Research Data Management Sustainability: Services, Infrastructure, Accountability, and Planning

Tina M. Griffin and Margaret Janz

abstract: Libraries at academic institutions have been involved in research data management services (RDMS) for more than ten years. Institutions have adapted their service models in response to funder, journal, and federal mandates that have consequently increased RDMS demand. However, institutions that hope to start or grow their services may have difficulty extrapolating from the published literature, as few studies recount RDMS development strategies.

This study aims to update on the status of RDMS service offerings, staffing and funding, and presents them according to the number of years a library has offered the service. This work also investigates RDMS service fulfillment, accountability in providing support, and planning strategies within the same institution sample. Updating the RDMS status, broadening the facets addressed, and presenting the data by cohort provides detail into how services have been maintained or developed so that institutions at a similar stage can make clearer decisions about how to keep RDMS sustainable.

Introduction

Libraries at academic institutions have been involved in research data management services (RDMS) for more than ten years. Research data management describes tasks such as:

- documenting data collection and project design,
- organizing research outputs and intermediary products,
- performing or documenting analysis steps and the tools or software used,
- managing active and archival storage, and
- sharing intermediate and final research results.

Services that support these activities include educating faculty, staff, and students on data management best practices; providing a data repository to curate, preserve, and share data; performing data analysis or data visualization and training others to do this; or providing a data management plan review service to assist with grant writing. As the demand for these services has increased due to funder, journal, and federal mandates, libraries have adapted their service models. Several studies have reviewed the state of RDM services, but these have only provided a snapshot of activities at a given time. Research Data Management activities—whether done by the researcher or supported by the institution through centers, the library, or other entities— require infrastructure, personnel, time, and money. Previous studies describe research data management support activities, like education efforts or repository services, individually, but there are few longitudinal studies. These studies also do not address the necessary decisions and changes that libraries must have made to make RDMS sustainable over time. This study will attempt to fill that gap by providing the status of RDMS and then broadening the facets addressed so that details about sustainability can be highlighted, and also presenting the data by cohort to help other institutions identify parallels to their RDMS.

Literature Review

Early this century, libraries started to offer research data management services in response to increasing federal mandates for sharing data, among other drivers. From 2007 through 2017, over fifty case studies reported research data management needs for individual institutions and therefore identified potential services. [1] Many larger studies have offered snapshots of library research data management services with the aim of summarizing service profiles and projecting trends.

The first large-scale study occurred in 2010 with the Association of Research Libraries (ARL) reporting on e-science and data support from a cross section of associates, baccalaureate, and research or doctorate-awarding institutions. This survey focused on understanding how ARL institutions were engaging with "e-science," broadly defined at the time as "big computational science, but also team science and networked science." [2] The ARL study results identified engagement through a combination of decentralized and centralized organizational groups (departments or centers versus campus). These initiatives included data centers, statistical analysis or computing centers, digital curation work, and high-performance computing. Library engagement was largely through campus collaborations, but also through consultation or reference service and some curatorial work. Specific service profiles were not yet described, and lack of time, staffing, funding, and leadership barriers were noted.

Shortly after the 2010 ARL study, two large surveys were published by the Association of College and Research Libraries (ACRL) and ARL. The ACRL survey of ARL

member-libraries outlined research data management service profiles and attempted to further delineate between "hands-on" and "informational services." [3] This survey also addressed various aspects of the research data lifecycle as defined by DataOne.[4] The ACRL researchers found that fewer than 25 percent of libraries were engaged in most aspects of RDMS, except for providing training, which nearly 45 percent reported doing. Most institutions had no plans to provide RDMS in the future. This study further analyzed their data by student body size, National Science Foundation grants awarded, and institution type. This study also looked at who was providing RDMS, finding that most subject librarians conducted library research data management services, as opposed to dedicated data specialists. In addition, RDMS responsibility was handled by an individual, a group or committee, or a department. Recommendations to develop staff capacity for RDMS entailed reallocation of existing staff.

The 2013 ARL *SPEC Kit 334*: *Research Data Management Services*, conducted by David Fearon et al, reported results, primarily from high level research institutions.[5] Most respondents (74 percent) reported that they were engaged in RDMS, but 23 percent noted no plan to offer RDMS within five years of the study. This marks a significant increase in both categories when compared to ACRL's results. *SPEC Kit 334* outlines three broad service areas: data management plan (DMP) support, other data management guidance, and data archiving, with no differentiation between technical and advisory services. Technical services being those that require hands-on work or participation in doing the data-related task, whereas advisory services provide training, resources, guidance, or referral. According to the survey results, subject liaisons provided most of the RDM services and the number of data

> **The most common organization structures for supporting RDMS were committees and partnerships between library departments.**

librarian positions was increasing. Overall, RDMS were being added to existing personnel responsibilities—most reported 25 percent or less of their time spent on RDMS work—rather than hiring data specialists. The most common organization structures for supporting RDMS were committees and partnerships between library departments. Partnerships with units external to the library were less common, but, when present, they occurred most often with central IT and research administration. General library funds almost exclusively financed these services. Forming partnerships, costs, staff hiring and training, and faculty awareness were listed as the prominent barriers to RDMS stability.

Andrew Cox and Stephen Pinfield surveyed UK institutions in 2014 to understand how they were involved in RDMS.[6] They reported that libraries were starting to engage in data management policy development and observed a cultural change toward acknowledgment of RDMS activities as a need. However, many institutions did not provide services across several data management categories but listed them as a top- or mid-level priority for development. Responses also reflected that staff did not have the skills needed to offer RDMS. Respondents also noted cost as a challenge to providing RDMS. When services were provided, they were a mix of technical and advisory in nature.

Carol Tenopir, in 2015, looked at RDMS over a three-year period to determine whether there were changes in levels of service, whether RDMS had increased, and

to identify barriers to service growth.[7] Overall, there were few differences reported. Tenopir found most respondents still did not offer, and did not plan to offer, RDMS. Most who did offer RDMS provided advisory-level services and only some technical ones. Who provided those services had not changed either, as most were provided by subject librarians and only a few data librarians had been hired. RDMS organizational structures were still handled by either individuals, groups, or committees. Reassignment of duties was still the most common method to address service needs. Interview responses identified barriers including lack of demand for RDMS, lack of awareness that libraries provide RDMS, lack of library staff skills (but this was also seen as an opportunity to partner with other campus entities), lack of institutional support, and time. Tenopir noted an "inconsistency between librarians' feelings about the importance of the library's involvement in RDMS versus the motivation to move forward." They also noted that "responding to regulatory mandates is not a sufficient basis for a successful RDMS program."[8]

Shortly after, Holly Yu reviewed RDMS through library website evaluation, limiting the study to ARL institutions that participated in Fearon's work. Yu then compared information collected to selected questions from Fearon's study.[9] Yu's website review looked for RDMS generally as well as DMP support, metadata consultation and tools, data archiving, institutional repository presence, and data sharing/access. They reported that all institutions studied provided some RDMS and DMP support and noted an increase overall. Some libraries provided metadata consultation or tools and data archiving, with moderate to significant increases in those services as well. However, there was little change in providing an institutional repository or providing data sharing and access support. Yu indicated that the organizational structure supporting RDMS may have changed. Institution-wide service offerings appeared to be stable, but decentralized services declined. The authors did not review RDMS staffing, staff training, or policies, stating their methodology wasn't appropriate for those tasks.

Tenopir revisited RDMS again in 2019, comparing results to the data they collected previously.[10] Generally, there was a modest increase in those offering RDMS, but services were still likely to be advisory. Technical services offered included working directly on projects and identifying data for deposit. There were increases in offering data-specific repositories, as well as supporting data deposit in larger institutional repositories. Most libraries offering technical RDMS had started within three years of this survey (approximately 2016). The data showed that staffing had not changed significantly. Most RDMS was provided by a single individual, although they did note that larger institutions were more likely to have groups or committees with primary responsibility for planning and programming. Reassigning staff remained the predominant approach to addressing capacity issues, but for those libraries planning for future RDMS, training existing staff and hiring staff was now of equal importance. Research libraries overall were planning to hire dedicated data services librarians, while other library types had subject librarians provide RDMS. Collaboration across units occurred consistently with offices of research and campus IT.

In 2020, Jane Radecki and Rebecca Springer conducted a web inventory on services.[11] Their work looked at a sample of institutions (R1, R2, and baccalaureate) for webpages that supported research data services. Their findings showed that R1 institutions pro-

vided an average of 7.6 services, while R2 universities provided 2.6, and baccalaureate institutions provided about 1.8 services. At R1 and baccalaureate institutions, libraries were the primary service providers, offering 2.4 and >1 services, respectively, while at R2 RDMS was provided by other entities. Advisory services were predominant while technical services were not common regardless of institution type. The findings described staffing and other technical services offered by providers outside of the library. With that consideration, the authors identified that RDMS tailored to clinical data, social sciences, business, or digital humanities were still lacking with less than 20 percent of their sample across any institution type providing this support.

In contrast, Elise Gowan and John Meier in the same year published a study identifying changes in RDMS staffing and services from 2014 through 2019 by using strategic planning documents, prior interviews, and a public dataset on policy documents.[12] They showed that of the 60 libraries affiliated with the Association of American Universities, there were "gains" in the number that employed data librarians and offered data services, but these were not significant. The types of data services were not described but libraries were the predominant providers. They also report slight "losses" in the number that had data housed in their repositories, but significant "gains" in the number of institutions that had a repository dedicated to retaining data or repositories that allowed data. Interestingly, the authors noted that institutions that identified RDMS as a priority employed fewer data librarians over time, but those that did not hired significantly more over the same time. The same pattern was seen regarding whether having a dedicated data repository was a priority. With data services generally, however, there was an increase in the number of services offered, regardless of priority. Looking further they found that formal planning for RDMS was uneven over time. It was unclear whether RDMS was becoming a foundational library service and therefore no longer being included in planning efforts.

