

The International Journal of Digital Curation

Volume 8, Issue 2 | 2013

Disciplinary Differences in Faculty Research Data Management Practices and Perspectives

Katherine G. Akers,
eScience Librarian,
Emory University Libraries

Jennifer Doty,
Data Management Specialist,
Emory University Libraries

Abstract

Academic librarians are increasingly engaging in data curation by providing infrastructure (e.g., institutional repositories) and offering services (e.g., data management plan consultation) to support the management of research data on their campuses. Efforts to develop these resources may benefit from a greater understanding of disciplinary differences in research data management needs. After conducting a survey of data management practices and perspectives at our research university, we categorized faculty members into four research domains – arts and humanities, social sciences, medical sciences, and basic sciences – and analyzed variations in their patterns of survey responses. We found statistically significant differences among the four research domains for nearly every survey item, revealing important disciplinary distinctions in data management actions, attitudes and interest in support services. Serious consideration of both the similarities and dissimilarities among disciplines will help guide academic librarians and other data curation professionals in developing a range of data management services that can be tailored to the unique needs of different scholarly researchers.



Introduction

Research data need not merely serve as material underlying conference papers, journal articles, and books. Rather, if documented, preserved and made accessible, datasets can stand alone as scholarly products with the potential to impact future research (Williford and Henry, 2012). Academic librarians are increasingly becoming engaged in data curation by providing infrastructure and developing services to support the management of research data on their campuses (ACRL Research Planning and Review Committee, 2012; Heidorn, 2011; Monastersky, 2013; Olendorf & Koch, 2012; Reznik-Zellen, Adamick and McGinty, 2012; Soehner, Steeves and Ward, 2010; Starr, Willett, Federer, Horning and Bergstrom, 2012; Tenopir, Birch and Allard, 2012). However, the data management needs of researchers vary substantially across disciplines. Not only do researchers in the humanities, social sciences and natural sciences create datasets that differ in size and content, they are also enmeshed in diverse research cultures and communities of interest with different attitudes toward and expectancies for data sharing and archiving (Digital Curation Centre, 2012; Borgman, 2012; Palmer and Cragin, 2008). Effective data curation, therefore, requires an understanding of these disciplinary distinctions and the development of services that are tailored to different populations of academic researchers (Cragin, Palmer, Carlson and Witt, 2010).


The Electronic Data Center¹ at the Emory University Libraries has a long history of helping researchers locate, acquire and prepare data for analysis during the early stages of data collection. Recently, we have begun to expand this support to encompass data management planning as well as the documentation, sharing and preservation of research data. To help guide the development of new services to support the management of data across all phases of the research lifecycle, we followed the lead of other academic libraries (e.g., Bardyn, Resnick and Camina, 2012; Gu, Averkamp, Walton, Saylor and Soderdahl, 2012; Parsons, Grimshaw and Williamson, 2013; Raggett, 2012; Scaramozzino, Ramirez and McGaughey, 2012; Wells Parham, Bodnar and Fuchs, 2010; Westra, 2010) and conducted a campus-wide survey of research data management practices and perspectives. As Emory University comprises several colleges and professional schools spanning a wide range of disciplines, we were particularly interested in whether faculty members in different fields of study have varying approaches to research data management or preferences for particular data management-related services. Therefore, we categorized faculty members into four broad research domains – the arts and humanities, social sciences, medical sciences, and basic sciences – and analyzed differences in their patterns of survey responses.

Methods

Survey Administration

In the fall of 2012, the Emory University Libraries – in cooperation with the Emory University Office of Institutional Research, Planning, and Effectiveness – sent an

¹Emory University Libraries Electronic Data Center: <http://edc.library.emory.edu/>



email invitation to all employees with faculty status, as classified by Human Resources, to voluntarily complete a brief online survey. The survey contained 13 possible questions (see Table 1 in the Appendix) and was administered using Qualtrics software. It was open for four weeks, and three email reminders were sent at one-week intervals. Survey respondents were allowed to move forward in the survey without requiring an answer to each question. For multiple choice questions, respondents were given the choice of selecting ‘other’ and writing in a text response.

Survey Data Cleaning

In some cases (25 out of 1,650 possible cases, < 0.02%), survey respondents did not select a pre-defined answer to multiple choice questions but instead selected ‘other’ and provided a text response that clearly matched one or more of the pre-defined answers. In these cases, survey responses were overwritten to reflect the information provided in the text response.

