

In Aggregate: Trends, Needs, and Opportunities from Research Data Management Surveys

Abigail Goben, MLS
Associate Professor
Data Management Coordinator
Library of the Health Sciences, University of Illinois at Chicago
1750 W Polk St.
Chicago, IL 60612
agoben@uic.edu
312-996-8292

Tina Griffin, MLIS
Assistant Professor
Information Services Librarian/Liaison
Library of the Health Sciences, University of Illinois at Chicago
1750 W Polk St.
Chicago, IL 60612
tmcg@uic.edu
312-996-8344

Preliminary data presented at:

In Aggregate: Trends, Needs, and Opportunities from Faculty Research Data Management Surveys
[presentation]. IASSIST Annual Conference, May 23-26, 2017, Lawrence, KS.

Abstract

A popular starting point for libraries engaging in research data management (RDM) services is a needs assessment (NA); and a preliminary count identified more than fifty published NA case studies.

However, no overarching analysis has yet been conducted. The authors compared assessments to characterize the case study institution types; establish the target population assessed; discover cross-institutional trends both in the topics covered and the issues identified; and determine remaining gaps in the literature. Thirty-seven studies conducted in the United States were included. Twenty-five were at public, Doctoral, Highest Research institutions. The most frequently assessed respondents were faculty (n= 3847). The most frequent topics involved storing, sharing, and maintaining long-term access to data. Gaps include assessing students, staff, and non-faculty researcher needs; determining needs at various sized and degree granting institutions; and investigating RDM needs for non-STEM disciplines.

Introduction

As awareness of research data management has grown due to federal funding requirements, university libraries have explored the service needs of their institutions and researchers. A popular starting point for many libraries has been a campus needs assessment to determine where researchers encounter barriers and, accordingly, where services might be most appropriately tailored. In the library literature, the overwhelming majority of published needs assessments are individual institution case studies and to date, no overarching analysis has been performed. The object of this review is to detail the institution types where needs have been assessed; establish the target populations assessed; discover cross-institutional data management needs both in the questions asked and the responses given; and determine what gaps in the literature yet remain.

Literature Review

Academic librarians have developed research data management (RDM) services and research over the past decade as researcher demands have derived from the implementation of the National Science Foundation Data Management Plan requirement¹ and the Office of Science and Technology Policy memorandum for increased access to federally funded research.² Librarians have undertaken reskilling³ and there has been extensive research into the preparation and engagement of librarians with research data management.⁴ This has led to several papers describing opportunities for librarians to engage with researchers to meet these emerging needs and requirements.⁵

Simultaneously, as libraries have looked internally to determine capacity and skill sets, there has been a desire to identify current RDM practices, issues, and needs. This has resulted in general cross-institutional surveys,⁶ discipline-focused bibliographic reviews,⁷ and discipline-wide surveys.⁸

By far, however, the most common evaluation of researcher needs has come in the form of case studies at individual institutions. These have proliferated over the past decade, with so many having been done that, in their recent article on starting an RDM program, Henderson and Knott explicitly argue that no further surveys are needed.⁹ While the majority of these case studies have been performed at US institutions, several case studies are also available from the England, Canada, and Australia.¹⁰ Yet, no articles exist that comprehensively review the depth and breadth of what has been surveyed. As a result, needs and practices remain inconsistently identified across institutional type, researcher level, etc.

Methods

Library literature indexes such as LISTA and general indexes such as Google Scholar were searched to populate an initial publication list. In addition, references and citations of initially identified articles were reviewed to discover more studies.

Case studies were included if they focused specifically on the RDM needs and behaviors at a specific institution or set of institutions. To retain homogeneity regarding research institution classification and funding mandates, only studies about United States universities were considered. The case study had to primarily focus on determining researcher needs as opposed to general library services or evaluations of RDM library service implementation. Accepted methodologies included surveys, focus groups, and interviews or combinations of those methods. Case studies could be discipline-specific or agnostic. Finally, case studies had to be published as a journal article, white paper, or in a completed manuscript by December 2017. Presentations and posters were excluded due to the challenges of comparing details across these format types.

A preliminary categorization for articles was developed by the authors. Categorical data included bibliographic information, instrument metadata, assessment target populations, question topics,

responses, and identified actionable outcomes. Question topics and responses were recorded from case study instruments, when available, and gleaned from narrative text in the methods, results, and discussion sections. Only questions or responses that were mentioned three or more times across all studies are reported in this analysis. Question topics and responses included: backup, cost, data format, data life cycle stage, data management plans, data size, description/metadata, finding data, funding source, organizing data, person responsible for data management, policy, storage, preservation, privacy, requirements for data management, research statement, retention, reuse, security, sharing, time and training/education. Question topics and responses were coded according to author agreed upon definitions, which are given in Appendix A. Institution classification was determined by the Carnegie Classification public data file.¹¹ The authors divided the included case studies in half for initial review and categorization. Ambiguous or multivalent results were discussed until a consensus was reached and coded under a single definition.

Results

A total of fifty-five publications were identified by the authors between initial searching and citation mining. After limiting for publication type, US only, and methodology, forty documents representing thirty-seven unique case studies were retained and categorized. Two case studies that had been reported in multiple articles¹² and one case study synthesizing two phases of a previously reported research were collapsed¹³ to avoid duplicate data and skewing representation or impact.

The forty documents were published between 2007 and 2016. The complete bibliography is available in Appendix B. The two highest years of article publication were 2015 (n=10) and 2014 (n=5). The papers were published in twenty-three different journals, by three universities as white papers, and one presented but unpublished manuscript.