Clearly there is a sustained interest in developing RDMS at academic institutions. The literature documents service profiles that remain mostly advisory over technical, despite offering a broader service range. Support structures like funding, staffing, and hiring do not appear to have stabilized even after more than ten years. Similarly, when reported, there is little consistency in organizational structure behind RDMS. Institutions that hope to start or grow their services may have difficulty extrapolating from the published literature as few studies recount RDMS development strategies. The most recent studies are web inventories that review only public information about service availability. Earlier studies focused on the existing or prospective service profiles, with some discussion of staffing or organization. Later studies addressed the depth of service profiles and some staffing, but as mentioned, only using publicly available data. Almost all studies record data according to type of institution, but only Fearon reported RDMS by years of service (up to three years) and that was over ten years ago. Except for Tenopir's work and Gowan and Meier's look at RDMS planning, there is little connection between disparate RDMS data over such a long time.

The purpose of this study is to take a deeper look at research data management support by connecting these disparate pieces. Drawing from common themes among the previous surveys, particularly the ARL *Spec Kit 334* and Tenopir from 2019, this survey provides an update on the current status of RDMS, particularly service offerings, staff-

ing and funding. This work also investigates RDMS service fulfillment, accountability in providing support, and planning strategies within the same institution sample. This data is presented by years in service rather than by institution type whenever possible, to show the RDMS progression relative to time. Follow-up questions allowed respondents to elaborate on challenges and changes in providing research data management support, an important detail rarely reported. Collectively updating the RDMS status, broadening the facets addressed, and presenting the data by cohort may give libraries more information so that they can make clearer decisions about how to keep RDMS sustainable for their institution.

Methods

The researchers developed a 62-question survey that asked about changes in service model, staffing and funding, accountability, and planning for research data management services since service inception (See Appendix A). Eight data librarians currently in the field reviewed the survey, which was then edited for bias, errors, and clarity. The Qualtrics survey was distributed over a period of six weeks during the fall of 2020 via data listservs and message boards. Lists reached included the DataCure Google Group, Research Data Access and Preservation discussion forum, International Association for Social Science Information Service and Technology listserv, Digital Library Federation listserv, and the Research Data Alliance forums. Reminders were sent three weeks after the initial announcement and again the day before the survey closed. IRB at both authors' institutions reviewed the protocol and it was determined to be exempt.

The survey campaign garnered 118 responses. Responses were deleted if they did not contain any data (21), or if the respondent only answered the question regarding service duration (9). An additional 16 responses were excluded because the services had been offered for less than one year at those institutions, and an additional two respondents did not consent and thus were exited from the study. Seventy responses remained, 32 of which were partially complete but offered enough data to remain in the study. None of the responses came from the same institution so no responses were merged for analysis. Finally, researchers removed any identifying information per the consent agreement before analysis began.

Results

Initial survey questions asked about years of service and leadership responsibilities to set the framework for potential subgroup analysis. Most respondents (31) indicated they had been offering research data services from 5 to 10 years. The remaining responses were almost evenly distributed, among 1-3 years (13), 3-5 years (14), and 10 or more years (12). Study data is presented according to these groups—1-3Y, 3-5Y, 5-10Y, and 10Y+—when meaningful. In addition, about one-third of respondents (25) reported that they consider themselves to be in a leadership position regarding data services, whereas thirteen did not, and the remainder did not report their level of responsibility.

Service offerings

The researchers first investigated the service scope for RDMS at participating institutions. Questions regarding the number, type, and level of services offered provide an update to previously published data and a baseline for further exploratory questions, while questions addressing the number of service requests filled, whether and how service profiles had changed, what services were easy or difficult to implement examine shifting service loads and provide potential indicators for long-term sustainability.

Among all institutions represented, the number of services offered ranges from two to 15. The average number of services from the 1-3Y cohort is fewest at 6.5, while the 3-5Y group averages almost 10 services. Those institutions offering services for five years or longer average just under eight (5-10Y offered 7.9; 10Y+ offered 7.8 services). Review of service profiles reveals that almost all respondents offer DMP planning, education or training, and data sharing or access services. The services offered least by each institution vary, but data analysis and data deidentification are commonly offered across most cohorts (See Table 1).

Table 1.

Survey responses were organized into cohort by number of years offering RDMS. The data in this table demonstrate the services most and least offered by each cohort.

	Most offered services		Least offered services	
Cohort	**Service**	**Percent of cohort offering**	**Service**	**Percent of cohort offering**
1-3 Years	DMP planning	91	Data purchase	27
(n=11)	Data Sharing/Access	91	Data preservation	27
	Education/Training	73	Metadata creation	27
			Data documentation (not metadata)	27
			Data analysis	18
			Data curation	18
			Deidentifying data	18
3-5 Years	DMP planning	100	Deidentifying data	23
(n=13)	Data Sharing/Access	92	Data analysis	15
	Data Curation	85	Other	15
	Data Storage	85		
	Data preservation	77		
	Finding data	77		

Table 1, continued.

Cohort	Most offered services		Least offered services	
	Service	Percent of cohort offering	Service	Percent of cohort offering
5-10 Years (n=29)	DMP planning	97	Provide resources (software/equipment)	24
	Education/Training	93	Deidentifying Data	10
	Data Sharing/Access	76	Other	7
	Data Storage	72		
10+ Years (n=8)	Education/Training	88	Deidentifying data	25
	DMP planning	88	Data documentation (not metadata)	25
	Data purchase	75	Other	13
	Finding data	75		

To clarify the level of service offered within these specific categories, the researchers asked whether the service itself was technical, advisory, or both. Across all cohorts, few institutions reported offering any service as only technical support and only data management planning was identified as primarily an advisory level service. Most respondents are offering both advisory and technical support for many services (sharing, storage, finding data, preservation, curation, and providing metadata), but about one-third of institutions still offer these as advisory only. Data citation, documentation, purchase, visualization, analysis, deidentification, and generally offering data resources are almost evenly split between technical and advisory models of support (See Figure 1).

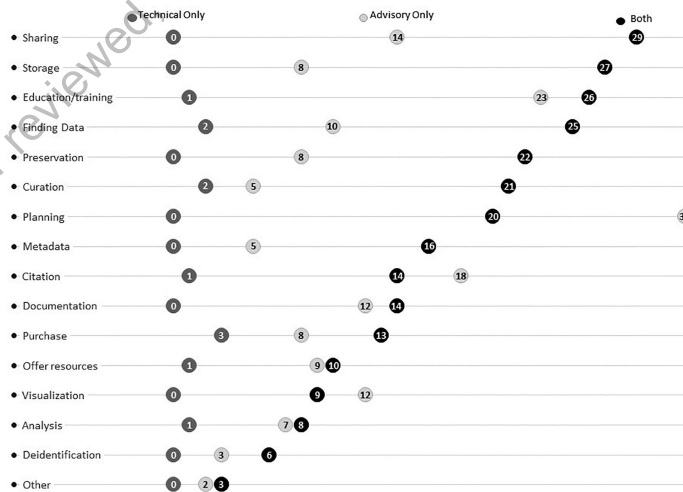Number of respondents offering RDMS by service type.

Figure 1. The total number of respondents reported as offering a specific research data management service, presented by service type.

Sorting this data by cohorts, the observed trend continues. The 1-3 years cohort offers predominantly advisory services related to sharing, finding data, storage, and preservation, while technical and advisory-type service is about equal when it comes to planning and education. The 3-5 years cohort predominantly advises on education, sharing, storage, preservation, curation, and documentation. DMP support and finding data for this group is offered at both the technical and advisory levels. Institutions in the 5–10-year cohort predominately offer advisory service for finding data, storage, preservation, curation, and metadata support. The cohorts are divided regarding whether they employ advisory or technical support for DMPs, education, sharing, citation, purchase, and data analysis. Interestingly, those offering services for ten years or more are split between technical and advisory models for any service offered (See Figure 2).

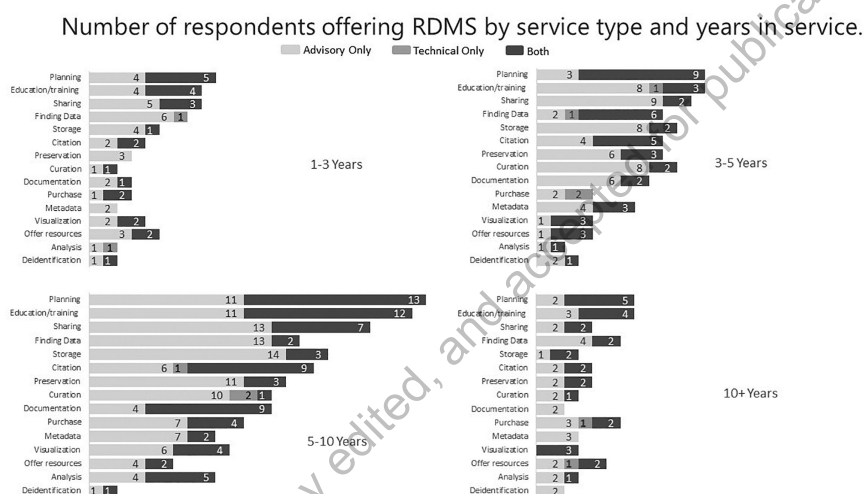Number of respondents offering RDMS by service type and years in service.