Categorization of Survey Respondents

To evaluate the differences in patterns of survey responses among researchers in different fields of study, we placed each survey respondent into one of four different categories: arts and humanities, social science, medical science or basic science. In some cases, respondents were categorized solely based upon their primary departmental affiliation (e.g., all respondents from the Art History department were assigned to arts and humanities). In other cases, respondents were categorized based on their specific research topics and methodology (e.g., some respondents from the Psychology department were assigned to social science and others to basic science).

Arts and humanities respondents were faculty members in area studies, art history, classics, creative writing, history, languages, liberal arts, literature, music, philosophy, religion, theater studies and theology. Social science respondents were faculty members in anthropology, business, economics, educational studies, journalism, law, linguistics, medicine, political science, psychology, public health, sociology and women’s studies. Medical science respondents were faculty members in biostatistics and medicine. Basic science respondents were faculty members in biology, biostatistics, chemistry, environmental studies, medicine, physics and psychology.

We considered medical science as research conducted in a clinical setting or otherwise ‘applied’ in nature, whereas basic science is research conducted in a laboratory or field setting or otherwise ‘experimental’ or ‘observational’ in nature.

Statistical Analysis

The representativeness of our samples was calculated for each Emory University liberal arts college (Emory College of Arts and Sciences, Oxford College) and professional school (Candler School of Theology, Goizueta Business School, Nell Hodgson Woodruff School of Nursing, Rollins School of Public Health, School of Law, School of Medicine) using Z-tests. Differences in survey responses among different fields of study (arts and humanities, social science, medical science, and basic science) were evaluated using chi-square (χ^2) tests. Statistical significance was set at $p < 0.05$. In the figures, overall responses regardless of field of study are shown

as gray bars. When statistically significant differences among groups were observed, the bars were overlaid by markers corresponding to the different fields of study. Complete results of all statistical analyses are provided in the Appendix (Tables 1 and 2). Survey data are available at the Emory University Dataverse (Akers and Doty, 2013).

Results

Survey Respondents

From a total of 5,590 Emory University faculty members, 456 initiated the survey (~8% response rate). Of these, 330 answered 'yes' to an initial question of whether they conduct research that generates some type of data and provided answers to at least one subsequent survey question. All analyses focused only on these 330 faculty members, who represented all of Emory University's major schools and colleges. Statistically representative samples of faculty members were obtained from the Candler School of Theology, Goizueta Business School, Oxford College, Nell Hodgson Woodruff School of Nursing and Rollins School of Public Health; an over-representative sample was obtained from the College of Arts and Sciences; and under-representative samples were obtained from the School of Law and School of Medicine (see Table 2 in the Appendix). The overall margin of error was $\pm 5\%$.

Depending on their research topics and methodology, faculty members were categorized as conducting research in arts and humanities ($n = 54$), social sciences ($n = 78$), medical sciences ($n = 124$), or basic sciences ($n = 74$).

Data Storage and Back-Up

Overall, the amount of digital research data currently stored by individual faculty researchers at Emory University mostly falls within the gigabyte range (Figure 1A). However, compared to researchers in other fields of study, basic science researchers are more likely to have larger quantities (i.e., terabytes) of data than researchers, and arts and humanities researchers are more likely to state that they do not know how much data they are storing.

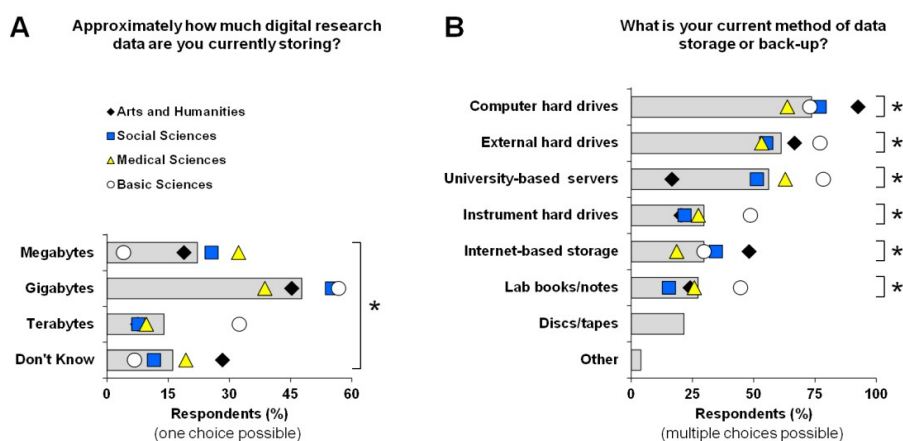


Figure 1. Data storage and back-up. * $p < 0.05$.