Table 1: Journal Titles

Journal Title	Count
Journal of eScience Librarianship	5
portal: Libraries and the Academy	4
Issues in Science and Technology Librarianship	3
College and Research Libraries	2
International Journal of Digital Curation	2
Journal of Librarianship and Scholarly Communication	2
Program Journal	2
Ariadne	1
Bulletin of the Association of Information Science and Technology	1
College and Research Libraries News	1
Educause	1
Journal of Academic Librarianship	1
Journal of Agricultural and Food Information	1
Journal of Library Administration	1
Journal of Professional Issues in Engineering Education & Practice	1
Journal of the Medical Library Association	1
Journal of Web Librarianship	1
Library Hi Tech	1
Library Trends	1
Oregon State (white paper)	1
Philosophical Transactions of the Royal Society A	1

Practical Academic Librarianship	1
Proceedings of the American Society for Information Science and Technology	1
Science and Technology Libraries	1
University of Iowa Staff Publications (white paper)	1
University of Minnesota Libraries (white paper)	1
(unpublished)	1

The range of authors was one to six, with over half of all papers having one (n=10) or two authors (n=15). Most of the case studies were done solely by librarians and sponsored by the University Library. Of the case studies where other campus partners were involved, collaborators included the Office of Research, Campus IT, the Dean of Graduate Studies, the Chief Information Officer, and campus grant specialists. Forty-three institutions were studied, most of which were Carnegie Classified as Highest Research, Doctoral institutions (n=33). The remaining studies were distributed among other institutions classified as Doctoral Higher Research (4); all levels of Master's institutions (5); and one Baccalaureate institution. Differences between the institution counts and the number of studies is the result of three publications studying multiple institutions.

Table 2: Institution classification

Carnegie classification	Count
Highest Research (R1), Doctoral	33
Higher Research (R2), Doctoral	4
Master's Larger Institutions (M1)	3
Master's, Medium Institutions (M2)	1

Master's, Small Institutions (M3)	1
Baccalaureate Institutions	1

Of the 43 institutions studied, 33 were public and 10 private. Public institutions account for seventy-six percent (33/43) of the represented institutions. Twenty-five institutions were both public and Highest Research activity and 4 were public higher research, representing fifty-eight percent (25/43) and 9 percent (4/43), respectively, of all institutions. Eight institutions were private with highest research activity. The remaining two private institutions have Master's small and baccalaureate classifications.

Assessment methods used included interviews, surveys, focus groups, and combinations of interviews with either surveys or focus groups. The most common method was interviews (n = 19), followed by surveys (n = 17), and focus groups (n = 6). The survey response rate was frequently unavailable but ranged from 5%-65.6% of the targeted population, with the majority reporting a less than 10% response. Focus group participant counts were 8, 10, 15, 18, 31, and one not reported. Interview participant counts ranged from 5-56. About 60% of all instruments were available with the papers, as appendices or supplemental documents. The most frequent instrument provided was interview instruments at 70% (12/17).

Recruitment methods were almost entirely emails (24/40), with preliminary emails targeting faculty listservs and secondary recruitment via emails from liaison librarians. Several studies relied on a personal invitation targeting researchers where a relationship already existed or where a researcher was identified as having funding (n = 5). The other method of recruitment was presentations (n = 3), and six studies did not report a recruitment method.

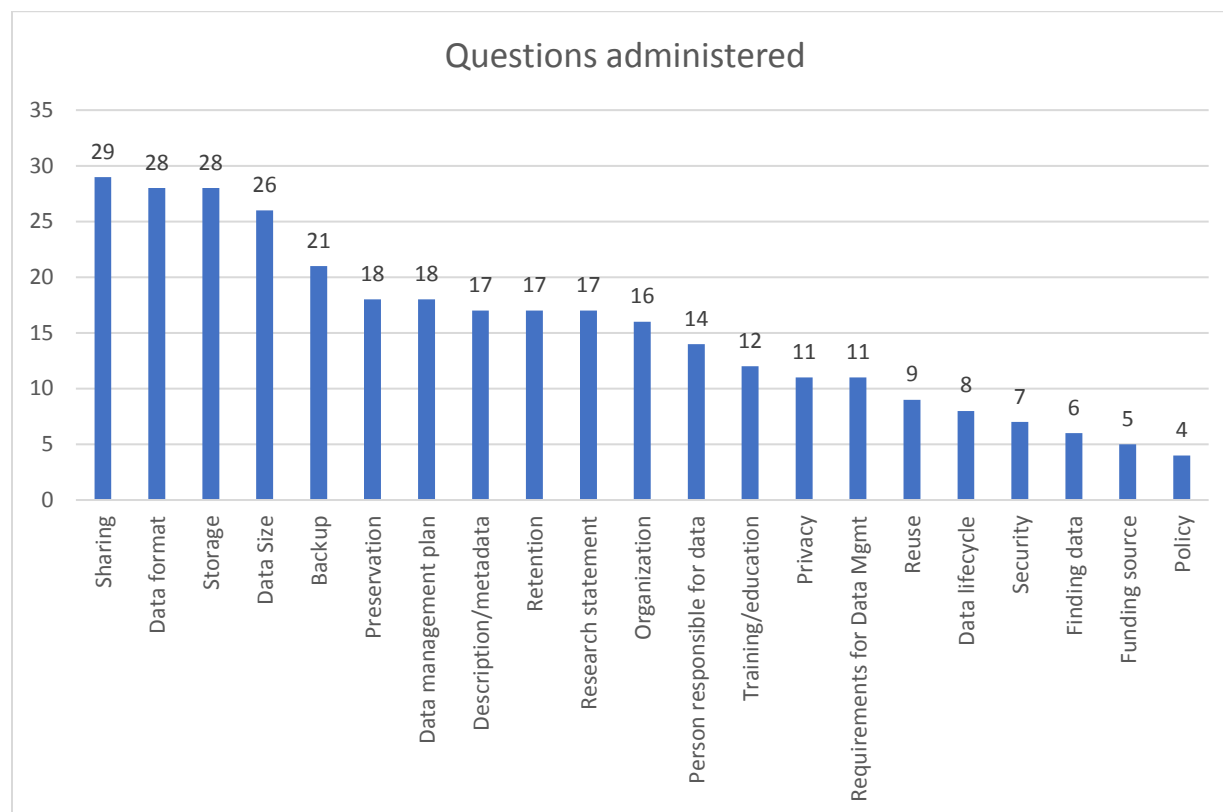
The total number of participants across all studies was 5359. Most respondents identified as faculty (n = 3847, 71.79% of all reported); followed by staff (n = 590, 11.01%); graduate students (n = 582, 10.86%);

undergraduate students (n = 143, 2.68%); post-doctoral researchers (n = 121, 2.26%); other (n = 70, 1.31%); and administration (n = 6, 0.11%). Twenty-three case studies interacted only with faculty.