Figure 2. RDMS offered, sorted by preferred service model and the number of years the institution has engaged in RDMS.

Even though services are offered, that does not mean they are used. Of the service options presented in the survey, most institutions fulfill only 1-10 requests of a given service per year, regardless of the number of years they have provided RDMS support, with one exception—eight institutions reported that they fulfill between 11 and 20 data visualization requests per year. The collective responses show that liaison relationships and electronic communication are the most popular methods to market RDMS. Many institutions also employ print materials and leadership advocacy.

Collectively, the data show that the longer an RDMS has been offered, the larger the number of requests that service has fulfilled. The 1-3Y group reported filling 609 requests per year across all types, with the most being for education or training (205). The 5-10Y group reported filling just over 1700 annual requests, with 335 for data storage. The 5-10Y group filled over 1800 requests, with the most for education and training (439). Finally, the group offering services for over ten years filled just over 2300 requests in a year, with the most for finding data at 810 (600 of which were from one respondent). Removing that outlier, this group otherwise fulfilled over 400 education and training requests.

When asked whether services have changed since initially offered, 69 percent (38 of 55) reported that they had and most (65 percent, or 25 of 38) reported that they now provide more of both technical and advisory services. For all cohorts, service demands led to these changes, except for those in the 10Y+ cohort who indicated leadership decisions led to changes. In the 3-5Y cohort changes were also affected by employee transitions and strategic planning initiatives in addition to service demand.

Service changes can be a challenge. The researchers wanted to know which service participants felt had been the easiest to implement. The most frequent response to this question was education or training (7). A few respondents expanded on what factors made this service easier to implement. They mentioned hiring someone, not needing external help, having existing buy-in from colleagues or researchers, having autonomy to offer training, having existing curriculum to use, and high demand for the service. The noted trade-offs associated with offering education and training services were shifting workload and effort to accommodate teaching. Specifically, increased time demands for education services meant a decrease in time spent on other services, a general increase in workload, and sometimes collegial tension around priorities.

> **The noted trade-offs associated with offering education and training services were shifting workload and effort to accommodate teaching. Specifically, increased time demands for education services meant a decrease in time spent on other services, a general increase in workload, and sometimes collegial tension around priorities.**

A few respondents reported that there were no trade-offs when adding education and training services. Other services reported as "easy" to implement included finding data (3); data storage, data visualization, advisory services generally (2 each, respectively); and offering consultations, data purchase or subscription, data management planning, digital humanities work, and offering resources generally (1 each, respectively). Elaborations on implementing these other services are aggregated in Table 2.

Similarly, participants were asked which service was the most difficult to implement. Seven respondents rated data storage, curation, and preservation as the most challenging. Five of these mentioned data repositories specifically. Reasons these services were identified as difficult to implement included cost; the number of stakeholders involved; campus politics regarding consensus-building; buy-in from users; resistance to sharing; opposition from library administration; researchers being "unaware of [the] library role," and having limited staff. Trade-offs that were made to support this service included abandoning shared governance over the project, increased staff responsibilities, time, and accepting soft-money funding. Five respondents also identified training and education as difficult to implement. Challenges noted included the breadth of topics to cover; faculty failure to identify the library for this service; difficulty achieving curricular integration without liaison collaboration; lack of colleague capacity, time, or interest; and general buy-in. The trade-offs required to implement education and training were

Table 2.

Free-text elaborations on why a given service was identified as easy to implement and what trade-offs, if any, had to be made in order to implement that service. Reported by all data cohorts in aggregate. Number of responses by category indicated in left column.

Type of Services	Factors that contribute to ease of implementation	Trade-offs
Education/ Training (7)	• Hiring someone • not needing external help • existing buy-in from colleagues or researchers • having autonomy to offer training • having existing curriculum to use • high demand for the service	• Shifting or increased workload • increased time and demands meant decreased time spent on other services • incomplete willingness by others created collegial tension going forward
Finding data/ data citation (3)	• "best known service... and the oldest data service offered. People were already familiar and experienced with the service" • hired a person • extension of reference/ instruction service	• Less time for data curation activities • lack of visibility regarding demand
Data storage (2)	• Hosted by consortia partner • migrated to platform that allowed data	• Lack of control (development schedule, branding)
Data visualization (2)	• Hiring • "easy to add technical workshops that are tool-based"	• Change team dynamics • none
Advisory services (2)	• More liaisons • no technical issues	• Less centralized, more management, more coordination • personnel policy
Consultations (1)	• Referrals	• Less time to market instruction service
Data purchase (1)	• Because of personal growth within job. Past the "learning curve"	
Data management planning (1)	• Service fit with need and expertise	• Redefining existing job parameters
digital humanities (1)	• Already providing informally	
Offering resources (1)	• Open source software	

Table 3.

Free-text elaborations on why a given service was identified as difficult to implement and what trade-offs, if any, had to be made to implement that service. Reported by all cohorts in aggregate. Number of responses by category indicated in left column.

Type of Services	Factors that contribute to difficulty of implementation	Trade-offs
Storage, curation, preservation (7)	• "Researchers aren't aware of library role" • Costs, # of stakeholders • Campus politics regarding consensus building, limited staff • Buy in from users • Resistance to sharing • Opposition from library administration • Implementation with multiple stakeholders and vendor.	• Abandoned shared governance • Increased responsibilities • Time, only soft money funded
Education/ Training (5)	• Breadth of topics • No infrastructure for teaching courses • Faculty don't look to the library for this support • Difficult to get curricular integration without liaison collaboration • Colleagues don't have capacity, time, or interest • Buy-in	• Loss of momentum for implementing this service • Decreased quality and quantity of instruction • Asynchronous mode • Need to "keep pushing"
Data visualization (2)	• Questions about scope of library role, learning curve associated • Lack of expertise and software	• Lack of staff, lack of willingness • Lack of time
Data portal (2)	• Lack of staff and resources • High buy-in required	• Lack of features were identified • Less time for curation activities

Type of Services	Factors that contribute to difficulty of implementation	Trade-offs
Data management plan advising (1)	• Lack of demand	
Finding data (1)	• "Don't have formal training in data collection ... Understanding data from a technical standpoint is difficult"	
Policy advising (1)	• "Integrating across the institution requires political know-how, continual contact with key stakeholders, and frequent last minute meetings. "	• Time on other tasks
Technical Services (generally)(1)	• Lack of staff	• Training for new position

identified as: a loss of momentum for the service overall; decreased quality and quantity of teaching and asynchronous delivery; and a "need to keep pushing." Other services identified as difficult to implement were data visualization and creating a data portal (2 each); data management planning, finding data, data policy advising and technical services generally (1 each). Elaborations and trade-offs related to the services that are difficult to implement are shown in Table 3.

Staffing and Funding

The researchers next explored the institutional support required in the dynamic RDMS landscape. They began with the required skills, staffing levels, and roles, as has been reported in previous literature. Next, the study delved further into personnel issues such as required skills and staffing of shifts, the percentage of employee effort allocated to RDMS, and funding.

Across all respondents, 69 percent (32/46) noted that there has been a shift in the skills needed to offer research data management service. By cohort, 1-3Y and 3-5Y 75 percent and 81 percent, respectively, reported that there has been a shift in what skills are needed, while 65 percent 5-10Y and 57 percent of the 10Y+ responded similarly. The 1-3Y cohort expressed needing a broad skill set, including general technical and specific software knowledge to support qualitative data or text analysis; the ability to analyze social media data; understanding general data management principles; understanding funders; locating data; finding repositories, and data science skills. Cohort 3-5Y needs coding and programming skills (4), followed by data visualization, curation, workflow optimization, interpersonal skills, human subject and IRB knowledge, and general data analysis (1 each). Cohort 5-10Y is similar with coding and programming (5), but follows with the broadest skill set similar to the 1-3Y group listing general data management and analysis skills; repository, storage, and preservation knowledge; digital humanities or social science knowledge, human subject and IRB knowledge, qualitative data or text mining analysis, the ability to teach, IT knowledge, advocacy and policy development

knowledge, the ability to perform data citation activities; the ability to handle data purchase and licensing; data management plan drafting and revision; data security and encryption knowledge; GIS skills, and metadata creation ability. The 10Y+ cohort differs. Their needs are for project management, political skills, understanding funders, and security or encryption knowledge.

Respondents also reported on skills that had been lost or no longer needed to support RDMS. Across all cohorts these include GIS, data visualization, high-performance computing, bioinformatics, clinical research data management, general data management, data analysis, or curation and archiving. Ten respondents indicated "nothing" fits this category of skills, or they replied that they "just keep adding" skills. Three respondents reported that skills lost were attributable to staff departures.

With the variety of skills needed to offer RDMS, the next questions focused on staffing levels, roles, and the percentage of staff effort allocated to RDMS. Most institutions indicated between one and five people providing research data management support. Starting with cohort 5-10Y and onward, the data show a small increase in the number of personnel providing RDMS (See Figure 3). Liaison, subject or reference librarians, and data librarians are the primary RDMS providers across all cohorts. In the institutions that have offered RDMS for between 3 and 10

> **Most institutions indicated between one and five people providing research data management support.**

years, the data show a significant number of data curators among personnel. The 10Y+ cohort reported several informationists—information specialists typically embedded in the clinical or biomedical research environment—providing RDMS (See Figure 4).