The most common methods of storing or backing-up data are via desktop or laptop computer hard drives, external hard drives (including USB drives), and university- or department-based servers (Figure 1B). Basic science researchers are more likely to rely on external hard drives, university-based servers, the hard drives of the instruments used to collect data, and lab books, field notes, or other printed/handwritten materials. By contrast, arts and humanities researchers are more likely to rely on computer hard drives and internet-based storage services, such as Dropbox and Google Drive. There were no significant differences among fields of study in use of CDs, DVDs, tapes or ‘other’ methods of data storage and back-up.

Data Management Planning

Overall, most (~82%) faculty researchers are only somewhat or not at all familiar with requirements for data management or data sharing plans as components of grant applications for federal funding agencies, such as the National Science Foundation (NSF), National Institutes of Health (NIH) and the National Endowment for the Humanities (NEH) (Figure 2). Furthermore, arts and humanities researchers are most likely to be completely unfamiliar with these funding agency requirements for data management plans.

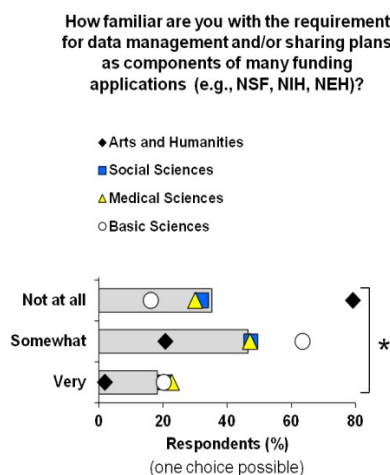


Figure 2. Data management planning. * $p < 0.05$.

Data Sharing

Most faculty researchers at Emory University do not currently share their research data with people outside of their research group (Figure 3A), although researchers in basic sciences were more likely to share their data than researchers in other fields of study.

Considering researchers who do share their data (i.e., those who answered ‘yes’ to the question shown in Figure 3A), emailing data upon request is the most common method of sharing data (Figure 3B). However, social science researchers are least likely to share data via email, and medical science researchers are least likely to post data on personal websites. Basic science researchers are most likely to share data via supplementary material linked to journal articles (e.g., supplementary datasets hosted

by the Public Library of Science (PLOS) journals or links to datasets deposited in the Dryad repository) or posted on department or university websites. No arts and humanities researchers share their data via data repositories or databanks. There were no significant differences among researchers in ‘other’ ways of sharing data, the most frequently noted of which were internet storage services (i.e., Dropbox, Microsoft SkyDrive, Amazon Web Services) and sponsored accounts or restricted access to university-based servers.

