

2019-10-22

Research Data Management Among Life Sciences Faculty: Implications for Library Service

Kelly A. Johnson
Cornell University

Et al.

Let us know how access to this document benefits you.

Follow this and additional works at: <https://escholarship.umassmed.edu/jeslib>

 Part of the [Life Sciences Commons](#), [Scholarly Communication Commons](#), and the [Scholarly Publishing Commons](#)

Recommended Citation

Johnson KA, Steeves V. Research Data Management Among Life Sciences Faculty: Implications for Library Service. *Journal of eScience Librarianship* 8(1): e1159. <https://doi.org/10.7191/jeslib.2019.1159>. Retrieved from <https://escholarship.umassmed.edu/jeslib/vol8/iss1/7>

Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in *Journal of eScience Librarianship* by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.



Full-Length Paper

Research Data Management Among Life Sciences Faculty: Implications for Library Service

Kelly A. Johnson¹ and Vicky Steeves²

¹ Cornell University, Ithaca, NY, USA

² New York University, New York City, NY, USA

Abstract

Objective: This paper aims to inform on opportunities for librarians to assist faculty with research data management by examining practices and attitudes among life sciences faculty at a tier one research university.

Methods: The authors issued a survey to estimate actual and perceived research data management needs of New York University (NYU) life sciences faculty in order to understand how the library could best contribute to the research life cycle.

Results: Survey responses indicate that over half of the respondents were aware of publisher and funder mandates, and most are willing to share their data, but many indicated they do not utilize data repositories. Respondents were largely unaware of data services available through the library, but the majority were open to considering such services. Survey results largely mimic those of similar studies, in that storing data (and the subsequent ability to share it) is the most easily recognized barrier to sound data management practices.

Conclusions: At NYU, as with other institutions, the library is not immediately recognized as a valuable partner in managing research output. This study suggests that faculty are largely unaware of, but are open to, existent library services, indicating that immediate outreach efforts should be aimed at promoting them.

Correspondence: Kelly A. Johnson: kaj98@cornell.edu

Keywords: research data management, RDM, library services, data services

Rights and Permissions: Copyright Johnson and Steeves © 2019

Disclosures: The authors report no conflict of interest.



All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Introduction

As scientific research becomes more data-informed and collaborative, the data management practices, needs, and attitudes among research faculty are examined with increased frequency and focus. In conjunction with changing formats and volume of collected data and the emphasis on open access, funding agencies and publishers are demanding evidence of more rigorous consideration and planning for research output (SPARC, n.d.; Diekema, Wesolek, and Walters 2014). The resultant emphasis on stricter data management provides an opportunity for librarians to provide meaningful and substantial services along the entirety of the research life cycle, including educating faculty on the mandates themselves. The importance of preparing for and providing these services is evidenced by the Association of College and Research Libraries' (ACRL) inclusion of Data Curation (2012), Data (2014), and Research Data Services (2016) in their biennial "Top Trends in Academic Libraries," as well as by the growing number of institutional efforts to create them.

Despite the widespread belief that librarians, if given the opportunity to strengthen their familiarity and comfort with best practices, are well positioned to assist with research data management (RDM) (Antell et al. 2014; Brown, Wolski, and Richardson 2015; Tenopir et al. 2013), implementing sustainable and scalable service models is difficult. The most fundamental, but not the only, barrier to service stems from confusion surrounding the definition of data management and its component parts (Carlson 2012; Hoy 2014). However, inconsistencies extend further than a definition of terms. Previous studies evidence wide variation in practices surrounding research data: differences in how data are collected, stored, and shared (Bishoff and Johnston 2015; Kennan and Markauskaite 2015; Tenopir et al. 2011; 2015); among disciplines, institutions, departments, and individuals (Akers and Doty 2013; Buys and Shaw 2015; Weller and Monroe-Gulick 2014); and in researchers' experience, knowledge, and perceptions (Akers and Doty 2012; Jahnke and Asher 2012; Scaramozzino, Ramírez, and McGaughey 2012). When viewed collectively, it is evident that librarians must recognize and address those needs that are relevant to their specific users.

As such, and recognizing that generalizations could not be assumed even among life sciences faculty at other institutions, the authors focused on gathering data specific to their patron base. At the time of writing, the first author had served as the Life Sciences Librarian at New York University (NYU) for approximately two years. During that time she had neither offered nor been asked to assist faculty with any aspect of their research data management, leaving a gap in service to the departments within her purview. The authors designed this study to guide them in strengthening their data management service to life sciences faculty, as well as to prepare them for future research.

Many librarians report a need for additional training to gain knowledge and confidence in providing RDM assistance (Antell et al. 2014; Bresnahan and Johnson 2013), and approaches to service implementation are widely shared (Tenopir, Birch, and Allard 2012). However, the coauthor, who serves as the Librarian for Research Data Management and Reproducibility at NYU, has expertise and comfort with best practices surrounding data management, and is already available to provide these services. The authors are thus less interested in exploring potential services and preparation, and are more focused on identifying researcher needs and understanding why the library's existent RDM services are underutilized.

In order to capture the broad practices, needs, and perceptions of NYU life sciences faculty, and mimicking similar efforts at other institutions (Akers and Doty 2012; 2013; Scaramozzino, Ramírez, and McGaughey 2012; Weller and Monroe-Gulick 2014), the authors surveyed faculty from the Center for Neural Science, the College of Global Public Health, and the Departments of Nutrition & Food Studies, Biology, and Environmental Studies. While respondents were asked to identify their departmental affiliation, there is no effort to identify differences among them. This study serves only as an exploratory examination of the general RDM needs of those departments, in aggregate, under the first author's purview. Information gained from the survey was intended to guide follow-up interviews that would then allow the authors to collect more detailed information about participants' circumstances, further contextualizing individual responses. The survey included questions relating to funder and publisher mandates, data management (including storage, backup, and sharing), and library services (see Appendix A). The results represent a needs assessment of a distinct population at a specific point in time; the authors recognize that the immediate relevancy of this study is largely in planning for further study and effective outreach at their own institution. However, this paper adds to the growing body of literature examining research scientists' data practices and needs that collectively allow general trends and subsequent strategies to emerge.

Literature Review

The White House Office of Science and Technology Policy released the Increasing Access to the Results of Federally Funded Research policy memorandum in February 2013 (Holdren 2013). In it, federal agencies with over \$100 million in annual research and development expenditures are directed to develop "clear and coordinated policies" to ensure public access to publications and digital data from federally funded research. The Open Access Directory website (oad.simmons.edu) maintains a list of publishers that also have begun formalizing policies for open access to their journals and associated raw data. Researchers are tasked with being aware of, understanding, and abiding by all mandates applicable to their research in order to maintain eligibility for funding and publication. While compliance is, at least in theory, relatively straightforward when considering manuscripts, the new and ill-defined standards associated with best data practices "place a burden on researchers to manage, preserve, and share digital data" (Buys and Shaw 2015, 3). Many researchers lack the awareness, time, confidence, or skills (and in some cases the drive) to adequately manage their data, making true compliance unlikely (Bishoff and Johnston 2015; Diekema, Wesolek, and Walters 2014).

In many cases, barriers to controlling data and to developing library services are shared. At a most basic level, "...funding organizations, academic institutions, researchers, and librarians continue to struggle towards a shared vocabulary with commonly understood definitions and to develop strategies to support these new initiatives" (ACRL Research Planning and Review Committee 2014, 294). But even if terminology were uniform, the broad diversity of practices, types, expectations, and attitudes surrounding data management makes standardization difficult. Tenopir et al. (2015) describe differences in data sharing practices and attitudes with relation to age, geographical location, and disciplines across the world, but variation exists across much narrower populations such as institutions and the departments within them.

In order to guide the development of library services, and recognizing that researchers have unique data support needs, institutions have developed and implemented a range of assessment strategies to identify their own patrons' practices and needs. Surveys, historically

the most popular instrument used in library science research (Malliari and Togia 2017), have been effective in this context, as have interviews, focus-groups, literature review, and bibliometric analysis.

Akers and Doty surveyed faculty at Emory University to find that the practices, needs, and perceptions of data management varied among faculty ranks (2012) and disciplines (2013). Through literature review and original research, Research Libraries UK discovered differences in data handling across the cultures underlying disciplines, subjects, and career stages in the UK and beyond (Aukland 2012). Librarians at the University of Minnesota reviewed data management plans (DMP) submitted with NSF grant applications, and found variations in how researchers approach data sharing (Bishoff and Johnston 2015). Herold (2015), also at the University of Minnesota, used recent faculty papers from three branches of the life sciences to track data accessibility, uncovering differences in rates and methods of data sharing. At the University of Kansas, Weller and Monroe-Gulick (2014) found that the types of data and the methods used to collect them varied not only across groups and disciplines, but among individual researchers as well.

