

JLSC

ISSN 2162-3309 | JLSC is published by the Iowa State University Digital Press | <http://jisc-pub.org>

Volume 10, 1 (2022)

Research Data Management needs assessment of Clemson University

Megan Sheffield & Karen B Burton

Sheffield, M. & Burton, K.B. (2022). Research Data Management needs assessment of Clemson University. *Journal of Librarianship and Scholarly Communication*, 10(1), eP13970. <https://doi.org/10.31274/jlsc.13970>

This article underwent semi-anonymous peer review in accordance with JLSC's peer review policy.



© 2022 The Author(s). This is an open access article distributed under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

RESEARCH ARTICLE

Research Data Management needs assessment of Clemson University

Megan Sheffield

Clemson University

Karen B Burton

University of South Carolina School of Medicine Greenville

ABSTRACT

Research data management (RDM) is a growing field of practice within academic librarianship and information management. Research data are generated by researchers investigating and describing new information; often, the data that are generated are digital in nature, for example, in spreadsheets or computer code. Researchers are experts in their fields but may not possess the same skillset as librarians and other information professionals when it comes to organizing, preserving, and sharing information. As a field, RDM encompasses a wide range of activities that include documenting and managing research data during a research project as well as sharing and preserving data after the research project is completed. Academic libraries can offer a variety of services that support researchers during the research life cycle; these services vary among institutions. The faculty, staff, and graduate students at Clemson University were surveyed by the library about their RDM needs in the spring of 2021. The survey was based on previous surveys from 2012 and 2016 to allow for comparison, but language was updated, and additional questions were added because the field of RDM has evolved. Survey findings indicated that researchers are overall more likely to back up and share their data, but the process of cleaning and preparing the data for sharing was an obstacle. Few researchers reported including metadata when sharing or consulting the library for help with writing a Data Management Plan (DMP). Researchers want RDM resources; offering and effectively marketing those resources will enable libraries to both support researchers and encourage best practices. Understanding researcher needs and offering time-saving services and convenient training options makes following RDM best practices easier for researchers. Outreach and integrated partnerships that support the research life cycle are crucial next steps for ensuring effective data management.

Keywords: data management, research data management, environmental scan, data services, data repository, research life cycle

Received: 03/16/2022 Accepted: 07/22/2022



© 2022 The Author(s). This is an open access article distributed under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

IMPLICATIONS FOR PRACTICE

1. Knowing the research data management (RDM) needs and practices of an institution can help libraries identify opportunities for education and outreach.
2. Researchers are often reluctant to share data for a variety of reasons, and, when they do share data, it rarely has sufficient metadata.
3. Institutional data repositories can provide a way for researchers to share their data and for libraries to be involved in the data deposit process (Bull & Schultz, 2018).
4. Faculty and graduate student education is a vital part of library research data management practices. Cultivating campus partnerships can lead to more robust RDM offerings and avoid duplication of services.

INTRODUCTION

Research data management (RDM) encompasses a wide range of activities that include documenting and managing data during a research project as well as sharing and preserving data after the research project is completed. Academic libraries can offer a variety of services that support researchers during the research life cycle, which vary among institutions. Basic services often include Data Management Plan (DMP) consultations and LibGuides for RDM best practices, whereas more advanced services may include data curation and deposit. Many federal agencies now require a DMP as part of their grant application and have mandated that the data be shared by depositing it in a publicly accessible repository. Supporting researchers in complying with these mandates is a vital part of research data services (RDS). Academic libraries are uniquely positioned to offer RDS that supports RDM throughout the research life cycle, but researcher needs will differ among institutions. Surveying the faculty, staff, and graduate students will inform current library services and guide strategic plans for future services. Whitmire et al. (2015) pointed out that “while the results of faculty surveys often reveal common themes, there is no substitute for having an understanding of local research practices when investing in the development of research support services” (p. 383).

LITERATURE REVIEW

A common theme in RDM is that researchers often “don’t know what they don’t know.” They know that they are supposed to do something to manage their research data, especially because many federal funding agencies require planning for data management and sharing in a DMP, but researchers often view figuring out what to do and how to do it as an unnecessary and time-consuming activity. Shen (2016) points out that “it is unclear to what extent and how much of the activity in data management, sharing, and reuse among faculty researchers involve

productive engagement and how much is just fulfilling government mandates” (p. 500). Morgan et al. (2017) suggests that the library is uniquely positioned to ease the burden of data management on researchers for the following reasons:

“RDM can be understood as a natural extension of the role of the Academic Librarian. Libraries have always managed research outputs, albeit in the form of books and journals, so librarians have many of the intellectual frameworks already in place. Librarians are familiar with the fundamental concept of data vs. metadata; they are skilled in working with information storage and retrieval within a digital environment; and have experience with pragmatic requirements such as managing embargoes to satisfy publisher and other commercial requirements” (p. 304).

Unfortunately, Wright et al. (2014) observed that “most researchers do not automatically think of the library when they need help with data management” (p. 19) and proposes the solution that the library should “become the resource rather than just point researchers to resources” (p. 21).

Reduce researcher barriers

Resources on RDM should be easy to find, be easy to access, and be in a format familiar to the user when possible. Perrier & Barnes (2018) found that researchers also wanted familiar tools that are easy to use, and “when participants spoke about the tools for data management, familiarity was an essential attribute.” Hickson et al. (2016) also found that researchers preferred familiar tools that provided a good user experience, and researchers were more open to data management strategies that integrated these familiar tools. Libraries already strive to provide a good user experience, and resources for RDM should follow this model.

Kenyon et al. (2020) found the following:

“Many faculty indicated the need to reach out to others for assistance and reported that they often sought help from their own colleagues as well as their larger campus or disciplinary communities. Even though faculty were willing to seek help from others, many reported that a lack of information about who had expertise in specific data-related topics, what support and services were available, and how to gain access to these services hindered their ability to do so.”

In other words, the faculty did not know what help was available or how to find out who to ask for help. Libraries should be able to easily fill this need by creating easily findable resource documents with this information.

Time

Saving time was a theme that appeared in several different studies. Researchers did not feel that they had the time to plan for and implement effective RDM procedures. Perrier & Barnes (2018) found that “researchers expressed interest in the University working with them to develop practical solutions that minimized the burden on their time” (p. 7). They also found that researchers wanted a pipeline for sharing their data, i.e., they submit their data and someone else does the rest because researchers felt that the time it takes to adequately describe data sets was a barrier to sharing. The willingness of researchers to practice effective RDM is not always the problem but, rather, the time it takes to learn best practices, set up procedures, train everyone working on a project, and ensure that they follow procedures. Morgan et al. (2017) agrees and says that “even though many academics acknowledge the importance of RDM, it is not surprising that the time investment required to engage with services designed to do just that can be challenging in itself. Even when initial contact is welcomed and encouraged, the pace of change and adoption of new practices in the RDM space takes some time and requires ongoing engagement” (p. 303).

Data storage and backup

Perrier & Barnes (2018) found that researchers had several main categories of needs. They wanted “clear guidance on appropriate backup tools” and “expressed a universal interest in being exposed to best practices” (p. 10). An institutional solution for long-term data storage and backup was something that would address potential privacy and obsolescence issues with third-party storage solutions. Although secure storage and backup options differ between institutions, providing information on what an institution already offers and how to use it could meet this need.