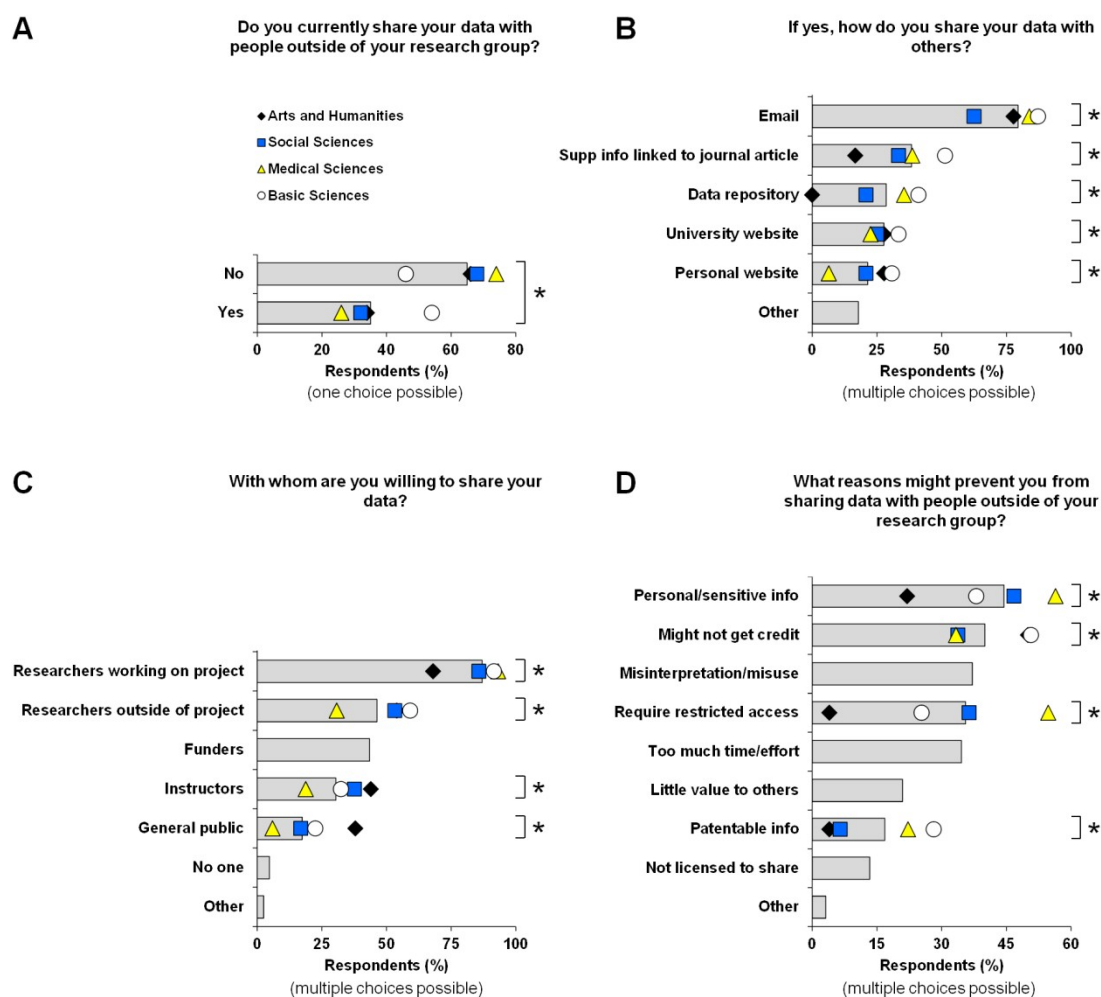


Figure 3. Data sharing. * $p < 0.05$.

Considering all faculty (i.e., those who answered ‘yes’ or ‘no’ to the question shown in Figure 3A), the vast majority are willing to share their data with other researchers (e.g., principal investigators, students, staff) working on the same projects (Figure 3C), although arts and humanities researchers are least willing to do so. However, fewer faculty researchers are willing to share their data with a wider audience. Most medical science researchers are not willing to share data with researchers outside of their projects or with instructors interested in using the data as a teaching tool. Arts and humanities researchers, however, are more willing to share their data with the general public than researchers in other fields of study. Interestingly, nearly half of all faculty members are not willing to share their data with project funders. There were also no significant differences among fields of study in

the proportion of researchers willing to share data with ‘other’ individuals or not willing to share data with anyone. For those who selected ‘other’, the most common write-in response was that the individuals with whom they are willing to share their data depend on whether manuscripts related to the data have been published and whether the data contain sensitive information. A few faculty members stated they are willing to share their data with the Institutional Review Board (IRB) or university administrators.

We next asked what reasons might prevent faculty researchers from sharing data with people outside of their research group (Figure 3D). We found that the top three reasons for not sharing data more widely are:

1. The data contain personal or sensitive information;
2. Researchers might not get credit for their data in terms of acknowledgement, citation, or authorship;
3. The data might be misinterpreted or misused.

Medical and social science researchers are most likely to not share data because they contain personal or sensitive information, or require secure and/or restricted access, whereas researchers in the arts and humanities are least likely to share these concerns. Researchers in basic sciences and arts and humanities are more concerned that they might not get credit for their data than researchers in medical and social sciences. Basic and medical science researchers are more likely to withhold data because the outputs of their research could be patentable or commercialized. There were no differences among fields of study in concerns that data might be misinterpreted or misused, sharing would require too much time and effort, data may be of little value to others, researchers are not licensed to share their data, or ‘other’ reasons preventing data sharing. Those who selected ‘other’ stated that data sharing is prevented by the IRB or the Health Insurance Portability and Accountability Act (HIPAA), that data would not be shared if still in the collection or analysis phases or if manuscripts related to the data had not yet been published, or that data stored on university servers are not easily accessible by others.