Within those roles, as research services begin, most staff are spending 30 percent or less of their time on RDMS, unless they are data librarians or, in some cases, informationists (See Figure 5, cohort 1-3Y). The 3-5Y cohort reported a similar percent effort except that there were more data curators with full-time RDMS responsibilities. In the cohorts that have offered RDMS for 5 or more years, there is a move toward 50 percent or more of a staff member's time spent on RDMS by liaisons and curators, while some data librarians are spending less time (Figure 5).

Irrespective of role, most respondents reported an increase in staff with primary or partial RDMS responsibilities (83 percent, 36/43), but about 25 percent (11/43) reported a decrease in staff and 11 percent (5/43) had no change. By cohort, most reported an increase (1-3Y, 83 percent, 3-5Y, 91 percent; 5-10Y, 90 percent). Only the 10Y+ cohort reported a decrease in staff positions (62.5 percent, 5/8) rather than an increase (37.5 percent, 3/8) in staff. The 3-5Y cohort reported more staff with RDMS as a primary responsibility. Within other cohorts, numbers of staff with RDMS as a primary responsibility were nearly equal to those with partial responsibility.

Across all cohorts, factors that influenced staffing changes included leadership decisions and strategic planning (24/37, 65 percent), followed closely by service demands (22/37, 59 percent) and employee transitions (14/37, 38 percent). Other influences were funding changes (9/37, 24 percent); fluctuation in supervisory capacities, targeted marketing (each 2/37, 5 percent); and miscellaneous other factors (3/37, 8 percent). The cohort data show that the 1-3Y, 3-5Y, and 5-10Y cohorts were primarily influenced by

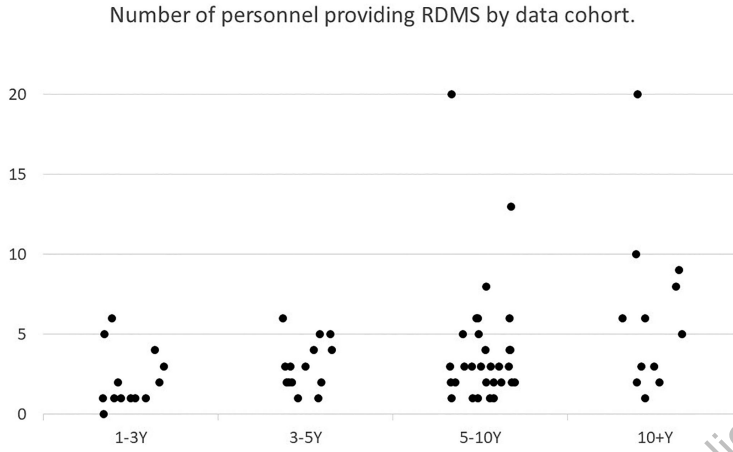Number of personnel providing RDMS by data cohort.

Figure 3. The number of personnel offering research data management services at their institution, presented by data cohort.
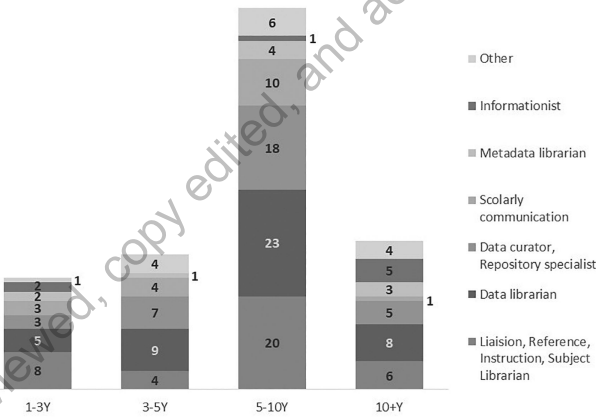
Number and titles of personnel providing RDMS.

Figure 4. Number of RDMS personnel, by job title, presented by data cohort.

leadership decisions or planning, as well as the overall trend of service demands, funding, and employee transitions. For the 10Y+ cohort, these three factors seemed to be less influential, with just two of the five respondents indicating an impact. The 3-5Y cohort indicated more changes due to marketing, as compared to other groups (1/11, 9 percent).

Most respondents indicated that they were unsure of their library budget (34/61). Of those who were able to report, most had budgets under 25 million dollars (22/61) and about half of those were under seven million (12/22). The 1-3Y cohort reported about half of their budgets under 7 million dollars (5/11); the 3-5Y cohort were mostly unsure of their budgets (11/13), while about one quarter of the 5-10Y group reported between
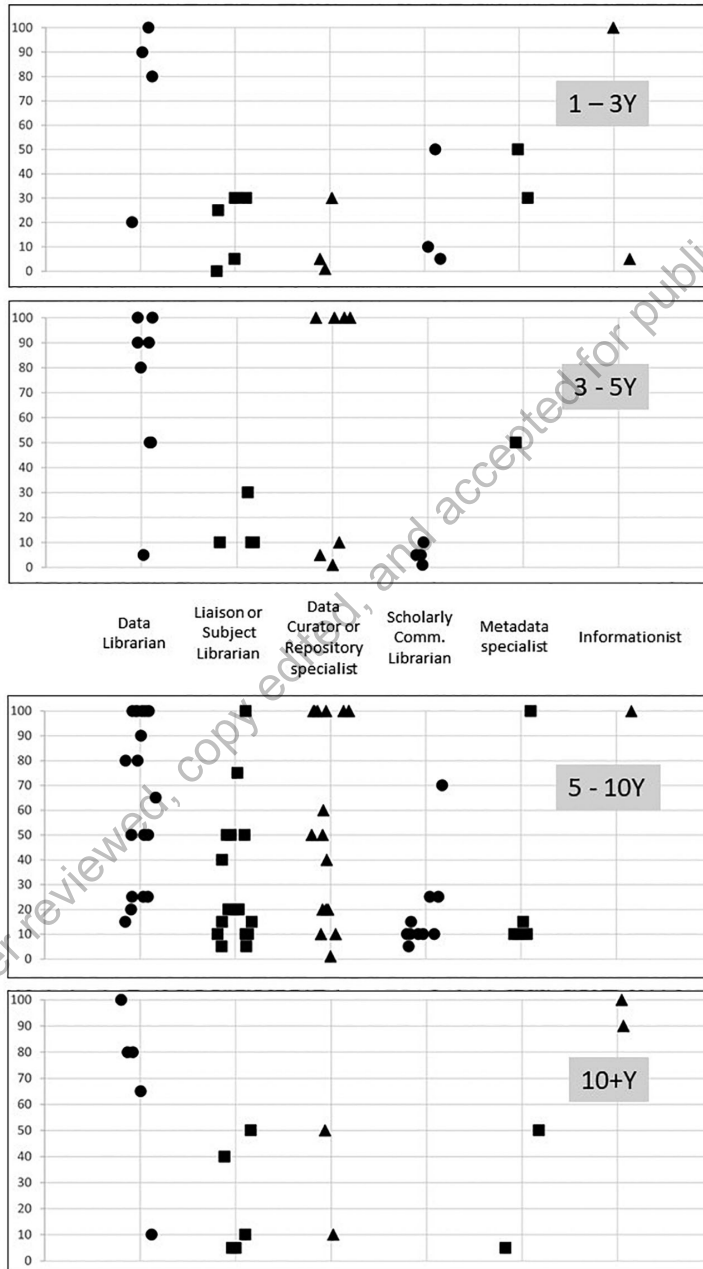
Figure 5. Percentage of personnel time spent on RDMS by job title, presented by data cohort.

8 and 25 million-dollar budgets (7/29). In that same cohort, twenty percent reported having under 7 million dollars (5/29), and the 10+Y cohort was the only one to include reports of 25-50 million-dollar budgets (2/8), but they also included two reports of under 25 million dollars, while the rest were unsure.

Overall, most RDMS is funded by library general funds (35/76, 46 percent), followed by special project or grant funds (11/76, 14.5 percent), library funds specifically designated for RDMS (8/76, 10.5 percent), or other campus entity funding (13/76, 17 percent). (Other campus entities include campus IT (4); a research office, academic units (3 each); the university provost (2); an institute, a grant, the Center for Translational Research, the research computing unit, and a consortium agreement (1 each). This trend carries across all cohorts. When asked how funding has changed, 26 reported increases (26/49, 53 percent), 13 reported decreases (13/49, 26.5 percent), and 10 reported no change (10/49, 20 percent). Of those that reported increases, half noted that the additional funding came from either library general funds or special project or grant increases (10 each). Five noted increases from other funders. Of those reporting decreases, 10 reported decreases from library funds (either general or dedicated), and one reported a decrease in funds from other entities.

Accountability

In addition to funding support and staffing, RDMS also needs support through leadership and service assessment. To this end, survey respondents were asked about organizational structures and their changes, as has been done in previous literature. Further questions explored assessment efforts.