Communication

Communicating the need for RDM was a theme found in many publications. Cox et al. (2017) observed that “the current absence of a clear value-chain link between good RDM practice by researchers, and incentives and reward mechanisms (such as assessment, credit, and tenure decisions), does not help to frame RDM activity as a researcher imperative. Rather, RDM is often viewed as something of a chore bringing little value to the data producer and most benefits to the consumer” (p. 2193). Libraries can help with educating researchers on how RDM benefits them by beginning any instructional session, resource documents, or individual communications with this information. Mannheimer (2014) puts it more eloquently by saying “we want to warm their academic hearts with topics like tenure and promotion, increasing citations and advancing science” (p. 44). Leveraging library resources to

communicate the connection between RDM and long-term goals, both for the individual researcher and the institution, is crucial for getting researchers to buy into RDM. Coates et al. (2018) agrees and says that “since the value of RDS may be unclear to faculty and administrators, we must be prepared to clearly describe how these services support institutional goals and demonstrate the value of library expertise in stewarding the scholarly record” (p. 29). Libraries must help researchers see that data that have been properly managed, curated, and preserved are more valuable to them and will enable an overall increase in research productivity (Shen, 2016).

RDM training

As mentioned previously, researchers want to save themselves time and effort by being provided clear guidance and best practices for RDM, and this can be provided through training that comes to researchers either in-person or virtually. Bishop & Borden (2020) found that “around 70 percent of scientists in the survey had no prior RDM training” (p. 688). There are many people involved in research products who need different levels of RDM training. Graduate students and research assistants are often the people who actually work with the data every day, and faculty need to encourage or require them to have introductory RDM training as part of onboarding to ensure that everyone is following the same procedures. Bishop & Borden (2020) found that many students have to learn RDM skills on the job, presumably from their principal investigator or other graduate students. When RDM training is not available, faculty often consult their support staff’s expertise for RDM (Kenyon et al., 2020). This practice can lead to inconsistent RDM procedures, thus highlighting the importance of training in best practices. Faculty may also need training in introductory RDM skills, but they have additional needs for higher level training because they are usually the ones who set up the project initially and are responsible for data deposit and sharing (Shen, 2016; Bull & Schultz, 2018).

RDM training topics

There are several topics that researchers were interested in accessing training on. Whereas DMPs are often where libraries begin when supporting RDM, Shen (2016) found that help with short-term storage, long-term storage, and metadata were what faculty actually wanted more than DMP consults. Because many journals now require an author to deposit their data in a repository, libraries need to offer support for that process that includes identifying, choosing, and depositing data. Making that process as easy and user friendly as possible will encourage researchers to deposit their data.

In a survey of academic librarians, Cox et al. (2019) found that there is a perceived skill gap for expertise in metadata creation for research data repositories, which perhaps indicates a need for

training and education in this area beyond the researchers themselves. Joo & Peters (2020) found that few researchers use any kind of metadata standard (although some have developed their own standards and have found that the resulting poor or lack of file-naming conventions makes retrieval a challenge). Again, finding, understanding, and using metadata standards is time-consuming for researchers, and libraries can help educate about how metadata standards can benefit the researcher. Joo & Peters (2020) go on to suggest that “highlighting the importance of metadata for preservation and access functions, such as data archiving, sharing, and reuse is a good place to start” (p. 644). Librarians are intimately familiar with the benefits of metadata and how lack of metadata impedes indexing and retrieval, but this is probably not an area of familiarity for researchers. Metadata helps researchers to avoid duplicating data, better share information, and promote their work in various fields of study. It provides users with the ability to search, retrieve, and evaluate data sets. Metadata allows users to find data and decide whether data meet a particular information need. It also allows users to discover, process, and use a data set. Metadata provides value to an organization and institution because it helps protect the organization’s investment in the data. It creates an institutional memory and advertises an institution’s research efforts, thus creating partnerships and collaborations through data sharing (Wiley, 2014, p. 40).

Marketing services

Whitmire et al. (2015) found that researchers often do not know what RDM services are already offered; therefore, the first step for libraries is to effectively market their services. Mannheimer (2014) suggests that promotional materials should be simple and use pictures to tell a story, and that people should be able to get the main idea in a glance. Nicholson & Bennett (2017) suggest using psychology to direct attention to your promotional materials. They found that negative information was noticed more than positive information and suggest that promotional materials contain “messages that explain the essentials (including the negative outcomes that can arise from not having, or not adhering to, a well-constructed DMP), acknowledge time constraints, and recommend reasonable and actionable protocols to nudge behaviors” (p. 157). In their study, Nicholson & Bennett (2017) also found that “the instance of strong negative messaging presented imagery of data management “gone wrong.” The impact of this message was enhanced by being placed in opposition to some of the positive benefits of having a strong data management plan” (p. 156). These findings suggest that an effective marketing strategy may start with the negative consequences of bad or no RDM planning and then integrate how libraries can help into the positive outcomes.

Partnerships

RDM encompasses many campus entities, and connecting these through partnerships can make RDM easier for researchers to navigate, thus resulting in better practices. Yu (2017) says that

campus partnerships are both the key and the biggest challenge in providing RDM support and that “engaging in partnership was found to be the foremost opportunity in providing RDS, and libraries could partner with campus units with larger technical capacity, and offices involved with sponsored programs, grants and research” (p. 786). With RDS spread through so many different campus entities, Faniel & Connaway (2018) point out that “having leadership from the top would make it easier to support RDM needs” (p. 110). Whereas each institution will, of course, have a slightly different structure, Rinehart (2016) offers some general suggestions for where libraries could look for opportunities to partner with other campus entities: Have the graduate school include an RDM training course in orientation or during the first semester. Find out what services the Office of Research has and identify opportunities for collaboration. Make sure the library is not duplicating RDM services offered elsewhere. Learn the terminology around RDM that other campus entities use to communicate effectively.

Wright et al. (2014) suggest conducting outreach to build relationships to market RDM services. Inviting people for coffee allows you to have one-on-one conversations and establish the library as a resource. Mannheimer (2014) agrees with this philosophy and adds that “in order to achieve community buy-in, we have to establish ourselves within the community” (p. 43). The University of Washington took a unique approach to this by hiring a data services coordinator who acted as a concierge service and provided referrals for researchers to the appropriate sources for RDM services (Wright et al., 2014).

RDS integration and maturity

RDM is not something that happens once for each research project but something that should be integrated and embedded in the research workflow (Yu et al., 2017; Morgan et al., 2017). Libraries should also be involved in the whole research workflow, not just the beginning through consulting on a DMP and the end when a researcher deposits their research products in the institutional repository (Choudhury, 2017). Each institution will offer a different level of RDS according to the needs of faculty and students and, of course, library budgets. For libraries that do not have a large budget for RDM support, Johnston’s (2017) opinion that “more mature RDS programs are not necessarily those that offer the longest menu of services or employ the largest number of staff, but rather those whose activities are more deeply embedded in the mission and activities of the library and the broader institution” (p. 10) should offer hope that they can still support RDM.

