

Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions

Karen Antell, Jody Bales Foote, Jaymie Turner, and Brian Shults

As long as empirical research has existed, researchers have been doing “data management” in one form or another. However, funding agency mandates for doing formal data management are relatively recent, and academic libraries’ involvement has been concentrated mainly in the last few years. The National Science Foundation implemented a new mandate in January 2011, requiring researchers to include a data management plan with their proposals for funding. This has prompted many academic libraries to work more actively than before in data management, and science librarians in particular are uniquely poised to step into new roles to meet researchers’ data management needs. This study, a survey of science librarians at institutions affiliated with the Association of Research Libraries, investigates science librarians’ awareness of and involvement in institutional repositories, data repositories, and data management support services at their institutions. The study also explores the roles and responsibilities, both new and traditional, that science librarians have assumed related to data management, and the skills that science librarians believe are necessary to meet the demands of data management work. The results reveal themes of both uncertainty and optimism—uncertainty about the roles of librarians, libraries, and other campus entities; uncertainty about the skills that will be required; but also optimism about applying “traditional” librarian skills to this emerging field of academic librarianship.



In January 2011, the National Science Foundation (NSF) began requiring researchers to submit a two-page data management plan as part of each funding proposal. NSF guidelines specify that this document should include information about the types of data to be gathered in the course of the research, the metadata standards to be used, the policies and provisions for reuse of the data by others, and plans for long-term data archiving.¹ This mandate affects

Karen Antell is Head of Outreach & Special Initiatives, Jody Bales Foote is Geology Librarian, Jaymie Turner is Serials & Electronic Resources Librarian, and Brian Shults is Interim Coordinator of Digital Initiatives, all in the University of Oklahoma Libraries; e-mail: kantell@ou.edu, jbfoote@ou.edu, jturner@ou.edu, bcsults@ou.edu. © 2014 Karen Antell, Jody Bales Foote, Jaymie Turner, and Brian Shults, Attribution-NonCommercial (<http://creativecommons.org/licenses/by-nc/3.0/>) CC BY-NC

scientists at practically every large research university, and it affects science librarians as well. Some libraries and some librarians have been engaged in data management for years, and a handful of institutions have already developed robust institutional repositories (IRs), data repositories (DRs), and data management strategies. But due to the recent NSF mandate, data management is no longer a peripheral topic for most science librarians. As this study will show, the vast majority of science librarians are aware of the NSF mandate. In other words, data management is entering the mainstream of academic library work.

The NSF is not the first funding agency to initiate a data management mandate. In 2003, the National Institutes of Health implemented a similar requirement, and some other government agencies and private foundations have done so as well. However, because of the NSF's size and influence, it is likely that many other funding organizations will follow suit. In short, the "data management mandate" appears to be at or near its tipping point: From now on, researchers applying for funding should expect to be required to plan for data management. Concomitantly, academic librarians probably should expect to be called upon to provide them with data management assistance.

Researchers from all disciplines work on funding proposals, of course, but researchers in the sciences constitute the majority of grant recipients at most research universities. Moreover, many of the data management requirements involve the kind of work in which librarians already have expertise—organizing information, applying metadata standards, and providing access to information. For these reasons, science librarians in particular are mobilizing to meet the needs of researchers faced with the challenge of developing data management plans. This is a new role for many science librarians, and, like many new ventures, it presents itself as both an opportunity and a challenge. By taking on responsibility for data management assistance, science librarians gain the opportunity to work more closely with research faculty, to provide patrons with access to vital data, and to develop expertise in the "data universe" that almost certainly will become increasingly important in the coming years. But the challenge of the unknown looms as well. Science librarians may well wonder whether they are prepared for this new role. Do they have the skills they need to help researchers with data management, or do they need to develop an entirely new vocabulary and skill set?

This study presents the results of a survey of science librarians at institutions affiliated with the Association of Research Libraries (ARL). The survey explores science librarians' awareness of and involvement in their institutions' IRs, DRs, and data management support services. In addition, the survey reveals science librarians' current job duties related to data management and their opinions about which job skills they think are necessary for librarians involved in data management work.

Literature Review

The library literature on data management is fairly new and, thus far, limited mainly to case studies from particular institutions. A few articles focus more on philosophical issues, providing context and supplying definitions of terms such as "data curation" and "e-science." To date, however, no published research has shed light on science librarians' evolving roles in the wake of data management initiatives at large research libraries.

Two articles by authors at Purdue University discuss the Distributed Data Curation Center (D2C2), created by the Purdue University Libraries in 2006 "to serve as a mechanism to bring researchers together to investigate ways in which optimal dataset management can be achieved at Purdue and throughout the research world."² Mullins (2007) provides a definition of e-science, "the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet," and explains possible differences in the definition of data curation between scientists and

librarians. The scientist might define data curation as “the process of examining, testing and selecting information to be deposited into a database.” For a librarian or archivist, on the other hand, data curation is “the intent ... to store, provide access, preserve, and carry forward into the future with assurance that the data will be accessible and retrievable for future verification or use.”³ Witt’s 2012 article discusses the Purdue University Research Repository (PURR), an institutional data repository.⁴ One of the goals of the PURR Working Group, chaired by the Purdue Libraries’ Interdisciplinary Research Librarian, is to bring librarians and researchers together in the data curation process. Training is provided to librarians to encourage data-related outreach, and a LibGuide, “Supporting Information for Data Services,” has been produced as a resource for librarians. Witt concludes by stating that “Working with data will become a mature component of librarianship when it is accepted into regular library practices: when terms like ‘data reference’ become simply ‘reference’ and datasets are not given any more specific or specialized treatment than other library collections.”⁵

The Georgia Tech Library took a proactive stance in 2009, well in advance of the 2011 NSF mandate, by surveying faculty and researchers to assess their data outputs and learn how they are managed, stored, shared, and preserved.⁶ Analysis of the survey is ongoing, but preliminary results have shown great interest from the faculty. Half of the 63 faculty and researchers surveyed also participated in interviews about data curation. As a result, “by the time the NSF data management plan requirement went into effect, the library was positioned to take a leading role in campus efforts to address the requirement, having already begun an institute-wide conversation about managing research data.”⁷

Choudhury (2008) investigated data curation with a focus on the institutional repository at Johns Hopkins University,⁸ where the “IR is being developed as a ‘gateway’ to the underlying digital archive that will support data curation.”⁹ According to him, “institutional repositories did not inspire changes in scholarly communication, but they could play an important role in supporting new forms of data-intensive scholarship.... [D]ata have become a new form of publication, which are critical for [scientists’] research and teaching purposes.”¹⁰ Choudhury also addresses the new roles of “data scientist” or “data humanist”: “They act as the human interface between the library and the eScience projects. In a fundamental sense, they may represent the future of subject librarianship and help craft a new relationship between the library and scientists.”¹¹

At the University of Houston, science librarians conducted a pilot study in 2010 to assess current data management practices on their campus.¹² They interviewed ten principal investigators of grant-funded projects, inquiring about what types of data were used, who managed the data, where the data were stored, and how long the data would be kept. The pilot study showed that more than one unit on campus was providing data management support to faculty, but the authors found little clarity on the specific services offered, leading them to suggest that the library take on the role as campus facilitator for assistance with data management. They proposed establishing a library Data Working Group, hosting meetings of data service providers on campus, and creating a series of “data management 101” instruction sessions for all liaison librarians.