Faculty numbers were reported as identified in the study; and, depending on the institutional appointment hierarchy, it is possible that some post-doctoral researchers and clinical staff have faculty status and therefore are reported in these numbers. One study focused on graduate students, but the participant numbers were low (n=6), and one study primarily interacted with post-doctoral trainees. No study actively focused on staff, undergraduate students, or administration.

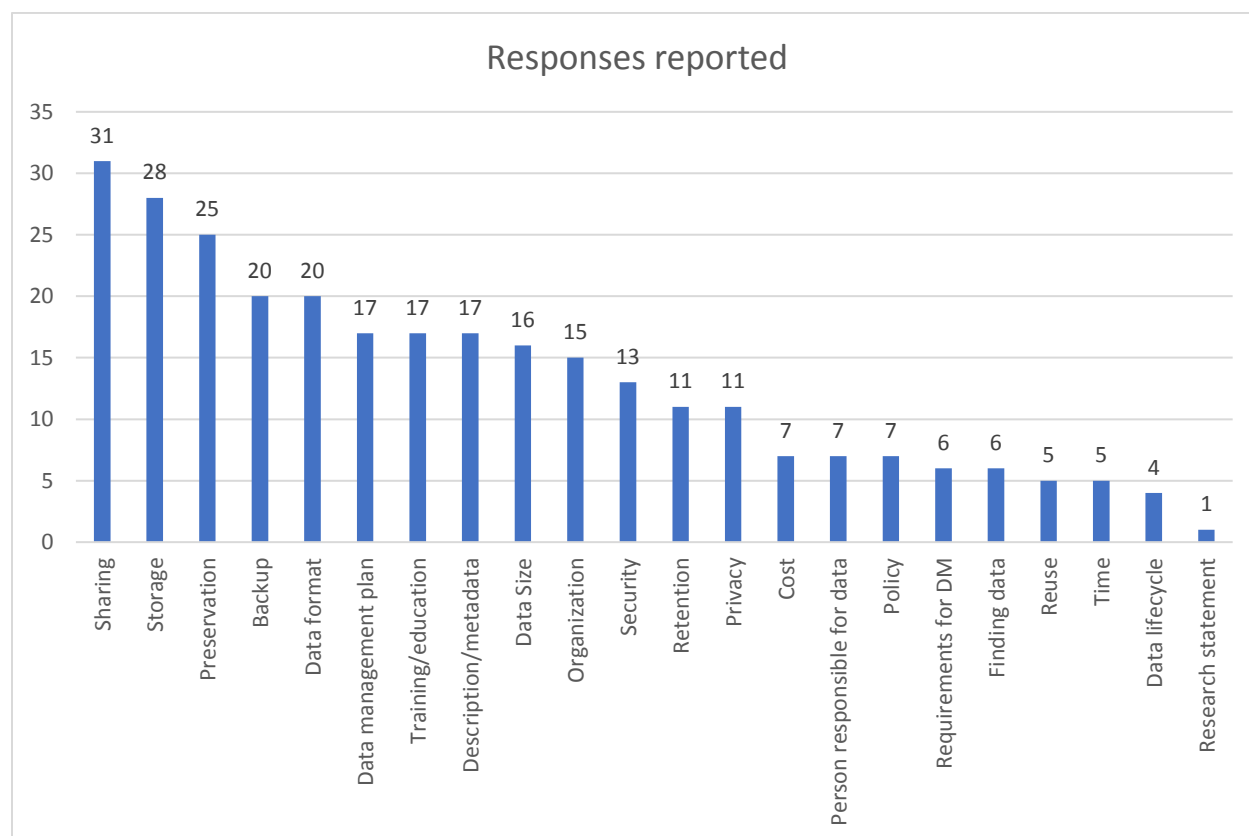
The most common question topics were: sharing, data format, storage, data size, and backup. These question topics were asked in greater than 50% of the evaluated case studies. The following topics appeared in case study questions 25-50% of the time: preservation, data management plans, description/metadata, retention, research statement, organization, personnel responsible for data, training/education, privacy, requirements for data management activities. The following topics were asked about in less than 25% of the case studies: reuse, data life cycle, security, finding data, funding source, policy.