To inform development of the library's data services at the Purdue College of Agriculture, a survey was used to identify the types of data researchers were collecting, the tools used to handle them, and the ways in which data literacy skills are learned (Pouchard and Bracke 2016). Librarians at the University of Illinois and Purdue University interviewed faculty from several departments to create Data Curation Profiles (DCPs), representative collections that illustrate how different data types are created and handled across the research life cycle (Witt et al. 2009). Further, a "DCP Toolkit" was created to help "launch discussions between librarians and faculty and aid in the planning of data services that directly address the needs of researchers" ("Data Curation Profiles" n.d.). Williams (2012) examined recent faculty publications to identify types of data collected and sharing practices of crop scientists at the University of Illinois. To better understand their practices and to identify areas for library service, she interviewed those faculty who publicly shared their data, finding that they had ideas for how the library could participate more fully in their data handling (2013). Herold (2015), using methods similar to those of Williams' initial study, discovered that data-sharing practices among ecology, evolution, and natural resources scientists at the University of Minnesota were quite variable, concluding that "data support service providers should not follow a one-size-fits-all approach to outreach" (p. 19). Zhao (2015) also evaluated users' needs without directly asking for their input by analyzing geospatial data requests at McGill University. She was then able to strengthen data discovery instruction and management services.

Researchers commonly report, either directly or indirectly, issues surrounding storage as their most pressing data need. Not only is the amount of available space problematic, researchers struggle with storage practices: selecting a space (e.g. laboratory computers, external hard drives, institutional or other repositories, etc.), metadata application, continued short- and long-term access, sharing, and ease of collaboration. At the California Polytechnic State University, San Luis Obispo (Cal Poly) College of Science and Mathematics, the vast majority of faculty felt personally responsible for managing their own data, but "...appear(ed) to need additional guidance for creating metadata (and) preserving and sharing data" (Scaramozzino, Ramírez, and McGaughey 2012, 361). Buys and Shaw (2015) found that faculty at Northwestern University identified "long term data access and preservation, services for data

storage and backup during active projects, information regarding best data practices...and tools for sharing research” (p. 15) among the data services that would be of most value to them. Westra (2010) described the “sub-optimal” data storage practices among natural sciences faculty at the University of Oregon, noting that “storage is a baseline requirement for other curation activities” (p. 3). These studies are representative, but far from exhaustive.

Scaramozzino et al. (2012) candidly assert that “scientists should not be left to manage digital data on their own” (p.351), and that “academic libraries have a remarkable opportunity to apply traditional strengths toward collecting and organizing digital research content” (p. 350). Bresnahan and Johnson (2013) reported that although there were training needs and anxiety surrounding provision of data management assistance, librarians at CU Boulder recognized the importance of offering such services through the library. Antell et al. (2014) observed that the majority of institutions affiliated with the Association of Research Libraries (ARL) are offering RDM services, and that while science librarians were uncertain of their data management skills they were optimistic about translating traditional librarian skills of “organizing information, applying metadata standards, and providing access to information” (p. 558) into RDM service development. The ACRL Research Planning and Review Committee (2014) stated that “increased emphasis on open data, data-plan management, and ‘big data’ research are creating the impetus for academic institutions...to develop and deploy new initiatives, service units, and resources to meet scholarly needs at various stages of the research process” (p. 294), and many libraries have already begun offering or planning these services. On their webpage dedicated to RDM service resources, the ARL provides links to the “structured comprehensive data management plans” of 12 affiliated libraries (Hswe and Holt n.d.), and in 2013 published a SPEC Kit dedicated to such services (Fearon et al. 2013). In it, they offer access to many services already established at ARL member libraries, including web pages, data archives, DMP tools, data policies, and more. Briney, Goben, and Zilinski (2015) gathered data on library RDM services from 206 research universities in the United States, finding that 50% offered some level of service, and 37% employed a data librarian. A similar 2013 study across Australia, New Zealand, Ireland, and the United Kingdom revealed that the majority of the RDM service types they examined were currently available or were planned at over half of the libraries included in the study (Corrall, Kennan, and Afzal 2013). Of 87 top worldwide research universities evaluated by Si, Xing, Zhuang, Hua, and Zhou (2015), 50 offered research data services. The Virginia Commonwealth University has even created a new position of Director of Research Data Management (Henderson and Knott 2015), and Oregon State University now offers a credit-bearing graduate course in Research Data Management (Whitmire 2015).

Despite their availability, libraries’ RDM services are underutilized. In many cases this is explained by either of two factors: researchers are unaware of these services, or do not view librarians as the best group to provide them. Research provides evidence of both. When she joined a research team at the University of California-Los Angeles, Federer (2013) found that “the researchers were interested in the possibility of having an informationist, (but) most of them had a limited awareness of the types of services that librarians can offer” (p. 299). A key finding of Jahnke and Asher’s (2012) investigation into data curation practices of university researchers was that “few researchers are aware of the data services that the library might be able to provide and seem to regard the library as a dispensary of goods (e.g., books, articles) rather than a locus for real-time research/professional support” (Executive Summary).

In other cases, researchers are aware of library services, but are hesitant to use them. In their study of faculty needs and attitudes at Cal Poly, Scaramozzino et al. (2012) found that while researchers recognize their need for RDM guidance and are open to instruction, they “do not perceive libraries as a source of data management expertise or as the best place to store academic research data” (p 362). Following interviews with biomedical researchers, Read et al. (2015) concluded that “the biggest challenge that libraries face in building data management services is the researchers’ perception that librarians do not understand research data and have no role to play in data management” (p. 133). Atmospheric science and engineering faculty at the University of Illinois Urbana-Champaign evidenced a “disconnect between awareness of services and their actual use” (Wiley and Mischo 2016, sec. Implications and Conclusions para. 3); at the time of their writing, faculty were aware of services but had not yet requested them.

The studies mentioned here join many others in indicating that, across the profession, “libraries must advocate their existing services and infrastructure to their faculty users, while making known the fact that they are in an appropriate position” (Diekema, Wesolek, and Walters 2014, 327) to assist them in managing their data across the research life cycle. It is critical for librarians to identify their patrons’ specific research needs and perceptions as they conceptualize relevant services and the outreach to promote them.

Methodology

Survey Development

The authors aimed to broadly assess the research data management practices of life sciences faculty at New York University. A brief (11 question) and simple (multiple-choice) survey was developed (see Appendix A) and, in order to maximize participation, the authors offered a five dollar donation to the NYU Community Fund for each participant (up to a maximum of \$500). Survey questions were designed based on similar studies (Akers and Doty 2012; 2013; Scaramozzino, Ramírez, and McGaughey 2012; Weller and Monroe-Gulick 2014) and with input from the NYU Qualitative Research Specialist in order to capture easily-interpretable information surrounding researchers’ awareness of funder and publisher mandates, their data management practices, and their likelihood to utilize related services offered through the library. Examples, but no definitions, were included where the authors felt additional context would assist researchers in understanding the question. In order to maximize the number of true participants, the survey was not tested with a subpopulation. Instead, question comprehension and data collection methods were evaluated using library staff who are familiar with the research life cycle.

The final question asked if the respondent were willing to participate in an interview, and was included in preparation for more detailed data collection following the survey. Together, the results of this exploratory survey and follow-up interviews will be used to focus service development and outreach efforts at New York University Division of Libraries.

Target Population

Requests for survey participation were emailed to faculty members served by the first author as NYU’s Life Sciences Librarian, including those affiliated with the Center for Neural Science,

the College of Global Public Health, and the Departments of Nutrition & Food Studies, Biology, and Environmental Studies. Email addresses were collected from each unit's webpage, provided they were available within two links of the faculty listing.

Other than those listed as Emeritus or affiliated with any campus or organization other than NYU's New York City Campus, all faculty were eligible for participation. Assistant, associate, adjunct, and affiliated faculty and faculty fellows were selected to receive the initial survey request. Because many NYU faculty are associated or affiliated with more than one academic department, the 304 collected email addresses were combined and deduplicated. From this list, all email addresses that did not include an NYU domain (e.g., @nyu.edu) were removed, leaving 277 faculty members to receive the first request for survey participation. Following the original call for participation, addresses returning delivery failures or auto-generated indications the faculty member was no longer affiliated with NYU were removed. Two hundred seventy two email addresses were used for all future email correspondence.

Survey Distribution

Three email requests for survey participation were distributed: the initial (July 25, 2016), first reminder (August 2), and final reminder (August 8). The survey was closed on August 10.

Data Collection and Analysis

The authors created and administered the survey using Qualtrics software. Calculations and visualizations for response totals, percentages, and cross-tabulations were done in Python 3 using key libraries such as Pandas and numpy. See Supplementary Materials for a link to the openly available data and analysis code from this study.

The authors recognize several limitations of this study. Survey respondents were self-selected; faculty with current RDM needs and those already familiar with the library or librarians may have been more likely to complete the survey. Using a survey did not allow the authors to provide context or clarify their questions, and the provision of examples within questions may have had the opposite effect of leading or confusing respondents. Closed-ended questions may also have limited the respondents' ability to accurately relay their intended response. Subsequent interviews suggest that survey questions were not misunderstood or misinterpreted, but the small number of interview participants precludes any assumptions of clarity with all respondents.

