



Full-Length Paper

Research Data Services in Academic Libraries: Data Intensive Roles for the Future?

Carol Tenopir¹, Dane Hughes¹, Suzie Allard¹, Mike Frame², Ben Birch¹, Lynn Baird³, Robert Sandusky⁴, Madison Langseth², and Andrew Lundeen⁵

¹University of Tennessee, Knoxville, TN, USA

²United States Geological Survey, Oak Ridge, TN, USA

³University of Idaho, Moscow, ID, USA

⁴University of Illinois at Chicago, Chicago, IL, USA

⁵Michigan State University, East Lansing, MI, USA

Abstract

Objectives: The primary objectives of this study are to gauge the various levels of Research Data Service academic libraries provide based on demographic factors, gauging RDS growth since 2011, and what obstacles may prevent expansion or growth of services.

Methods: Survey of academic institutions through stratified random sample of ACRL library directors across the U.S. and Canada. Frequencies and chi-square analysis were applied, with some responses grouped into broader categories for analysis.

Results: Minimal to no change for what services were offered between survey years, and interviews with library directors were conducted to help explain this lack of change.

Conclusion: Further analysis is forthcoming for a librarians study to help explain possible discrepancies in organizational objectives and librarian sentiments of RDS.

Correspondence: Carol Tenopir: ctenopir@utk.edu

Keywords: research data service, data management, libraries and data sets



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

The growth of data intensive science, coupled with funding mandates for data management plans and government open data, has led to a growing emphasis on data management across all academic disciplines. At the same time, the roles of academic libraries have changed dramatically within the last decade. Academic librarians are now often integrated as partners in all aspects of the research process, from data collection to publication and preservation of research output. A suite of research data services (RDS) needed by academic communities is emerging in academic libraries in response to the growth of data intensive research, changing roles of libraries, and the recognition of a need for research data management.

In 2011, an assessment of research data services (RDS) in North American academic libraries by the Usability & Assessment Working Group of the NSF-funded DataONE project found that many research academic libraries have added research data services and many more were planning to offer a variety of research data services in the future. The respondents came from a panel of library directors from all types of academic libraries in Canada and the United States that was put together by the Association of College & Research Libraries (ACRL). Results from the 2011 baseline survey were published as an ACRL white paper (Tenopir, Birch & Allard 2012). In order to measure changes in practice over time, a second survey was planned at the outset and conducted in 2014 and reported here. There have been other assessments of the role that libraries play in providing RDS; but few have been conducted with the intent of long-term, periodic assessment of RDS within libraries (Soehner, Steeves & Ward 2010; Antell et al. 2013).

These assessments of RDS in libraries are an integral part of the DataONE project (dataone.org), which has a mission to ensure the preservation and accessibility of science data. DataONE is multi-scale, multi-national, and cross-disciplinary in scope, and provides cyberinfrastructure for data discovery, as well as tools and educational materials to help build a culture and capability of research data management. One of the tasks of the Usability & Assessment Working group of DataONE is to build an understanding of perceptions, attitudes, behaviors, and requirements with respect to data and data activities within stakeholder communities (Tenopir, Birch & Allard 2012). Academic libraries are an important stakeholder community who are playing a part in building a culture and infrastructure for RDS.

The second assessment of RDS within academic libraries sought to address and answer three main research questions:

RQ1: What RDS are offered in academic libraries in the U.S. and Canada in 2014, and are there different levels of service based on demographic factors such as student population, number of tenure track faculty, and amounts of external funding?

RQ2: Have RDS increased among academic libraries since 2011 and will they increase in the future?

RQ3: What obstacles may prevent expansion or growth of services?

Note that the term data “curation” is often used to define the processes involved in managing data throughout the data life cycle, from data collection, to description, and preservation

(Soehner, Steeves & Ward 2010)¹. Except where we are describing the terms used by other authors in their research, in this article we use the term data “management” to refer to the broad suite of services or processes involving data, including services that assist with data management planning, finding repositories for both accessing and depositing data, metadata description, and preservation.

Related Research

Recognition of the need for good data management is now widespread, and the resource requirements to accomplish these tasks are broadly discussed in scholarly and popular literature (Tenopir et al. 2011). As is the case with other cultural and scientific artifacts, academic libraries can play a role in helping researchers find, describe, and preserve data, and to help implement these data management requirements. Some believe that libraries have the mandate and may have the capacity to curate data to serve the advancement of science and society as a whole (Heidorn 2011). The Association of Research Libraries (ARL) E-Science Working Group conducted a survey in 2010 to determine how ARL libraries have been approaching this task of providing RDS to their patrons (Soehner, Steeves & Ward 2010). Although the majority of institutions lacked designated departments for providing RDS, over one-third had conducted assessments of their researchers’ needs for data services. The DataONE 2011 baseline assessment of ACRL library directors identified that well under half of libraries surveyed offered some form of RDS; however, more libraries were planning to offer such services within the next two years (Tenopir, Birch & Allard 2012). Most of the libraries already offering RDS were in research-focused/PhD-granting institutions, such as those who are likely members of ARL.

Institutional assessments of RDS needs are plentiful and continue to be performed and published. Many libraries have conducted surveys, (Parsons, Grimshaw & Williamson 2013; Parham et al. 2012) while others have taken more personal approaches, such as interviews and focus groups (Carlson 2012; Mclure et al. 2014). Regardless of assessment methods employed, academic librarians are seeking to understand the needs of the researchers at their institutions and develop targeted strategies to meet those needs. Studies show that researchers’ needs vary from discipline to discipline, so one solution may not be sufficient (Akers & Doty 2013; Weller & Monroe-Gulick 2014).

Libraries are also discovering the need for workforce development and training regarding RDS. The baseline assessment discovered that most libraries are shifting current staff into data positions instead of hiring new data professionals (Tenopir, Birch & Allard 2012). A recent survey of ARL science librarians found that less than one-quarter of these librarians felt that they have the skills to help scientists with data management (Antell et al. 2013). Training for current staff may be on-the-job, or more formalized (MacMillan 2015; Tenopir et al. 2013). Library and information science (LIS) programs are also developing courses and programs to provide information professionals with a solid foundation in data management and curation (Varvel, Bammerlin & Palmer 2012).

In some scientific disciplines, data managers emerged decades ago to support the need to organize, preserve, and curate data on behalf of field and bench researchers (Tayi & Ballou

¹ <https://www.dataone.org/best-practices>

1998; Gray et al. 2005); however, it is only more recently that academic libraries have recognized the need to have staff with research data skillsets (Mayernik et al. 2014). Needless to say, librarians with the full spectrum of RDS skills may be difficult for academic libraries to find and retain. Collaboration with groups outside of the library or with other institutions will be essential for fully developing RDS for an academic community (Norman & Stanton 2014; Deards 2013). However, the 2010 ARL survey found that just under half of responding libraries had built collaboration with RDS at other institutions (Soehner, Steeves & Ward 2010). Data generated by academic researchers are varied and require a wide range of knowledge about data and metadata standards (Heidorn 2011). Collaborations are also recommended for developing stable cyberinfrastructure to support data curation in the long term (Tansley & Tolle 2009).