Figure 1: Questions administered



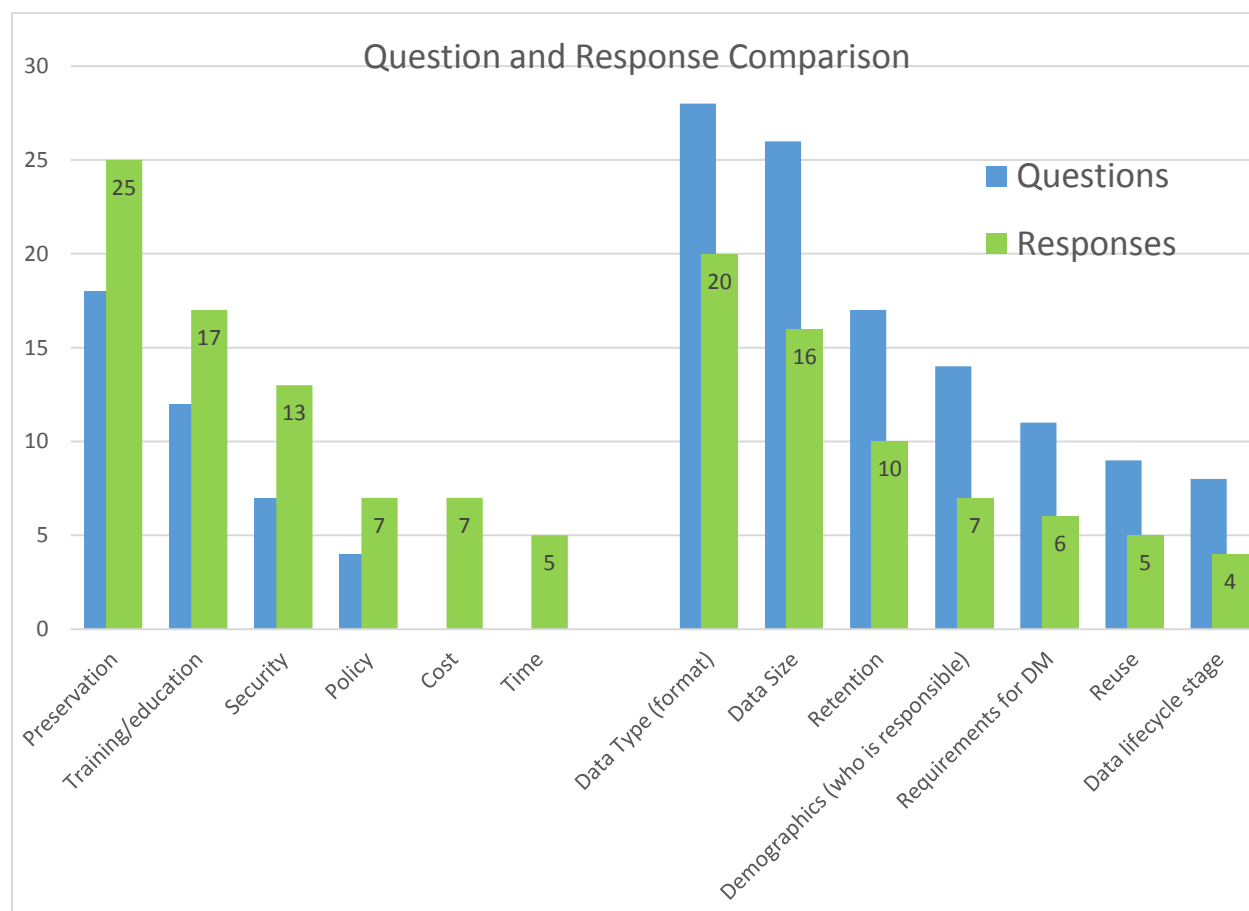
The most common responses and areas of interest reported by the participants were sharing, storage, preservation, backup, and data format. These appeared in greater than 50% of the case studies. The topics appearing in case studies 25-50% of the time included data management plans, training/education, description/metadata, data size, organization, security, retention, and privacy. The following topics appeared as responses in fewer than 25% of the case studies: costs, personnel responsible for data, policy, requirements for data management, finding data, reuse, time, and data life cycle.

Figure 2: Responses reported



Comparing question topics to responses, the case studies asked about data format, size, retention, persons responsible for data management, requirements for data management, data reuse, and life cycle more often than was indicated a response. Conversely, preservation, training/education, security, policy, cost, and time were indicated as a response more often than was asked.

Figure 3: Question and Response comparison

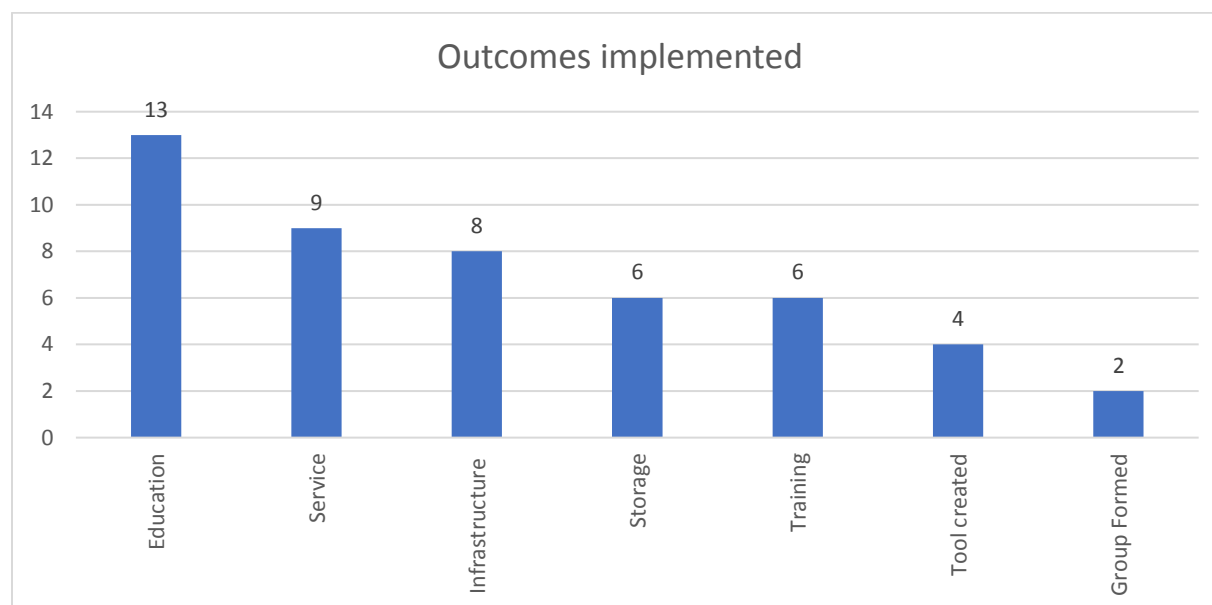


The remaining topics were asked and reported on equally. Over all studies, only 22% report out on 75% or more of the questions they asked, 38% report on 50-75% of their questions, and 35% of cases report out less than 50% of the questions.

Twenty-three case studies (66%) included outcomes, defined as activities that the library implemented following the completion of the needs assessment. Of the institutions that did provide outcomes, the most commonly reported were education ($n=13$), creation of a service ($n=9$), and creation of infrastructure ($n=8$). Of the eleven remaining studies that did not have concrete outcomes, five

proposed a variety of intentions, usually services, education, and outreach. Unanticipated but reported outcomes included hiring data librarians or specialists. Several institutions also mentioned creation of a LibGuide, but this was not tied to a specific service or job responsibilities.