Results

Requests for survey participation were delivered to 272 faculty members. The survey was initiated by 84 and completed by 79 individuals, for an effective response rate of 29%. Results immediately revealed confusion surrounding participants' unit affiliations. Ten respondents reported no affiliation with any of the units included in the study, despite all having been identified via departmental websites. The authors opted to include these survey responses in their evaluation; while those faculty may not self-identify as affiliates of the life sciences at NYU they had been designated as such, and their RDM practices were likely still representative. Due to inconsistent interpretation, dual appointments across departments, and survey anonymity, it was not possible to calculate response rate by affiliation. Further, while it would

be interesting to explore differences across other parameters such as academic rank, tenure status, or current research efforts, this study was designed to consider responses in aggregate. As a result, and in conjunction with the relatively small number of respondents per unit, we report here the number of responses per category (Table 1) without further analysis surrounding unit affiliation (percentages rounded to nearest integer): Center for Neural Sciences (17, 21%); College of Global Public Health (19, 24%); Department of Biology (15, 19%); Department of Environmental Studies (11, 14%); Department of Environmental Studies and College of Global Public Health (2, 3%); Department of Nutrition and Food Studies (4, 5%); Department of Nutrition and Food Studies and College of Global Public Health (1, 1%); None of These (10, 13%). Because collected responses were considered to represent the survey population as a whole, the authors did not investigate response bias, although they do recognize it as a potential limitation.

Table 1: Respondents' Unit Affiliations

Affiliation	Count
Center for Neural Sciences	17
College of Global Public Health	19
Department of Biology	15
Department of Environmental Studies	11
Department of Environmental Studies, College of Global Public Health	2
Department of Nutrition and Food Studies	4
Department of Nutrition and Food Studies, College of Global Public Health	1
None of These	10
Total	79

Publisher and Funder Mandates

This study aimed to determine only if faculty were aware of whether funder and publisher mandates applied to their research, not their level of familiarity or comfort in satisfying them, and suggests that they are. In response to "Do any of your funders require a data management plan (DMP) with grant applications?," most respondents indicated awareness by selecting "Yes" (45, 57%) or "No" (24, 30%), with few (10, 13%) indicating they were "Unsure."

Similarly, when asked "Do any of your funders or journal publishers require that you share your data upon publishing an associated paper?," the majority selected "Yes" (52, 66%) or "No" (18, 23%), while only 9 (11%) chose "Unsure."

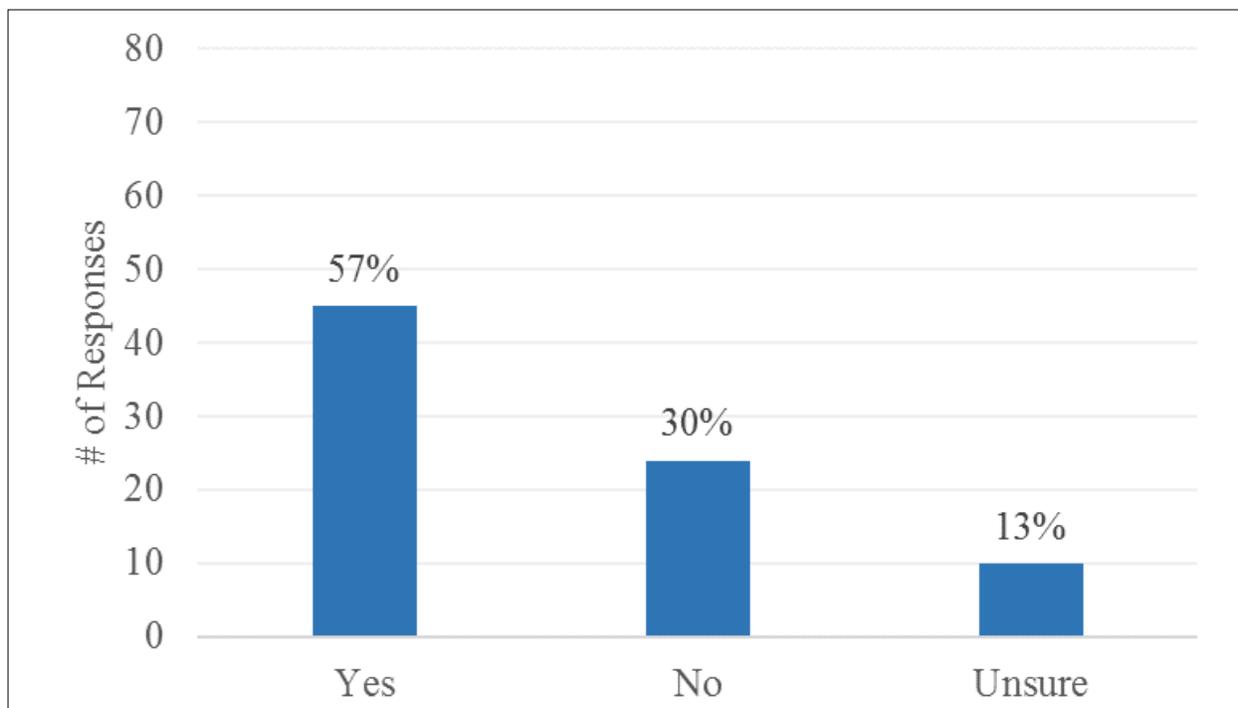


Figure 1: Awareness of DMP mandates (responses to "Do any of your funders require a Data Management Plan (DMP) with grant applications?")

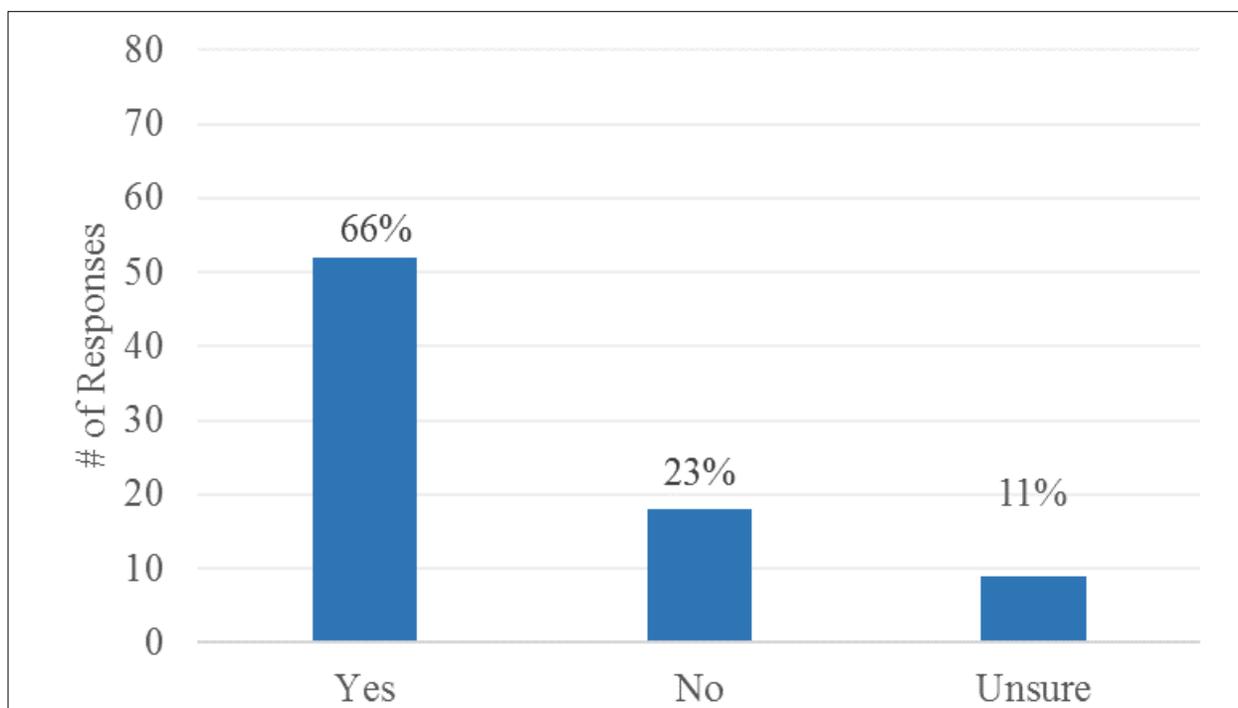
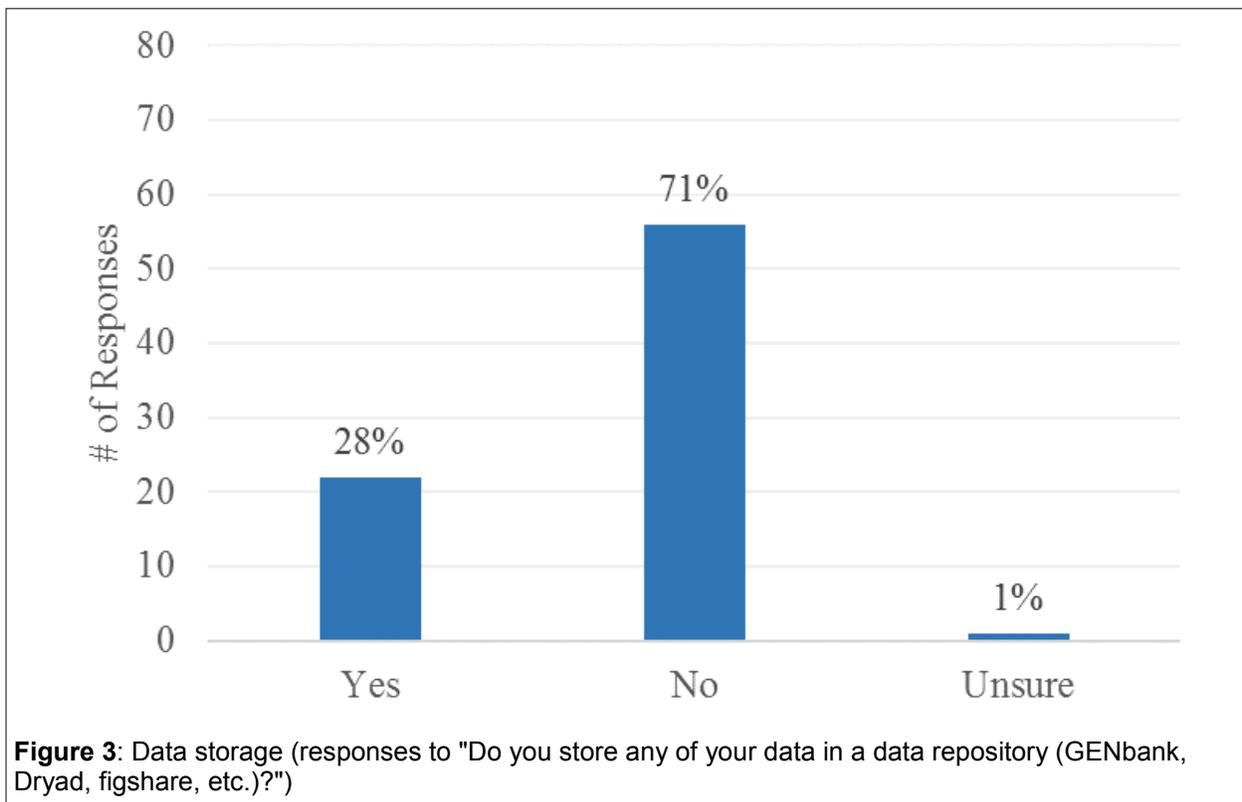