When asked about the organizational structure for RDMS, most answered that a single position within the library was responsible (13/42, 30.1 percent), followed closely by a committee comprised of members from two or more library departments, or members of two or more units at the institution (9/42, 21.4 percent each). Six responded that RDMS organization was based in a single department. Sixty-nine percent of respondents indicated that there has been a shift in organizational structure. This has mostly occurred within the 3-5Y and 5-10Y cohorts (72% and 73% respectively). Sixty-five percent of the 1-3Y cohort experienced organizational restructuring, while half of the 10Y+ cohort did. A few within each cohort elaborated on specifically how the structure has changed. Some in the 1-3Y cohort reported changes in responsibilities, including dedicated roles and department creation (3 responses). Cohort 3-5Y saw more changes in group stakeholders and reporting lines, increased personnel, and more positions with direct leadership responsibilities (8 responses). In the 5-10Y cohort, several respondents indicated that a focused group involved with RDMS had been formed. A few others reported that such groups had dissolved. Several also reported that leadership transitions or changes had both positive and negative effects, such as changes in visibility, stability, and advocacy (13 responses overall). The 10Y+ cohort responses were almost completely the opposite of the 1-3Y group's in that they reported groups and collaborations dissolving and fewer people with RDMS responsibilities (3 responses).

Not surprisingly, across all respondents a majority reported that accountability for RDMS lay with the library department head or director (30/42, 71.4 percent). This was

reflected in the individual cohorts as well. There were six responses that indicated there was no direct accountability (the next most frequent response) and these came from those in the 10Y+ group (4) and from the 5-10Y (2). Three from the 5-10Y cohort reported that the university librarian or dean was accountable. No one indicated that another academic unit was accountable for service provision.

About half of the respondents (23/43, 53 percent) stated that they have assessed their research data management services. This remains true for institutions that have been in service from 3 to 10 years. However, the majority of newer (1-3Y) and longstanding (10+Y) service-offering institutions have not assessed their services. For the ones that did assess services, the primary method was general statistics or counts (22/23, 95.6 percent), followed by user surveys (16/23, 69.6 percent), reflective practice through sum-maries or reports (15/23, 65.2 percent), and lastly by benchmarking (6/23, 26.1 percent) or evidence-based practice (5/23, 21.7 percent). Again, these methods hold true for most cohorts, except for 10Y+ whose responses were evenly distributed among all assessment types. How assessment affected each RDMS was different for each cohort. It affected the education format and amount offered for members of the 1-3Y cohort. For the 3-5Y and 5-10Y cohorts this was still true, but assessment also influenced staff number and time allocation, marketing and external engagement, internal collaboration, and hiring. For the 10Y+ cohort members, assessment informed general planning and hiring.

Planning

Strategic planning is required to ensure proper alignment with larger institutional goals. This, in turn, enables ongoing support, stability, and sustainability for RDMS. The last set of survey questions asked participants about the scope and alignment of strategic planning, if it occurred.

First, respondents indicated whether there is a strategic plan for RDMS at their institution. Over half indicated that there was not one (24/43, 55.8%), with all cohorts following the trend. Data for those that did note a strategic plan show that most plans span from one to four years (12/18 respondents). Five respondents noted having a five-year strategic plan, and only one institution planned beyond that timeframe. Strategic plan goals were likely to be linked to either the library or institutional missions (16 and 14 responses, respectively). A few goals were linked to a specific library initiative (6) or campus initiative (3). Again, the cohorts individually aligned with the overall distribution.

The survey also asked whether the strategic plan had been assessed and by what means. Most respondents indicated that they had not assessed their plans (12/18, 66.6 percent). Of those who had, there was an even distribution between use of general statistics and reflective reports (2 each) and formative assessment or evidence-based practice (1 each) as the chosen assessment methods. Even if respondents did not note a formal plan assessment, most respondents indicated that they were able to complete their strategic goals at least partially (11/16, 68.7 percent). Factors that affected goal completion included time, engagement with external departments and external sup-porters, funding, "de-siloing," position realignment, perseverance, and cooperation. Most of the goals that weren't completed were postponed (8/14, 57.1 percent), with two respondents indicating pandemic-related delays. The rest of the goals were altered

(4/18) or eliminated (2/18). Reasons for these changes or eliminations included facing unrealistic expectations, hiring freezes, staff changes, lack of professional development funds, or switching away from a discipline-specific approach.

For those that didn't have a formal strategic plan (24/43, 55.8 percent), the survey asked if there was another method used for guiding RDMS growth and development. Five respondents indicated that there was nothing else linked, and two responded that RDMS was guided by stakeholder or committee decisions. The majority of these institutions indicated that they linked their RDMS goals to either a specific job description or library mission (5 each). Others indicated that they were linked to either an institution mission (3), library (3), or campus initiative (2). None of these respondents reported any formal assessment measures, but seven indicated that goals were at least partially completed. Factors that affected goal completion were staffing, staff motivation, demonstrated need, stakeholder coordination, technological infrastructure, funding, and policy development. One respondent noted that having technical infrastructure, coordination, and policy development happening concurrently helped. Goals that weren't completed among this group were equally likely to be altered (4) or postponed (3). Only one respondent noted that they had eliminated incomplete goals. Reasons given for the non-completion of these goals included general scaling-down of services, loss of oversight or vision, position elimination or a change in positions to become less data specific, or a lack of repository support.

Discussion

Service offerings

Past service profiles have shown that institutions primarily offered what is currently described as advisory services.[13] In the cohort data presented here, the 1-3Y group offered fewer services than the other cohorts and filled fewer service requests per year, which is unsurprising as they are likely to have the least capacity in terms of staff or resources. It is interesting that the 10Y+ cohort members also reported offering fewer services and addressing fewer requests. It is unclear whether their service offerings contracted or if they stayed limited throughout their service history. Respondents from each cohort reported an increase in technical services, particularly related to work that involves repositories, such as preservation, curation, metadata creation, and documentation. These types of service offerings were most robust for the 5-10Y cohort. There have been

Respondents from each cohort reported an increase in technical services, particularly related to work that involves repositories, such as preservation, curation, metadata creation, and documentation.

fewer services offered in applied technical services like data visualization or analysis, and this is likely related to staff skill gaps. Despite the stratification in service profile, the longer the service had been established, the more service requests were fulfilled. In addition, service scope for those institutions with long-established service was closer to covering the entire research life cycle.

With education and training being a service that has been consistently offered over time, as documented in the literature, and reported by most cohorts here, it is unsurprising that it was reported as the easiest service to implement. There has been a strong effort to develop open RDMS training materials and there is a robust body of literature that documents education efforts. One of the earliest rationales presented RDMS education as an extension of information literacy and that was seen in the response explanations of why this was easier. Other reasons mentioned in the responses included that there was some autonomy in offering education and lower costs to be able to provide this service. Further study would show whether this correlates to the early service providers who may have less autonomy or funding.

But despite these seemingly low barriers, for other respondents, education or training is considered difficult, enough so that it is listed as the second most challenging barrier overall. Those that found it difficult reported barriers in personal relations and marketing the service. In addition, low cost is not no cost, and cost was mentioned as a trade-off. Despite the alignment with information literacy, education on data management is still mostly elective. Since many of the RDMS providers are in liaison roles, this appears to add to existing instruction responsibilities, especially since staffing does not keep pace with RDMS demands.

The remainder of the "easy" services to implement were a mix of technical and advisory services. Most of the advisory services (DMP planning, consultations, finding and citing data, and data purchase) were reported to be easy to offer because of appropriate staffing, existing staff skills and alignment with other services already present. Technical services (data storage, digital humanities, and data visualization) were made easier by partnerships and hiring, in some cases also alignment with other work. Trade-offs for both technical and advisory revolved around shifting workloads, time allocation, and job descriptions.

The most difficult service to implement was storage, curation, or preservation—most often associated with repositories. Difficulties arose around campus politics and stakeholders, but also with costs and cultural resistance to sharing. It is unsurprising that large, expensive projects like repositories face challenges. The numbers of people involved, technical specifications, and prospective or current demand are factors difficult to align. Staffing, skills, and costs are constant barriers to this type of project. Many of the early surveys advocated for developing "buy-in" from stakeholders early and to address RDMS as collaborative groups.[14] For projects on this scale, it would appear to be the efficient path, but the current survey responses indicate otherwise, noting trade-offs including changes to shared governance and collegial tensions. These types of challenges appear in the responses related to other services deemed hard to implement as well. For example, most respondents reported that there had been changes to their organizational structure

> **The most difficult service to implement was storage, curation, or preservation—most often associated with repositories. Difficulties arose around campus politics and stakeholders, but also with costs and cultural resistance to sharing.**

that impacted RDMS, with the 3-5Y and the 5-10Y cohorts reported most organizational changes, including changes in stakeholders, reporting lines, and governance groups forming or dissolving. These organizational changes are likely the contributing factors behind the difficulty in implementing repository and storage-related services.

Staffing and Funding

Predictably, institutions from all cohorts indicated a shift in needed skills among their RDMS staff. This demonstrates that RDMS providers are not being complacent in understanding their institutional research needs. Having a skilled workforce to support RDM was identified early in the literature as a requirement for libraries to be able to support evolving modes of academic scholarship.[15] According to the data presented here, the longer an institution offers RDMS, the more diverse the skills required to support that service. This aligns with the expanding service profiles observed over time. However, the profile of those skills appears to have shifted from subject- or discipline-specific skills such as GIS and informatics to data science skills like programming, coding, and understanding specific software. The only cohort that differed greatly were ones that offered RDMS for over ten years. The desired skills for this group were administrative or leadership in nature. With staff capacity repeatedly mentioned as a barrier, it is logical that the skill sets evolved to ones that could be used broadly across disciplines and applied to a variety of research needs.[16]

With this constant service demand, it doesn't appear that the number of dedicated staff providing RDMS has followed proportionally. There are more people with dedicated responsibilities for RDMS and more that have part-time responsibility, but only in institutions that have offered services for at least five years. For institutions starting RDMS, it appears that many positions are either solo with 100 percent responsibility or held by existing positions that have RDMS added on top of other work (30 percent RDMS time or less). As RDMS becomes more established at an institution, particularly starting at the five-year mark, there are likely to be more staff with at least half of their job dedicated to RDMS activities. Most of this work across any cohort still resides with data librarians and subject or liaison librarians, as was also true in previous studies.[17] One might expect that the longer RDMS is offered, the greater amount of time there would be to secure positions, but the data do not support this notion. The 10Y+ cohort has similar staffing, titles, and roles to the 1-3Y and 3-5Y cohorts. The 10Y+ cohort would have been in service at the time of Tenopir's 2012 survey, which indicated that most respondents had or were planning to reassign staff to RDMS rather than hire new personnel. This approach is born out in the data for the present study.