Figure 4: Outcomes implemented



Discussion

Institutional Representation in the Literature

It is unsurprising to find that most of these studies were done at public universities with high research output. These institutions receive a large amount of federal funding, which dictates data sharing and access to research outputs, and have a perceived additional accountability and requirements for transparency to state governments and other public funding agencies. This raises the question of whether these institutions are overrepresented across these studies, skewing the reports of availability or need for infrastructure. We report that public institutions account for a majority of published studies at 76% (33/43), and public, highest research account for 58% (25/43). The current Carnegie Classification data file ¹⁴ reports a total of 1225 institutions of the same classes represented in this study, of which 456

institutions are public, and 81 public, highest research output. In percentages, the total institution population is only 37% public (456/1225) and 6% (81/1225) public, highest research, indicating that that public overall and public highest research institutions are significantly overrepresented in the needs assessment literature. While it is not expected that the general research data management needs will dramatically differ at private institutions, not understanding how these things may be resourced across the two types creates the potential for uneven ground. The 2018 report on higher education revenues shows that private institutions revenues rely more tuition dollars while public institutions rely on grants and contracts. Private institutions also receive more in other revenues such as gifts, private grants, capital, etc. than public ones.¹⁵ Because of this, some of the more well-endowed private institutions may be able to provide more robust infrastructure such as ongoing storage, a data repository, campus-wide access to costly visualization software or provide services such as data visualization and curation assistance as their funding sources may have more flexibility when determining allocations. This could cause disparities if publicly-funded researchers need to write infrastructure, visualization, or preservation costs into their grants, resulting in overall higher costs in the grant and/or creating tension between funding their own infrastructure and executing their research.

Additionally, the data management needs predicated on highest research institutions distorts needs at institutions where promotion and tenure may not have the same research requirements and where graduate assistants are not as readily available. Stamatoplos, Neville, and Henry comment in their characterization of researcher understanding of data management that there are gaps in the literature where non-research intensive institutions are concerned. This study reports similarities between their Master's level institutions and the reported literature, and reports on an additional five institutions, but this is not adequate to determine if data management needs are truly comparable.¹⁶ Further, Clement et al. note that liberal arts colleges may privilege funding towards curricular and teaching needs as opposed to research infrastructure, further limiting the resources available to the faculty.¹⁷

One potential solution for the infrastructure disparity has been identified in a recent National Science Foundation report, which calls for dedicated funding to provide sustained mid-scale research infrastructure development and maintenance support.¹⁸ Other possibilities may involve collaboration between institutions to pool resources as proposed by the Data Curation Network¹⁹ or expanded support for disciplinary repositories and aggregators like DataONE.²⁰

An additional correlate may be that the large public research institutions are where librarians are tenure-track or are otherwise under an onus to publish, or where a data librarian may have been hired in the past decade. Needing to publish original research or justifying new hires or services are likely to skew what is researched and presented in the library literature.

Comparison and Continuity Between Studies

The structure of the individual studies and the format of their specific questions made it impossible for the authors to speak beyond generalities in this review. One flaw of many studies is that they are not true needs assessments. Needs assessment are the in-depth statistical analysis of “what should be” compared to “what is”.²¹ Most of the studies reported here take a discovery stance, where the focus is on one but not both of those aspects. As a result, there is an inconsistent reporting between questions and response. This disparity may have arisen from a lack of beta testing of the instruments, a disconnect between the library and researchers, or a failure to incorporate previous literature when developing the instruments. While study instruments were frequently provided, the raw data underlying the study were not and, therefore, could not be used to make our own determinations.

As noted by Altschuld and Kumar, the wording structure in assessment instruments also affects results.²² The specific example they give of framing a Likert scale consistently or meaningfully across studies exists as an issue here, but there also was inconsistent terminology both within and across studies. For example, the terms security (as defined by the authors, focused on IT security and access to the data)

and privacy (subject privacy for audio recordings, sensitive data, HIPAA information, etc.) were conflated in reporting questions, responses, or in the discussions. Together, these issues prevent the authors from making direct comparisons between the studies. An opportunity for qualitative analysis on these topics exists and may lead to a more nuanced interpretation of the data.

Other issues that limited analysis were: the participation results usually had a low response rate; personal invitation to participate in the research study will have introduced selection bias; and many institutions did not identify the specific colleges or departments surveyed or gave only very general explanations such as covering all the major colleges at a university. Specifically regarding low response rate, many studies mentioned this as a weakness of the publication; however, several articles asserted getting a representative sample of their institutions despite most not providing a power analysis to corroborate.

Demographic Characterization of Respondents

When mentioned, these studies indicated disciplinary responses from the STEM and STEM adjacent fields. Several reported that humanities or non-STEM scholars claimed that they did not collect or work with data, despite potentially performing text and image analysis, gathering artifacts or recordings, or needing to archive other products of research. As documented by Partlo, when speaking with humanities or those outside of traditional STEM fields, using inclusive language like ‘products of research’ or focusing on research objects as opposed to ‘data’ or other science-centric language is likely to improve response rate from these other disciplines.²³ Because of this limitation and the self-selecting nature of most of the case study invitations, it is unclear if the workflows, curation, or data management needs are significantly different in these areas.

An impressive number of faculty members were contacted and participated across all thirty-seven studies. However, considering that research groups consist of a varying array of post-doctoral

researchers, staff, and graduate and undergraduate students, the research needs of these other groups are likely underrepresented. This is important because in many research groups, these ancillary support people have the highest contact with research data. Frequently they are the ones generating, analyzing, and annotating, more so than the faculty investigators who were the subject of most of the case studies. More research is needed to see if the needs of non-faculty groups differs or if their perception of needs changes as they progress through academic ranks.

Based on the counted responses, the results above most likely reflect the needs of the faculty researchers. If faculty are focused on and responding based on their immediate personal data management needs, while the librarians performing the research bringing a broader institutional view across the full spectrum of data management activities, then this may potentially explain the disconnect between the questions and reported responses. It also complicates the challenges faced by the libraries to identify and provide scalable services.

Overarching Faculty RDM Needs and Gaps

Looking at the questions topics individually, the highest number of both questions and responses relate to concerns with sharing data. Understandably, the early literature reflected uncertainty regarding the then-new National Science Foundation and National Institute of Health data management mandates. The early focus remained on the creation of data management plans as the infrastructure for sharing data was in its infancy. Data management plan help is listed as a need in almost 50% of the studies reviewed here and has been consistent across the timespan of published literature. A decade later, these unfunded mandates still exist, and this issue has come to light again as increasingly high-impact journals are accepting, if not requiring, data associated with publication.²⁴ More recently, the conversation has begun to shift from whether one can and should share data to when, where, and how a researcher can share their data, either because of personal interest, disciplinary standards, or funder

and journal requirements. Repository identification and selection was intermittently mentioned as a possible solution addressing these issues. This was also mentioned as a “service” outcome of several studies. Interestingly, there is a potential expectation for the library to house data as well. This has led to some institutions implementing data-specific repositories, prompting interesting collaborations at the institution level with many campus entities.