Methodology

The survey instrument was built in the Qualtrics software and housed on the servers at the University of Tennessee, Knoxville. The survey, approved by the University of Tennessee Institutional Review Board for Human Subjects, was administered to a stratified random sample panel of ACRL library directors across the United States and Canada. Three hundred and fifty invitations to participate with a link to the questionnaire were sent on March 11, 2014 via Qualtrics Mailer. Thirteen email addresses were unreachable, yielding a final total of 337 valid invitations. The survey closed on April 12, 2014 with 146 responses, 128 of which were usable, for a 38% response rate. The participants were asked to respond on behalf of their institutions.

Responses were imported from Qualtrics and analyzed in SPSS. In order to run chi-square analysis, survey questions asking for several date choices of when future services were planned were recoded into “yes,” “no, but plan to,” and “no” responses. Some demographic choices were recoded and consolidated in a similar manner, such as full time equivalent (FTE) recoded into “fewer than 5,000 students,” and “5,000 or more students,” and tenure/tenure track faculty (fewer than 100 faculty/100 faculty or more).

The questionnaire included demographic questions as well as a range of items on current and planned RDS and workforce development (see Appendix A).

To better understand the survey results, semi-structured interviews of five academic library directors were conducted after the survey. Respondents from the ACRL panel were invited to participate based on how knowledgeable they were about RDS. Responses were imported into NVivo software to help determine common patterns in interview responses.

Results

Overall Results

Of the identifiable institution types, over half of the responding institutions (53.7% of n=128) are research universities. Two-year institutions represented 17.9% of identifiable institution types, while four-year institutions represented 28.4%.

Well over half (65%) of the respondents have fewer than 5,000 full-time equivalent students (Table 1), and 43% of institutions have fewer than 100 tenure or tenure-track faculty under their employment (Table 2). Three-quarters of campuses (76%) receive less than \$10 million in external research grants, with noticeably fewer receiving more than \$10 million annually (Table 3).

Table 1: Frequencies and percentages for survey participants by full time equivalent (FTE) stu-

FTE Students	Frequency (Percent)
Up to 1,999	41 (32.5%)
2,000-4,999	41 (32.5%)
5,000-9,999	18 (14.3%)
10,000-24,999	16 (12.7%)
25,000 or more	10 (7.9%)
Total	126 (100%)

dents

Table 2: Frequencies and percentages for survey participants by the number of tenure-track and

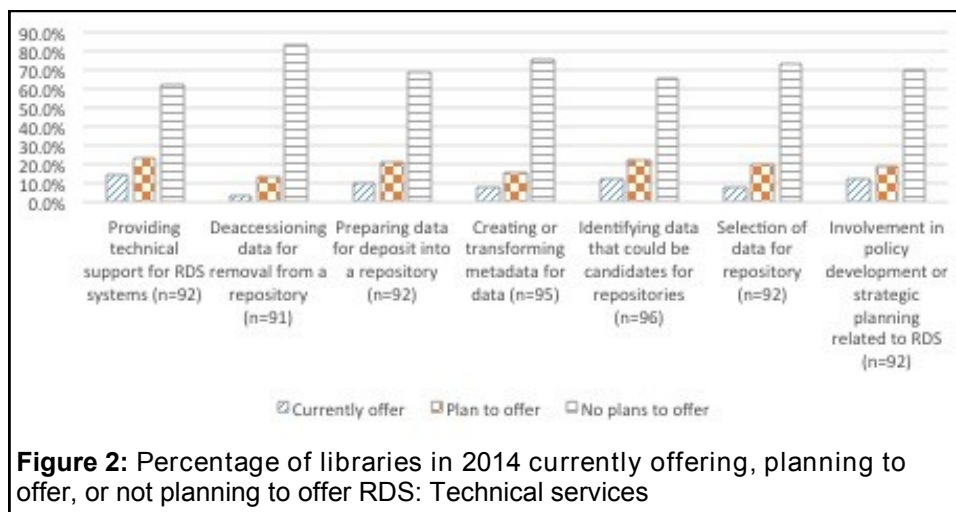
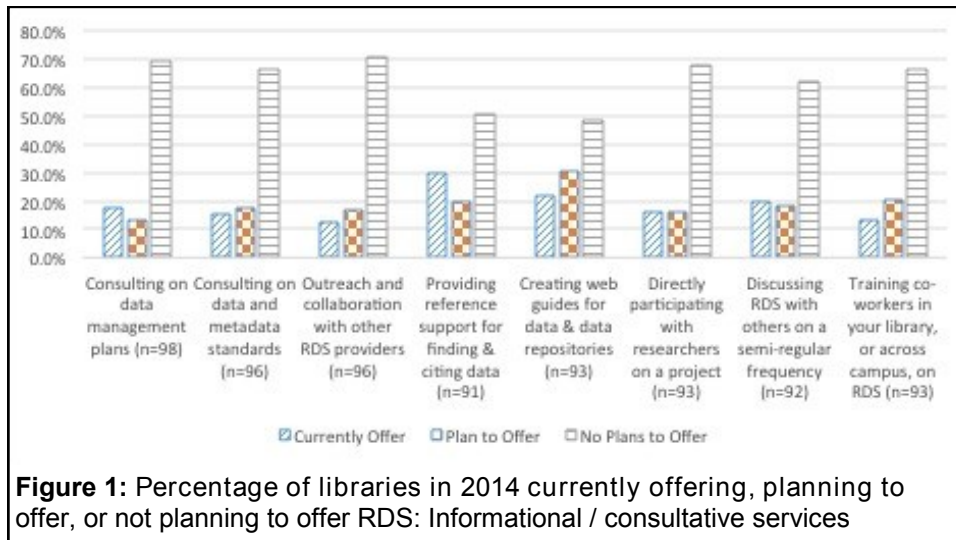
Tenure Track/Tenure Faculty Employed	Frequency (Percent)
Less than 100	53 (43.4%)
100-249	38 (31.1%)
250-499	13 (10.7%)
500-999	9 (7.4%)
1,000+	9 (7.4%)
Total	122 (100%)

tenured faculty employed

Approx. Value	Frequency (Percent)
Less than \$1 million	22 (37.9%)
\$1-9.9 million	22 (37.9%)
\$10-49.9 million	2 (3.4%)
\$50-100 million	4 (6.9%)
More than \$100 million	8 (13.8%)
Total	58 (100%)

Table 3: Approximate value for all external research grants awarded on campus annually

The majority of institutional respondents do not currently offer, nor do they plan to offer, most types of RDS (Figure 1 & Figure 2). In general, more institutions currently offer or plan to offer what could be defined as informational /consultative services (Figure 1) rather than what could be defined as technical services (Figure 2). Providing reference support for finding and citing data (29.7%) and creating web guides for data and data repositories (21.5%) are the most frequently offered services. This is not surprising, as consultative RDS align well with traditional reference or liaison librarian services.