Figure 2: Awareness of data sharing mandates (responses to "Do any of your funders or journal publishers require that you share your data upon publishing an associated paper?")

Data Storage

Based on both their literature review and anecdotal experience, the authors suspected that data storage would prove problematic (and of high concern). Faculty were asked if and where their data were stored (Figure 3) and backed up (Figure 4) in order to identify potential issues and to provide context for future interviews.

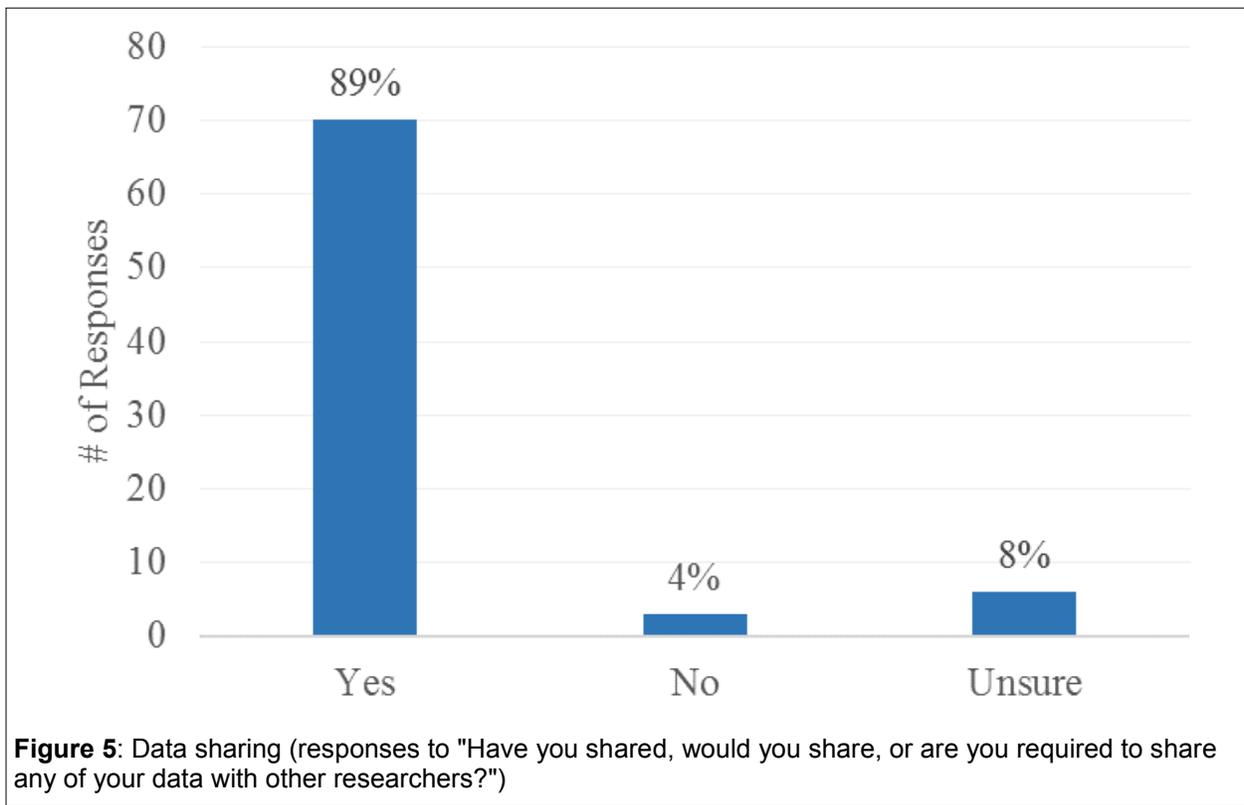
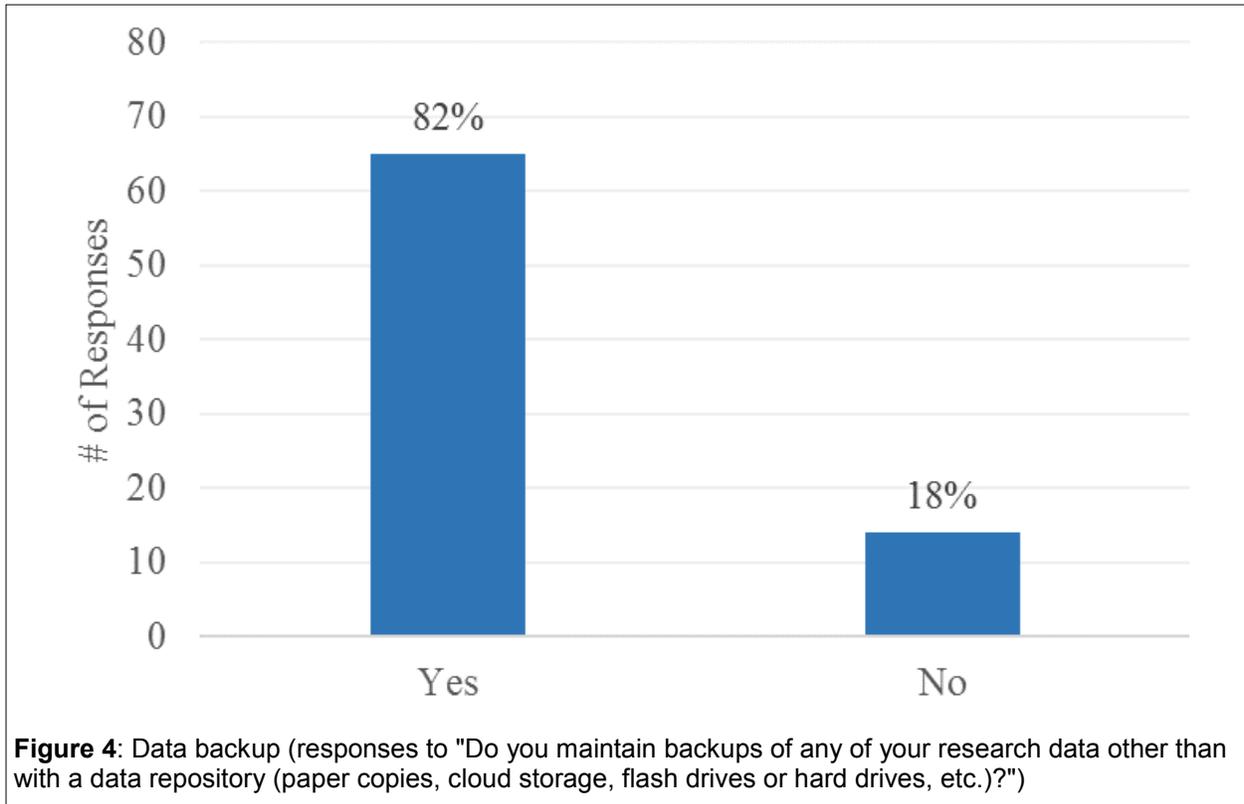
Most researchers indicated that they either did not (56, 71%) or were unsure (1, 1%) if they stored their data in a data repository, while 22 (28%) indicated they did so (Figure 3).



The majority (65, 82%) indicated they did back their data up someplace other than in a data repository, but 14 (18%) did not (Figure 4).