Cox et al. (2017) provide a maturity model for RDS that details services and support at each level. They point out that advisory services such as DMP consultations and LibGuides about RDM are commonly offered at the basic level but technical services are not. The technical services they list include curating data, selecting a repository, preparing data for deposit,

carrying out long-term preservation of research data, and creating or transforming metadata for data. These RDS require librarians to learn additional skills and often necessitate hiring additional staff. Many institutions have increasing support for RDM in their strategic plans, and libraries are positioned to embed themselves in the research workflow and become the resources that faculty and students need to properly manage their research data.

History of RDM at Clemson University

As RDM has increased in importance in the last decade, the librarians at our (and many other) institutions attempted to meet our users' needs in this area. In mid-2012, a library staff member conducted a campus-wide survey for an internal report; because it was only intended to be used for planning purposes within the library, the survey did not get institutional review board (IRB) approval, and detailed results cannot be publicly shared. Later that year, the hire of an "E-Science Librarian" was the first in a series of coordinated efforts to focus library resources toward research data. This position was a reference liaison for several science departments as well as a resource that could advise in general terms on things such as DMPs, data repositories, and data management practices. Working from the findings of the 2012 survey, they created resources such as a LibGuide for writing DMPs and workshops on open data, and they worked with other liaisons to offer internal professional development opportunities regarding disciplinary data.

At the time of the second environmental scan in 2016, the library had hired additional positions such as a scholarly communications librarian and implemented an institutional repository (IR). The 2016 survey did go through proper channels for IRB approval but, owing to difficulties with a skewed sample size and complex survey design, it was not shared publicly beyond a conference presentation. This also made accurate direct comparisons with the previous and subsequent surveys neither accurate nor feasible. However, the library continued to use the findings for strategic planning. Using the results of the 2016 survey, additional resources and services were aimed at helping researchers manage their data. The library tested adding data sets to the IR but found it an imperfect solution and instead focused on advising researchers about the more prominent general-purpose and discipline-specific repositories. This timeframe coincided with gradual changes to funding agency guidelines (particularly National Institutes of Health [NIH], National Science Foundation [NSF], and United States Department of Agriculture [USDA]) that put data management in the spotlight for researchers as well as the institutional research compliance office; as a result, the library began to form partnerships with other offices on campus and served as a resource for these offices as well.

By 2018, the E-Science Librarian was spending a considerable amount of time on RDS initiatives (such as general RDM workshops, reviewing DMPs, advising the Research Compliance office, assisting researchers with selecting data repositories/submitting their data, and

working with research groups to implement their DMPs), and the library decided to create an entire position dedicated to this work. The Data Services Librarian was hired and began in mid-2019. This position does not have a collections budget associated with it the same way that the liaison positions do, but this person frequently receives requests to purchase data sets. Although there is demand for the position, the global pandemic derailed some aspects of the library's strategic plans in support of RDM, especially those such as implementing an institutional data repository that had significant costs attached. Since the 2016 survey, the library has offered significantly more trainings and workshops on data-related topics, such as open data, Python, data visualization, and several on writing DMPs for various funding agencies and disciplines. Although these workshops tend to be highly attended and are publicized through relevant avenues such as the Office of Research, the Research Compliance Office, and the Graduate School, the library still relies on the results of this survey to learn more about what is happening in the larger campus community and identify areas for growth.

METHODS

The previous RDM-needs surveys from 2012 and 2016 were reviewed and updated. In many cases, previous participants demonstrated misunderstandings about terminology regarding metadata and the definition of “research data;” as a result, the wording of many questions were edited for clarity. Previous environmental scans also suffered from complicated survey design that made statistical analysis difficult; for this iteration, answer-choice structure of most questions was changed. Changes in the larger field of RDM informed many of the changes and the addition of new questions. For example, widespread adoption of ORCID (which was introduced in 2012) and their integration with other registries and digital object identifiers (DOIs) have made it much easier to track citations to data sets (Cousijn & Lammey, 2018). Additionally, the evolution of online data repository services has made them much more accessible to the average researcher. Many repository services now offer data curation and metadata generation in addition to storage space, which makes the process of sharing data less burdensome; new tools for data-level metrics mean that researchers are increasingly able to track citations of the data, which provides incentives beyond goodwill and an interest in open research (Singh, 2019; Vannan et al., 2020). As a result, questions that were added beyond those asked in previous years focused on respondents' use of data sets, data set citation, and solicited comments about a possible data repository at Clemson. Qualtrics software was used to design and administer the survey. No questions were required, and respondents could submit the survey without answering every question.

To ensure accurate tracking of college and departmental demographics, a pre-populated drop-down menu that contained all of the colleges and departments at Clemson was created, and an “other” option with a free text box was included for respondents who were not affiliated with a college or department. The file types were condensed from previous surveys to represent

more broad categories. A new question about what data management services people want to see the libraries offer was added, including a free text response option. The survey was approved by Clemson’s IRB (approval number IRB2021-0072).

It was arranged for the survey to be sent out through the Office of Research Compliance via the all-faculty and all-graduate-student listservs to increase survey participation. Previous surveys were sent out to distribution lists maintained by subject liaisons; the authors of this article believe this distribution method had a negative effect on the response rate in previous iterations and that it potentially biased the results. For example, if a specific liaison didn’t send the email or didn’t have an updated list, whole departments potentially missed the call to participate. Additionally, the 2016 survey was skewed heavily toward graduate students (148 out of 172 or 86% of respondents identified primarily as graduate students); although the reason for this is not known with certainty, it is likely that sending the survey out over the summer prevented it from reaching many research faculty, and, instead, it reached mostly graduate students working on campus over the summer break. Emails inviting participation in the survey were also sent to individuals the researchers had professional relationships with to increase participation. The survey was open April 8, 2021, through May 9, 2021, and received 259 complete responses.

RESULTS

Demographics

Survey responses were collected from researchers in each of Clemson’s seven degree-granting colleges as well as from library faculty members (Figure 1). Although there are an additional