Yakel’s 2007 article enumerates five aspects of data curation: lifecycle management of materials; active long-term involvement by data creators and managers; appraisal and selection of materials; provision of access; and preservation.¹³ This article also describes recent initiatives and conferences that bring together academic and governmental bodies for the purpose of maintaining perpetual access to datasets. In addition, Yakel notes that digital curation is being addressed in academic library science programs. For example, the University of Michigan School of Information has established a preservation specialization with a digital curation focus.

In their 2011 article, Hswe and Holt investigate several issues related to the NSF data management mandate.¹⁴ The authors note the need for both “inreach,” or educating librarians about data management concepts, and “outreach,” or interacting with faculty researchers about the role libraries can play in data management. They suggest that data literacy be integrated into graduate courses in research methodology and emphasize that “the emergence of collaboration as a requirement itself in this enterprise of response cannot be underestimated.”¹⁵ While acknowledging that researchers will not necessarily think to consult librarians as they create data management plans, they remain optimistic about this new collaborative opportunity: “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.”¹⁶

In a 2010 article, Brown reports the results of a survey about the level of involvement of academic librarians in the practice of data curation in New Zealand higher education institutions.¹⁷ The study’s sample size was small, and, although the study revealed that involvement in actual data curation projects remained very low, it also identified considerable potential for librarians to lend their expertise to collaborative data curation projects with researchers. Another study from New Zealand examined a data management project conducted in 2008 on biodiversity research at the University of Otago.¹⁸ One outcome of the project was that many researchers at the university were reminded of “the Library’s potential in the emerging e-research environment: ‘I’d forgotten about the Library, what a good idea.’”¹⁹

Morgan (2007) examines the use of institutional open access repositories for data and other nontext materials.²⁰ He focuses on the development of SPECTRA (Submission, Preservation, and Exposure of Chemistry Teaching and Research Data), a collaborative project of the university libraries and chemistry departments at the University of Cambridge and Imperial College London, in cooperation with the eBank-UK project. One of the premises of SPECTRA is that “chemistry as a discipline has been slower than the physical and biomedical sciences to adopt and exploit Open Access concepts in the handling of experimental data and research publications.”²¹ Thus, the main objective of the project was “to develop a set of customized software tools that would enable chemists routinely to deposit experimental data in open access repositories, employing the DSpace repository platform used by the two libraries.”²²

Heidorn’s 2011 article²³ makes the case for libraries to take on the role of curating digital data: “Libraries have the skill sets, longevity, and most of the infrastructure needed to accomplish this task for many types of data. If libraries do not actively engage in the task, then society may choose to create a new type of institution to curate digital data.”²⁴ Heidorn also notes that several schools of library and information science are offering courses and certificates in data curation.

In a 2009 article, Garritano and Carlson of Purdue highlight some of the skills that subject librarians need if they are to assist researchers in setting up a data management plan.²⁵ These include being aware of the scholarly communication trends in the discipline and having knowledge of different data formats and metadata standards. Furthermore, in a 2010 book chapter, these authors describe new models of research support at Purdue University Libraries, citing librarians’ ability “to become directly involved in the development of cyberinfrastructure and to provide support for e-science research.”²⁶ They report that a new library position, “data research scientist,” has been created at Purdue, with the goal of “building interdisciplinary research initiatives in data management, curation, and preservation and related areas.” However, involvement in data management is shared among many library employees: “22 librarians

have participated in more than 47 multidisciplinary grant proposals.... [These] activities give librarians an opportunity to reclaim our status as the central provider and steward of research information, no matter in what form it is captured."²⁷

A 2012 article by Dietrich et al. investigates the data management requirements of several major U.S. funding agencies and how they affect libraries that offer data management services to researchers.²⁸ The authors note that requirements are vague, funding for data preservation is a concern to researchers, and many researchers are unsure where they can deposit their data. This uncertainty indicates researchers' need for assistance and guidance—roles that librarians with the requisite skills can readily fill.

Methodology

The researchers developed a 16-question online survey and e-mailed it in September 2012 to 507 science librarians at the 116 academic libraries affiliated with the ARL. These 507 librarians were identified by searching the web pages of ARL academic libraries. In most cases, science librarians' e-mail addresses were available on their institutions' web pages. For six institutions, e-mail addresses were not posted online, so the researchers contacted these six libraries to request the science librarians' contact information. All six of these libraries responded. Seven of the 507 e-mails were returned to the researchers as permanently undeliverable, so the survey was delivered to exactly 500 science librarians. One hundred seventy-five responses were received, for a response rate of 35.0 percent. The survey instrument is shown in the Appendix.

The survey included five multiple-choice questions, three questions inviting narrative comments, and eight questions containing both multiple-choice and narrative elements. Thus, both quantitative data and qualitative data were collected. None of the questions required a response; participants were free to skip any questions if they so desired. Data analysis is reported in the Results section. For the narrative-response questions on the topic of job skills and duties, two researchers coded the narrative responses into relevant categories. (The coding categories are shown in the legends to figures 5, 6, 7, and 9.)