Data Sharing

Here, too, the authors wished to gauge both the risk of mandate non-compliance and general attitudes about sharing research output (Figure 5). Because data accessibility is dependent on adequate storage and findability, managing data is most critical where there is a need or desire to share it. Faculty interest was therefore presumed to correlate with the need for stricter management and accessibility of their data. Most (70, 89%) had shared, would share, or were required to share their data, with only a few indicating they would not (3, 4%) or were unsure (6, 8%).



Current Issues with Data Management

Before exploring whether they recognize the library's potential to assist with data management, the authors wished to determine if faculty recognize such a need exists (Figure 6). Forty three (54%) revealed that they had experienced difficulty with some aspect of managing their data, but 36 (46%) felt they had not.

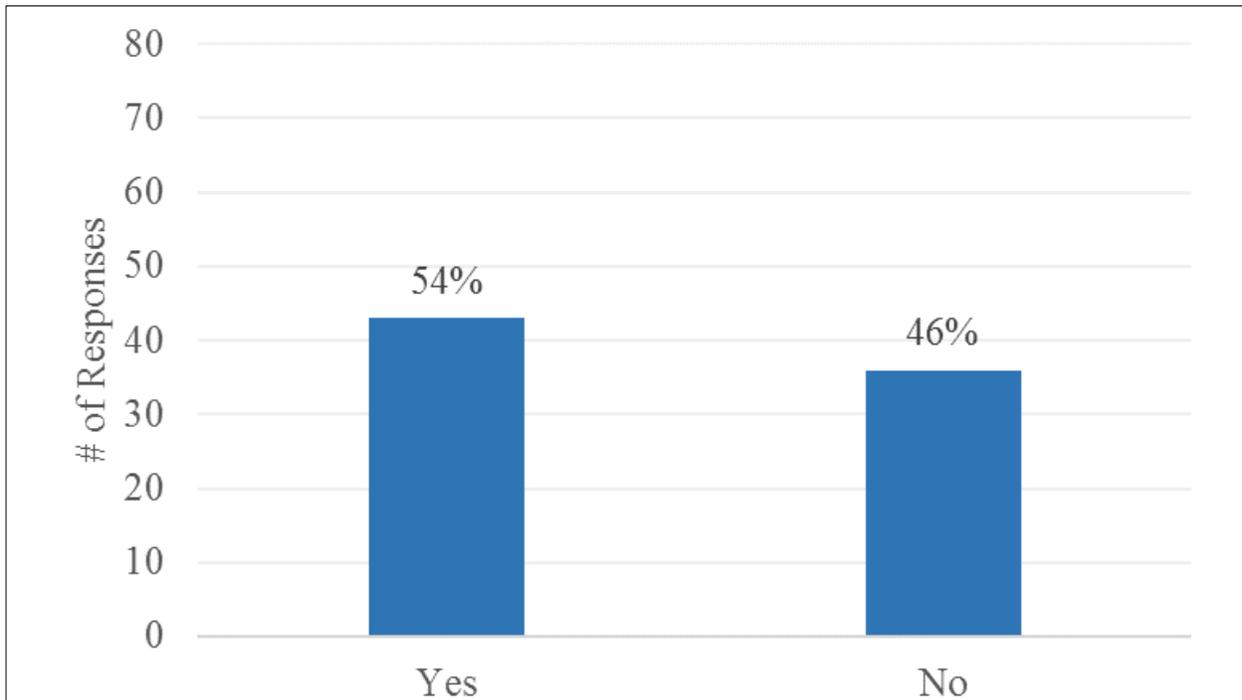


Figure 6: Difficulties with data management (responses to "Have you encountered difficulties with any aspect of data management (writing Data Management Plans, selecting repositories, sharing or reusing your own data, etc.)?")

Library Services

The greatest incentive for completing this study was to determine where and how the library can assist research faculty with their data management needs. Two distinct contributors to successful service were considered when designing the survey: awareness of library services (Figure 7) and confidence that the library can adequately provide them (Figure 8). The vast majority (69, 87%) of respondents were unaware of any RDM services currently offered through the library; only 10 (13%) were familiar with at least one.

It follows that the majority (66, 84%) had not utilized library services, with only 13 (16%) having taken advantage of them.

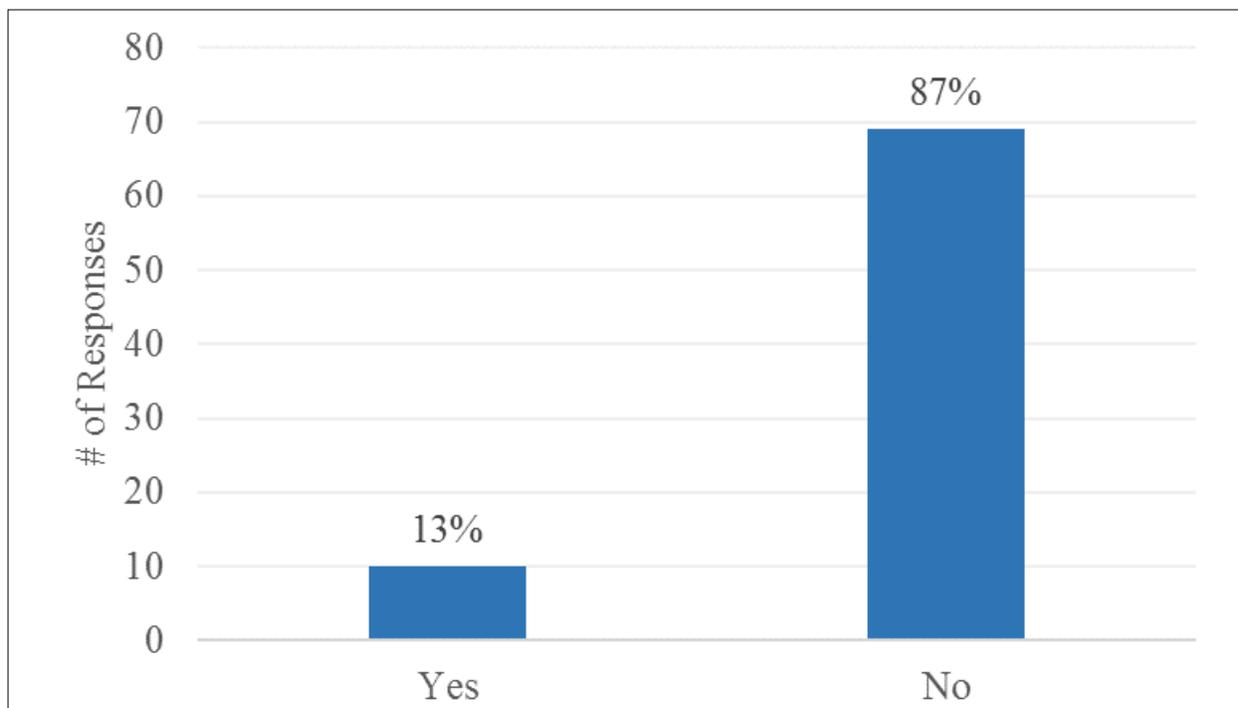


Figure 7: Awareness of library services (responses to “Are you aware of any data management services currently offered through the library?”)