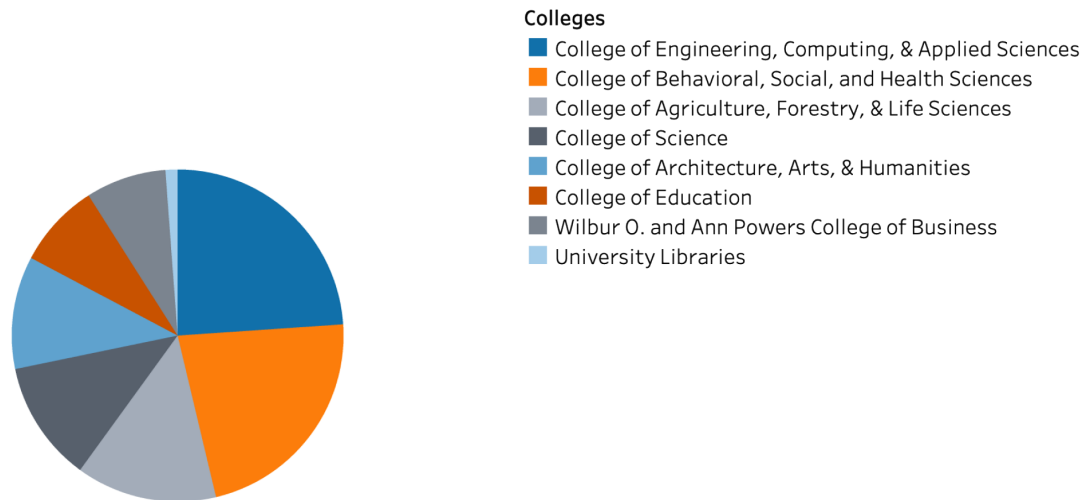


Figure 1. Breakdown of Responses by College

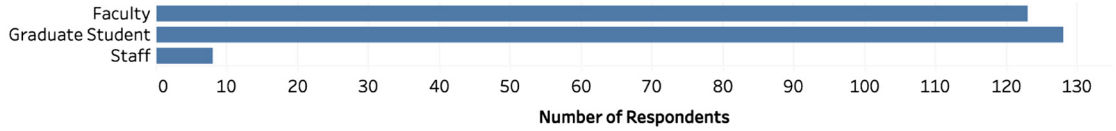


Figure 2. Breakdown of Survey Respondents by Status

three colleges with no respondents, these (Honors College, Emeritus College, and Graduate School) researchers are typically also members of another college within Clemson that serves as their primary affiliation.

Of the total number of respondents, 123 identified as faculty (47.5%), 128 were graduate students (49.4%), and 8 were staff (3.1%), which typically came from the university’s central information technology (IT) department (Figure 2).

Research data produced at Clemson University

The most common type of research data produced by respondents was “digital text and numeric data” (24.7%), followed closely by “presentations” (16.8%), “statistical data (SAS, MatLab, etc)” (13.6%), and “audiovisual data” (the full list can be found in Figure 3).

Respondents were also asked to think about their most recent research product and estimate the total amount of data produced (Figure 4). The majority of respondents reported producing less than 100 GB of data (57.8%), although many responded that they were not sure (26.6%).

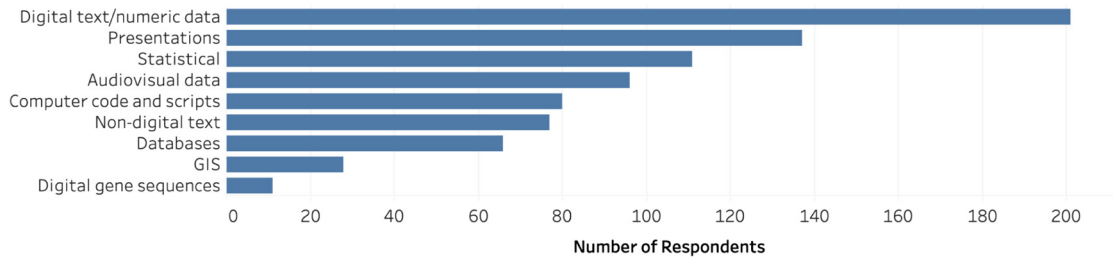


Figure 3. Research Data Produced at Clemson University. GIS, Geographical Information System.

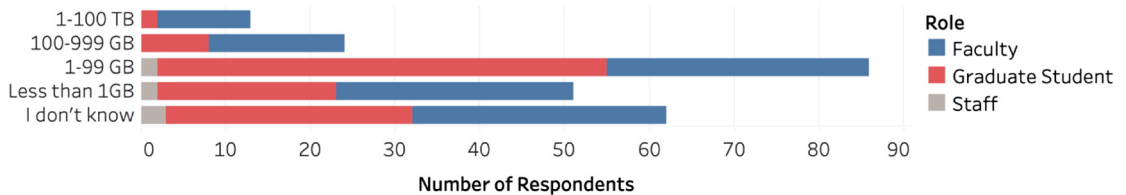


Figure 4. Amount of Data Produced

Data management

Several questions on the survey dealt with general data management practices as well as the DMPs typically required by federal funding agencies. When asked how they protect their files from corruption or loss, the most popular option was to store on physical media such as a universal serial bus (USB) or external hard drive (161 responses, 27.4%). In descending order, some managed and protected their data by using cloud-based options (156 responses, 26.6%), by restricting access to files (57 responses, 9.7%), by saving files to a local (usually departmental) server (55 responses, 9.4%), manually creating backup files (46 responses, 7.8%), automatically creating backup files (39 responses, 6.6%), saving to a centralized campus server (32 responses, 5.5%), and storing in a data repository or archive (27 responses, 4.6%). Only nine respondents (1.5%) indicated that they did not use any of the aforementioned methods to manage or secure their data (Figure 5).

When asked whether they had ever been involved in grant-funded research that required a DMP, a majority of respondents (106, 54.7%) indicated “No,” with 64 (32.7%) responding “Yes” and 26 (13.2%) responding “I’m not sure.” However, once these numbers were broken down to show the status of the respondents, it became clear that about half of the faculty said “Yes” (54 responses, 50.4%) and most graduate students said “No” (58 responses, 69.0%) (Figure 6).

When asked in a follow-up question whether they had ever personally written a DMP, the affirmative numbers shrank considerably, with only 45.3% of faculty (49 respondents) and

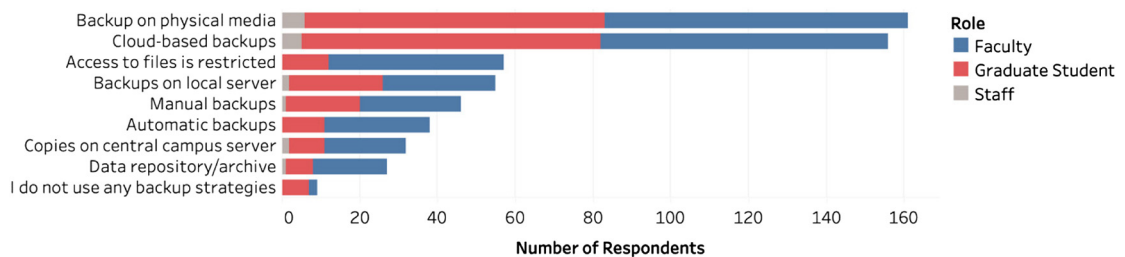


Figure 5. Methods of Data Backup

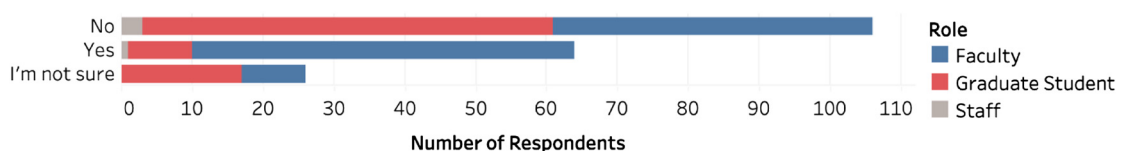


Figure 6. Does Your Research Require a DMP?

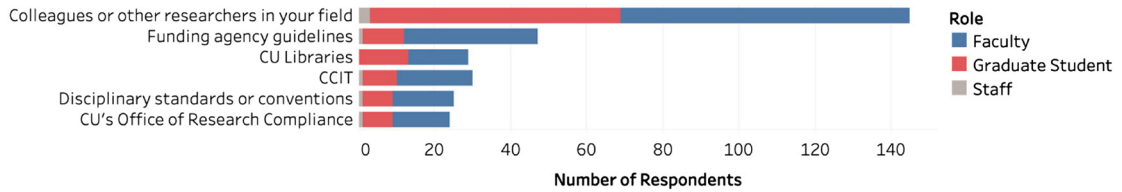


Figure 7. DMP Assistance. CCIT; Clemson Computing and Information Technology; CU, Clemson University.