The survey also gathered information about *who in the library, if anyone, provides reference/consultation/instructional RDS to researchers* and *who has primary leadership responsibility for plans and programs for RDS*. Individual discipline librarians (i.e., subject specialists or liaison librarians) or staff members are the largest group providing RDS informational/consultative services in the responding libraries (Table 4). Only a small percentage (6.7%) of libraries have dedicated data librarians to provide RDS.

Almost three-quarters of respondents indicated that their library is not involved in RDS. Of those libraries that are involved in RDS, responsibility for planning and developing programs differs from library to library. Some libraries have a single individual who is responsible, some have a group or committee that is responsible, and others have a combination of individuals, committees, and/or departments responsible for RDS planning (Table 5).

Table 4: Who in the library provides research data reference/consultation/instruction services to researchers?

	Percentage (n=90)
Individual Discipline Librarians / Staff	61.1%
Dedicated Data Librarian(s) / Specialist(s)	6.7%
Other (please specify)	32.2%
<i>Various Staff or Others</i>	10.0%
<i>Reference, Instructional, or Information Services Librarian</i>	6.7%
<i>No One</i>	14.4%
<i>None specified</i>	1.1%

Table 5: If your library is involved in any RDS, who in the library has primary leadership responsibility for plans and programs for RDS? (Q11 in Academic_Libraries_RDSFollow_Up_2014 SPSS data

	Percentage (n=89)
A Single Individual	7.9%
A Group/Committee/Team	6.7%
A Department/Unit	1.1%
A Combination of the Above	10.1%
Other	1.1%
My Library is not Involved in RDS	73.0%

set)

Respondents were also asked *how their libraries have developed staff capacity for RDS*. A little over half of the 25% of those offering RDS (54.5% or 12 of 22) indicated that they have reassigned existing staff (Table 6). Only a few (6) have hired new staff.

	Frequency	Percentage
Hired staff specifically to support RDS	6	27.2%
Reassigned existing staff	12	54.5%
Planning to hire staff	6	27.2%
Planning to reassign existing staff	6	27.2%
Other	7	31.8%

Table 6: How has your library developed staff capacity for RDS? (Check all that apply). n=22
Although their libraries may not yet offer RDS, many of the library directors agree that issues of research data are important. Respondents were asked to report how much they agree or disagree with a number of statements relating to their opinion on library involvement in RDS (Table 7). Level of agreement was based on a five-point Likert scale where one equals strongly agree, to five, strongly disagree. The majority of respondents agree or strongly agree that *losing data jeopardizes the future of scholarship* (59.3%) and that *librarians should be stewards of all types of scholarship, including data sets* (75.6%). Nearly half of the respondents (46.5%) believe the *library needs to offer RDS to remain relevant to the institution*; however, only about a third (36.1%) of respondents agree or strongly agree their institution’s *researchers will be at a disadvantage for grants if the library does not provide RDS*. Only a few (10.4%) believe their *library will see decreases in funding if it does not offer RDS*.

	Mean Rating (S.D.)	Strongly Agree & Agree	Strongly Disagree & Disagree	Neither Agree nor Disagree
The library needs to offer research data services (RDS) to remain relevant to the institution.	2.74 (± 1.219)	46.5%	30.2%	23.3%
The library will see decreased funding if it does not offer research data services (RDS).	3.78 (± 1.100)	10.4%	54.6%	34.9%
Losing data/data sets jeopardizes future scholarship.	2.19 (± 1.090)	59.3%	9.3%	31.4%
Librarians should be stewards of all types of scholarship, including data sets.	1.92 (± 0.948)	75.6%	8.1%	16.3%
Researchers at my institution will be at a competitive disadvantage for grants if the library does not provide research data services (RDS).	2.91 (± 1.102)	36.1%	29.0%	34.9%

Table 7: Statements about which respondents were asked to agree or disagree regarding their opinion about library involvement in RDS. Level of agreement was based on a five-point Likert scale with one equal to strongly agree and five equal to strongly disagree. n=86
Differences by Institution Type

The survey revealed some statistically significant differences among two-year, four-year, and research institutions with respect to the RDS offered or planned to be offered, the staff responsible, and the development of RDS capacity. Not surprisingly, librarians at four-year and research universities are more likely to consult with faculty, staff, or students on data management plans (Table 8). Libraries at four-year and research universities are also more likely than those from two-year institutions to have plans to discuss RDS with other librarians, RDS professionals, or others on campus (Table 9).

Table 8: The percentage of libraries, by type of institution (Two-Year Institution, Four-Year Institution, Research University) that are currently offering, planning to offer, or not planning to offer consultation services for data management planning. * indicates significant difference based on standardized residual $\geq |1.6|$ n=78

	Currently Offering	Planning to Offer	Not Planning to Offer
Two-year (n=16)	0.0%*	6.3%	93.8%
Four-year (n=22)	31.8%	13.6%	54.5%
University (n=40)	17.5%	22.5%	60.0%
Chi Square		$X^2(4)=9.755,$ $p=0.045$	

tion, Research University) that are currently offering, planning to offer, or not planning to offer consultation services for data management planning. * indicates significant difference based on standardized residual $\geq |1.6|$ n=78

Table 9: The percentage of libraries, by type of institution (2-YR: Two-Year Institution, 4-YR: Four-Year Institution, UNIV: Research University) that currently, plan to, or do not plan to discuss RDS with other librarians, RDS professionals, or others on campus. * indicates significant difference based on standardized residual $\geq |1.6|$

	Currently Discussing	Planning to Discuss	Not Planning to Discuss
Two-year (n=16)	0.0%*	6.7%	93.3%*
Four-year (n=22)	23.8%	14.3%	61.9%
University (n=40)	30.8%	28.2%	41.0%
Chi Square		$X^2(4)=12.807,$ $p=0.010^3$	

Year Institution, UNIV: Research University) that currently, plan to, or do not plan to discuss RDS with other librarians, RDS professionals, or others on campus. * indicates significant difference based on standardized residual $\geq |1.6|$

Based on the institution type, there were statistically significant differences in the levels of agreement with the following statement: *The library needs to offer research data services (RDS) to remain relevant to the institution* ($F(2)=5.026, p=0.009$). A Tukey post-hoc test

² Monte Carlo simulations were used due to small sample sizes

³ Monte Carlo simulations were used due to small sample sizes

revealed the level of agreement was significantly greater for libraries at research universities (2.33 ± 1.146 , $p=0.007$), compared to libraries at two-year institutions (3.46 ± 0.877).

Differences by Annual External Funding Amount

Just as type of institution reflects a research focus, and therefore a greater focus on research data, institutions that receive more grant money can be expected to focus more on research data. Not surprisingly, significantly higher percentages of libraries at institutions receiving \$50 million or more in external funding currently offer more RDS (Figure 3). Additionally, a higher percentage of these institutions indicated that dedicated data librarians, as opposed to individual discipline librarians, provide research data reference, consultation and instruction services. Institutions with more research funding are also more likely to provide opportunities for their staff to develop RDS skills, have policies and procedures in place for RDS, and collaborate with others, both within their institution and with other institutions (Table 10 & 11).