Data Preservation

The vast majority of faculty researchers do not deposit their data in data repositories or databanks (Figure 4A). However, researchers in basic sciences are more likely to do so than researchers in other fields of study. The most commonly used data repositories/databanks (reported by those who answered ‘yes’ to the question shown in Figure 4A) are those provided by the National Center for Biotechnology Information – including GenBank, Sequence Read Archive (SRA), Gene Expression Omnibus (GEO), and Database of Genotypes and Phenotypes (dbGaP) – and Protein Data Bank (PDB). Less commonly used data repositories/databanks include Yale University’s NeuronDB and ModelDB, the Cambridge Crystallographic Data Centre (CCDC), Mouse Genome Informatics (MGI), Alzheimer’s Disease Neuroimaging Initiative (ADNI), HIV Drug Resistance Database (HIVdb), Rutgers University’s Cell and DNA Repository, Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD), the Dataverse Network, and the Inter-university Consortium for Political and Social Research (ICPSR).

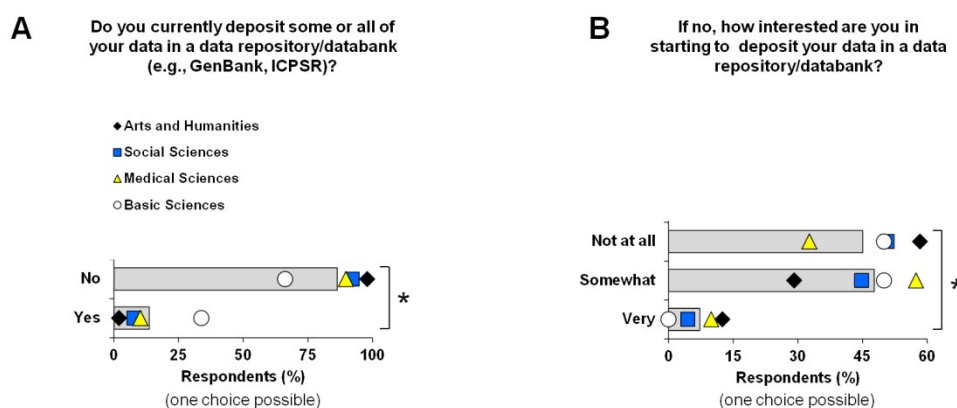


Figure 4. Data preservation. $*p < 0.05$.

Considering those faculty who do not currently deposit their data in a data repository/databank (i.e., those who answered ‘no’ to the question shown in Figure 4A), slightly over half (~55%) stated they are somewhat or very interested in starting to deposit their data (Figure 4B). In particular, medical science researchers are most interested in starting to deposit their data in a data repository/databank.

Data Documentation

As expected, most faculty researchers are not at all familiar with documenting and/or creating metadata for their data so the contents of their datasets can be understood by others and/or computer-readable (Figure 5), with no significant differences among fields of study.

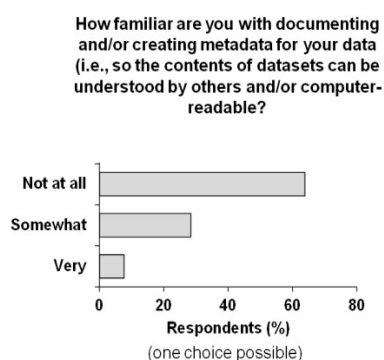


Figure 5. Data documentation. $*p < 0.05$.

Interest in Data Management-Related Services

Finally, we asked faculty members to indicate whether they would use a range of data management-related services if they were offered by Emory University (Figure 6). Regardless of field of study, the two services that garnered the most interest are:

1. Faculty workshops on data management practices,
2. Assistance preparing data management plans for grant applications.

Other services in which faculty expressed interest are:

3. Assistance with data-related confidentiality, privacy, legal, or intellectual property-related issues;
4. Personalized consultation on data management for specific labs or research groups;
5. An institutional data repository;
6. Assistance with data documentation/metadata creation;
7. Workshops on data management practices for students, technicians, administrative assistants, or postdoctoral fellows;
8. Digitization of print or other types of physical records, and
9. Assistance with identifying or using appropriate data repositories/databanks.