The next prevalent need reported was storage. While there is interest in data repositories at some institutions, for many this is less about a permanent, curated home for data associated with publications and grants, and more about active storage for ongoing research work. This is arguably beyond the scope of library research data services at most institutions and more within the domain of campus IT to offer storage solutions. This is reflected in the lack of outcomes beyond finding or creating a data repository. However, this does speak to a need for the library to have a working relationship with campus level entities like IT to effectively communicate support and handoff. As indicated in the outcomes, some libraries are already taking these steps by forming working groups with entities such as IT, clinical and translational science award (CTSA) recipients, and supercomputing centers. These partnerships also create an opportunity to educate campus partners on library goals and practices such as backup and preservation, the needs listed immediately after and closely related to sharing and storage.

A frequent refrain among these studies is that data management skills are acquired through the culture of the research group, rather than through formal training. However, training and education came up as a resultant need in almost 50% of the case studies, suggesting the need for more formal education as students and early career researchers are beginning to work with data or are developing their skills sets. This was also the most highly reported outcome, which is unsurprising as education can frequently be implemented by individuals or small teams at institutions, without the need for extensive campus funding or collaboration. The success or scalability of those educational interventions is an opportunity for further research. As research funding continues to become more competitive, graduate students,

post-doctoral researchers, and junior faculty will need these skills when applying for grants. For those students moving beyond academia, data management skills may become a marketing point for industry, government, or other non-tenure track jobs.

The need for description or metadata was ranked next. However, from the case study instruments available or referred to, it is unclear whether the survey respondents understood what was intended by this phrase, and this was noted in the discussion of many papers. When further defined, the case study usually mentioned discipline specific standards as opposed to more general metadata that might be included with any project. This library-centric language likely caused confusion leading to potential misunderstandings of the intent of the question. Interestingly, despite respectable response number, addressing metadata or description was not identified as a targeted outcome in any of the case studies.

There were several topics that were of lower interest in the responses. One of these was the persons responsible for the data. Few of the case studies, particularly earlier in the decade of publication reviewed, asked about who was creating the data, doing data entry, performing analysis, or holding other aspects of responsibility. This became a more frequent question over time. Another was the data life cycle. This was mentioned in earlier studies, likely coinciding with the advent of the DataOne and New England Consortium data management curricula,²⁵ but has almost entirely disappeared from papers published in 2015 and later. A further low question/result was data policy. This was generally unexamined in the library literature prior to 2015 when a preliminary landscape on the topic was published,²⁶ but then starts to become more common as a question and a response. Finding data and data reuse were also not frequently asked. Interestingly, one would expect that with the data sharing mandate becoming established, we would see an increase in this question and response, but this is not yet the case. Two questions that were never directly asked by any of the studies was cost or time associated with data management activities, although funding source was asked. In many of the studies that allowed for researcher input, respondents mentioned time and cost as infrastructure or support

needs in order to perform any data management, particularly sharing. The outcomes reported by the case studies for the implementation of tools, services, and data management groups come as a likely response to these articulated needs. Strong campus partnerships could lead to the establishment of data pipelines within an institution, greatly reducing the workload and time involved in data management for researchers.²⁷

Conclusion

This study confirmed that researchers are most worried about storage, sharing, and issues that revolve around long term access to data regardless of context asked. Issues of intermediate concern were data management plan assistance, security/privacy, data organization, and who the responsible party regarding data management is. Of little interest, either asked or remarked upon, were finding data, the data life cycle, and data policy. In aggregate, library outcomes have focused on education, service, and infrastructure as opposed to larger community partnerships and policy development. Overall, studies are minimally discussing metadata, curation services, or cost recovery for data management activities, although these topics were brought up. In some cases, disambiguation of and/or education on these terms to research participants will be necessary in order to gain useful knowledge from future research. We concur with Henderson and Knott²⁸ that myriad case studies have been done at highest research institutions. The authors' interest in the subject was to generally evaluate whether more research was

needed, and if so in what areas. Based on the gaps identified in this study, we would recommend future research be targeted to staff and post-doctoral researchers that institutionally are not classified as faculty; graduate students; and undergraduate students because they have the highest contact with research data and have been the least studied in terms of need. Needs assessments at liberal arts colleges, community colleges, smaller research institutions, and private institutions may also be beneficial as these are underrepresented compared to research that has already been done. Lastly it is likely that more data is needed regarding data management needs in non-STEM disciplines like humanities, business, and social science. A more thorough qualitative analysis could be done on this data to determine if a STEM bias exists and which disciplines may deserve more attention. We expect that because of the difference in workflow and working with different research outputs that classify as data that there might be a significant difference in applying data management best practices as all levels of the data life cycle for other disciplines

Additionally, inconsistency between instruments made it difficult to correlate topics from one study to another. It also speaks to a need within the greater data librarian community to structure our studies in such a way as to consistently build upon each other and advance our research topics, rather than repeatedly revisiting the same material at each of our respective institutions. We recommend reuse or combination of instruments rather than recreation or newly created tools. We did see this in some measure in the use of selected modules or modification of the Data Curation Profiles. However, the full description of the modules used or modifications made often were not available. Release of full survey instruments, either with articles or in clearly marked repositories, would help future libraries reuse the data already created from this body of work and prevent duplication of effort. Release of de-identified raw data would also help in quantifying and provide an opportunity for more nuanced study.