	Approximate Annual External Funding	
	< \$50 Million (n=32)	\$50 Million or more (n=9)
Individual Discipline Librarian	84.4%	22.2%*
Dedicated Data Librarian	0.0%*	44.4%
Other	15.6%	33.3%
Chi Square	$X^2(2)=19.187$, $p=0.001$	

Table 10: Individual responsible for providing research data reference, consultation and instruction

	< \$50 Million	\$50 Million or more	Fisher's Exact P value
Provide Opportunities for Staff to Develop RDS Skills (n<50M=33, n>50M=8)	21.2%	100.0%*	P=.001
Have RDS Policies and Procedures in Place (n<50M=33, n>50M=9)	0.0%	33.3%†	P=.007
Collaborate with Other Departments within Institution (n<50M=33, n>50M=9)	18.2%	100.0%•	P=.001
Collaborate with Other Institutions (n<50M=33, n>50M=8)	3.0%	50.0%‡	P=.003

⁴ Monte Carlo simulation was used due to small sample size

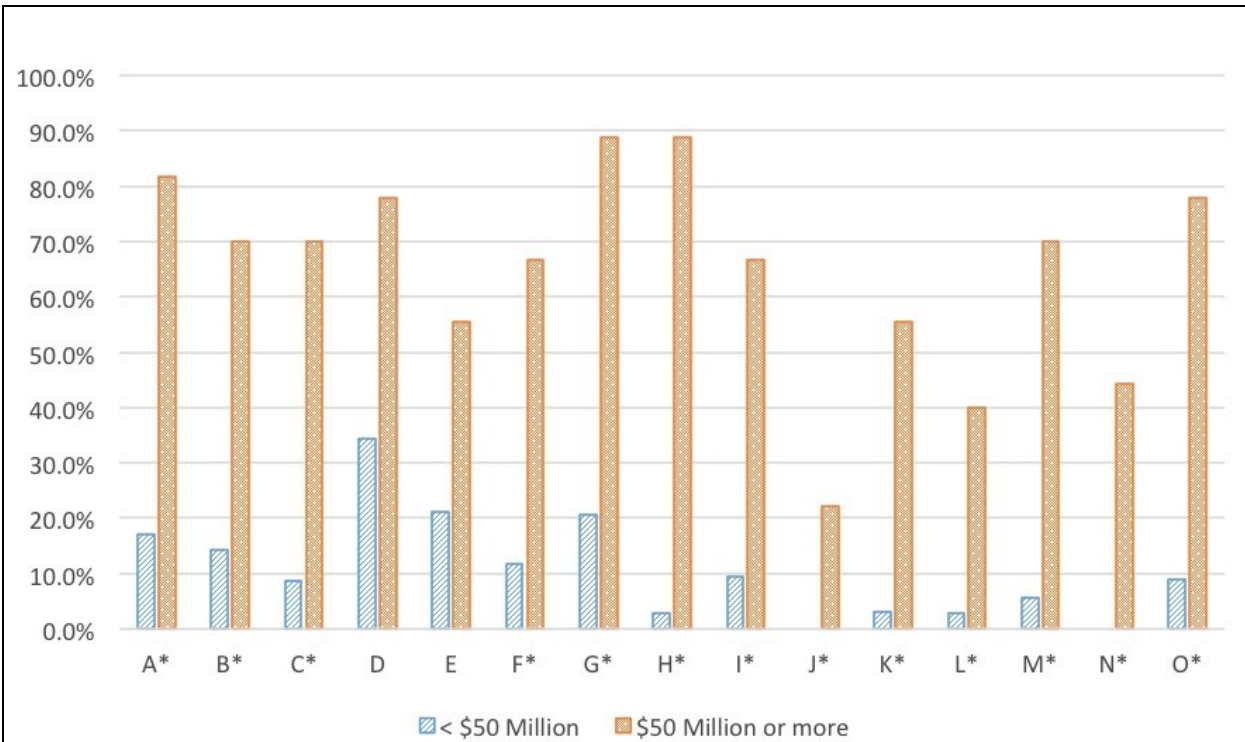


Figure 3. The percentage of libraries currently offering RDS whose institutions receive less than \$50 million in annual external funding versus those whose institutions receive \$50 million or more in annual external funding. * indicate significant differences between funding categories and are based on standardized residuals $\geq |1.6|$ ⁵

- A. Consulting on data management plans ($X^2(2)=15.969$, $n=46$, $p<0.001$) ($n<50M=35$, $n\geq 50M=11$);
- B. Consulting on data and metadata standards ($X^2(2)=12.632$, $n=45$, $p=0.003$) ($n<50M=35$, $n\geq 50M=10$);
- C. Outreach and collaboration with other RDS providers ($X^2(2)=18.964$, $n=45$, $p<0.001$) ($n<50M=35$, $n\geq 50M=10$);
- D. Providing reference support for finding & citing data ($X^2(2)=6.025$, $n=41$, $p=0.051$) ($n<50M=32$, $n\geq 50M=9$);
- E. Creating web guides for data & data repositories ($X^2(2)=5.671$, $n=42$, $p=0.057$) ($n<50M=33$, $n\geq 50M=9$);
- F. Directly participating with researchers on a project ($X^2(2)=12.532$, $n=43$, $p=0.003$) ($n<50M=34$, $n\geq 50M=9$);
- G. Discussing RDS with others on a semi-regular frequency ($X^2(2)=14.831$, $n=43$, $p=0.001$) ($n<50M=34$, $n\geq 50M=9$);
- H. Training co-workers in your library, or across campus, on RDS ($X^2(2)=31.904$, $n=43$, $p<0.001$) ($n<50M=34$, $n\geq 50M=9$);
- I. Providing technical support for RDS systems ($X^2(2)=14.906$, $n=41$, $p=0.001$) ($n<50M=32$, $n\geq 50M=9$);
- J. Deaccessioning data for removal from a repository ($X^2(2)=8.273$, $n=42$, $p=0.040$) ($n<50M=33$, $n\geq 50M=9$);
- K. Preparing data for deposit into a repository ($X^2(2)=16.379$, $n=41$, $p=0.001$) ($n<50M=32$, $n\geq 50M=9$);
- L. Creating or transforming metadata for data ($X^2(2)=12.754$, $n=45$, $p=0.002$) ($n<50M=35$, $n\geq 50M=10$);
- M. Identifying data that could be candidates for repositories ($X^2(2)=20.201$, $n=45$, $p<0.001$) ($n<50M=35$, $n\geq 50M=10$);
- N. Selection of data for repository ($X^2(2)=18.749$, $n=43$, $p<0.001$) ($n<50M=34$, $n\geq 50M=9$);
- O. Involvement in policy development or strategic planning related to RDS ($X^2(2)=18.964$, $n=43$, $p<0.001$) ($n<50M=34$, $n\geq 50M=9$);

