reorganization of infrastructure, may have to occur, but many of the essentials for assisting faculty researchers on data management issues have long been available in academic libraries. In addition, the cross-specialization making up these collaborations could be transferrable for the creation of a broader range of research services. As Dorothea Salo suggests in a posting at the *Book of Trogoon* blog, “I encourage libraries and IT shops building data-management services

on the strength of the NSF’s plan requirement to diversify, and that quickly. Find non-NSF people to help. Do a survey or focus-group study to demonstrate non-NSF-related data-management needs. Pay some attention to the digital humanities.”⁸

The emergence of collaboration as a requirement itself in this enterprise of response cannot be underestimated, and it continues to be born out in ways suggestive of communities of practice—knowledge networks of people sharing common interests and commonly created intellectual resources.⁹ In a community of practice, everyone contributes to the whole, as members share

The emergence of collaboration as a requirement itself in this enterprise of response cannot be underestimated, and it continues to be born out in ways suggestive of communities of practice—knowledge networks of people sharing common interests and commonly created intellectual resources.

information, seek collective wisdom, and learn from each other. The “ARL Data Sharing Support Group,” a mailing list based in Google Groups, the creation of which coincided with the launch in December 2010 of ARL’s web-based *Guide for Research Libraries: The NSF Data Sharing Policy*,¹⁰ exemplifies such a community. With more than 200 members, the group has informally discussed a variety of questions, ranging from inquiries

regarding video archiving and storage solutions; to challenges surrounding the cultivation of a data services program needing the support of intra-campus alliances; to the idea of sharing data management plans—or not—among researchers belonging to the same institution. Members also post announcements about developments in the data management sphere, and ARL uses the list as a vehicle for alerting librarians to new content in the *Guide for Research Libraries: The NSF Data Sharing Policy*. Issues and ideas that have arisen in this forum include the following:

- How will data management plans help federal funding agencies in the future? The NSF data management plan requirement arguably enables the agency to do an environmental scan, in the sense of finding out what is being accomplished in this problem space across US institutions. The knowledge resulting from NSF’s review of these plans could inform the development of baseline best practices and policies concerning the future curation of scientific research data.
- There has been some brief discussion about surveys, including survey models to follow, whom to survey, and what to survey. While there are

institutions that have distributed surveys to faculty for identifying requirements for data curation services, others are wary of burdening faculty with these requests and incurring “survey fatigue.” List members have reported that data interviews and informal evaluations of need in areas such as information technology have helped fill the survey void.

- The especially basic guidelines for these plans (except in cases where NSF directorates or divisions provide more specific instructions) and the two-page limit seem appropriate for the first year of a requirement, but also imply a “wait and see” approach. Plans submitted this year could potentially shape what the requirement looks like next year.
- It is clear that in the wake of the NSF mandate, libraries are doing their best, within their current means, to meet researchers’ needs for assistance in fleshing out data management plans. But without additional resources (whether from NSF or from research universities submitting proposals) it will be difficult to expand on these efforts toward the development of necessary services—not to mention sustain them.
- While the requirement has compelled librarians to give serious thought to how to help faculty develop these plans, there is preliminary concern about researchers as yet not sharing them, or making them available, within a community of interest—essentially as examples for peers to view and learn from. Many librarians believe that a culture of sharing these plans, particularly in cases where no sensitive information is evident, should be fostered. Of additional interest is seeing whether NSF itself will provide model data plans in the months following January 2011.
- Similarly, there is as much concern that researchers will *not* think to consult librarians—which could lead to inaccuracies, misinformation, or unrealistic expectations in the plans themselves.

Since the ARL Data Sharing Support Group list began, much of the posting activity has convened around matters of practice and the search for advice and solutions. The list has the makings of a resource that documents the efforts various institutions are applying toward data planning and that captures a host of use cases, or ideas for case studies to pursue, in support

and awareness of new processes and services to develop for assisting faculty researchers.

In this enterprise of response, however, it is early days yet. While at some institutions, research data assessment activities are well underway, at others the value of data as a long-term and reusable asset, the management of which is worthwhile funding, still has traction to gain among high-level university administrators, who have a vested interest in continuing to receive significant research funds. Questions such as how institutions will curate research data in a centralized way, make them discoverable, findable, and usable, and ensure long-term preservation and access will depend much on identifying relevant stakeholders, arriving at a common ground and obtaining buy-in (which includes stakeholder participation in the process and not hand waving), and forging the right working relationships to get things done. As librarians work increasingly across units and departments both within and beyond their libraries, it will be energizing for the profession to see what models for agility, collaboration, communication, program development, process management, and workflow design come into play that can be adapted for local environments. Foremost, what results externally from this enterprise, in terms of innovative user services, must be evident and of benefit to the faculty and students whom academic libraries serve, especially if we want to foster “new users of data.” As Meredith Farkas notes in a posting to her blog, *Information Wants to Be Free*, “We need to understand how they [our patrons] do research, how they use our current resources, why some of them don’t use the library, and what they want from the library that they’re not currently getting.”¹¹ When this understanding is achieved, the “last mile” will be completed.