9.5% of graduate students (8 respondents) saying that they had. When asked where they would go for help with a DMP, responses indicated they would be most likely to approach colleagues (145 responses, 45.2%), followed by (in descending order) their funding agency guidelines (47 responses, 14.6%), the University’s centralized IT department (30 responses, 9.4%), the Libraries (30 responses, 9.4%), other disciplinary standards (25 responses, 7.8%), or the University’s Office of Research Compliance (24 responses, 7.5%) (Figure 7).

Data sharing

When asked how they had shared their data in the last year, less than half (43.6%) said they had not made their data publicly available, whereas 28.7% provided data upon request. Only 22.0% openly shared their data in a repository or other publicly accessible space. When asked what prevented them from sharing, respondents’ most common response (29.6%) was that they were waiting to publish. Other popular responses were “My funder didn’t require me to” (10.8%), “It takes too much time to prepare data for sharing” (9.1%), and privacy concerns such as the Health Insurance Portability and Accountability Act (HIPPA), Family Educational Rights and Privacy Act (FERPA), and other personally identifiable information (PII) (10.8%) (Figure 8).

Respondents were asked whether they provided metadata with their data sets. This question caused confusion for respondents to the 2016 survey; therefore, to avoid confusion, two

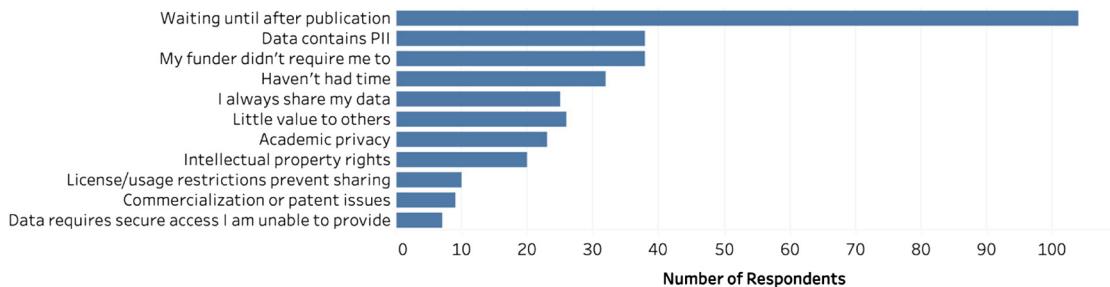


Figure 8. Reluctance to Share Data

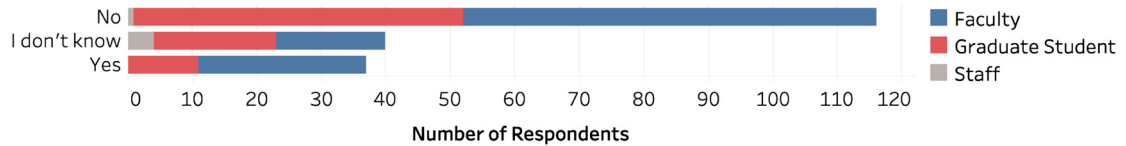


Figure 9. Do You Provide Metadata for Your Data Sets?

examples were given of what metadata could look like in different disciplines (these examples were preserved in the version of the survey seen in [Appendix A](#)). After this clarification, 117 respondents (57.4%) said “No,” and a further 40 (19.6%) said “I’m not sure.” Only 37 (18.1%) responded that they provided metadata (Figure 9).

Data set acquisition and citation

Respondents were asked whether they had ever cited a data set in their research, and 83 (43.7%) said “Yes.” A little over half (99 respondents, 52.1%) said “No,” and 8 (4.2%) were not sure. When those who had answered affirmatively were asked where they found the data set, the free text responses indicated researchers are getting their data sets in a variety of ways, including direct sharing with peers (which includes advisors sharing data with their graduate students), searching government websites, following citations in their disciplinary literature, and via library data bases.

When asked whether their data sets cost money, 111 respondents (73.5%) said “No.” Meanwhile, 6 respondents (4.0%) said that they paid for it personally, and 15 (9.9%) said that the university or their grant paid for it. Another 19 respondents (12.5%) were unsure whether their data sets cost money.

When asked whether they would use data sets through the Libraries if that were an option, the majority (111, 73.5%) said “Yes,” and 72 (37.9%) said “I’m not sure.” Only seven respondents (3.7%) said “No.” When asked a follow-up question regarding whether the Libraries should proactively purchase data sets that would be applicable for a variety of uses or have a fund to purchase specific data sets on demand, slightly more than half (98 respondents, 51.6%) preferred a data set-purchasing fund. A total of 42 (22.1%) preferred for the Libraries to proactively purchase data sets, and 50 (26.3%) were unsure of which option they preferred (Figure 10.).

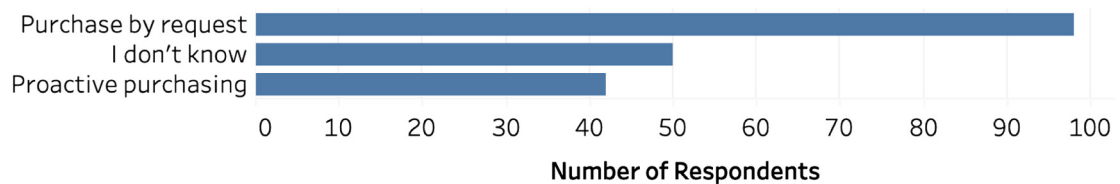


Figure 10. How Do You Want the Library to Purchase Data Sets?

Repository reactions

When researchers were asked about the possibility of a data repository for Clemson University, it was abundantly clear through the variation in responses that not all respondents were familiar with what a data repository is and how it works. Responses that related to a data repository were collected from the two questions that asked directly about the benefits of and concerns about data repository, the question asking what services researchers would like to see the libraries offer, and the final survey question that invited researchers to share any additional thoughts and concerns regarding their data needs. Although these questions were intended to elicit different answers, many responses had overlapping concepts. In an attempt to accurately represent the intent of the survey respondents, all responses from these four questions were read, and then they were sorted into the three categories (regardless of which question they were originally associated with) of repository concerns (Table 1), repository benefits (Table 2), and library services that researchers would like to see offered (Figure 11).

When asked whether Clemson University should have a data repository, 80% of respondents said yes, but many also voiced concerns. With so many responses reflecting uncertainty about exactly what the functions of such a data repository would be, any further solicitation of

Category	Number of Responses
Long-term concerns (sustainability, upgrades, capacity, obsolescence)	9
Privacy	6
Usability	5
Getting scooped	4
Access	4
Data will not be used	3
Timeline	2

^a 14 responses were not an answer to this question and were moved to the appropriate category.

Table 1. Categories of Data Repository Concernsa

Category	Number of Responses
Sharing and collaboration	36
Answer not related to repository concerns	15
Reproducibility and transparency	10
Compliance	9
Short- and long-term storage	4

^a 15 responses were not an answer to this question and were moved to the appropriate category.