Faculty expressed less interest in data citation services (e.g., assignment of permanent digital object identifiers). Researchers who selected ‘other’ wrote in that they are interested in receiving support for setting up their own servers, reliably storing terabytes of data, creating and managing databases, designing data collection tools, and more easily sharing data with other researchers or research groups. A couple of faculty members explicitly stated they are not interested in any services.

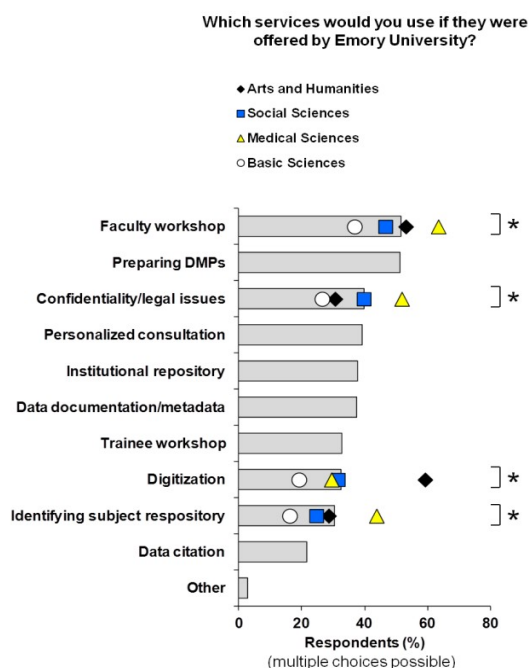


Figure 6. Interest in data management-related services. * $p < 0.05$.

Compared with researchers in other fields of study, researchers in medical science are more interested in faculty workshops on data management, assistance with data-related confidentiality/legal issues, and identifying appropriate data repositories. Also, arts and humanities researchers are most interested in digitization of research materials in physical formats.

Discussion

Different disciplines vary widely in their research funding, technical infrastructures, collaboration networks, source materials, subject populations, methodologies, ethical considerations and types of research outputs. Therefore, to be most effective, data curation requires discipline-specific approaches. Although disciplinary differences in data management issues have been noted by previous reports (Connaway and Dickey, 2009; Digital Curation Center, 2010; Jahnke and Asher, 2012; Pryor, 2009; Witt, Carlson, Brandt and Cragin, 2009), there has been little quantitative analysis of these variations. By categorizing our survey respondents into arts and humanities, social sciences, medical sciences, and basic sciences, we revealed important disciplinary distinctions in data management actions, attitudes, and interest in support services. Notably, our results complement those obtained by another survey recently conducted by the University of Oxford, which had a similar sample size and equivalent breakdown of respondents into four research domains (Wilson, Jeffreys, Patrick, Rumsey and Jefferies, 2013). Studies such as these advance our understanding of the data management needs of different populations of academic researchers.

Compared to other fields of study, the curation of scientific data receives a disproportionate amount of concern, in part due to the large quantities of data produced and amounts of federal funding dedicated to this area of research. As such, the data management practices and perceptions of scientists have been documented by several previous studies (Borgman, Wallis and Enyedy, 2006; Piwowar, 2011; Karasti, Baker and Halkola, 2006; Tenopir et al., 2011; Westra, 2010; Whyte and Pryor, 2011; Williams, 2013). The results of our survey are largely consistent with this existing body of information. As expected, basic science researchers at our university tend to store more digital data than other researchers, with approximately one-third of these researchers storing terabytes of data. For storage and back up, basic science researchers use university-based servers more than other researchers, which may reflect either the preferential allocation of campus resources to basic scientists or their greater familiarity with or need for these resources. As the basic sciences rely heavily on specialized instruments for data collection, the hard drives of these instruments are also key data storage locations. Further, basic scientists often use laboratory notebooks or other handwritten documents to record data or metadata, which presents the challenge of integrating physical and digital records to preserve the meaningfulness of data over time (Briney, 2012). Not surprisingly, basic scientists are the most familiar with funding agency requirements for data management plans. They are also the most likely to share data with people outside of their research groups and to deposit some or all of their data in data repositories or databanks. As noted by another study (Tenopir et al., 2011), the primary reason why basic scientists choose not to share their data is that they might not get credit for their work. However, at the same time we did not find a high degree of interest in services related to data citation, perhaps because researchers are relatively unfamiliar with this concept. Therefore, basic science researchers in particular may benefit from increased awareness of

methods to receive more credit for their data via the publication of data papers (Chavan and Penev, 2011), the archival of datasets in data repositories, the assignment of DOIs to datasets, and the consideration of datasets as products of scholarly research in NSF funding applications (National Science Foundation, 2013).