Regardless of the number of years providing RDMS, all cohorts' staffing changes have been equally influenced by both leadership decisions and service demands as shown in this study. It will also be interesting to see whether those in the 5-10Y cohort are able to maintain their staffing levels as they continue their service. Most respondents reported that RDMS is supported by general library funds, followed by grant or project soft funds, irrespective of cohort. One quarter of this group also reported funding decreases in these lines. Early literature reported that most RDMS funding came from general library funds.[18] Dedicated funding lines may be necessary to secure RDMS positions and

protect service offerings. Tenopir's follow-up study indicated that staffing, skills, and funding are still barriers to providing RDMS and this data corroborates that assertion.[19]

Accountability

In this study, the most noted organizational structure for supporting RDMS was a single position, followed by a group or committee. This finding reflects the results of the earliest RDMS studies. [20] However, all cohorts in this study had experienced organizational changes, with the 3-10Y group documenting the most. The cohort with the youngest programs reported the most changes in job responsibilities and department creation. This aligns with activities needed to offer a standardized service profile. It is difficult to maintain services when they rely on one person or if the responsibilities aren't specifically written into job descriptions. The longer-standing RDMS programs identified changes in stakeholders, personnel number, and leadership. As services become more established, it makes sense that the increased service numbers, and therefore visibility, encourages stakeholder interaction. This helps to ensure that their needs are addressed in a meaningful way. It is notable that the 10Y+ cohort saw collaborations dissolving and employed fewer people with RDMS responsibilities. Across all cohorts a strong majority indicated that accountability lay with a department head or director but not necessarily the library dean (if there was one). No one indicated that another academic unit was held responsible for RDMS, but this study didn't evaluate services, infrastructure, or accountability offered by parallel units within an institution.

> **It is difficult to maintain services when they rely on one person or if the responsibilities aren't specifically written into job descriptions.**

Regardless of leadership, internal accountability might be inferred through service assessment. About half the respondents assessed services primarily through statistics or counts, followed by user surveys and summaries or reports. The effect of these assessments varies by cohort. Like the responses about staffing and funding, the 1-3Y cohort assessment realigns the format education takes and amount of education that can be offered. However, for the 3-10Y cohort assessment affects the service profile but also affects staffing and funding. It is likely that these assessments are used to justify increasing the service profile as well as the staff numbers seen in this group. The 3-10Y cohort is also the group with the greatest number of staff with partial RDMS responsibilities, which could also reflect shifting responsibilities of existing staff. For the longest-standing cohort, assessment reportedly affects general planning and hiring, but as previously seen, this cohort is staffed at the same levels as those with newer RDMS.

Planning

Within the literature, particularly within the abundance of RDMS needs assessments, RDMS provision appears to be reactive rather than proactive.[21] The purpose of these initial institutional scans was to discover the existing needs so that services could be developed to address them. The early service snapshots show that many planned to offer a variety

of RDMS but did not yet do so.[22] However, by 2019, Tenopir reported "libraries have yet to fully respond to the growth of RDMS that do not fit with their existing structure."[23]

As mentioned previously, most funds come from general budget lines and not dedicated funding lines. This allows flexibility in addressing the variety of needs libraries serve, but one could argue that the demand for RDMS is growing and requires dedicated funding. Strategic planning is required to effectively staff these services and provide adequate funding, and to manage the limited resources institutions can devote to RDMS amid competing priorities.

Increasingly, partnership is required to accomplish some RDMS needs. Early studies indicated that common partners include the university office of research, information technology centers, discipline-specific departments or institutes, and similar entities.[24] Strategic planning will strengthen these partnerships (whether internal or external) through setting a clear vision, defining responsibilities and boundaries, outlining shared goals, and providing measurable outcomes. It is notable that Gowen and Meier reported that half of the strategic plans that mentioned RDMS in 2014 no longer mentioned it by 2019.[25] Their supposition is that they are no longer mentioned because the "goal has been achieved" and RDMS is rolled into library operations. However, libraries might do well to keep RDMS as part of their strategic plans, so that they can develop into their full capabilities for RDMS, as Tenopir proposed.[26]

Limitations

The data in this study may not be representative of all academic institutions as respondents self-selected to participate in this research. In addition, the data presented here was collected just ahead of the COVID 19 pandemic. The resulting change in academic operations, particularly library services, has yet to be fully realized and this will likely affect RDMS provision. However, research on this topic still focuses on large, high research institutions, as they are the primary RDMS providers. A study of smaller institutions may yield different results.

Future directions

RDMS may be approaching the level of a standard service in libraries; however, across institutions, support that would make this work sustainable appears immature. Many of the studies on research data management services, this one included, report that libraries and other RDMS providers are creating broad service profiles that appear to be "just in case" they are needed rather than right-sized for the institution. Research on this topic focuses on service provision and not the underlying trade-offs and costs of offering that service. The impact of RDMS work has yet to be determined. There are far more studies that describe RDMS service implementation than describe outputs or outcomes. Asking questions such as:

- Do these services affect grant funding success rates?
- Do RDMS assist in preparing graduates for their careers? or
- Do RDMS support community partners directly?

could more effectively calibrate RDMS scope and therefore determine staffing, funding, and skill levels for long-term success. Additionally, there are a few institutions that have been able to support RDMS for a decade or more. Detailed descriptions of how they have sustained their services over that time would be valuable.

More work centering on accountability, planning, and funding is needed as well. Most respondents indicated assessment was conducted through statistics, counts, or user surveys. Studies mapping services to curricula, discipline competencies, research objectives, and funder requirements (like compliance with the NIH Data Management and Sharing Policy) could help determine which stakeholders should be contributing to these efforts. In addition, these efforts would help to specify support needs to provide a more robust rationale for long-term investment. Libraries are almost always considered cost centers and as such, RDMS currently appears to be supported with general funds rather than dedicated funding lines. A deeper look into funding sources, changes, and options would be useful to clarify how RDMS is supported amid other competing financial priorities.

Research that addresses RDMS support from smaller institutions would be welcome. While these institutions have different missions and values because research is not always their priority, research is conducted there and RDMS profiles are likely to be different. It follows that the approach to staffing, funding, accountability, and planning will also be different, and there are few studies that describe any of these aspects at non-research-intensive institutions.

Irrespective of institution size, development of standardized instruments for studying RDMS work would help everyone plan for RDMS. Several disruptors such as the NIH Data Management and Sharing Policy requirements; the introduction of artificial intelligence large language models like ChatGPT which affect data privacy, security, and use; and COVID-19 severity decreases will affect RDMS provision and operations. Studies examining the intersection between RDMS and these topics could support institutions as they manage changes driven by external factors like these.

Conclusion

In 2013 Minglu Wang reported support for RDMS "is still limited, national-level coordinated infrastructure is still very weak, and … researchers lack incentive to adopt new tools and software." [27] Library directors believe that libraries need to offer RDMS to be relevant, that losing data jeopardizes future scholarship, and that librarians should be stewards of all types of scholarship including data. [28] Research data service provision is growing and evolving within libraries, but the data presented here align with that from previous studies. Services are expanding but not necessarily in a concerted manner, with intentional planning and support.

Libraries in their initial years of offering data services have fewer RDMS staff, offer fewer services, and favor the advisory service model. The data presented here suggests that those in this stage of RDMS use the time to engage in strong strategic planning to identify partners, funding, and chart a path for growth or stability. This is also the time to develop and integrate assessments mapped to curriculum or institutional priorities so that it is easier to justify continued investment. In-depth assessments would also clarify what services would be useful and what scale would be feasible to achieve in the next growth stage. Professional development or hiring should align with the projected

service profile and include developing leadership and management skills, as the shift in organizational structure and workloads that comes with the middle service years will need them.

Libraries that are past their initial RDMS years have critical decisions to make. The data presented shows that these groups tend to expand services the most, but uptake of— and support for— those services is not proportional. RDMS leaders in these institutions should consider an intentional, focused service profile, possibly reducing service options. If services are to expand, evaluation of potential trade-offs needs to be carefully considered before implementation, as few institutions can sustain exhaustive services for a decade or longer. Converting from general statistics-counting to more meaningful assessments is useful here as well. More robust assessment can help to clarify vision and direction, as strategic partnerships may struggle with the employee and organizational transitions that were reported in the 3-10Y cohort. Leadership responsibilities were greater in this cohort also. Investing heavily into professional development that emphasizes vision, planning, and effective conflict or change management should also contribute to RDMS sustainability.

The responses from libraries that had passed ten years of RDMS show that growth is not always sustainable or desired. This cohort reported data that looks like those starting as RDMS providers in many respects. This cohort has the advantage of being able to review their history and learn from it. Strategies from the early- and middle- service years are applicable here and, combined with their experience, institutions that have sustained RDMS for this long should be able to accelerate their growth, if warranted. If their RDMS is stable, succession planning may be necessary at this stage as well.