This study reported twenty-three topics of data management needs across studies conducted over approximately ten years. This is a relatively short time and broad view of data management needs

considering the growing advancements that the academic research enterprise has encountered in the same time. As such, we suggest that all of these topics would be worth further in-depth study as the data librarianship profession and library-provided research data services increasingly specialize in response to these advancements. Specific opportunities areas for research include assessing the scope of technical services like security and privacy, examining backup implementations, accommodating formats and sizes, providing storage or preservation services; or alternately assessing the scope of advisory services such as data management plan review, and educational best practices for organization, documentation, metadata, finding, sharing, and reusing data. The authors anticipate the need for further evaluation of institutional needs, by both qualitative and quantitative measures, against disciplinary needs to more comprehensively determine the trends and correlate these with research data management library service maturity.

Acknowledgements

We gratefully thank our first readers for this manuscript, Kristin Briney, Paula Dempsey, and Amy Nurnberger and our colleague Rebecca Raszewski for her ongoing support.

¹ National Science Foundation, "Dissemination and Sharing of Research Results," *US NSF - About*, November 30, 2010, <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.

² John P. Holdren, "Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research," *Office of Science and Technology Policy, Executive Office of the [US] President, The White House*, February 22, 2013, http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

³ Andrew Cox, Eddy Verbaan, and Barbara Sen, "Upskilling Liaison Librarians for Research Data Management," *Ariadne: A Web & Print Magazine of Internet Issues for Librarians & Information Specialists*, no. 70 (2012): 26–26; Dorothea Salo, "Retooling Libraries for the Data Challenge," *Ariadne* 64 (2010), <http://www.ariadne.ac.uk/issue64/salo>.

⁴ Andrew M. Cox and Stephen Pinfield, "Research Data Management and Libraries: Current Activities and Future Priorities," *Journal of Librarianship and Information Science* 46, no. 4 (2014): 299–316; Catherine Soehner, Catherine Steeves, and Jennifer Ward, "E-Science and Data Support Services: A Study of ARL Member Institutions," *Association of Research Libraries*, 2010, <http://eric.ed.gov/?id=ED528643>; Carol Tenopir et al., "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?," *Journal of EScience Librarianship* 4, no. 2 (December 21, 2015): e1085, doi:10.7191/jeslib.2015.1085; Carol Tenopir et al., "Research Data Management Services in Academic Research Libraries and Perceptions of Librarians," *Library & Information Science Research* 36, no. 2 (April 2014): 84–90, doi:10.1016/j.lisr.2013.11.003; Carol Tenopir et al., "Academic Librarians and Research Data Services: Preparation and Attitudes," *IFLA Journal* 39, no. 1 (March 2013): 70–78, doi:10.1177/0340035212473089.

- ⁵ K. Antell et al., "Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions," *College & Research Libraries* 75, no. 4 (July 1, 2014): 557–74, doi:10.5860/crl.75.4.557; Marianne Stowell Bracke, "Emerging Data Curation Roles for Librarians: A Case Study of Agricultural Data," *Journal of Agricultural & Food Information* 12, no. 1 (January 31, 2011): 65–74, doi:10.1080/10496505.2011.539158; Sheila Corral, Mary Anne Kennan, and Waseem Afzal, "Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research," *Library Trends* 61, no. 3 (2013): 636–674; Susan Hickson et al., "Modifying Researchers' Data Management Practices: A Behavioural Framework for Library Practitioners," *IFLA Journal* 42, no. 4 (2016): 253–265.
- ⁶ Katherine G. Akers et al., "Building Support for Research Data Management: Biographies of Eight Research Universities," *International Journal of Digital Curation* 9, no. 2 (October 30, 2014), doi:10.2218/ijdc.v9i2.327; Anne R. Diekema, Andrew Wesolek, and Cheryl D. Walters, "The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories," *The Journal of Academic Librarianship* 40, no. 3–4 (May 2014): 322–31, doi:10.1016/j.acalib.2014.04.010; Youngseek Kim and Jeffrey M. Stanton, "Institutional and Individual Factors Affecting Scientists' Data-Sharing Behaviors: A Multilevel Analysis: Institutional and Individual Factors Affecting Scientists' Data Sharing Behaviors: A Multilevel Analysis," *Journal of the Association for Information Science and Technology* 67, no. 4 (April 2016): 776–99, doi:10.1002/asi.23424; John Ernest Kratz and Carly Strasser, "Researcher Perspectives on Publication and Peer Review of Data," *PLoS One* 10, no. 2 (2015): e0117619.
- ⁷ Erin Kerby, "Research Data Practices in Veterinary Medicine: A Case Study," *Journal of EScience Librarianship*, 2015, e1073, doi:10.7191/jeslib.2015.1073.
- ⁸ N. R. Anderson et al., "Issues in Biomedical Research Data Management and Analysis: Needs and Barriers," *Journal of the American Medical Informatics Association* 14, no. 4 (July 1, 2007): 478–88, doi:10.1197/jamia.M2114.
- ⁹ Margaret Henderson and Teresa Knott, "Starting a Research Data Management Program Based in a University Library," *Medical Reference Services Quarterly* 34, no. 1 (January 2, 2015): 47–59, doi:10.1080/02763869.2015.986783.
- ¹⁰ Neil Beagrie, Robert Beagrie, and Ian Rowlands, "Research Data Preservation and Access: The Views of Researchers," *Ariadne*, no. 60 (2009), <http://www.ariadne.ac.uk/issue60/beagrie-et-al>; G Knight, "Research Data Management at LSHTM: Web Survey Report," *London School of Hygiene and Tropical Medicine, London* 8 (2013); Linda D. Lowry, "Bridging the Business Data Divide: Insights into Primary and Secondary Data Use by Business Researchers," *IAASSIST Quarterly*, 2015, 14; Christina Sewerin, "Research Data Management Faculty Practices: A Canadian Perspective," 2015, <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2098&context=iatul>.
- ¹¹ Indiana University Center for Post-Secondary Research, "Carnegie Classifications 2015 Public Data File," 2016, <http://carnegieclassifications.iu.edu/downloads/CCIHE2015-PublicDataFile.xlsx>.
- ¹² Leslie Delserone, "At the Watershed: Preparing for Research Data Management and Stewardship at the University of Minnesota Libraries," *Library Trends* 57, no. 2 (2008): 202–10, doi:10.1353/lib.0.0032; University of Minnesota Libraries, "Understanding Research Behaviors, Information Resources, and Service Needs of Scientists and Graduate Students: A Study by the University of Minnesota Libraries," June 2007, <http://purl.umn.edu/5546>; Travis Weller et al., "Differences in the Data Practices, Challenges, and Future Needs of Graduate Students and Faculty Members," *Journal of EScience Librarianship*, 2015, doi:10.7191/jeslib.2015.1070; Travis Weller and Amalia Monroe-Gulick, "Understanding Methodological and Disciplinary Differences in the Data Practices of Academic Researchers," *Library Hi Tech* 32, no. 3 (September 9, 2014): 467–82, doi:10.1108/LHT-02-2014-0021.
- ¹³ Elizabeth Berman, "An Exploratory Sequential Mixed Methods Approach to Understanding Researchers' Data Management Practices at UVM: Findings from the Quantitative Phase," *Journal of EScience Librarianship*, March 31, 2017, e1098, doi:10.7191/jeslib.2017.1098; *ibid*.
- ¹⁴ Indiana University Center for Post Secondary Research, "Carnegie Classifications 2015 Public Data File."
- ¹⁵ J McFarland, "The Condition of Education 2018: Post Secondary Institution Revenues" (National Center for Education Statistics, May 2018), <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018144>.
- ¹⁶ Anthony Stamatoplos, Tina Neville, and Deborah Henry, "Analyzing the Data Management Environment in a Master's-Level Institution," *The Journal of Academic Librarianship* 42, no. 2 (March 2016): 154–60, doi:10.1016/j.acalib.2015.11.004.
- ¹⁷ Ryan Clement et al., "Team-Based Data Management Instruction at Small Liberal Arts Colleges," *IFLA Journal* 43, no. 1 (March 2017): 105–18, doi:10.1177/0340035216678239.