Results

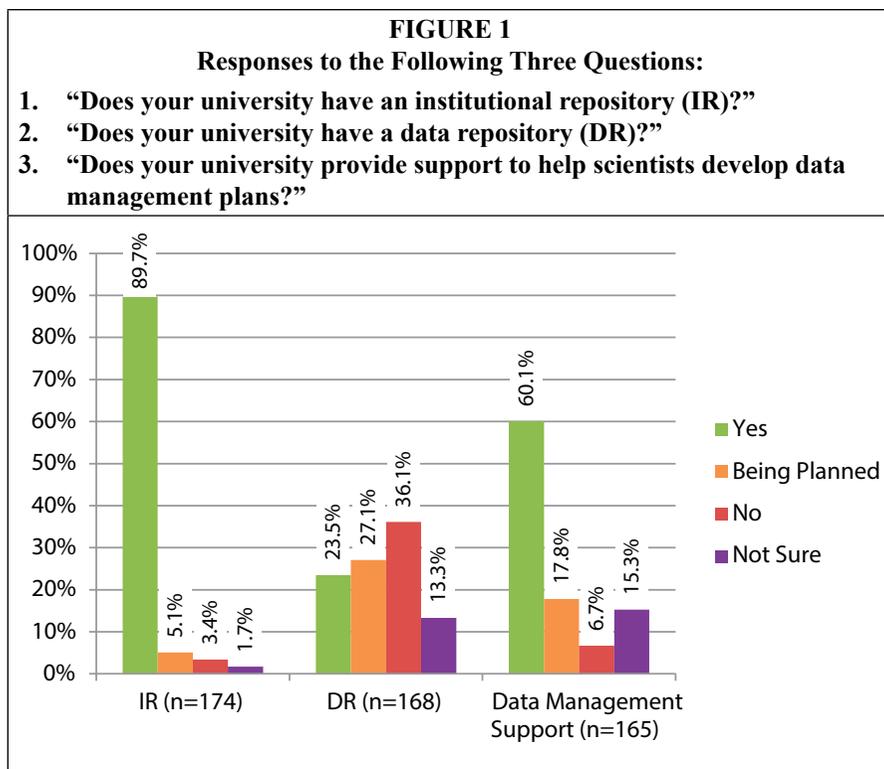
Awareness of NSF Mandate

The first survey question asked whether respondents are aware of the NSF mandate that went into effect in January 2011 requiring grant proposals to include a two-page data management plan. Only nine respondents (5.1%) answered "no" to this question, indicating that they were unaware of the mandate; the remaining 166 (94.9%) answered "yes."

Institutional Repositories, Data Repositories, and Data Management Assistance

The next set of questions asked about respondents' universities' IRs and DRs and the provision of data management assistance to researchers. (See figure 1.) Almost 90 percent of respondents reported that their university has an IR, and 5.1 percent indicated that an IR is being planned. By contrast, just 23.5 percent of respondents reported that their university has a DR, and 27.1 percent indicated that a DR is being planned. Only 3.4 percent reported that their institution has no IR, whereas 36.1 percent indicated that their university has no DR. A majority of respondents (60.1%) reported that their university provides data management assistance to researchers, and another 17.8 percent indicated that such assistance is being planned.

Participants who indicated that their institutions have (or are planning to have) both an IR and a DR were asked whether the DR is (or will be) part of the IR. Thirty-four percent of respondents answered "yes," and 13.2 percent answered "no." Fully 52.8 percent answered "not sure."



The survey invited respondents to add comments if they wished. With regard to the existence of IRs, DRs, and data management assistance at their institutions, respondents made a total of 119 comments, many of which clarified or expanded upon their yes-or-no responses. For instance, 8.4 percent of comments indicated that the institution’s IR also accepts data, and 10.9 percent noted that the institution is in the very beginning stages of implementing an IR or a DR. Other comments mentioned details of institutions’ IRs and DRs. For example, seven comments revealed that the institution participates in a consortial IR or DR, and four comments indicated that the institution has multiple subject-specific IRs or DRs. The provision of data management assistance was mentioned in 31.9 percent of comments, and 24 percent of these reveal that librarians are available for consultation on data management plans. Two comments noted that researchers do not necessarily trust the library with their data or do not believe that librarians are qualified to work with research data.

In the next section of the survey, participants were asked about which campus entities operate their IRs and DRs and provide data management assistance. (See figure 2.) For these questions, multiple responses were allowed, so the totals equal more than 100 percent. Fully 90.1 percent of respondents indicated that the library operates the IR, whereas just 9.9 percent of IRs are operated by the campus information technology department and only 3.7 percent are run by the campus research office. Campus research offices and information technology departments appear more likely to be involved in DRs than IRs: According to survey results, 20.3 percent of DRs are operated by the campus information technology department, and 8.7 percent are operated by the campus research office. However, even for DRs, libraries have the strongest showing, with 36.2 percent of respondents indicating that their DRs are operated by the libraries. With regard to

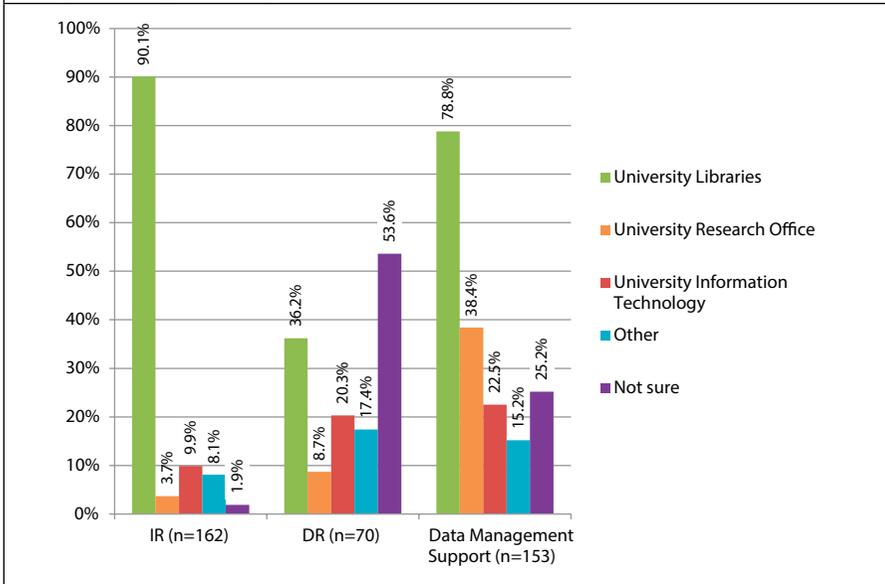
data management assistance, 78.8 percent of respondents indicated that the university library provides such assistance, followed by 38.4 percent for university research offices and 22.5 percent for university information technology departments. Only 1.9 percent of respondents reported that they were unsure about which campus entity operates the IR, whereas 53.6 percent were unsure about which entity operates the DR and 25.2 percent were unsure about which entity provides data management support.

Respondents provided 106 comments in this section of the survey, most of which simply elaborated upon the multiple-choice responses. Typical comments included “The California Digital Library [operates our IR]” and “Our [DR] is co-sponsored by the Libraries, campus IT, and the office of the vice president for research.” Fourteen comments expressed uncertainty, particularly with regard to data management assistance. For example, one respondent noted: “I know other groups will [assist with data management], but I’m not sure who. Individual departments have IT people and I think they often help.” According to another participant, “The library provides guidelines and examples for departments, but I am unaware of whether other campus service providers also provide support.”