The data presented here may help institutions navigate RDMS growth by better understanding the intersections between service profiles, staffing, funding, accountability, and planning at different points in time. Libraries that choose to engage in strategic planning for RDMS may more effectively allocate their time, personnel, and financial resources to realize their institutional vision and create stability for research data management services.

Acknowledgments

Data availability statement

De-identified data associated with this study will be available through INDIGO, the institutional repository for the University of Illinois Chicago. DOI:10.25417/uic.25017005

Tina M. Griffin is an Associate Professor and Liaison Librarian in the Information Services and Research Department at the University of Illinois Chicago. She may be reached via email at: tmcg@uic.edu.

Margaret Janz is an interpreter for Great Parks of Hamilton County in Cincinnati, Ohio. She may be reached via email at: mjanz@greatparks.org.

Appendix A

Survey Instrument

Note: Question numbers were not visible to the participants. Questions are displayed in the order they were presented to the participants. Participants may not have seen all questions due to the skip/display logic coded into the survey (not shown).

Research Data Management Service Sustainability

Start of Block: General Questions

Q1.1
We are inviting you to participate in a survey on data management services, with a focus on factors that affect sustainability of these services.

The survey is expected to take about 30-45 minutes to complete and has been approved by the University of Illinois at Chicago IRB Office and has been determined as exempt. You will not directly benefit from participating in this web survey today.

Participation in the study is voluntary. You may choose not to participate. If you decide to participate in this research study, you may withdraw at any time. If you decide to withdraw, you will not be penalized. Any identifiable information collected is for the purpose of aggregating data or for disambiguating responses from within the same institution. These identifiers will be removed before publication or dissemination of any data.

Thank you for participating in this important research. If you have any concerns about the survey, please contact the investigators below.

Q1.2 Please indicate your participation in this research.
- o I have read the consent form and agree to participate. (continue to questions) (1)
- o I have read the consent form and do not wish to participate. (exit survey) (2)

Q1.3 Welcome!

Thank you for your time in taking this survey. You will be able to move back and forward through the questions.

This survey has five sections. In this first section, please answer general questions about your library and data services.

Q1.4 How long has your library been offering research data services?
- o We are not currently offering services or have offered them less than one year. (1)
- o 1 - 3 years (2)
- o 3 - 5 years (3)
- o 5 - 10 years (4)
- o More than 10 years (5)

Q1.5 How many people are providing research data services in your library? (Please enter a whole number)

_____

Q1.6 Which of these positions are involved in data services in your library? (select all that apply)
- o Data librarian (1)
- o Informationist (2)
- o Liaison/Reference/Instruction/Subject librarian (3)
- o Data curator/Repository specialist (4)
- o Scholarly communication librarian (5)
- o Metadata librarian (6)
- o Other (7) _____

Q1.7 What percentage of their time does these positions spend on research data services? (Please enter a whole number)
- o Data librarian (1) _____
- o Informationist (2) _____
- o Liaison/Reference/Instruction/Subject librarian (3) _____
- o Data curator/Repository specialist (4) _____
- o Scholarly communication librarian (5) _____
- o Metadata librarian (6) _____
- o Other (7) _____

Q1.8 What is your library's annual budget (total expenditures)?
- o Less than 7 million (1)
- o 8 - 25 million (2)
- o 26 - 50 million (3)
- o 51 - 70 million (4)
- o Over 70 million (5)
- o Unsure (6)

End of Block: General Questions

Start of Block: Service Model part 1

Q2.1 In this second section, please answer questions about your service model for data services.

Q2.2 What research data services do you offer? (select all that apply)
- o Education or training (1)
- o Data purchase or subscription (2)
- o Data analysis (3)
- o Data visualization (4)
- o Data curation (5)

o Data preservation (6)
o Offer resources (software access or license, equipment, etc.) (7)
o Data management planning (8)
o Data storage (repository, local, or other storage option) (9)
o Finding data (10)
o Metadata creation (11)
o Data citation (12)
o Data sharing or access (13)
o Deidentifying data (14)
o Data documentation (not metadata creation) (15)
o Other (16) _____

Q2.3 What is the approximate number of service requests that you fulfill each year?
o Education or training (1) _____
o Data purchase or subscription (2) _____
o Data analysis (3) _____
o Data visualization (4) _____
o Data curation (5) _____
o Data preservation (6) _____
o Offer resources (software access or license, equipment, etc.) (7) _____
o Data management planning (8) _____
o Data storage (repository, local, or other storage option) (9)_____
o Finding data (10) _____
o Metadata creation (11) _____
o Data citation (12) _____
o Data sharing or access (13) _____
o Deidentifying data (14) _____
o Data documentation (not metadata creation) (15) _____
o Other (16) _____

End of Block: Service Model part 1

Start of Block: Tech vs advisory loop

Q3.1 Would you describe your level of service for ${lm://Field/1} as technical (perform tasks or analysis) or advisory (provide information, training, or guidance) or both?
o Technical (1)
o Advisory (2)
o Both (3)

End of Block: Tech vs advisory loop

Start of Block: Service model part 2

Q4.1 Has the menu of services you offer changed since they were implemented?
- o Yes (1)
- o No (2)

Q4.2 How has your service menu changed?
- o We shifted to more technical services than advisory services (2)
- o We shifted to more advisory services than technical services (1)
- o We offer more of both technical and advisory services (3)
- o We offer less of both technical and advisory services (4)

Q4.3 What services were added?

_____
_____
_____
_____
_____

Q4.4 What services were dropped?

_____
_____
_____
_____
_____

Q4.5 What led to these service changes? (select all that apply)
- o Leadership decisions (1)
- o Funding increase or decrease (2)
- o Demand for services (3)
- o Employee transitions (4)
- o Marketing response (5)
- o Strategic planning (6)
- o Other (7)_____

Q4.6 What service was the easiest to implement?

_____

Q4.7 What factors made this service easy to implement?

_____
_____
_____
_____
_____

Q4.8 What were the trade-offs that had to be made to implement this service?
_____
_____
_____
_____
_____

Q4.9 What service was the hardest to implement?
_____

Q4.10 What factors made this service difficult to implement?
_____
_____
_____
_____
_____

Q4.11 What trade-offs had to be made to implement this service?
_____
_____
_____
_____
_____

Q4.12 Has there been a shift in the skills present or needed to offer research data services?
   o Yes (1)
   o No (2)

Q4.13 What skills have been needed or are now present?
_____
_____
_____
_____
_____

Q4.14 What skills are no longer needed or are no longer present?
_____
_____
_____
_____
_____

End of Block: Service model part 2

Start of Block: Infrastructure and Support

Q5.1 In this third section, please answer questions about infrastructure and support for your data services.

Q5.2 How are your research data services funded? (select all that apply)
- o Library budget - general funds (1)
- o Library budget - specific allocation/line (2)
- o Special project/grant funds (3)
- o Other campus entity contributes (Please list who contributes) (4) _____
- o Fee for service (5)
- o Unsure (6)
- o Other (7) _____

Q5.3 How has funding for research data services changed?
- o Library funding has increased (1)
- o Library funding has decreased (2)
- o Special project/grant funds have increased (3)
- o Special project/grant funds have decreased or ended (4)
- o Other entity funding has increased (5)
- o Other entity funding has decreased or ended (6)
- vFees to users have increased (7)
- o Fees to users have decreased (8)
- o Funds have not changed (9)

Q5.4 How has staffing changed from the beginning of offering research data services? (select all that apply)
- o We have more staff dedicated to RDS as their primary responsibility (1)
- o We have less staff dedicated to RDS as their primary responsibility (2)
- o We have more staff dedicated to RDS as their partial responsibilty (3)
- o We have less staff dedicate to RDS as their partial responsibility (4)
- o There have been no staffing changes (5)

Q5.5 What influenced staffing changes? (select all that apply)
- o Funding increase or decrease (1)
- o Demand for services (2)
- o Employee transitions (3)
- o Supervisory capacity (4)
- o Marketing response (5)
- o Strategic planning/leadership decisions (6)
- o Other (7) _____

Q5.6 How has research data services been marketed or promoted? (select all that apply)
- o Print materials (posters, flyers, signs) (1)
- o Electronic communication (digital signs, websites, email or listerv) (2)
- o Liaison or staff outreach (3)
- o Leadership advocacy (4)

o Other (5) _____

o Unsure (6)

o We have not market or promoted these services (7)

Q5.7 Have you assessed your research data services?

o Yes (1)

o No (2)

Q5.8 How has your research data services been assessed?

o General statistics and counts (1)

o Reflective practice (summaries, reports) (2)

o Formative assessment (user surveys) (3)

o Developmental assessment (benchmarking) (4)

o Evidence based practice (5)

Q5.9 How has assessment efforts influenced your decision making for maintaining or changing services?

_____

_____

_____

_____

_____

End of Block: Infrastructure and Support

Start of Block: Accountability and Planning

Q6.1 In this fourth section, please answer questions about accountability and planning for data services.

Q6.2 What best describes your organizational structure regarding research data services?

o A committee/group composed of members from two or more departments within the library (1)

o A committee/group composed of members from the library and members from other departments/units at the institution (2)

o A single position within the library (3)

o A single department within the library (4)

o Other (5) _____

Q6.3 Has there been a shift in organizational structure?

o Yes (1)

o No (2)

Q6.4 How has the organizational structure changed?