A growing data-intensiveness of scholarly research applies not only to the basic sciences but also to the social sciences. Although social scientists are increasingly generating larger amounts of data, much of this data is not shared or archived, largely due to concerns about confidentiality or privacy of human research subjects (Digital Curation Centre, 2010; Gutmann et al., 2009; Jahnke and Asher, 2012; King, 2011). We found that only around a third of social science researchers at our university share their data with people outside of their research group. Compared to other types of researchers, social scientists are least likely to share their data via email, perhaps because their data contain private or confidential information that require a more secure method of transmission. Consistent with this possibility, two of the most pressing reasons that might prevent social science researchers from sharing their datasets are that they contain personal or sensitive information or require restricted modes of access. To support the specific needs of social science researchers, academic librarians could assist in the careful de-identification of data (DeWolf, 2002), champion new tools allowing protected storage of and controlled access to sensitive data (Jahnke and Asher, 2012), or create a ‘cold room’ for secure computing within the library, such as that provided by the University of Wisconsin-Madison Libraries² or the Johns Hopkins University Population Center³.

Many of the issues that prevent data sharing among social science researchers are also relevant to medical science researchers. Although federal funding agencies, such as the NIH, strongly encourage data sharing, only a fraction of clinical research data are shared beyond the original research teams to avoid potentially violating the Health Insurance Portability and Accountability Act (HIPAA) for the protection of patient privacy (Freyman, Kirby, Perry, Clunie and Jaffe, 2012). Compared to researchers in other fields of study, medical researchers at our university are least willing to share their data with researchers not working on the same projects, instructors interested in using data as teaching tools, or the general public. Similar to social scientists, medical researchers also state that the most critical reasons for not sharing datasets are because they contain personal/sensitive information or require restricted access. However, researchers in the medical sciences expressed high levels of interest in taking advantage of data management-related services on campus, particularly attending workshops on data management practices and receiving assistance with data-related confidentiality and legal issues. Furthermore, although most medical researchers do not currently deposit their data in data repositories or databanks, approximately two-thirds are interested in doing so, with assistance identifying or using appropriate data repositories being a highly desired service. Therefore, complementary to a previous study that identified a need for easier discovery and use of existing datasets that are relevant to medical research (Bardyn, Resnick and Camina, 2012), we found that medical researchers could also benefit from assistance finding appropriate places to archive the datasets that they produce.

²The ‘Cold Room’ Secure Computing Enclave at DISC:
<http://www.disc.wisc.edu/restricted/coldroom.html>

³ Hopkins Population Center Data Services:
http://web.jhu.edu/popcenter/hpc_associate_services/data_services/

←—————→

Interestingly, researchers in the arts and humanities were most likely to state that they do not know how much data they are storing, perhaps due to uncertainty regarding whether their research material meets the commonly accepted definition of ‘data’. Therefore, when considering the needs of researchers in the arts and humanities, academic librarians may want to adopt a view of data that is more expansive than that typically applied to the natural and social sciences (Muñoz and Renear, 2011; Williford and Henry, 2012). At our university, arts and humanities researchers tend not to store and back up their data using university-based servers but instead rely heavily on computer/external hard drives and internet-based storage. This may reflect a need to easily access data from off-campus locations during field or archival work. Compared to researchers in other fields of study, arts and humanities researchers are much less familiar with funding agency requirements for data management plans, which is not surprising considering that these researchers are typically less dependent on sources of federal funding that are largely skewed to the natural sciences. Practically no arts and humanities researchers who responded to our survey use data repositories to archive or share their data, most likely because fewer repositories exist for this type of data, such as the UK Data Archive or the Cultural Policy and the Arts National Data Archive (CPANDA). Although this population of researchers is least likely to share their data with other researchers working on their projects, perhaps because their studies are often performed without collaborators, they are the most interested in sharing their data with the general public. Of all of the potential data management-related services, arts and humanities researchers were most interested in the digitization of physical research materials. These survey responses highlight some of the ways in which academic librarians can be involved in developing new approaches for the curation of arts and humanities data (Muñoz and Renear, 2011; Svensson, 2011), including advocating for better data storage and computing facilities on campus, creating institutional or subject-specific data repositories to support data preservation and sharing with a wide audience, and providing more options for high quality digitization of a variety of physical research materials.