The next section of the survey asked respondents to indicate how many library employees work with their institutions’ IRs and DRs and provide data management assistance. (See figure 3.) For all three questions, a strikingly large proportion of respondents reported being unsure of the answer (30.8% for IRs, 54.8% for DRs, and 31.8% for data

FIGURE 2
Responses to the Following Three Questions:

1. “What campus entity operates (or will operate) the IR?”
 2. “What campus entity operates (or will operate) the DR?”
 3. “Which campus entities provide (or will provide) support for scientists developing data management plans?”
- (Because respondents were able to choose multiple responses, the percentages add up to more than 100%.)



management assistance). Of the respondents who indicated that their institutions have an IR, 61.0 percent reported that between one and five library employees work with the IR, and an additional 8.1 percent indicated that six or more library employees work with the IR. By contrast, only 35.5 percent of respondents whose institutions have DRs reported that between one and five library employees work with the DR, and just 5.7 percent reported that six or more library employees do so. Just under 4 percent reported that no library employees work with the DR. With regard to data management assistance, 11.3 percent of respondents reported that no library employees are involved in such work, 41.1 percent indicated that between one and five library employees provide such assistance, and 15.9 percent reported that six or more library employees are involved.

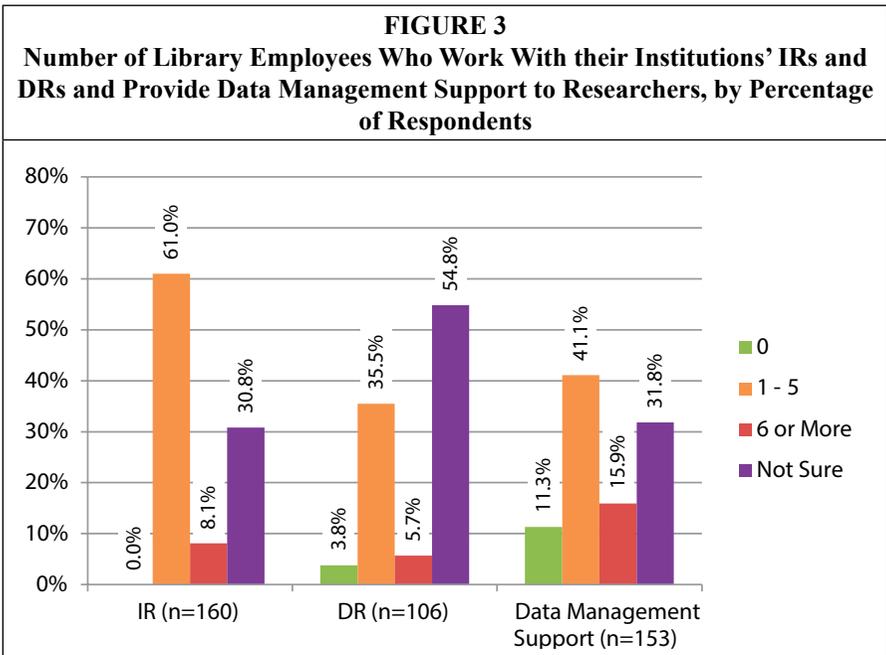
Job Duties and Skills

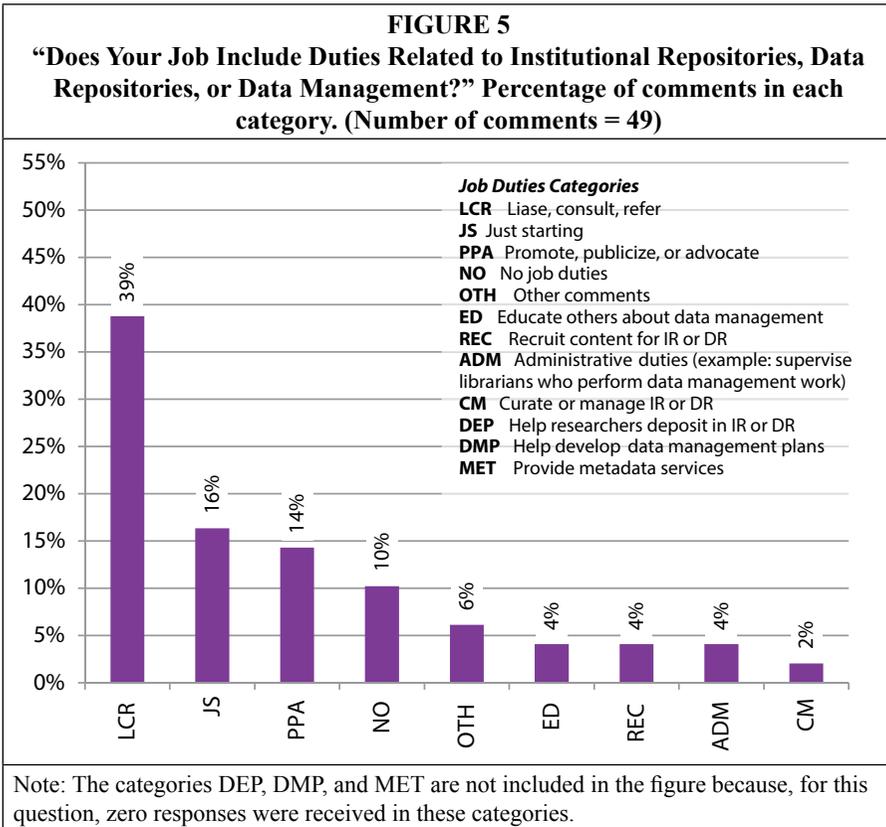
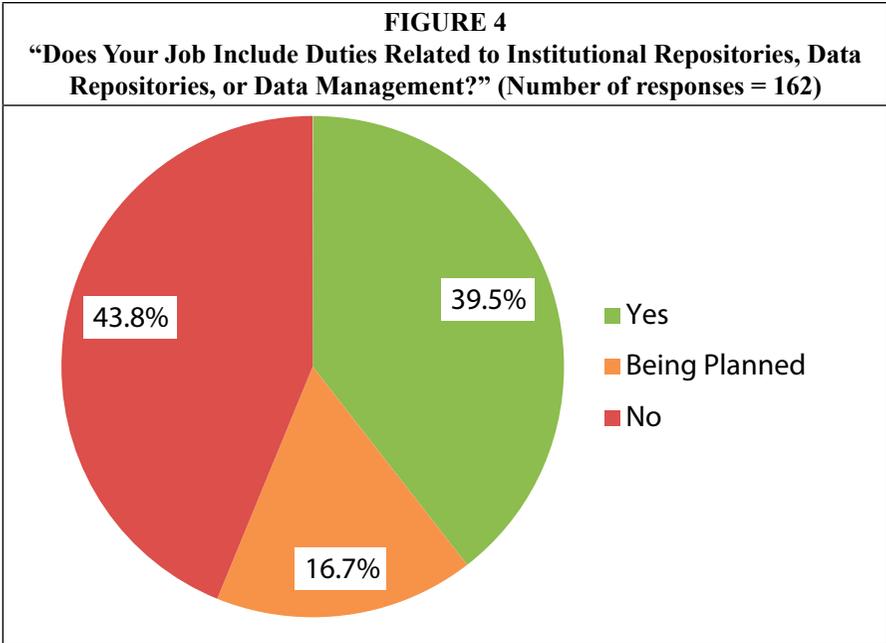
The next set of survey questions asked about job duties and skills related to data management. One hundred sixty-two respondents answered the question, “Does your job include duties related to institutional repositories, data repositories, or data management?” Almost 40 percent answered “yes,” and 16.7 percent answered “being planned,” whereas 43.8 percent answered “no.” (See figure 4.)