⁵ Monte Carlo simulations were used due to small sample sizes

services based on institution’s approximate annual external funding. * indicates significant difference based on standardized residual $\geq |1.6|$. $X^2(2)=19.187$, $n=41$, $p<0.001^4$

Table 11: Percentage of libraries at institutions with less than \$50 million and \$50 million or more in

annual external funding whose respondents answered in the affirmative to the survey statements shown. All differences are significant based on Fisher’s Exact Test. * $p<0.001$, † $p=0.007$, • $p<0.001$, ‡ $p=0.003$

Differences by Full-Time Equivalent Student Populations

More institutions with larger student populations currently offer RDS compared to institutions with smaller populations. More libraries at institutions with 5,000 or more students responded they currently provide outreach *and collaboration with other RDS providers either on or off campus* (23.5% compared to 6.5% (< 5,000 FTE students)), *directly participate with researchers on a project* (27.3% compared to 10.0%), and *train co-workers in their library, or across campus, on RDS* (25.0% compared to 6.6%). Libraries at institutions with larger student

Analyzed by FTE	Chi Square	p-value	N
Outreach and collaboration with other RDS providers either on or off campus	$X^2(2)=6.226$	$p=0.044$	$n=96$
Directly participate with researchers on a project	$X^2(2)=8.826$	$p=0.012$	$n=93$
Train co-workers in their library, or across campus, on RDS	$X^2(2)=6.394$	$p=0.041$	$n=93$
Who in library: dedicated data librarians	$X^2(2)=9.912$	$p=0.009^*$	$n=89$
Collaboration: within institution	$X^2(1)=4.623$	$p=0.032$	$n=89$

populations are also more likely to have dedicated data librarians (16.1% compared to 0%), and have a higher rate of collaboration with other departments or units within their institutions regarding RDS (35.5% compared to 15.5%).

Table 12: Differences by Full-Time Equivalent Student Population. * p-value based on Fisher’s Exact Test

Respondents from universities with larger student populations were more likely to agree with the following statement: *The library needs to offer research data services (RDS) to remain relevant to the institution* ($F(4,81)=3.097$, $p=0.020$). A Tukey post-hoc test revealed the level of agreement was significantly greater for libraries at institutions with 25,000 or more FTE students (1.20 ± 0.447) compared to libraries at institutions with 5,000 – 9,999 FTE students (3.00 ± 1.000 , $p=0.041$) and up to 1,999 students (3.07 ± 1.307 , $p=0.012$).

Differences by Faculty Size

Libraries at institutions with more than 100 faculty members are more likely to currently offer RDS or plan to offer RDS.

Analyzed by Faculty Members (< or >= 100)	Chi Square	p-value	n	n<100	n>=100
Consulting on data management plans	$X^2(2)=10.616$	$p=0.005$	n=97	n<100=43	n>=100=54
Consulting on data and metadata standards	$X^2(2)=7.635$	$p=0.022$	n=96	n<100=43	n>=100=53
Outreach and collaboration with other RDS providers	$X^2(2)=14.924$	$p=0.001$	n=96	n<100=43	n>=100=53
Providing reference support for finding & citing data	$X^2(2)=12.174$	$p=0.002$	n=91	n<100=41	n>=100=50
Creating web guides for data & data repositories	$X^2(2)=10.257$	$p=0.006$	n=93	n<100=42	n>=100=51
Directly participating with researchers on a project	$X^2(2)=18.115$	$P=0.001$	n=93	n<100=42	n>=100=51
Discussing RDS with others on a semi-regular frequency	$X^2(2)=13.966$	$p=0.001$	n=92	n<100=41	n>=100=51
Training co-workers in your library, or across campus, on RDS	$X^2(2)=6.308$	$p=0.043$	n=93	n<100=41	n>=100=52
Providing technical support for RDS systems	$X^2(2)=9.109$	$p=0.011$	n=92	n<100=42	n>=100=50
Deaccessioning data for removal from a repository	$X^2(2)=7.446$	$p=0.016^{\wedge}$	n=91	n<100=40	n>=100=51
Preparing data for deposit into a repository	$X^2(2)=2.651$	$p=0.278^{\wedge}$	n=92	n<100=42	n>=100=50
Creating or transforming metadata for data	$X^2(2)=6.831$	$p=0.036^{\wedge}$	n=95	n<100=43	n>=100=52
Identifying data that could be candidates for repositories	$X^2(2)=18.605$	$p<0.001$	n=96	n<100=43	n>=100=53
Selection of data for repository	$X^2(2)=7.155$	$p=0.024^{\wedge}$	n=92	n<100=41	n>=100=51
Involvement in policy development or strategic planning related to RDS	$X^2(2)=16.434$	$p<0.001$	n=92	n<100=41	n>=100=51

Table 13: Libraries currently offering, planning to offer, and not planning to offer RDS whose institutions have fewer than 100 faculty members versus those whose institutions have 100 or more faculty members.

* Indicate significant differences between faculty categories and are based on standardized residuals $\geq |1.6|$. ^ Indicate Monte Carlo simulations used due to small sample size.

In addition to differences in RDS offerings, respondents from institutions with more faculty members indicated their libraries have a higher rate of collaboration with other departments or units within their institutions (32.0% compared to 10.3%) ($\chi^2(1)=5.946, n=89, p=0.015$).

Based on the institutions' number of faculty members, there were statistically significant differences in the levels of agreement with the statement The library needs to offer research data services (RDS) to remain relevant to the institution ($F(4,81)=3.430, p=0.012$). A Tukey post-hoc test revealed that the level of agreement was significantly greater for libraries at institutions with 1,000 or more faculty members ($1.25 \pm 0.500, p=0.025$) compared to libraries at institutions with fewer than 100 faculty members (3.11 ± 1.158). There were also statistically significant differences in the levels of agreement with Librarians should be stewards of all types of scholarship, including data sets ($F(4,81)=3.277, p=0.015$). Additionally, there were statistically significant differences in the levels of agreement with Researchers at my institution will be at a competitive disadvantage for grants if the library does not provide research data

RDS	F	P value	Tukey post hoc and comparison
Library needs to offer RDS to remain relevant to institution	F(4,81)=3.430	p=0.012	1,000 or more faculty: 1.25 ± 0.500, p=0.025
			Less than 100 faculty: 3.11 ± 1.158
Librarians should be stewards of all types of scholarship, including data sets	F(4,81)=3.277	p=0.015	No significant post hoc tests
Researchers will be at a competitive disadvantage for grants if the library does not provide RDS	F(4,81)=2.989	p=0.024	500 - 999 faculty members: 1.88 ± 0.991, p=0.017
			Less than 100 faculty members: 3.18 ± 1.010

services ($F(4,81)=2.989, p=0.024$). A Tukey post-hoc test revealed that the level of agreement was significantly greater for libraries at institutions with 500-999 faculty members ($1.88 \pm 0.991, p=0.017$) compared to libraries at institutions with fewer than 100 faculty members (3.18 ± 1.010) (Table 14).