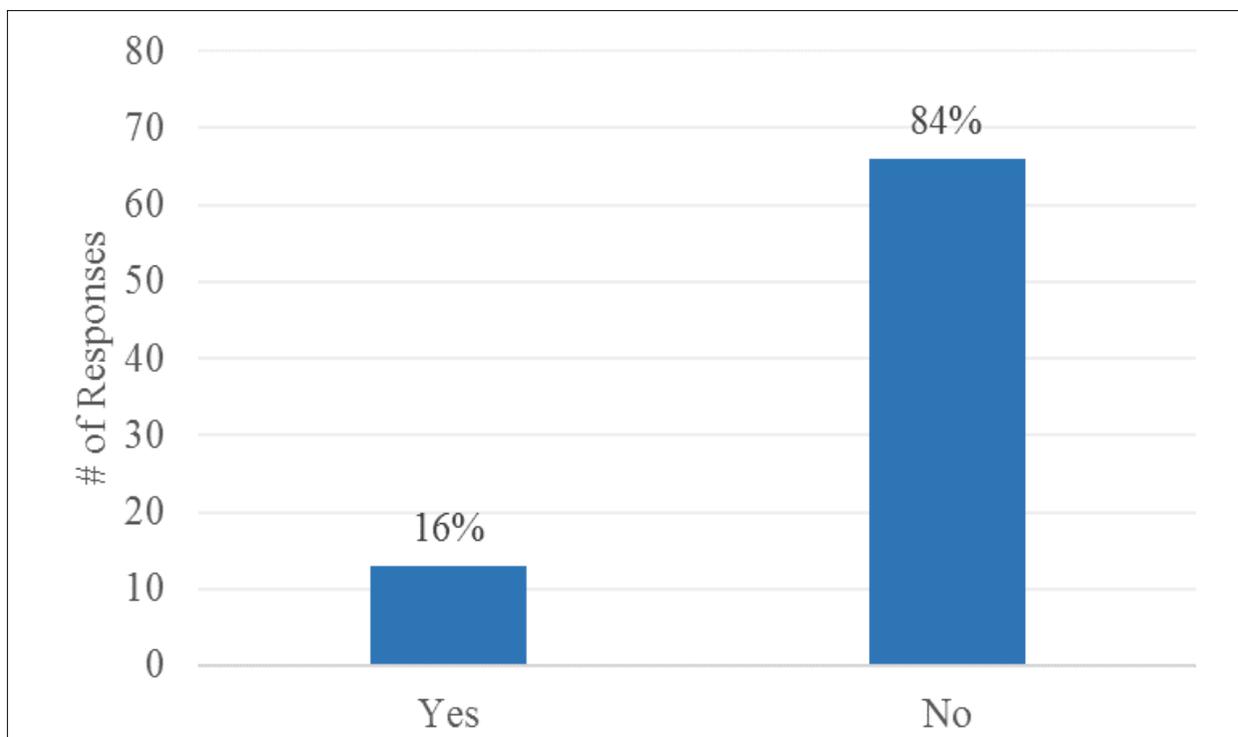
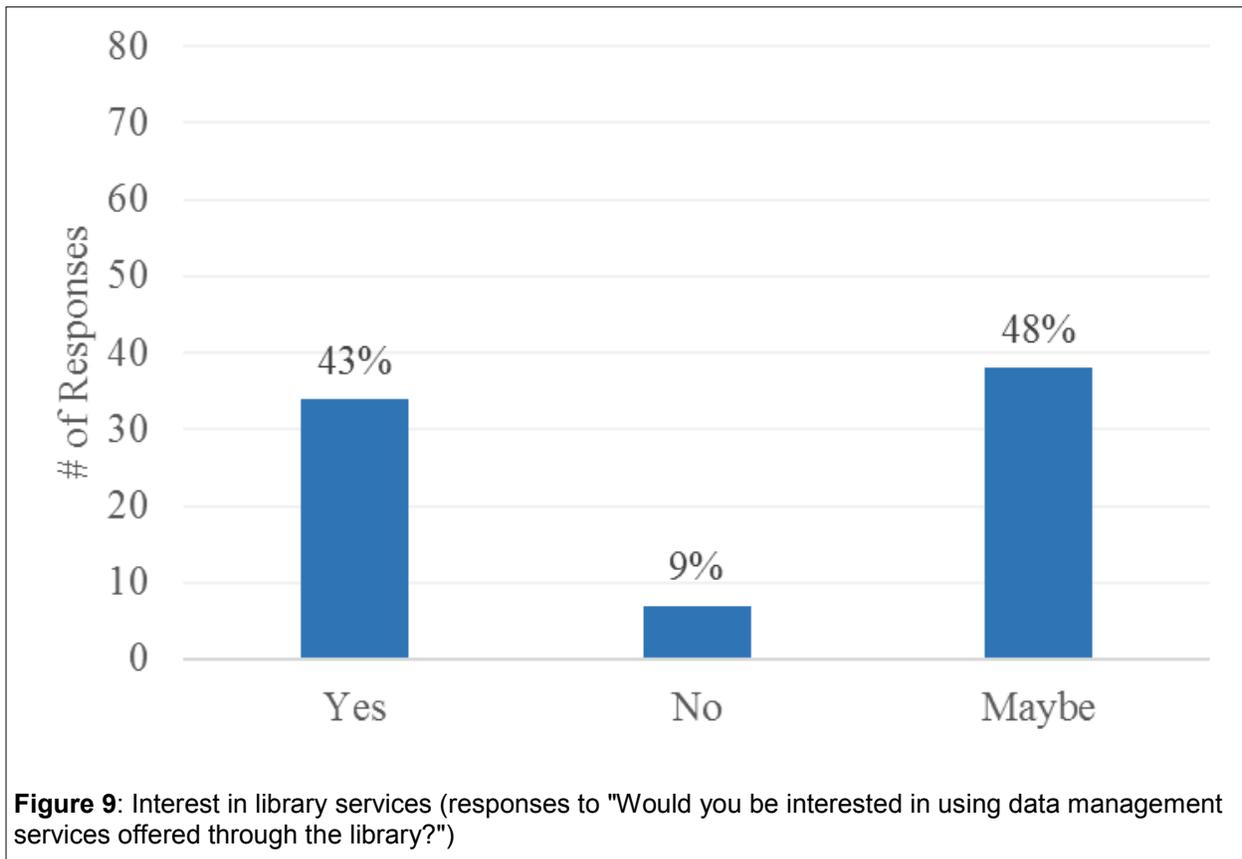


Figure 8: Requests for assistance (responses to "Have you requested assistance through NYU Libraries in preparing or implementing any aspect of your data management?")

However, most indicated they would (34, 43%) or may (38, 48%) be interested in using library services. Seven (9%) were not interested (Figure 9).



Cross-tabulated survey data provide insight into potentially problematic RDM practices as well as opportunities for library assistance.

Mandate Compliance

Of the 52 faculty who answered "yes" to whether they are required to share data, 31 (60%) indicate their data are not stored in a repository, and 1 (1%) indicated they were unsure. Further, 7 (13%) reported they neither utilize a repository nor keep backups, leaving the authors to question if sharing were even possible (Table 2).

It is possible, however, that these discrepancies arise not from poor practices, but from misunderstanding what is meant here by "repository" or "backup." The examples provided within the survey did not reflect the intended breadth of options. For example, institutional repositories were not specifically mentioned, and respondents may not have associated them with the questions as posed.

Also of note, but rare, are those faculty (2, 3%) who demonstrate they may not comply with data sharing requirements, reporting that their funders or publishers require them to share data while indicating that they are "unsure" whether they would do so (Table 3).

Table 2: Data sharing mandates and storage

				Do any of your funders or journal publishers require that you share your data upon publishing an associated paper?			
				Yes	No	Unsure	Total
Do you store any of your data in a data repository (GENBank, Dryad, figshare, etc.)?	Yes	Do you maintain backups of any of your research data other than with a data repository (paper copies, cloud storage, flash drives or hard drives, etc.)?	Yes	19	1	1	21
			No	1	0	0	1
	No		Yes	24	12	7	43
			No	7	5	1	13
	Unsure		Yes	1	0	0	1
			No	0	0	0	0
			Total	52	18	9	79

Table 3: Data sharing mandates and practices

		Do any of your funders or journal publishers require that you share your data upon publishing an associated paper?			
		Yes	No	Unsure	Total
Have you shared, would you share, or are you required to share any of your data with other researchers?	Yes	50	14	6	70
	No	0	3	0	3
	Unsure	2	1	3	6
	Total	52	18	9	79

Data Storage and Sharing Potential

Respondents were largely open to sharing data; only a few indicated they would not be willing (3, 4%) or were unsure (6, 8%). However, as noted above, these data may not be stored in a manner to allow sharing. While 70 respondents have shared or are willing to share data, only 22 (31%) stored their data in a repository, and only 59 (84%) had their data stored or backed

up somewhere. This leaves 11 (16%) respondents who were willing (or required) to share data but were likely unable to do so easily, if at all (Table 4).

Regardless of mandates or willingness to share data, storage is a problem. Thirteen (16%) respondents indicated they neither store their data in a repository nor maintain backups. If those responses reflect actual practices, those researchers are at serious risk for permanent data loss.

Table 4: Data storage and sharing practices

		Do you store any of your data in a data repository (GENBank, Dryad, figshare, etc.)?						
		Yes		No		Unsure		
		Do you maintain backups of any of your research data other than with a data repository (paper copies, cloud storage, flash drives or hard drives, etc.)?						
		Yes	No	Yes	No	Yes	No	Total
Have you shared, would you share, or are you required to share any of your data with other researchers?	Yes	21	1	36	11	1	0	70
	No	0	0	2	1	0	0	3
	Unsure	0	0	5	1	0	0	6
	Total	21	1	43	13	1	0	79

The Library’s Role

Over half of the respondents reported difficulty in managing their own data (43, 54%). However, only 7 (16%) of those have requested assistance through the library (Table 5).

This failure to take advantage of library services provides opportunity for improvement, and further examination leaves room for optimism. The vast majority (69, 87%) of respondents were unaware of library data services, yet 72 (91%) indicated “yes” or “maybe” when asked if they would be interested in using them. Further, of the 10 faculty members who were already aware of library services, 7 (70%) had requested them (Table 6).

These responses indicate tremendous potential to partner more closely with faculty in managing their data, and suggest that the biggest obstacle here is awareness rather than confidence in the library’s capabilities surrounding RDM.

Thirty (38%) respondents agreed to participate in an in-person interview with the authors. At the time of this writing, 15 (19%) had been interviewed, but information collected from interviews had not yet been quantified or analyzed. Other than brief reference to author impressions elsewhere in this paper, interview responses are not reported here.