In addition, 38 respondents added a total of 49 comments specifying which duties are included in their jobs. (See figure 5.) By far, the largest contingent (39%) indicated that the respondent’s job included tasks such as “liaise, consult, or refer,” defined by the researchers as follows:

- Includes the work of subject specialists who are the first point of contact for researchers with questions about all aspects of IR, DR, and data management.
- Includes helping researchers identify an appropriate depository outside their own university.

Furthermore, a substantial proportion of respondents indicated that they were “just starting” to perform data management tasks (16%) or that their work included “promoting, publicizing, or advocating for” the library’s data management services





or repositories (14%). Other comments included “data management has been added to our positions” and “[No, but] I wish it did.” Two respondents noted that their jobs currently did not “officially” include data management duties but that they were learning about data management in preparation for anticipated changes.

The next question asked respondents to “describe [their] job duties related to institutional repositories, data repositories, and data management.” Eighty-two respondents answered this question; all together, they provided 152 comments. (See figure 6.) Of these, as with the previous question, the largest proportion of respondents (25.0%) indicated that their job duties include “liaising, consulting, or referring.” The next most frequently cited job duty was “help researchers develop data management plans,” with 15.8 percent of respondents, followed closely by “promote, publicize, or advocate” for the library’s data management services (15.1%). In addition, several comments indicated that respondents are “just starting” to work in this area (9.2%) or are working to educate themselves about data management (also 9.2%). One participant noted, “We are in the planning stages, but hope the library staff will be involved.”

One hundred thirty-six respondents provided a total of 333 comments in response to the next question, “What skills do you think science librarians need in order to help scientists with data management?” (See figure 7.) The most frequently cited response category was “knowledge of the data lifecycle,” with 17.1 percent of comments. This was followed by “subject-specific knowledge or skills” (13.8%) and “communication, networking, and reference skills” (13.2%). Other frequently cited response categories were “metadata skills” (10.8%) and “software or computer skills” (9.9%). “Knowledge of the research process” was mentioned in 6.3 percent of the comments.

Many responses recorded for this question did not cite specific skills but instead recorded philosophical comments about librarians’ role in data management. For example, some librarians expressed the opinion that data management duties are a natural extension of the science librarian’s job, whereas others disagreed:

- “I think many of the necessary skills for the roles that make most sense for us to play are ones that librarians already have, including organization and

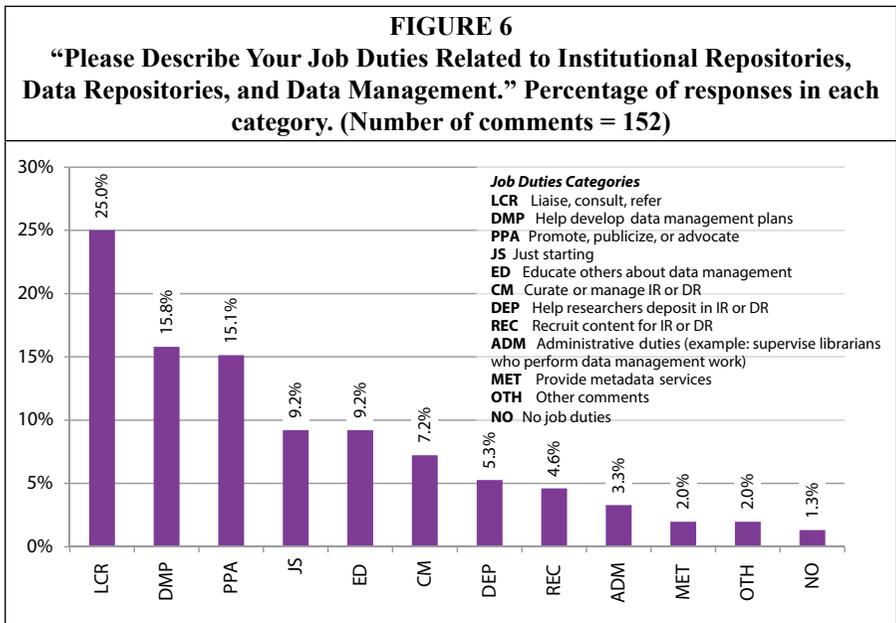
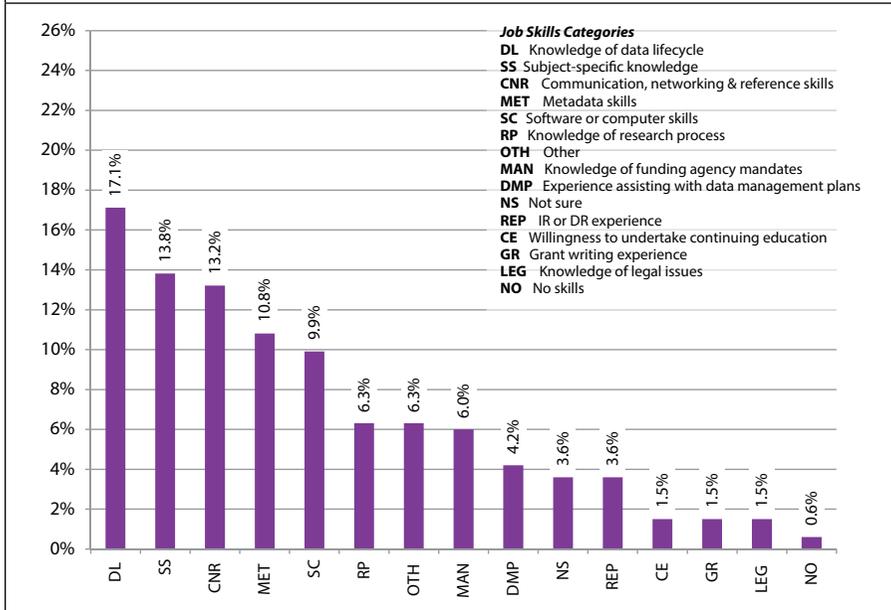