Table 14: Library involvement in RDS (Q25 on survey) by Faculty Population Size.

Post hoc tests include mean and standard deviation

Comparison of 2011 and 2014 Survey Responses

As stated above, a total of 128 ACRL libraries responded to the 2014 academic library survey compared to 221 in 2011 (Tenopir, Birch & Allard 2012), however, respondents were similar in their demographic characteristics in terms of the number of full time equivalent (FTE) students ($X^2(4)=1.258$, $p=0.868$), the number of tenured and tenure-track faculty members ($X^2(4)=1.427$, $p=0.839$), and the type of institution ($X^2(2)=5.400$, $p=0.067$). Based on a comparison of the

RDS	Status	2011	2014
Consulting on data management plans ($n_{2011}=220$, $n_{2014}=98$)	Currently Offering	20.5%	17.3%
	Planning to Offer	22.3%	13.3%
	Not Planning to Offer	57.3%	69.4%
Consulting on data and metadata standards ($n_{2011}=218$, $n_{2014}=96$)	Currently Offering	17.9%	15.6%
	Planning to Offer	23.9%	17.7%
	Not Planning to Offer	58.3%	66.7%
Outreach and collaboration with other RDS providers ($n_{2011}=218$, $n_{2014}=96$)	Currently Offering	11.0%	12.5%
	Planning to Offer	23.4%	16.7%
	Not Planning to Offer	65.6%	70.8%
Providing reference support for finding & citing data ($X^2(2)=7.965$, $n=311$, $p=0.019$) ($n_{2011}=220$, $n_{2014}=91$)	Currently Offering	44.1%	29.7%
	Planning to Offer	21.8%	19.8%
	Not Planning to Offer *	34.1%	50.5%
Creating web guides for data & data repositories ($n_{2011}=220$, $n_{2014}=93$)	Currently Offering	22.3%	21.5%
	Planning to Offer	33.6%	30.1%
	Not Planning to Offer	44.1%	48.4%
Directly participating with researchers on a project ($n_{2011}=219$, $n_{2014}=93$)	Currently Offering	21.0%	16.1%
	Planning to Offer	19.6%	16.1%
	Not Planning to Offer	59.4%	67.7%
Discussing RDS with others on a semi-regular frequency ($n_{2011}=218$, $n_{2014}=92$)	Currently Offering	18.8%	19.6%
	Planning to Offer	28.0%	18.5%
	Not Planning to Offer	53.2%	62.0%
Training co-workers in your library, or across campus, on RDS ($n_{2011}=219$, $n_{2014}=93$)	Currently Offering	11.4%	12.9%
	Planning to Offer	27.4%	20.4%
	Not Planning to Offer	61.2%	66.7%
Providing technical support for RDS systems ($n_{2011}=221$, $n_{2014}=92$)	Currently Offering	14.5%	14.1%
	Planning to Offer	27.1%	23.9%
	Not Planning to Offer	58.4%	62.0%
Deaccessioning data for removal from a repository ($n_{2011}=220$, $n_{2014}=91$)	Currently Offering	5.5%	3.3%
	Planning to Offer	17.3%	13.2%
	Not Planning to Offer	77.3%	83.5%
Preparing data for deposit into a repository ($n_{2011}=220$, $n_{2014}=92$)	Currently Offering	9.5%	9.8%
	Planning to Offer	26.4%	21.7%
	Not Planning to Offer	64.1%	68.5%
Creating or transforming metadata for data ($n_{2011}=218$, $n_{2014}=95$)	Currently Offering	11.9%	8.4%
	Planning to Offer	22.0%	15.8%
	Not Planning to Offer	66.1%	75.8%
Identifying data that could be candidates for repositories ($n_{2011}=218$, $n_{2014}=96$)	Currently Offering	11.0%	12.5%
	Planning to Offer	33.5%	21.9%
	Not Planning to Offer	55.5%	65.6%

survey responses, there has been little change in the percentages of libraries that currently offer, plan to offer, and do not plan to offer most of the specific RDS from 2011 to 2014 (Table

	Currently Offering		Planning to Offer		Not Planning to Offer	
	2011	2014	2011	2014	2011	2014
Two-year (n₂₀₁₁=68, n₂₀₁₄=15)	39.7%	33.3%	7.4%	13.3%	52.9%	53.3%
Four-year (n₂₀₁₁=53, n₂₀₁₄=21)	39.6%	33.3%	18.9%	19.0%	41.5%	47.6%
University (n₂₀₁₁=99, n₂₀₁₄=38)	49.5%	31.6%	33.3%	26.3%	17.2%	42.1%*

15). In 2014, however, significantly fewer libraries say they have plans to provide reference support for finding and citing data (Table 16).

Table 15: The percentage of libraries currently offering, planning to offer, and not planning to offer RDS in 2011 and 2014. * indicate significant differences between years and are based on standardized residuals $\geq |1.6|$.

Table 16: Percentage of libraries by type of institution currently offering, planning to offer, and not planning to offer reference support for finding and citing data in 2011 and 2014.* indicates significant difference between years based on standard residual $> |1.6|$. $X^2(2)=9.498$, $n=137$, $p=0.009$ (University)

Differences by Faculty Size between 2011 and 2014

While there are no significant differences between 2011 and 2014 for the entire survey

	2011	2014
<100 Faculty Members (n₂₀₁₁=81, n₂₀₁₄=43)	25.9%	4.7%*
More than 100 Faculty Members (n₂₀₁₁=134, n₂₀₁₄=53)	38.8%	35.8%

population regarding identifying data that could be candidates for repositories, differences exist for this service when the population is broken down by faculty size. Multi-dimensional chi-square tests based on the type of RDS offered, survey year, and faculty size indicate that, in 2014, a significantly smaller percentage of libraries at institutions with fewer than 100 faculty members reported that they plan to offer this service, compared to 2011 ($X^2(2)=8.452$, $n=124$, $p=0.015$) (7). Similarly, there was no significant change in percentages between 2011 and 2014 for libraries at institutions with more than 100 faculty members⁶.

Table 17: The percentage of libraries from institutions with small and large faculty planning to offer RDS identifying data that could be candidates for repositories in 2011 and 2014.* indicates significant difference between years based on standard residual $> |1.6|$. $X^2(2)=8.452$, $n=124$, $p=0.015$

⁶ Analyses were also conducted looking at differences between years based on the number of full-time equivalent students at the libraries' institutions; however, no significant differences were found.

Discussion & Conclusions

The role of the academic library must adapt over time to changes in the research landscape. As we move into a new paradigm of “data-intensive” research (Tansley, Stewart & Tolle 2009), academic libraries are working to determine what, if any, new research data services they will need to offer to continue to support their faculty, staff, and students. The majority of the ACRL institutions surveyed are not offering RDS, although research universities are more likely to offer a range of data services than libraries at other types of academic institutions. Plans for the future of RDS are still being considered by academic libraries.