Table 5: Requests for library assistance with data management issues

		Have you requested assistance through NYU Libraries in preparing or implementing any aspect of your data management?		
		Yes	No	Total
Have you encountered difficulties with any aspect of data management (writing Data Management Plans, selecting repositories, sharing or reusing your own data, etc.)?	Yes	7	36	43
	No	6	30	36
	Total	13	66	79

Table 6: Use of and interest in library services

		Have you requested assistance through NYU Libraries in preparing or implementing any aspect of your data management?						
		Yes			No			
		Would you be interested in using data management services offered through the library?						
		Yes	No	Maybe	Yes	No	Maybe	Total
Are you aware of any data management services currently offered through the library?	Yes	4	0	3	2	1	0	10
	No	3	0	3	25	6	32	69
	Total	7	0	6	27	7	32	79

Discussion

Funding agencies’ research data mandates are largely inconsistent. The Scholarly Publishing and Academic Resources Coalition (SPARC, n.d.) maintains a page dedicated to the sharing policies of federal agencies such as the National Science Foundation (NSF), United States Department of Agriculture (USDA), and National Institutes of Health (NIH), all sources of significant funding for research in the life sciences. The 16 agencies currently listed show

variation in when and how data are to be shared, how metadata are to be applied, and how data are to be preserved, among other things (SPARC, n.d.). What's more, these mandates are fluid. So much so that SPARC also dedicates space to each agency's terms of compliance "looking forward."

With such inconsistency it is understandable that the terms and execution of compliance are sometimes overlooked or misunderstood. Over half of the researchers surveyed here reported that their funders or publishers required a DMP and/or required that research data be publicly shared, but knowledge of mandates does not necessarily indicate they are understood or satisfied. If NYU researcher familiarity mimics that at other research universities (Akers and Doty 2013; Diekema, Wesolek, and Walters 2014; Jahnke and Asher 2012) their knowing that mandates exist is no indication they are prepared to comply with them. Perhaps more dangerous is the suspicion that many researchers are not aware of the consequences of non-compliance, such as ineligibility for future grant funding and delayed or rejected publications. Still, the authors have found both anecdotally and through subsequent interviews that some faculty treat DMPs and sharing requirements as tedious if not frivolous components of their funding. The time and effort necessary to maintain current knowledge and understanding of data requirements may best be handled by librarians, whether because they are better prepared to do so or because they believe such efforts are within their purview. Realistically, the responsibility falls to both researchers and the librarians who serve them as we collectively seek to steward research output in a way that ensures safe storage, continued accessibility, and mandate compliance.

When mandates dictate that research data be made available to others, it follows that those data should be stored in a repository (be it external, institutional, or possibly personal) in order to ensure retrievability. While this survey did not collect anything past rudimentary storage and data sharing practices, data storage among NYU life sciences researchers appears problematic, as it does for other academic institutions (Buys and Shaw 2015; Scaramozzino, Ramírez, and McGaughey 2012; Westra 2010). However, as previously mentioned, survey questions may not have been entirely clear to respondents. It is appropriate to mention here that "data sharing," "data repository," and even "data management" are, at best, nebulous concepts. Researchers may not, for example, have recognized storing their data in an open repository, providing data sets on request, or allowing colleagues to access their laboratory notebooks as sharing, but the authors would consider all to be appropriate examples. As such, these results may not adequately reflect past or future practices. Whether their willingness or reluctance stems from funder or publisher requirements or their own convictions (or, in some cases, issues surrounding confidentiality), most NYU life sciences faculty indicated they are open to sharing data. Accordingly, the authors are confident in saying that many if not most faculty would benefit from learning about realistic storage options to keep data safe and available for future use. One option is the Open Science Framework (OSF), an open-source tool that eliminates many of the concerns and roadblocks surrounding research data (Center for Open Science n.d.; Spies 2013). Users are able to utilize OSF throughout the entirety of the research life cycle, from manuscript writing, collaboration, data storage, and public or restricted access to any portion of their projects. These and other features of OSF significantly reduce the danger of data loss while allowing researchers to share data in whatever ways mandates and personal convictions dictate. Further, the sometimes significant cost of data storage is eliminated. OSF introduction and guidance represent a significant and far-reaching service to faculty, but previous workshops and instruction sessions at NYU (led by two data management

librarians, including one of the authors, who are highly skilled in using OSF) have been minimally attended. The library must find a way to not only provide guidance, but to make faculty aware of such tools while demonstrating their value.

Based on their previous experience with life sciences faculty, the authors were unsurprised that the vast majority of survey respondents were unaware of data services offered through the library. Research libraries often struggle to communicate their availability (and to make the breadth of their services known). According to Buys and Shaw (2015), “the greatest challenge facing many librarians is raising awareness among academic researchers that the library can provide guidance in data management” (p. 20-21). As a faculty researcher herself, Keil (2014) discusses the many data services the library is poised to offer, as well as her being previously unaware of them. While it should not be assumed that faculty are immediately prepared to involve the library once aware of services, it is encouraging that most respondents indicated they either would use or would consider using the library for data assistance. In light of these results, the most immediate strategy to increase service uptake by these departments should focus on outreach. However, the challenge is not in recognizing the need for outreach efforts, but in creating and implementing them successfully.

Needs assessments themselves may serve as effective outreach; Peters and Dryden (2011) report that their pilot study to evaluate data management practices at the University of Houston “turned into a phenomenal outreach opportunity” (p. 398); the researchers hope for the same at NYU. Questioning faculty about their awareness of library services has at least served as an introduction, with some faculty casually expressing their surprise (and pleasure) that such services are available. Follow-up interviews will not only allow the researchers to provide additional context and to collect more robust data, but will also provide for more direct marketing opportunities. It will be interesting to evaluate changes in service uptake following their completion.

As previously mentioned, NYU libraries already offer RDM services including workshops, presentations, consultations, and research guides to inform on RDM practices. In addition to finding effective means of advertising current services, developing new ones may be valuable. A variety of research libraries have initiated programs, or have begun to actively inform their development, and many of these are collected and described in the ARL’s SPEC Kit 334 (Fearon et al. 2013). As librarians continue to define and navigate their role in research data management, it is important to examine the progress and efforts across all institutions. While this study was primarily developed to assess the specific RDM practices of a specific population, it contributes to the wider body of research informing the profession’s approach to RDM services. The survey provided some idea of faculty’s practices and beliefs, but not why they adopted them. Subsequent interviews are necessary to gain a more complete understanding of RDM needs.

Conclusion

Because this exploratory study builds on similar studies, the results were in many ways unsurprising; survey responses demonstrate issues with adequate data storage, widespread difficulty along at least some part of the research life cycle, and an ignorance of library services. However, this project confirmed that the authors must improve and increase the use of data services by NYU life sciences research faculty. Although we are prepared to deliver

data services, we must advertise them. Because previous efforts (such as workshops, presentations, and research guides) have been ineffective in generating robust use of data management librarians, the challenge lies in creating and executing novel outreach efforts.

Fortunately, NYU libraries are well-positioned to provide and increase use of data services for their life sciences faculty, both because they employ two librarians who are prepared to execute them and because their faculty are largely open to library assistance. Issues of and surrounding data storage, in particular, present an immediate opportunity for library involvement, and near-future efforts should focus on increasing the visibility of related services.

Supplemental Content

Appendix A

An online supplement to this article can be found at <http://dx.doi.org/10.7191/jeslib.2019.1159> under “Additional Files”.

Data Availability

All data and supplementary materials are found on the Open Science Framework:

Johnson, Kelly, and Vicky Steeves. 2019. “Research Data Management Among Life Sciences Faculty: Implications for Library Service.” OSF. <https://doi.org/10.17605/OSF.IO/SM22Z>

References

- ACRL Research Planning and Review Committee. 2012. “2012 Top Ten Trends in Academic Libraries: A Review of the Trends and Issues Affecting Academic Libraries in Higher Education.” *College & Research Libraries News* 73(6): 311–320. <https://doi.org/10.5860/crln.73.6.8773>
- . 2014. “Top Ten Trends in Academic Libraries: A Review of the Trends and Issues Affecting Academic Libraries in Higher Education.” *College & Research Libraries News* 75(6): 294–302. <https://doi.org/10.5860/crln.75.6.9137>
- . 2016. “2016 Top Trends in Academic Libraries: A Review of the Trends and Issues Affecting Academic Libraries in Higher Education.” *College & Research Libraries News* 77(6): 274–281. <https://doi.org/10.5860/crln.77.6.9505>
- Akers, Katherine G., and Jennifer Doty. 2012. “Differences among Faculty Ranks in Views on Research Data Management.” *IASSIST Quarterly* 36(2): 16–20. <https://doi.org/10.29173/iq771>
- . 2013. “Disciplinary Differences in Faculty Research Data Management Practices and Perspectives.” *International Journal of Digital Curation* 8(2): 5–26. <https://doi.org/10.2218/ijdc.v8i2.263>
- Antell, Karen, Jody Bales Foote, Jaymie Turner, and Brian Shults. 2014. “Dealing with Data: Science Librarians’ Participation in Data Management at Association of Research Libraries Institutions.” *College & Research Libraries* 75(4): 557–574. <https://doi.org/10.5860/crl.75.4.557>
- Aukland, Mary. 2012. “Re-Skilling for Research: An Investigation into the Role and Skills of Subject and Liaison Librarians Required to Effectively Support the Evolving Information Needs of Researchers.” Research Libraries UK. <http://www.rluk.ac.uk/wp-content/uploads/2014/02/RLUK-Re-skilling.pdf>