Table 2. Categories of Data Repository Benefitsa

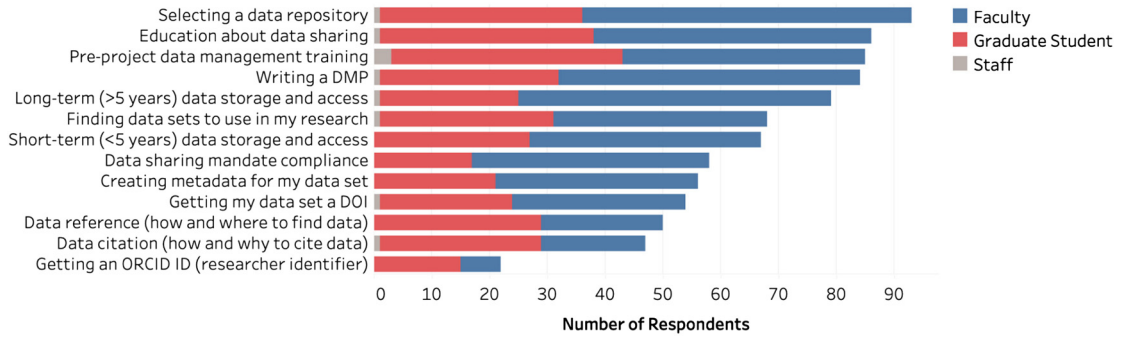


Figure 11. How Can Clemson University Help You With Your Research Data?

opinions on an institutional repository would need to begin with an explanation and definition of terms so that faculty, graduate students, and staff can give helpful feedback.

DISCUSSION

Research data produced at Clemson University

When compared with previous surveys from 2012 and 2016, researchers did not report large changes in the amount of data produced as seen in Table 3. One figure that does seem to be steadily increasing is the percentage of respondents unaware of the size of their data sets; this could be owing to a number of factors beyond the scope of this survey, but one possible explanation would be that researchers are less aware of files sizes if they are relying more on cloud storage and automatic backups as opposed to physical media.

Survey results confirmed suspicions that the majority of research data being produced at Clemson University does not fit the definition of “Big Data” as defined by Ricard & Urban

	2012	2016	2021
1 < GB	29.1%	26.8%	21.5%
1 GB to 100 GB	34.2%	34.7%	36.3%
100 GB to 1 TB	12.2%	10.2%	10.1%
1 TB to 100 TB	4.7%	6.3%	5.5%
100 TB to 1 PB	0.4%	0.8%	0%
> 1 PB	0%	0%	0%
I don’t know	19.4%	21.3%	26.6%

Table 3. Comparison of Size of Data Sets from 2012, 2016, and 2021 Surveys

(2015) as “massive amounts of data generated by a wide variety of sources.” The volume of data that Clemson University’s researchers are producing was generally more moderate in size, and the variety was reported to consist of common file types such as text, tabular data, audiovisual files, statistical data, and computer code. As data collection methods become more sophisticated and software records more detailed information, file sizes do seem to be gradually increasing, but more information would be necessary to determine this with certainty. Graduate students seem to be less aware of certain specifics of their research data, which makes sense given that, for many, it will be their first foray into scholarly research projects.

Data management

Given that Clemson University is a research-intensive school with a heavy emphasis on agriculture, engineering, and other applied sciences, it is somewhat surprising that only about half of the respondents have written a DMP before (Figure 6). However, it must be noted that survey responses came from all disciplines, and some disciplines (especially in the humanities and some social sciences) are much less likely to need a DMP with any grant proposal. Although results from 2016 have some limitations in terms of what can be extrapolated owing to heavy skew toward graduate student respondents, this is definitely an improvement over those survey results in which only 6.5% had written a DMP previously (the question was not asked in 2012, as the DMP was a relatively new part of the grant proposal process). It is also important to note the relatively small number of graduate students with any knowledge of DMPs (Figure 7); this is an easy area for the library to offer services and outreach, which can impact a key part of the grant funding process that other organizations on campus cannot always address. The relatively low ranking of the Libraries in the hierarchy of where researchers go for help with their DMPs highlights an area of possible growth for library services.

Data sharing

Clemson University’s researchers still show a knowledge gap around data sharing. In the 2012 survey, only 9% of respondents indicated that they always shared their research data; 56.7% said that they respond to individual requests, and 28.6% said they never shared beyond their research group. In 2016, these respective numbers changed to 15.1% always sharing, 29.0% sharing by request, and 51.4% never sharing. Although we believe the 2016 answers were skewed by the disproportionate numbers of graduate students in the sample, the percentage that always share definitely increased. In the 2021 survey, this number has increased again to 22%, and there appears to be a combination of factors contributing to their reluctance (Figure 8). One of these factors is a lack of expertise to make sure that personally identifiable data are not inadvertently included; although the Libraries are not equipped to offer a service of this magnitude, it does indicate an area that might benefit from some training. Another

factor is the lack of time and expertise to prepare metadata and optimize a data set for ingestion into a repository. Typically, in the flow of academic research, most researchers analyze their data and write it up and then move on to a new endeavor. Ideally, cleaning and organizing data should be part of that workflow, thus avoiding having to go back to clean and organize data for public consumption. However, this change would represent a new step in the process and cannot be expected to be implemented overnight.

One of the incentives to share data is the possibility that it will be found by another researcher and cited or used in further research. However, not taking the time to provide your shared data with appropriately descriptive metadata can doom your data to obscurity. Knowledge of and training on how to generate appropriately descriptive metadata is an area in which RDS can strive for growth. Data sharing and metadata creation are both areas in which librarians are uniquely suited to provide a new service, assuming we have the necessary infrastructure, staffing, and institutional support.

Data set acquisition and citation

Surprisingly, even though few researchers indicated that they regularly shared their data, many had cited or used data sets from other sources; in particular, there was a shift in understanding that can be seen in the two right columns of Table 4. Many more researchers are aware of whether or not they have cited a data set than in the previous survey. The last several years have seen an increase of researchers that use secondary data sources to augment their work (Khan et al., 2021), such as for a meta-analysis, and this is a trend that we feel is likely to continue as more data are produced and we find better ways to manage and analyze it. This is also an area that is directly tied to library services. If we begin to treat data sets as another resource just like journals, books, etc, then it becomes easy to see that librarians could assist researchers with finding data sets, acquiring them, and citing them. At the moment, the library does not have a formalized way to acquire data sets, but the survey results indicate that researchers would likely use data sets provided by the library, and that a combination of large data sets that could be used for multiple purposes as well as a fund to purchase specific data sets on demand would be beneficial.

Year	Yes, cited a data set	Not cited a data set	Not sure
2016	42.6%	30.6%	26.9%
2021	43.7%	52.1%	4.2%

Table 4. Researchers Citing Other Data Sets

CONCLUSION

Offering effective RDS encompasses a wide variety of areas. Understanding researcher needs and offering time-saving services and convenient training options makes following RDM best practices easier for researchers. Communicating how RDM benefits researchers and integrating RDS into the research life cycle through partnerships are crucial for ensuring both are actually done. Researchers want RDM resources, and offering and effectively marketing those resources will enable libraries to meet this need as well as to support researchers and encourage best practices.