FIGURE 7
“What Skills Do You Think Science Librarians Need in Order to Help Scientists with Data Management?” Percentage of responses in each category. (Number of comments = 333)



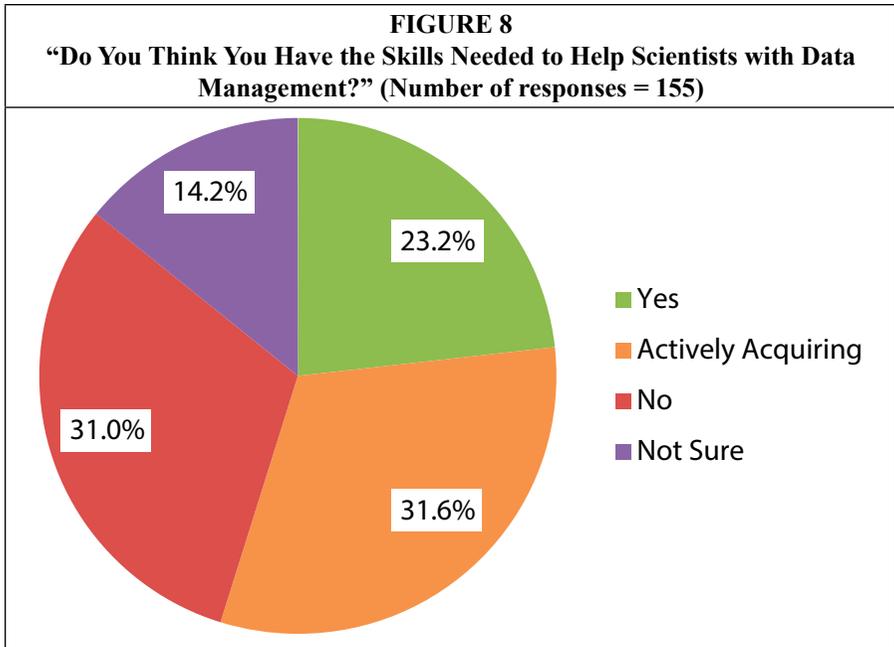
knowing where to find information. I don't think that training librarians in domain-specific metadata is really feasible given the diversity of types of data just within a single field."

- "Much of it is similar to reference skills."
- "I am not convinced that this is an appropriate role for librarians. I believe it should rest with the university research offices."
- "Data management may need to be a separate job from that of science librarian. I think this is a whole different skill set."

Next, respondents were asked, "Do you think you have the skills needed to help scientists with data management?" One hundred fifty-five respondents answered this question. (See figure 8.) For 23.2 percent of respondents, the answer was "Yes," and 31.6 percent answered, "I am actively working to acquire these skills"; 31.0 percent answered, "No," and 14.2 percent answered, "I am not sure."

The results of this question ("Do you think you have the skills needed to help scientists with data management?") were cross-tabulated with the results of other survey questions. This revealed that the respondents most confident in their skills—those who responded "yes" or "I am actively working to acquire these skills"—were also the most likely to work in libraries where at least one library employee was performing data management work. They were also the most likely to know how many other librarians in their institution were assisting researchers with data management, and, unsurprisingly, they were the most likely to report that their job duties included tasks related to IRs, DRs, or data management.

The next survey item asked respondents to "describe the data management skills that [they] have or are acquiring." For this question, 75 respondents provided a total of 214 comments. Of these, 80.0 percent indicated skills that the respondents already



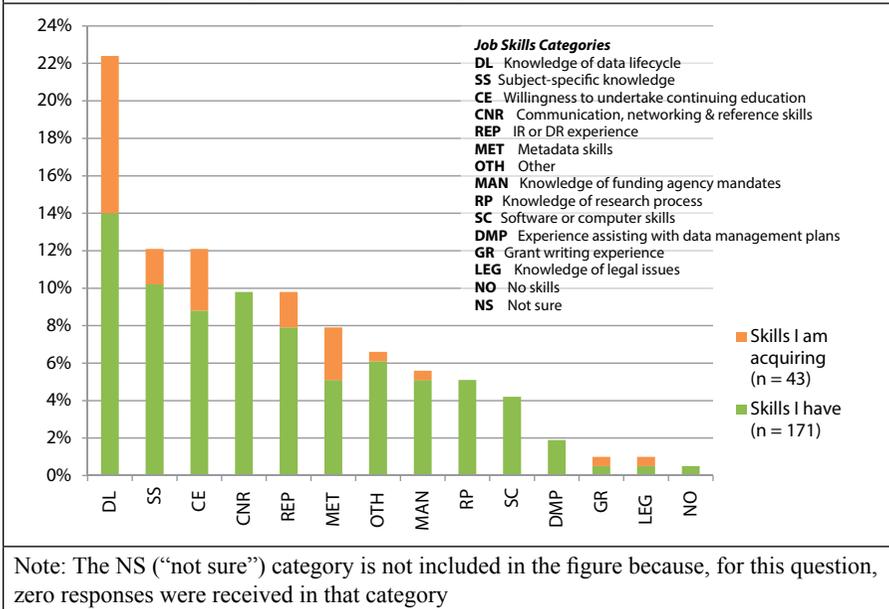
have, and the remaining 20.0 percent mentioned skills that the respondents are currently acquiring. (See figure 9.) By far, the most frequently mentioned skills were in the category “knowledge of the data lifecycle,” with 22.4 percent. This was followed by “subject-specific knowledge or skills” (12.1%), “willingness to undertake continuing education” (also 12.1%), “communication, networking, and reference skills” (9.8%), and “experience working with IRs or DRs” (also 9.8%). Skills mentioned less frequently include “metadata skills” (7.9%), “knowledge of the research process” (5.1%), “software or computer skills” (4.2%), and “experience helping researchers develop data management plans” (1.9%).