- Bishoff, Carolyn, and Lisa Johnston. 2015. "Approaches to Data Sharing: An Analysis of NSF Data Management Plans from a Large Research University." *Journal of Librarianship and Scholarly Communication* 3(2): eP1231. <https://doi.org/10.7710/2162-3309.1231>
- Bresnahan, Megan M., and Andrew M. Johnson. 2013. "Assessing Scholarly Communication and Research Data Training Needs." *Reference Services Review* 41(3): 413–433. <https://doi.org/10.1108/RSR-01-2013-0003>
- Briney, Kristin, Abigail Goben, and Lisa Zilinski. 2015. "Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies." *Journal of Librarianship and Scholarly Communication* 3(2): eP1232. <https://doi.org/10.7710/2162-3309.1232>
- Brown, Rebecca, Malcolm Wolski, and Joanna Richardson. 2015. "Developing New Skills for Research Support Librarians." *Australian Library Journal* 64(3): 224–234. <https://doi.org/10.1080/00049670.2015.1041215>
- Buys, Cunera, and Pamela Shaw. 2015. "Data Management Practices across an Institution: Survey and Report." *Journal of Librarianship and Scholarly Communication* 3(2): eP1225. <https://doi.org/10.7710/2162-3309.1225>
- Carlson, Jake. 2012. "Demystifying the Data Interview: Developing a Foundation for Reference Librarians to Talk with Researchers about Their Data." *Reference Services Review* 40(1): 7–23. <https://doi.org/10.1108/00907321211203603>
- Center for Open Science. n.d. "The Open Science Framework." Center for Open Science. Accessed September 12, 2017. <https://cos.io/our-products/open-science-framework>
- Corrall, Sheila, Mary Anne Kennan, and Waseem Afzal. 2013. "Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research." *Library Trends* 61(3): 636–674. <https://doi.org/10.1353/lib.2013.0005>
- "Data Curation Profiles." n.d. Accessed July 26, 2017. <http://datacurationprofiles.org/overview.php>
- Diekema, Anne, Andrew Wesolek, and Cheryl Walters. 2014. "The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories." *Journal of Academic Librarianship* 40(3/4): 322–331. <https://doi.org/10.1016/j.acalib.2014.04.010>
- Fearon, David Jr, Betsy Gunia, Sherry Lake, Barbara E. Pralle, and Andrew L. Sallans. 2013. "Research Data Management Services, SPEC Kit 334 (July 2013)." 334. <http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334>
- Federer, Lisa. 2013. "The Librarian as Research Informationist: A Case Study." *Journal of the Medical Library Association* 101(4): 298–302. <https://doi.org/10.3163/1536>
- Henderson, Margaret E., and Teresa L. Knott. 2015. "Starting a Research Data Management Program Based in a University Library." *Medical Reference Services Quarterly* 34(1): 47–59. <https://doi.org/10.1080/02763869.2015.986783>
- Herold, Philip. 2015. "Data Sharing among Ecology, Evolution, and Natural Resources Scientists: An Analysis of Selected Publications." *Journal of Librarianship and Scholarly Communication* 3(2): eP1244. <https://doi.org/10.7710/2162-3309.1244>

- Holdren, John P. 2013. "Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research." White House Office of Science and Technology Policy. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- Hoy, Matthew B. 2014. "Big Data: An Introduction for Librarians." *Medical Reference Services Quarterly* 33(3): 320–326. <https://doi.org/10.1080/02763869.2014.925709>
- Hswe, Patricia, and Ann Holt. n.d. "NSF Guide: Resources for Data Management Planning." Association of Research Libraries. n.d. <http://www.arl.org/focus-areas/e-research/data-access-management-and-sharing/nsf-data-sharing-policy/243-resources-for-data-management-planning#.WXz4HoqQyV5>
- Jahnke, Lori M, and Andrew Asher. 2012. "The Problem of Data: Data Management and Curation Practices among University Researchers." *Council on Library and Information Resources* pub 154. <http://www.clir.org/pubs/reports/pub154/problem-of-data>
- Keil, Deborah E. 2014. "Research Data Needs from Academic Libraries: The Perspective of a Faculty Researcher." *Journal of Library Administration* 54(3): 233–240. <https://doi.org/10.1080/01930826.2014.915168>
- Kennan, Mary Anne, and Lina Markauskaite. 2015. "Research Data Management Practices: A Snapshot in Time." *International Journal of Digital Curation* 10(2): 69–95. <https://doi.org/10.2218/ijdc.v10i2.329>
- Malliari, A., and A. Togia. 2017. "An Analysis of Research Strategies of Articles Published in Library Science Journals: The Example of Library and Information Science Research." *Qualitative and Quantitative Methods in Libraries* 5(4): 805–818. <http://qqml-journal.net/index.php/qqml/article/view/6>
- Peters, Christie, and Anita Riley Dryden. 2011. "Assessing the Academic Library's Role in Campus-Wide Research Data Management: A First Step at the University of Houston." *Science & Technology Libraries* 30(4): 387–403. <https://doi.org/10.1080/0194262X.2011.626340>
- Pouchard, Line, and Marianne Stowell Bracke. 2016. "An Analysis of Selected Data Practices: A Case Study of the Purdue College of Agriculture." *Issues in Science and Technology Librarianship* 85(1). <https://doi.org/10.5062/F4057CX4>
- Read, Kevin B., Alisa Surkis, Catherine Larson, Aileen McCrillis, Alice Graff, Joey Nicholson, and Juanchan Xu. 2015. "Starting the Data Conversation: Informing Data Services at an Academic Health Sciences Library." *Journal of the Medical Library Association* 103(3): 131–135 5p. <https://doi.org/10.3163/1536-5050.103.3.005>
- Scaramozzino, Jeanine Marie, Marisa L. Ramírez, and Karen J. McGaughey. 2012. "A Study of Faculty Data Curation Behaviors and Attitudes at a Teaching-Centered University." *College & Research Libraries* 73(4): 349–365. <https://doi.org/10.5860/crl-255>
- Si, Li, Wenming Xing, Xiaozhe Zhuang, Xiaoqin Hua, and Limei Zhou. 2015. "Investigation and Analysis of Research Data Services in University Libraries." *The Electronic Library* 33(3): 417–449. <https://doi.org/10.1108/EL-07-2013-0130>
- SPARC. n.d. "Browse Data Sharing Requirements by Federal Agency." SPARC: Setting the Default to Open in Research and Education. <http://datasharing.sparcopen.org/data>
- Spies, Jeffrey R. 2013. "The Open Science Framework: Improving Science by Making It Open and Accessible." Doctoral dissertation, University of Virginia. <https://thesiscommons.org/t23za>

- Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. "Data Sharing by Scientists: Practices and Perceptions." Edited by Cameron Neylon. *PLoS ONE* 6(6): e21101. <https://doi.org/10.1371/journal.pone.0021101>
- Tenopir, Carol, Ben Birch, and Suzie Allard. 2012. "Academic Libraries and Research Data Services." *Current Practices and Plans for the Future*. Chicago, IL: Association of College and Research Libraries. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.248.3251&rep=rep1&type=pdf>
- Tenopir, Carol, Elizabeth D. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. 2015. "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide." *PLoS ONE* 10(8): e0134826. <http://dx.doi.org/10.1371/journal.pone.0134826>
- Tenopir, Carol, Robert J. Sandusky, Suzie Allard, and Ben Birch. 2013. "Academic Librarians and Research Data Services: Preparation and Attitudes." *IFLA Journal* 39(1): 70–78. <https://doi.org/10.1177/0340035212473089>
- Weller, Travis, and Amalia Monroe-Gulick. 2014. "Understanding Methodological and Disciplinary Differences in the Data Practices of Academic Researchers." *Library Hi Tech* 32(3): 467–482. <https://doi.org/10.1108/LHT-02-2014-0021>
- Westra, Brian. 2010. "Data Services for the Sciences: A Needs Assessment." *Ariadne*, no. 64 (June). <http://www.ariadne.ac.uk/issue64/westra>
- Whitmire, Amanda. 2015. "Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lessons Learned." *Journal of Librarianship and Scholarly Communication* 3(2): eP1246. <https://doi.org/10.7710/2162-3309.1246>
- Wiley, Christie, and William Mischo. 2016. "Data Management Practices and Perspectives of Atmospheric Scientists and Engineering Faculty." *Issues in Science and Technology Librarianship* 85(1). <https://doi.org/10.5062/F43X84NJ>
- Williams, Sarah C. 2012. "Data Practices in the Crop Sciences: A Review of Selected Faculty Publications." *Journal of Agricultural & Food Information* 13(4): 308–325. <https://doi.org/10.1080/10496505.2012.717846>
- . 2013. "Data Sharing Interviews with Crop Sciences Faculty: Why They Share Data and How the Library Can Help." *Issues in Science and Technology Librarianship* 72(1). <https://doi.org/10.5062/F4T151M8>
- Witt, Michael, Jacob Carlson, D. Scott Brandt, and Melissa H. Cragin. 2009. "Constructing Data Curation Profiles." *International Journal of Digital Curation* 4(3): 93–103. <https://doi.org/10.2218/ijdc.v4i3.117>
- Zhao, Jennifer Congyan. 2015. "Identifying User Needs: A Case Study on Geospatial Data Requests at McGill University." *Journal of Map & Geography Libraries* 11(1): 37–59. <https://doi.org/10.1080/15420353.2015.1010057>