ACKNOWLEDGMENTS

The authors would like to thank Jennifer Groff and Andrew Wesolek for their efforts with previous iterations of the Research Data Management Survey, as well as their colleagues and families for generous and insightful comments and suggestions on this manuscript.

REFERENCES

- Bishop, B. W., & Borden, R. M. (2020). Scientists' research data management questions: lessons learned at a data help desk. *Portal: Libraries & the Academy*, 20(4), N.PAG-N.PAG. <https://doi.org/10.1353/pla.2020.0032>
- Bull, J., & Schultz, T. (2018). Harvesting the academic landscape: streamlining the ingestion of professional scholarship metadata into the institutional repository. *Journal of Librarianship and Scholarly Communication*, 6(1), eP2201. <https://doi.org/10.7710/2162-3309.2201>
- Choudhury, S. (2017). Data management and preservation of digital research data. *Against the Grain*, 29(5), 30–34. <https://doi.org/10.7771/2380-176x.7838>
- Coates, H. L., Carlson, J., Ryan, C., Henderson, M., Johnston, L. R., & Shorish, Y. (2018). How are we measuring up? Evaluating research data services in academic libraries. *Journal of Librarianship and Scholarly Communication*, 6. <http://dx.doi.org/pallas2.tcl.sc.edu/10.7710/2162-3309.2226>
- Cousijn, H., & Lammey, R. (2018, November 8). *Why data citation matters to publishers and data repositories*. Crossref. <https://www.crossref.org/blog/why-data-citation-matters-to-publishers-and-data-repositories/>
- Cox, A. M., Kennan, M. A., Lyon, L., & Pinfield, S. (2017). Developments in research data management in academic libraries: towards an understanding of research data service maturity. *Journal of the Association for Information Science & Technology*, 68(9), 2182–2200. <https://doi.org/10.1002/asi.23781>
- Cox, A. M., Kennan, M. A., Lyon, L., Pinfield, S., & Sbaffi, L. (2019). Maturing research data services and the transformation of academic libraries. *Journal of Documentation*, 75(6), 1432–1462. <https://doi.org/10.1108/JD-12-2018-0211>

- Faniel, I. M., & Connaway, L. S. (2018). Librarians' perspectives on the factors influencing research data management programs. *College & Research Libraries*, 79(1), 100–119. <https://doi.org/10.5860/crl.79.1.100>
- Hickson, S., Connor, M., Poulton, K. A., Richardson, J., & Wolski, M. (2016). Modifying researchers' data management practices. *IFLA Journal*, 42(4), 253–265. <https://doi.org/10.1177/0340035216673856>
- Johnston, L. (2017). *Curating research data*. Association of College and Research Libraries, a division of the American Library Association.
- Joo, S., & Peters, C. (2020). User needs assessment for research data services in a research university. *Journal of Librarianship & Information Science*, 52(3), 633–646. <https://doi.org/10.1177/0961000619856073>
- Kenyon, J., Attebury, R., Doney, J., Godfrey, B., Martinez, J., & Seiferle-Valencia, M. (2020). Help-seeking behaviors in research data management. *Issues in Science & Technology Librarianship*, 96, 1–1. <https://doi.org/10.29173/istl2568>
- Khan, N., Thelwall, M., & Kousha, K. (2021). Measuring the impact of biodiversity datasets: data reuse, citations and altmetrics. *Scientometrics*, 126(4), 3621–3639. <https://doi.org/10.1007/s11192-021-03890-6>
- Kouper, I., Fear, K., Ishida, M., Kollen, C., & Williams, S. C. (n.d.). Research data services maturity in academic libraries. *Curating Research Data: Practical Strategies for Your Digital Repository (Vol. 1, pp. 153-170)*. Association of College and Research Libraries.
- Mannheimer, S. (2014). Ready, engage! Outreach for library data services. *Bulletin of the Association for Information Science & Technology*, 41(1), 42–44. <https://doi.org/10.1002/bult.2014.1720410115>
- Morgan, A., Duffield, N., & Walkley Hall, L. (2017). Research data management support: sharing our experiences. *Journal of the Australian Library & Information Association*, 66(3), 299–305. <https://doi.org/10.1080/24750158.2017.1371911>
- Nicholson, S. W., & Bennett, T. B. (2017). The good news about bad news: communicating data services to cognitive misers. *Journal of Electronic Resources Librarianship*, 29(3), 151–158. <https://doi.org/10.1080/1941126X.2017.1340720>
- Perrier, L., & Barnes, L. (2018). Developing research data management services and support for researchers: a mixed methods study. *Partnership: The Canadian Journal of Library and Information Practice and Research*, 13(1), 1–23. <http://dx.doi.org/pallas2.tcl.sc.edu/10.21083/partnership.v12i2.4115>
- Ricardo, M., and Urban, S. (2015). *Databases Illuminated*. Jones & Bartlett Learning.
- Rinehart, A. (2016). Finding the connection: research data management and the office of research. *Bulletin of the Association for Information Science & Technology*, 43(1), 28–30. <https://doi.org/10.1002/bul2.2016.1720430107>
- Schwartz, M. (2014). Charleston conference report. *Library Journal*, 139(20), 1–1.

- Shen, Y. (2016). Strategic planning for a data-driven, shared-access research enterprise: Virginia Tech Research Data Assessment and Landscape study. *College and Research Libraries (Online)*, 77(4), 500. <https://doi.org/10.5860/crl.77.4.500>
- Singh, A. (2019, March 24). Enabling researchers to make their data count. <http://dx.doi.org/10.2139/ssrn.3359161>
- Vannan, S., Downs, R. R., Meier, W., Wilson, B. E., and Gerasimov, I. V. (2020). Data sets are foundational to research. Why don't we cite them? *Eos: Science News by AGU*. <https://doi.org/10.1029/2020EO151665>
- Whitmire, A. L., Boock, M., & Sutton, S. C. (2015). Variability in academic research data management practices. *Program*, 49(4), 382–407. <http://dx.doi.org.pallas2.tcl.sc.edu/10.1108/PROG-02-2015-0017>
- Wiley, C. (2014). Metadata use in research data management. *Bulletin of the Association for Information Science & Technology*, 40(6), 38–40. <https://doi.org/10.1002/bult.2014.1720400612>
- Wright, S., Whitmire, A., Zilinski, L., & Minor, D. (2014). Collaboration and tension between institutions and units providing data management support. *Bulletin of the Association for Information Science & Technology*, 40(6), 18–21. <https://doi.org/10.1002/bult.2014.1720400607>
- Yu, H. H. (2017). The role of academic libraries in research data service (RDS) provision. *Electronic Library*, 35(4), 783–797. <https://doi.org/10.1108/EL-10-2016-0233>
- Yu, F., Deuble, R., & Morgan, H. (2017). Designing research data management services based on the research lifecycle—a consultative leadership approach. *Journal of the Australian Library & Information Association*, 66(3), 287–298. <https://doi.org/10.1080/24750158.2017.1364835>

Appendix A

RESEARCH DATA MANAGEMENT SURVEY

1. To begin, please tell us your primary role at this university
 - Faculty
 - Staff
 - Graduate Student
 - Other

2. Please tell us what college you are mainly affiliated with:
 - College of Agriculture, Forestry, and Life Sciences
 - College of Architecture, Arts, and Humanities
 - College of Behavioral, Social, and Health Sciences
 - Wilbur O. and Ann Powers College of Business
 - College of Education
 - College of Engineering, Computing, and Applied Sciences
 - College of Science
 - University Libraries
 - Other