Discussion

Although the vast majority of science librarians (94.9%) are aware of the NSF mandate for applicants to submit data management plans, “uncertainty” is perhaps the strongest theme that emerges from the survey results. Substantial percentages of respondents are unfamiliar with the details of data management assistance and initiatives on their own campuses or even within their own libraries. For example, 53.6 percent are unsure about which entity on campus operates the DR, and 31.8 percent are not aware of how many library employees are involved in providing data management support. Furthermore, 15.3 percent are not sure about whether such support is provided on campus, and, among those who say that their campus offers data management support, 25.2 percent do not know which campus entity provides this service. Also, as shown in figure 3, a large proportion of respondents are unsure about how many library employees work with IRs and DRs at their institutions (“not sure” was chosen by 30.8 percent of respondents for IRs and 54.8 percent for DRs). Moreover, respondents’ comments also reflect the theme of uncertainty, as expressed in comments such as the following:

- “Does our institution have an [IR]? It depends on your definition.”
- “I know the Libraries [provide data management assistance]. Other offices may as well, and I’m just not aware.”

FIGURE 9
“Please Describe the Data Management Skills that You Have or Are Acquiring.”
Percentage of responses in each category. (Number of comments = 214)



- “I am not sure [what data management skills I need]. I have not seen this articulated clearly.”
- “I have no idea [what data management skills I need].”
- [With regard to data management skills needed by librarians:] “Unknown training is coming.”
- “I am not sure that science librarians should be required to have [data management] skills. It depends on your definition of a science librarian vs. a data management librarian.”

Perhaps it is no surprise that uncertainty is so prevalent. After all, formal data management is still an emerging field within librarianship: Only a handful of libraries have been involved in data management for more than a few years. In addition, even the definitions of general terms such as “data management,” “data curation,” “institutional repository,” and “data repository” are still in flux.

A second overarching theme relates to the job skills that participants say are necessary for science librarians engaged in data management. Specifically, respondents’ comments about job skills and job duties reflect an interesting disagreement about the kinds of skills librarians will need if they are to provide assistance with data management. On the one hand, many participants indicated that librarians will need specific new skills, such as knowledge of the data lifecycle, computer and software skills, grant-writing expertise, or discipline-specific knowledge. On the other hand, a large number of respondents named more traditional “librarian” skills, such as an understanding of metadata; knowledge of the research process; and the ability to liaise with researchers, refer researchers to appropriate resources, and educate researchers and graduate students about repositories and data management requirements.

The survey did not ask specifically, “Do you think that traditional ‘science librarian’ skills are sufficient preparation for librarians who plan to assist researchers with

data management?" or "Do you think that science librarians need specific continuing education to prepare themselves to take on data management duties?" Therefore, it is impossible to state what the prevailing attitude is about these questions. However, the comments about job skills and duties demonstrate an interesting diversity of opinion. The questions about job skills elicited hundreds of comments expressing a wide range of responses, but with no clear consensus. This in itself is an interesting finding, and it may be related to the "uncertainty" theme mentioned above: Science librarians, as a group, are not certain (yet) about whether they will need significant additional training to be ready to take on data management work. In other words, as noted in the Results section above, science librarians have not come to a consensus about whether the data management role is a natural extension of their jobs, or a set of duties that would be better suited to librarians holding a different job title and assuming a different role within their respective organizations.

However, the sheer number of comments made about job skills and current job duties—748 comments in response to four questions—shows that science librarians at ARL institutions are thinking about data management duties and have opinions about the kinds of skills necessary to enter this new area of academic librarianship, even if they have not formed a consensus. It should be remembered that none of the survey questions required a response; participants were able to skip questions if they desired. The fact that so many participants chose to provide so many responses indicates that the question of job skills for data management is on the collective minds of science librarians at research institutions.

Of course, it is possible that consensus will be neither forthcoming nor necessary: The most likely outcome might well be that some science librarians will perform basic data management work (such as liaising, consulting, and referring) using their existing skills while maintaining their roles as science librarians, while others will undertake additional training and become data management specialists. The recent proliferation of new job titles suggests that this kind of shift might already be occurring: Over the past few years, research libraries have added roles such as "Data Management Librarian" (Oregon State University), "Research Data Librarian" (Cornell University), "Data Curation Librarian" (Northeastern University), "Science Data Librarian" (Stanford University), and "Data Services Specialist" (New York University; Purdue University). It would not be surprising to learn that many such positions have been assumed by former science librarians.

The large number of responses about job skills also raises the question of whether library and information science graduate programs are preparing today's students for the new positions likely to await them upon graduation. A recent study found that only 31 percent of graduate programs offered courses in data management in 2012, and many of these were very new "special topics" courses that had not yet made their way into the regular curriculum.²⁹ However, over the last few years, continuing education opportunities have emerged outside the traditional degree programs. Examples include "Introduction to Data Science," an open online course offered by Syracuse University, and "Applied Data Science: Managing Research Data for Re-Use," a short summer course at the Inter-University Consortium for Political and Social Research, a data repository affiliated with the University of Michigan.

Although this survey gathered information only from science librarians at large research institutions, it is clear that smaller academic libraries also are facing the challenge of providing data management support to their researchers. Sarah Goldstein and Sarah Oelker of Mount Holyoke College, a small liberal arts college in Massachusetts, note that faculty members at their institution are heavily involved in federally funded research. Although they note some "disadvantages for a small college in dealing with

data curation,” such as smaller staffs and budget constraints, they also highlight some characteristics that may help smaller academic libraries provide data management support. For instance, at smaller institutions, libraries are more likely to be merged with information technology services. This can make it easier for personnel to work together flexibly and quickly. In addition, many small colleges already have functional consortial arrangements in place for other library services, such as collection development and preservation, and this existing framework can easily be deployed to apply to cooperative work in data management.³

Conclusion

Data management work in research libraries clearly is still in its emergent phase. Funding agency mandates are relatively new, and only a few research libraries have developed robust strategies to assist researchers with creating data management plans and preparing their data for deposit. However, this study reveals that the vast majority of science librarians are aware of the funding agency mandates, and many of them are actively preparing to do data management work by educating themselves and cultivating new skills. In addition to learning new skills, though, science librarians also report that they are applying their “traditional” librarian skills to data management tasks. Reference skills—the ability to liaise, refer, consult, and teach—are among the competencies that survey respondents cite most frequently as being necessary for science librarians who plan to assist researchers with data management.

As evidenced by the sheer number of comments made by survey respondents, science librarians clearly have given some thought to questions about data management. However, they also express a great deal of uncertainty about many aspects of the data management initiatives under way on their campuses and in the library world. Some of this uncertainty is undoubtedly due simply to the fact that data management roles are new to librarians. However, some of the uncertainty also may stem from a lack of clarity about the roles played by various entities on campus or even by various departments within the library: Between 30 and 55 percent of respondents reported they are “not sure” about which entity on campus operates their university’s DR, which entity provides data management assistance, or even whether library employees are involved in these ventures (see figures 2 and 3). Efforts to increase communication among campus offices and library departments might well be beneficial in reducing librarians’ uncertainty and, more important, in promoting more efficient coordination of data management initiatives.