¹⁸ National Science Board, "Bridging the Gap: Building a Sustained Approach to Mid-Scale Research Infrastructure and Cyberinfrastructure at NSF" (Alexandria, VA: National Science Foundation, October 1, 2018), <https://www.nsf.gov/nsb/publications/2018/NSB-2018-40-Midscale-Research-Infrastructure-Report-to-Congress-Oct2018.pdf>.

¹⁹ Lisa Johnston et al., "Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services," *Journal of EScience Librarianship* 6, no. 1 (February 28, 2017), doi:10.7191/jeslib.2017.1102.

²⁰ "DataONE," accessed November 2, 2018, <https://www.dataone.org/>.

²¹ James W Altschuld and David D Kumar, *Needs Assessment: An Overview* (Thousand Oaks, Calif.: SAGE Publications, 2010), 4, <http://public.eblib.com/choice/publicfullrecord.aspx?p=996658>.

²² Ibid., 5.

²³ Kristin Partlo, "From Data to the Creation of Meaning Part II: Data Librarian as Translator," *IASSIST Quarterly* 38, no. 2 (2014).

²⁴ "Availability of Data & Materials : Authors & Referees @ Npg," accessed February 7, 2016, <http://www.nature.com/authors/policies/availability.html>.

²⁵ DataONE, "DataONE Education Modules," November 12, 2012, <https://www.dataone.org/education-modules>; Lamar Soutter Library - University of Massachusetts Medical School, "New England Collaborative Data Management Curriculum," 2015, <http://library.umassmed.edu/necdmc/index>.

²⁶ Kristin Briney, Abigail Goben, and Lisa Zilinski, "Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies," *Journal of Librarianship and Scholarly Communication*, Summer 2015.

²⁷ Lisa Zilinski, Abigail Goben, and Kristin Briney, "Going Beyond the Data Management Plan," in *The Medical Library Association Guide to Data Management for Librarians* (Lanham, MD: Rowman and Littlefield, 2016).

²⁸ Henderson and Knott, "Starting a Research Data Management Program Based in a University Library."

Appendix A: Question Topics and Response Definitions.

Category	Definition
Backup	Identifying whether a copy of the data is stored either with or separately from the original data file. This may be remote or cloud storage or may be a locally kept external hard-drive for digital files.
Cost	Financial resources required to perform data management tasks. This includes paying for personnel, storage, software, etc.
Data Format	File types of the data. This may include spreadsheets, video, images, or proprietary data types off equipment.
Data Lifecycle Stage	Using one of the established life cycle diagrams or definitions, generally verbs that talk about aspects of data as it is collected, processed, analyzed, and published.
Data Management Plan	Specifically looking at the National Science Foundation and other funding agencies requirements for data management plans in the grant process
Data Size	The size of data generated, as individual units or in aggregate. This may be both analog for lab notebooks or specimens or electronic, described in MB, GB, etc.
Description/Metadata	The process which creates context between and/or describes various data objects. This includes wayfinding objects (table of contents, indexes), established or developed metadata schema, narrative documents (read-me files), or standards/ontologies.

Finding Data	Locating data for reuse, whether that is in disciplinary repositories, on peer websites, or other relevant locations.
Funding Source	Which agency provided funds for the research to be performed
Organizing Data	Organization of the data. Differing from preservation as this is referring to data currently being used. Examples would be file naming conventions and folder organization.
Persons Responsible for Data Management	Identifying who has access to or control over the data –whether that be creating data, performing data entry, data analysis, etc.
Preservation	Curation of the data beyond the life of the research project; not just storage/ backup. May include modification and description activities.
Privacy	Focusing more on the privacy of research subjects. May include HIPAA compliance, anonymization, etc.
Policy	Focused on institutional or research team data policy.
Requirements for Data Management	Does the institution, funder, or journals used by the faculty require data management? Does their discipline?
Research Statement	A general description of the research project or a faculty member's overall work.
Retention	Specifies how long data is kept after a project is completed, at the end of a grant, etc.
Reuse	Making data available for replication, education, or other uses.
Security	Identifying physical and electronic security of the data. May include encryption, password access, etc.

Sharing	Identifying whether data is shared internally such as with other members of the research team, externally with other researchers, externally to meet funder requirements, openly to the public etc.
Storage	Reporting where the digital files or analog specimens are kept either during or after the project. E.g. Local computer, department shared drive, cloud-based storage, etc.
Time	Time associated with performing data management tasks
Training/Education	Formal and informal education and training for members of the research team

For Appendix B – please see the separate document in INDIGO