3. Please tell us what department you are mainly affiliated with:
 - Department of Agricultural Sciences
 - Department of Animal and Veterinary Sciences
 - Department of Food, Nutrition and Packaging Science
 - Department of Forestry and Environmental Conservation
 - Department of Plant and Environmental Sciences
 - School of Architecture
 - Department of Art
 - Department of City Planning and Real Estate Development
 - Nieri Family Department of Construction Science and Management
 - Department of English
 - Department of History and Geography
 - Department of Landscape Architecture

- Department of Languages
- Department of Performing Arts
- Department of Philosophy and Religion
- Department of Communication
- School of Nursing
- Department of Parks, Recreation and Tourism Management
- Department of Political Science
- Department of Psychology
- Department of Public Health Sciences
- Department of Sociology, Anthropology and Criminal Justice
- School of Accountancy
- Department of Aerospace Studies
- John E. Walker Department of Economics
- Department of Financial Management
- Department of Graphic Communications
- Department of Management
- Department of Marketing
- Department of Military Leadership
- Master of Business Administration
- Department of Education and Human Development
- Department of Education and Organizational Leadership Development
- Department of Teaching and Learning
- General Engineering Program
- Department of Automotive Engineering
- Department of Bioengineering
- Department of Chemical and Biomolecular Engineering
- Glenn Department of Civil Engineering
- School of Computing
- Holcombe Department of Electrical and Computer Engineering
- Department of Engineering and Science Education
- Department of Environmental Engineering and Earth Sciences

- Department of Industrial Engineering
 - Department of Materials Science and Engineering
 - Department of Mechanical Engineering
 - Department of Biological Sciences
 - Department of Chemistry
 - Department of Genetics and Biochemistry
 - School of Mathematical and Statistical Sciences
 - Department of Physics and Astronomy
 - University Libraries
 - University IT
 - Other
4. Think about your most recent research project. What kind of data did you produce? Please tell us the types of files your research produced by selecting those boxes below. If you are unsure, please select “Other” and provide a brief description of your data.
- Non-digital text (handwritten notes/sketches/paper lab notebooks)
 - Digital text and numeric (data,.csv,.xls,.docx,.pdf, etc.)
 - Audiovisual data (images, video, audio)
 - Computer code and scripts
 - Statistical (SAS, MatLab, etc.)
 - Databases
 - GIS
 - Digital gene sequences
 - Presentations
 - Other (please describe)
5. Again, think about your most recent research project. Approximately how much digital data did you produce?
- $1 \leq$ GB (gigabyte)
 - > 1 GB but < 100 GB
 - > 100 GB but < 1 TB (terabyte)
 - > 1 TB but < 100 TB
 - > 100 TB but < 1 PB (petabyte)
 - > 1 PB

- I don't know
6. How do you protect your files from corruption or loss? Please select all that apply
- Copies are saved on a disk, USB drive, or computer hard drive
 - Copies are saved on a local server
 - Copies are saved on a central campus server
 - Copies are saved on a web-based or cloud server
 - Copies are stored in a data repository or archive
 - Backup files are automatically generated
 - Backup files are manually generated
 - Access to files is restricted
 - I have not adopted any strategies
 - Other (please specify)
 - How have you shared your data in the past year? Please select all that apply
 - An appropriate version of the data i.e. cleaned, de-identified, etc. is made publicly available on my website, a publisher's website, or at a public data archive or repository
 - The data is not made publicly available, but I respond to individual requests for access on a case-by-case basis
 - The data is not made publicly available beyond the members of the research team
 - Additional comments:
7. Why have you not shared your data when your research was complete? Please select all that apply
- Little value to others
 - It takes too much time to get an appropriate version ready to share i.e. cleaned, de-identified, etc.
 - My funder didn't require me to
 - Confidentiality or privacy of participants (HIPPA, FERPA, classified data, vulnerable populations, etc.)
 - Academic privacy
 - Intellectual property rights
 - I'm waiting until I publish a paper using my data
 - Commercialization or patent issues

- Data has license or usage restrictions that prevent sharing
 - Data requires secure access I am unable to provide
 - I always share my data
 - Other (please specify)
8. Metadata is descriptive information or documentation about your data that enables your data to be found online by other researchers. There are two examples of metadata linked below. Have you provided metadata about your data so that others can use it?
- Example 1
 - Example 2
 - Yes
 - No
 - I'm not sure
 - Additional comments:
9. Have you been involved in grant-funded research that required a Data Management Plan (DMP)? Note: This is not the Administrative Data Management Plan you may have submitted to University IT.
- Yes
 - No
 - I'm not sure
10. Have you ever written a DMP? Note: This is not the Administrative Data Management Plan you may have submitted to University IT.
- Yes
 - No
 - I'm not sure
11. Where do you go for help with data management? Please select all that apply. Note: This is not regarding the Administrative Data Management Plan you may have submitted to University IT.
- Colleagues or other researchers in your field
 - Funding agency guidelines
 - University Office of Research Compliance
 - University IT

- University Libraries
 - Disciplinary standards or conventions
 - Other (please specify)
12. Have you ever cited a data set in your research? For example, running an analysis using openly available data or another researcher's data in a research article.
- Yes
 - No
 - I'm not sure
13. If yes, how did you find the data set?
14. Did the data set cost money?
- Yes, I paid for it personally
 - Yes, research funding or the university paid for it
 - No
 - I'm not sure
15. Would you be more likely to use a data set from another researcher if you did not have to pay for access?
- Yes
 - No
 - I'm not sure
16. If Clemson University Libraries purchased access to data sets similar to the way they currently provide access to other resources (databases, journals, articles, books, etc), would you use those resources?
- Yes
 - No
 - I'm not sure
17. Would you prefer Clemson University Libraries to proactively purchase access to a variety of data sets, or would you prefer to request Clemson University Libraries purchase a specific data set you want to use?
- Proactively purchase access
 - Tell Clemson University Libraries what data set I want
 - I don't know

18. To your knowledge, has your data ever been used or cited by another researcher?
 - Yes
 - No
 - I'm not sure
19. How can Clemson University Libraries help you via workshops or an individual consultation? Please select all that apply
 - Planning for and setting up data management practices prior to beginning my research
 - Writing a Data Management Plan (DMP)
 - Understanding and complying with my funding agency's data sharing mandate
 - Understanding my data sharing options
 - Finding and/or choosing a data repository
 - Describing and documenting my data for online discoverability (creating metadata)
 - Getting my data set a DOI (digital object identifier)
 - Getting myself an ORCID ID (researcher identifier)
 - Finding datasets to use in my research
 - Short-term (< 5 years) data storage and access
 - Long-term (> 5 years) data storage and access
 - Data citation (how and why to cite data)
 - Data reference (how and where to find data)
 - Additional comments
20. What benefits do you think a research data repository would have for University researchers?
21. What concerns would you have about a research data repository?
22. Do you think this university should have a research data repository?
 - Yes
 - No
 - Other (please explain)
23. Please share any additional thoughts or concerns you have regarding your research data needs