Thus far, research on the topic of data management in research libraries has been limited mainly to case studies from specific institutions; little has been written about science librarians’ adoption of new skills and new roles. Data management initiatives are creating new opportunities for librarians to engage with researchers on their campuses, but they are also driving important changes in librarians’ career development. This study marks a first step toward understanding the profession’s evolving roles.

Appendix: Survey

1. (Item 1 is the informed consent document.)
2. Are you aware of the National Science Foundation mandate that went into effect in January, 2011, requiring that grant proposals include a two-page Data Management Plan? **Yes** **No**
3. Does your university have an institutional repository?
 Yes **Being planned** **No** **Not sure**
 Comments:
4. What campus entity operates (or will operate) the institutional repository? (Check all that apply.)
 University Libraries University Research Office
 University Information Technology Not sure
 Other (please specify)
 Comments:
5. How many library employees work with the institutional repository?
 0 1–2 3–5 6–10 11 or more Not sure
6. Does your university have a data repository?
 Yes **Being planned** **No** **Not sure**
Comments:
7. Is the data repository part of an institutional repository (or will it be)?
 Yes **No** **Not sure**
 Comments:
8. What campus entity operates (or will operate) the data repository? (Check all that apply.)
 University Libraries University Research Office
 University Information Technology Not sure
 Other (please specify)
 Comments:
9. How many library employees work with the data repository?
 0 1–2 3–5 6–10 11 or more Not sure
10. Does your university provide support to help scientists develop data management plans?
 Yes **Being planned** **No** **Not sure**
 Comments:
11. What campus entities provide (or will provide) support for scientists developing data management plans? (Choose all that apply.)
 University Libraries University Research Office
 University Information Technology Not sure
 Other (please specify)
 Comments:

12. How many library employees provide support for scientists working on data management plans?
 0 1 - 2 3 - 5 6 - 10 11 or more Not sure
13. Does your job include duties related to institutional repositories, data repositories, or data management?
 Yes Being planned No
Comments:
14. Please describe your job duties related to institutional repositories, data repositories, and data management.
Comments:
15. What skills do you think science librarians need in order to help scientists with data management?
Comments:
16. Do you think you have the skills needed to help scientists with data management?
 Yes No
 I am actively working to acquire these skills
 I am not sure
17. Please describe the data management skills that you have or are acquiring.
Comments:

Notes

1. National Science Foundation, "Proposal Preparation Instructions," Chapter II in *Grant Proposal Guide*, available online at www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp [accessed 5 December 2012].
2. James L. Mullins, "Enabling International Access to Scientific Data Sets: Creation of the Distributed Data Curation Center (D2C2)," *Libraries Research Publications*. Paper 85 (2007), available online at http://docs.lib.purdue.edu/lib_research/85 [accessed 5 December 2012].
3. *Ibid.*
4. Michael Witt, "Co-designing, Co-developing, and Co-implementing an Institutional Data Repository Service," *Journal of Library Administration* 52, no. 2 (2012): 172–88.
5. *Ibid.*, 186.
6. Susan Wells Parham, Jon Bodnar, and Sara Fuchs, "Supporting Tomorrow's Research: Assessing Faculty Data Curation Needs at Georgia Tech," *C&RL News* 73, no. 1 (2012): 10–13.
7. *Ibid.*, 13.
8. G. Sayeed Choudhury, "Case Study in Data Curation at Johns Hopkins University," *Library Trends* 57, no. 2 (2008): 211–20.
9. *Ibid.*, 211.
10. *Ibid.*, 215.
11. *Ibid.*, 217.
12. Christie Peters and Anita Riley Dryden, "Assessing the Academic Library's Role in Campus-Wide Research Data Management: A First Step at the University of Houston," *Science & Technology Libraries* 30, no. 4 (2011): 387–403.
13. Elizabeth Yakel, "Digital Curation," *OCLC Systems & Services* 23 no. 4 (2007): 335–40.
14. Patricia Hswe and Ann Holt, "Joining in the Enterprise of Response in the Wake of the NSF Data Management Planning Requirement," *Research Library Issues*, no. 274 (2011): 11–17.
15. *Ibid.*, 16.
16. *Ibid.*, 14.
17. Ellouise Brown, "'I Know What You Researched Last Summer': How Academic Librarians Are Supporting Researchers in the Management of Data Curation," *New Zealand Library & Information Management Journal* 52, no. 1 (2010): 55–69.
18. Gillian Elliot, "Biodiversity Data Management Project: Extending the Boundaries of Information Management in Collaboration with Life Scientists at the University of Otago," *New Zealand Library & Information Management Journal* 51, no. 2 (2009): 104–20.
19. *Ibid.*, 116.
20. Peter Morgan, "Facilitating the Deposit of Experimental Chemistry Data in Institutional Repositories: Project SPECTra (Submission, Preservation, and Exposure of Chemistry Teaching and Research Data)," *IATUL Conference Proceedings* 17 (2007): 1–8.
21. *Ibid.*, 1.
22. *Ibid.*
23. P. Bryan Heidorn, "The Emerging Role of Libraries in Data Curation and E-science," *Journal of Library Administration* 51, no. 7/8 (2011): 662–72.
24. *Ibid.*, 670.
25. Jeremy R. Garritano and Jake R. Carlson, "A Subject Librarian's Guide to Collaborating on e-Science Projects," *Issues in Science and Technology Librarianship* 57 (Spring 2009), available online at www.istl.org/09-spring/refereed2.html [accessed 5 December 2012].
26. Jake R. Carlson and Jeremy R. Garritano, "E-Science, Cyberinfrastructure and the Changing Face of Scholarship: Organizing for New Models of Research Support at the Purdue University Libraries," in *The Expert Library: Staffing, Sustaining, and Advancing the Academic Library in the 21st Century*, eds. Scott Walter and Karen Williams (Chicago: ACRL, 2010), 236.
27. *Ibid.*, 265–66.
28. Dianne Dietrich, Trisha Adamus, Alison Miner, and Gail Steinhart, "De-Mystifying the Data Management Requirements of Research Funders," *Issues in Science and Technology Librarianship* 70 (Summer 2012), available online at www.istl.org/12-summer/refereed1.html [accessed 5 December 2012].
29. Rebecca L. Harris-Pierce and Yan Quan Liu, "Is Data Curation Education at Library and Information Science Schools in North America Adequate?" *New Library World* 113, no. 11/12 (2012): 598–613.
30. Sarah Goldstein and Sarah K. Oelker, "Planning for Data Curation in the Small Liberal Arts College Environment," *Sci-Tech News* 65, no. 3 (2011): 5–11.