In order to fully offer technical RDS, libraries need to have technologically skilled staff or greatly increase opportunities for technology training for their existing staff, which might not be feasible due to resource constraints.

Interviews provided deeper insights into the challenges of offering technical RDS. There is some sentiment that these services should be offered by other institutional departments with larger technical capacity. Some interviewees suggested partnering with their institution’s Information Technology department, as well as offices that are involved with sponsored programs, grants, and research. Another example of collaboration to grow RDS is to create a “research data network,” including researchers with varied expertise, to address RDS issues based on multiple discipline-specific needs. One such discipline-specific network mentioned is a digital humanities working group, which provides a platform for exchanging ideas, keeping up with trends in the fields and focusing on RDS issues. Similar sentiments about collaboration were reported in an earlier study (Pinfield, Cox & Smith 2014). If libraries collaborate with other departments or other institutions, they may be able to offer a more complete suite of RDS for their institution.

As academic libraries begin to decide on which types of RDS they plan to provide in-house, they also need to decide who exactly will be providing those services. Library leaders need to decide whether they will invest in developing current librarians and staff, or if they will hire data librarians. There are many data-related skills that librarians and library staff need to acquire to keep pace with the changing landscape and the costs related to workforce development are not insignificant. For many reasons, few academic libraries have hired data librarians. For example, there may not be enough of a perceived demand for RDS to warrant a full-time data librarian. In the interviews, some library directors commented on the lack of demand at their institution for RDS, which support the findings of an earlier study that found there is little or no demand for RDS from patrons at many institutions (Tenopir, Birch & Allard 2012). In some instances, patrons may be unaware that the library could offer these services. Some interviewees mentioned that faculty, staff, and students wished they had known that the library offered RDS earlier as they mistakenly believed RDS were beyond the suite of services associated with libraries.

To improve awareness of RDS, some interviewees suggested marketing library-offered RDS by meetings with various campus groups. Some also mentioned the importance of meeting with university leaders about implementing RDS. Until libraries and their institutions see more demand, they may not feel that it is worth investing in RDS. Many academic libraries may not have the resources or administrative support to hire a research data services librarian.

Institutions with larger student populations and higher levels of external funding are more likely to have full-time data librarians.

The 2014 survey also found that some institutions have hired data librarians and are also reassigning others within the library to take on data-related responsibilities. A single person may not be able to keep up with all of the different standards and best practices for data across all disciplines. According to Heidorn, data at research and teaching institutions are very heterogeneous, thereby making it difficult for any one person to have all the necessary skills to provide researchers at his/her institution with RDS (2011). Instead, a team approach may be best for implementing RDS. Pinfield, Cox and Smith describe how important it is for libraries to maintain awareness of the various disciplinary cultures and practices with respect to RDS within their institutions (2014). Therefore, it may make sense to have subject librarians or liaisons take on the role of RDS provider for their discipline because, ultimately, they will have the best understanding of the cultures within that domain. These librarians will likely understand the best way to approach researchers about sensitive topics such as data sharing.

Despite the low percentage of ACRL academic libraries surveyed that are actively providing RDS, many respondents agree that libraries should be involved in data curation and that data are important for future scholarship. There may be an inconsistency between librarians' feelings about the importance of the library's involvement in RDS versus the motivation to move forward. If the library's leadership does not perceive a disadvantage to their institutions' research capabilities without library-led RDS, there may be little incentive to start offering these services. The need for RDS may be externally imposed. Many libraries that have started to offer RDS are doing so in response to funding agency Open Data (OSTP)⁷ mandates surrounding research data.

However, interviewees noted that it is important to portray RDS as not just a means of compliance, but also as services that will directly benefit the researchers themselves. This is especially true because there is still some doubt about the enforcement of funder requirements and the consequences of not complying with those requirements. Therefore, solely relying on the "compliance" argument may make it difficult for the library to "sell" RDS to researchers (Pinfield, Cox & Smith 2014). This means responding to regulatory mandates is not a sufficient basis for a successful RDS program. Libraries have the responsibility to advocate for good research data management practices.

Development of an RDS program requires both a top-down and bottom-up approach. Leadership and engagement from the librarians and staff who will provide and promote the services is essential; however, leadership and resources are also required from the top-level administrators. Interviewees mentioned that this support is a necessity for RDS implementation. Some indicated that campus-wide support is important for securing funding for RDS since libraries do not always have the resources to implement such a program on their own. Similarly, authority figures who are interested in creating partnerships to implement campus-wide RDS have secured more institutional funding for the library.

A lack of institutional support may be one of the reasons there has not been a faster library adoption rate of RDS (Pinfield, Cox & Smith 2014). In 2011, 15 to 35 percent of libraries

⁷ <https://www.whitehouse.gov/administration/eop/ostp/initiatives>

surveyed indicated they were planning to offer most types of RDS. Therefore, we anticipated that we would find an increase in the percentage of libraries offering RDS in 2014; however, this was not the case. Instead of an increase, the percentage of libraries offering RDS remained fairly stable and in many cases declined slightly. The interviews helped us understand why libraries did not follow through on their plans.

Without top-level support from their institutions, even if library directors had wanted to implement RDS, it might not have been possible. Academic institutions are slow-moving. They require a lot of time to make substantial changes, and different departments move at different paces. In cases where libraries are coordinating RDS efforts with other departments, difficulties in enacting change may occur. Some respondents stated that time is a very limited resource when it comes to implementing RDS. Specifically, library staff, researchers, and office partners have limited time to either implement or use RDS. Interviewees also suggested that libraries need to find ways to make the RDS process more automatic and less time-consuming for researchers. For example, a central repository could be created where researchers could both compile and deposit data, as opposed to moving data at a later time.

Libraries are proceeding cautiously. Some researchers have “massive expectations” when it comes to RDS. Librarians and upper-level administrators realize that the amount of work could quickly get out of control. With concerns over resources and support from various institutional levels, some are hesitant to get involved. During the interviews, some library directors provided their experiences of curbing the expectations of researchers on what and how much the library can do to help. Some researchers may assume that the library is going to do data management for them; whereas RDS may just mean to assist researchers with data management. Additionally, if the RDS offered become too successful, there is concern the library will not have resources to support the demands. For example, one interviewee described concerns about data storage and the library not being able to keep up with demand for server space. The interviewee also mentioned repository traffic concerns if users start “taking out all of our data left and right.”

Implementing RDS in an academic library requires resources, be it personnel, time, skills, money, or support. One solution for dealing with limited resources is to for libraries to partner with others on campus or with other universities and share resources. One example of collaboration within an institution is where the institutional office of research provides data management training, while the library provides a repository for completed data management plans. An intra-institutional example is the Data Management Planning tool developed by the University of California’s California Digital Library in collaboration with other universities (<https://dmp.cdlib.org>).