_____
_____
_____
_____
_____

Q6.5 Where does the accountability lie for research data service provision?
    o Library Dean or University Librarian (1)
    o Library department head or director (2)
    o Other academic unit (3) _____
    o Other (4) _____
    vThere is no direct accountability (5)

Q6.6 Has there been a strategic plan for research data service growth and development
    o Yes (1)
    o No (2)

Q6.7 What were the broad goals stated in the plan?

_____
_____
_____
_____
_____

Q6.8 What was the time frame for the plan?
    o 1 - 2 years (1)
    o 3 - 4 years (2)
    o 5 years (3)
    o More than 5 years (4)

Q6.9 What were the goals for research data services linked to?
    o Library mission or values (1)
    o Institution mission or values (2)
    o Library initiative (3)
    o Campus initiative (4)
    o Other (5) _____
    o The goals were not directly linked (6)

Q6.10 Have the goals been assessed?
    o Yes (1)
    o No (2)

Q6.11 How have these goals been assessed?
- o General service statistics (1)
- o Reflective practice (summaries or reports) (2)
- o Formative assessment (user surveys) (3)
- o Developmental assessment (benchmarking) (4)
- o Evidence based practice (5)

Q6.12 Were goals able to be completed?
- o Yes - all of them (1)
- o Yes - some of them (2)
- o No (3)

Q6.13 What was the key factors in being able to complete the strategic goals?

_____
_____
_____
_____
_____

Q6.14 What happened to goals that didn't get completed?
- o They were postponed (1)
- o They were altered / changed (2)
- o They were eliminated / abandoned (3)

Q6.15 Please elaborate on what goals were postponed, altered, or eliminated.

_____
_____
_____
_____
_____

Q6.16 What was the method used to plan for research data service growth and development?

_____
_____
_____
_____
_____

Q6.17 What was the time frame for the plan?
- o 1 - 2 years (1)
- o 3 - 4 years (2)
- o 5 years (3)
- o More than 5 years (4)

Q6.18 What was the plan linked to?

    Job description (1)
- o Library mission or values (2)
- o Institution mission or values (3)
- o Library initiative (4)
- o Campus initiative (5)
- o Other (6) _____
- o Was not linked to anything (7)

Q6.19 Have the goals for this plan been assessed
- o Yes (1)
- o No (2)

Q6.20 How were the goals assessed?
- o General statistics (1)
- o Reflective practice (summaries or reports) (2)
- o Formative assessment (user surveys) (3)
- o Developmental assessment (benchmarking) (4)
- o Evidence based practice (5)
- o Employment performance review (6)

Q6.21 Were the goals able to be completed?
- o Yes - all of them (1)
- o Yes - some of them (2)
- o No (3)

Q6.22 What were the key factors in being able to complete the goals?

_____
_____
_____
_____
_____

Q6.23 What happened to the goals that didn't get completed?
- o They were postponed (1)
- o They were altered/changed (2)
- o They were eliminated/abandoned (3)

Q6.24 Please elaborate on what goals were postponed, altered, or eliminated.

_____
_____
_____
_____
_____

End of Block: Accountability and Planning

Start of Block: Closing

Q7.1 In this forth section, we give you an opportunity to offer your advice on data services.

Q7.2 If you could offer advice on starting and/or maintaining research data services, what would be your top three points?

   o Point 1 (1) _____
   o Point 2 (2) _____
   o Point 3 (3) _____

Q7.3 Do you consider yourself to be in a leadership position regarding research data services at your library?

   o Yes (1)
   o No (2)

End of Block: Closing

Start of Block: General Demographics

Q8.1
To close, please answer these few demographic questions so that we may aggregate and analyze the data.

These answers will be used to collapse responses from within the same institution and to group data among similar institutions, but will be deleted before analysis.

   o Your Name (1) _____
   o Your institution (2) _____
   o Your title/position (3) _____
   o Your department or unit (4) _____

End of Block: General Demographics

Appendix B

Disclosures

Human Subject Research

The survey instrument used for this study was reviewed by the Institutional Review Board at the University of Illinois Chicago and the University of Pennsylvania and were determined to be exempt (#2020-0938 UIC and #844332 UPenn). Informed consent was obtained from all respondents.

Author contribution statement

TG conceived the project, planned the study design, and managed the project. TG and MJ distributed the survey, collected and analyzed the data. TG wrote the manuscript

with input from MJ. Questions about this work should be directed to TG as corresponding author.

Notes

1. Abigail Goben and Tina Griffin, "In Aggregate: Trends, Needs, and Opportunities from Research Data Management Surveys," *College & Research Libraries* 80, 7 (January 1, 2019): 903–24, https://doi.org/10.5860/crl.80.7.903.

2. Catherine Soehner, Catherine Steeves, and Jennifer Ward, *E-Science and Data Support Services: A Study of ARL Member Institutions*, (Washington, D.C.: Association of Research Libraries, August 2010), http://files.eric.ed.gov/fulltext/ED528643.pdf.

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

4. DataOne Community Engagement & Outreach Working Group, "Introduction to Data Management," 2017, https://dataoneorg.github.io/Education/lessons/01_management/01_management.pdf

5. David Fearon et al., *SPEC Kit 334: Research Data Management Services* (Washington, D.C.: Association of Research Libraries, July 2013), https://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/.

6. Andrew M. Cox and Stephen Pinfield, "Research Data Management and Libraries: Current Activities and Future Priorities.," *Journal of Librarianship & Information Science* 46, 4 (December 2014): 299–316, https://doi.org/10.1177/0961000613492542.

7. Carol Tenopir et al., "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?" *Journal of eScience Librarianship* 4, 2 (2015): e1085, https://doi.org/10.7191/jeslib.2015.1085.

8. Tenopir et al., "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?"

9. Holly Yu, "The Role of Academic Libraries in Research Data Service (RDS) Provision," *Electronic Library* 35, 4 (July 2017): 783–97, https://doi.org/10.1108/EL-10-2016-0233.

10. Carol Tenopir et al., "Research Data Services in Academic Libraries: Where Are We Today?" (Choice White Paper no. 5, ACRL/Choice, 2020), https://www.choice360.org/research/research-data-services-in-academic-libraries-where-are-we-today/.

11. Jane Radecki and Rebecca Springer, "Research Data Services in US Higher Education" Ithaka S&R, last modified November 18, 2020, https://doi.org/10.18665/sr.314397.

12. Elise Gowen and John J. Meier, "Research Data Management Services and Strategic Planning in Libraries Today: A Longitudinal Study," *Journal of Librarianship and Scholarly Communication (JLSC)* 8, 1 (April 18, 2020): 2336, https://doi.org/10.7710/2162-3309.2336.

13. Soehner, Steeves, and Ward, "E-Science and Data Support Services"; Tenopir, Birch, and Allard, "Academic Libraries and Research Data Services"; Tenopir et al., "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?"; Tenopir et al., "Research Data Services in Academic Libraries: Where Are We Today?"; Fearon et al., "SPEC Kit 334."

14. Fearon et al., "SPEC Kit 334"; Soehner, Steeves, and Ward, "E-Science and Data Support Services."

15. Tenopir, Birch, and Allard, "Academic Libraries and Research Data Services"; Tenopir et al., "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?"; Andrew Creamer et al., "Data Curation and Management Competencies of New England Region Health Sciences and Science and Technology Librarians," (poster, April 6, 2011), http://escholarship.umassmed.edu/escience_symposium/2011/posters/8.

16. Tenopir et al., "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?"; Tenopir et al., "Research Data Services in Academic Libraries: Where Are We Today?"; Soehner, Steeves, and Ward, "E-Science and Data Support Services"; Carol Tenopir et al., "Research Data Services in European Academic Research Libraries," *Liber Quarterly: The Journal of European Research Libraries* 27, 1 (January 2017): 23–44, https://doi.org/10.18352/lq.10180.

17. Fearon et al., "SPEC Kit 334"; Tenopir et al., "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?"; Tenopir et al., "Research Data Services in Academic Libraries: Where Are We Today?"; Gowen and Meier, "Research Data Management Services and Strategic Planning in Libraries Today: A Longitudinal Study"; Radecki and Springer, "Research Data Services in US Higher Education."

18. Fearon et al., "SPEC Kit 334."

19. Tenopir et al., ""Research Data Services in Academic Libraries: Where Are We Today?"

20. Tenopir, Birch, and Allard, "Academic Libraries and Research Data Services"; Tenopir et al., "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?"

21. Goben and Griffin, "In Aggregate: Trends, Needs, and Opportunities from Research Data Management Surveys."

22. Tenopir, Birch, and Allard, "Academic Libraries and Research Data Services"; Tenopir et al., "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?"

23. Tenopir et al., ""Research Data Services in Academic Libraries: Where Are We Today?," 3.

24. Tenopir, Birch, and Allard, "Academic Libraries and Research Data Services"; Soehner, Steeves, and Ward, "E-Science and Data Support Services"; Fearon et al., "SPEC Kit 334."

25. Gowen and Meier, "Research Data Management Services and Strategic Planning in Libraries Today: A Longitudinal Study."

26. Tenopir et al., ""Research Data Services in Academic Libraries: Where Are We Today?," 3.

27. Minglu Wang, "Academic Library, e-Science/e-Research, and Data Services in a Broader Context," Paper presented at Imagine, Innovate, Inspire, the ACRL 2013 Conference, Indianapolis, IN, April 2013, https://doi.org/10.7282/T3222RT1.

28. Tenopir et al., "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?," 8.