RDS requires a skilled workforce and motivation from library leadership and library staff to transform the library’s role in support of academic scholarship. The results of this survey of academic library directors reveals the degree to which their libraries are engaged in RDS. To understand individual librarians’ attitudes towards these new roles, we also surveyed librarians who work in academic libraries. The results of the 2011 baseline survey of academic librarians showed that many academic librarians do not feel prepared to take on these new roles in spite of plans by their libraries to offer RDS (Tenopir, Birch & Allard 2012). To see if this mismatch between organizational RDS objectives and the readiness of individual librarians continues, we conducted a follow-up survey of academic librarians in 2014; results are forthcoming.

Supplemental Content

Appendices A and B

An online supplement to this article can be found at <http://dx.doi.org/10.7191/jeslib.2015.1085> under "Additional Files".

Acknowledgement

Special thanks to University of Tennessee doctoral students, Priyanki Sinha, Danielle Pollack, Ivanka Pjesivac and Betsy Dalton and the University of Tennessee Statistical Consultants.

Funding Statement

This study was funded by NSF DataONE, William Michener, Principal Investigator. It is part of the activities of the DataONE Usability & Assessment Working Group and various members of that group assisted with the survey activities at times.

Disclosure

The authors report no conflict of interest.

References

- Akers, Katherine G. and Jennifer Doty. 2013. "Disciplinary Differences in Faculty Research Data Management Practices and Perspectives." *International Journal of Digital Curation* 8:5-26. <http://dx.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* 75:557-574. <http://dx.doi.org/10.5860/crl.75.4.557>
- Carlson, Jake. 2012. "Demystifying the Data Interview: Developing a Foundation for Reference Librarians to Talk with Researchers about Their Data." *Reference Services Review* 40:7-23. <http://dx.doi.org/10.1108/00907321211203603>
- DataONE. 2013. "Best Practices." Accessed October 8. <https://www.dataone.org/best-practices>
- Deards, Kiyomi. 2013. "Why, How, and Where We're Going Next: A Multi-Institution Look at Data Management Service." *CLIR Report: Research Data Management, Principles, and Prospects* 160:43-63. <http://www.clir.org/pubs/reports/pub160/pub160.pdf>
- Gray, Jim, David T. Liu, Maria Nieto-Santisteban, Alex Szalay, David J. DeWitt, and Gerd Heber. 2005. "Scientific Data Management in the Coming Decade." *ACM SIGMOD Record* 34:34-41. <http://dx.doi.org/10.1145/1107499.1107503>
- Heidorn, P. Bryan. 2011. "The Emerging Role of Libraries in Data Curation and E-science." *Journal of Library Administration* 51:662-672. <http://dx.doi.org/10.1080/01930826.2011.601269>
- MacMillan, Don. 2015. "Developing Data Literacy Competencies to Enhance Faculty Collaborations." *Liber Quarterly* 24:140-160. <http://doi.org/10.18352/lq.9868>
- Mayernik, Matthew S., Lynne Davis, Karon Kelly, Bob Dattore, Gary Strand, Steven J. Worley, and Mary Marlino. 2014. "Research Center Insights into Data Curation Education and Curriculum." In *Volume 416 of Communications in Computer and Information Science: Theory and Practice of Digital Libraries — TPDL 2013 Selected Workshops*, edited by Lukusz Bolikowski, Vittore Casarosa, Paula Goodale, Nikos Houssos, Paolo Manghi, and Jochen

Schirrwagen, 239-248. Springer International Publishing. <http://dx.doi.org/10.1007/978-3-319-08425-1>

Mclure, Merinda, Allison V. Level, Catherine L. Cranston, Beth Oehlerts, and Mike Culbertson. 2014. "Data Curation: A Study of Researcher Practices and Needs." *portal: Libraries and the Academy* 14:139-164. <http://dx.doi.org/10.1353/pla.2014.0009>

Michener, William K., Suzie Allard, Amber Budden, Robert B. Cook, Kimberly Douglass, Mike Frame, Steve Kelling, Rebecca Koskela, Carol Tenopir, and David A. Vieglais. 2012. "Participatory Design of DataONE — Enabling Cyberinfrastructure for the Biological and Environmental Sciences." *Ecological Informatics* 11:5-15. <http://dx.doi.org/10.1016/j.ecoinf.2011.08.007>

Norman, Belinda and Kate Valentine Stanton. 2014. "From Project to Strategic Vision: Taking the Lead in Research Data Management Support at the University of Sydney Library." *International Journal of Digital Curation* 9:253-262. <http://dx.doi.org/10.2218/ijdc.v9i1.316>

Parham, Susan Wells, Jon Bodnar, and Sara Fuchs. 2012. "Supporting Tomorrow's Research: Assessing Faculty Data Curation Needs at Georgia Tech." *College & Research Libraries News* 73:10-13. <http://crln.acrl.org/content/73/1/10.full>

Parsons, Thomas, Shirley Grimshaw, and Laurian Williamson. 2013. "Research Data Management Survey: Report." *The University of Nottingham ePrints*. <http://eprints.nottingham.ac.uk/id/eprint/1893>

Pinfield, Stephen, Andrew M. Cox, and Jen Smith. 2014. "Research Data Management and Libraries: Relationships, Activities, Drivers and Influences." *PLoS ONE* 9:7-34. <http://dx.doi.org/10.1371/journal.pone.0114734>

Soehner, Catherine, Catherine Steeves, and Jennifer Ward. 2010. "E-Science and Data Support Services: A Study of ARL Member Institutions." *Association of Research Libraries*. <http://www.arl.org/storage/documents/publications/escience-report-2010.pdf>

Tansley, Stewart and Kristin Michele Tolle, eds. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery, vol. 1*. Redmond, WA: Microsoft Research. <http://research.microsoft.com/en-us/collaboration/fourthparadigm>

Tayi, Giri Kumar and Donald P. Ballou. 1998. "Examining Data Quality." *Communications of the ACM* 41:54-57. <http://dx.doi.org/10.1145/269012.269021>

Tenopir, Carol, Ben Birch, and Suzie Allard. 2012. "Academic Libraries and Research Data Services: Current Practices and Plans for the Future." *Association of College and Research Libraries (ACRL) White Paper*. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Tenopir, Carol, Robert J. Sandusky, Suzie Allard, and Ben Birch. 2013. "Academic librarians and research data services: preparation and attitudes." *IFLA Journal* 39:70-78. <http://dx.doi.org/10.1177/0340035212473089>

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." *PLoS ONE* 6:e21101. <http://dx.doi.org/10.1371/journal.pone.0021101>

University of California. 2015. "DMPTool." Accessed October 6. <https://dmp.cdlib.org/>

Varvel Jr, Virgil E., Elin J. Bammerlin, and Carole L. Palmer. 2012. "Education for Data Professionals: A Study of Current Courses and Programs." *Proceedings of the 2012 iConference, ACM*: 527-529. <http://dx.doi.org/10.1145/2132176.2132275>

Weller, Travis and Amalia Monroe-Gulick. 2014. "Understanding Methodological and Disciplinary Differences in the Data Practices of Academic Researchers." *Library Hi Tech* 32:467-482. <http://dx.doi.org/10.1108/LHT-02-2014-0021>