Conclusions

A primary objective of conducting this survey was to gather information on researchers’ practices and perspectives to guide our development of library services to support the management of research data at Emory University. By obtaining survey responses from representative samples of faculty researcher populations, we were able to identify some of the most appropriate research data services for our campus and recognize areas in which stronger partnerships and collaborations between the library and other campus units are essential for the successful implementation of such services.

As our survey was in progress, we worked to set up Shibboleth authentication access for Emory University researchers to the DMPTool⁴, a free online tool created by the University of California Curation Center and California Digital Libraries that walks researchers through creating data management plans for grant proposals. Soon afterward, we also held a workshop that introduced the DMPTool and general concepts in research data management to faculty and graduate students. When our

⁴DMPTool: <https://dmp.cdlib.org/>

survey was complete, we were happy to find that the two potential services receiving the most interest were faculty workshops on data management practices and assistance with preparing data management plans, meaning that our initial efforts to provide wider support for research data management were in line with the needs of our faculty researchers. To further this line of support for data management plan preparation, we are now working to customize the DMPTool to provide sample language and links to resources that are relevant and specific to our university.

Our survey results have also helped us determine appropriate solutions to research data storage at our university. Our institutional repository, OpenEmory⁵, currently accepts only peer-reviewed journal articles authored by faculty members and not research data. However, the somewhat low level of interest by faculty members in an institutional repository for research data suggests that expanding OpenEmory to accept research data – which would require development costs, more storage space, and additional staff – might not be warranted. Instead, we are exploring other ways of supporting the preservation and sharing of research data that might be more useful to our faculty members, such as facilitating deposit of data into disciplinary repositories or setting up an instance of the Dataverse Network⁶ for our university. These alternative options could allow us to address researchers' concerns with sharing confidential or sensitive information, while also increasing awareness of the benefits of releasing datasets to a wider audience and making them citable.

By categorizing our survey respondents into the arts and humanities, social sciences, medical sciences, and basic sciences, we were able to reveal key distinctions among different populations of researchers in their research data management needs. For example, we discovered that arts and humanities researchers (as well as librarians) may benefit from participating in conversations about definitions of arts and humanities research data and how they may differ from those for other research domains. Arts and humanities faculty may also benefit from a greater awareness of campus resources aimed at helping researchers manage their digital files. Therefore, we are now targeting arts and humanities scholars on campus through widely advertised workshops that address methods for managing one's digital assets. Furthermore, we are engaged in formal efforts to develop greater support for managing the data used in ongoing and emerging digital humanities projects as part of the newly formed Emory Center for Digital Scholarship⁷. It is important to mention, however, that although we considered only differences in data management actions and attitudes among four broad research domains, further distinctions are certain to exist among specific disciplines (i.e., astronomy vs. ecology, psychology vs. economics, art history vs. literature). Nonetheless, serious consideration of both the similarities and dissimilarities among disciplines will help guide academic librarians in developing a range of data management-related services that can be tailored to the unique needs of different researchers, thereby resulting in more effective and comprehensive approaches to research data curation.

⁵OpenEmory: <https://open.library.emory.edu/>

⁶The Dataverse Network: <http://thedata.org/>

⁷Emory Center for Digital Scholarship: <http://digitalscholarship.emory.edu/>

Acknowledgements

We thank Vincent Carter for his generous assistance with survey design and administration and analysis of sample representativeness. We also thank members of the Emory University Libraries Research Data Management team for feedback on survey design and presentation of the results.

References

- ACRL Research Planning and Review Committee. (2012). 2012 top ten trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News*, 73, 311-320. Retrieved from http://crln.acrl.org/content/73/6/311_full
- Akers, K.G. & Doty, J. (2013). Emory University research data management faculty survey. *Emory University Dataverse*. Retrieved from <http://hdl.handle.net/1902.29/11578>
- Bardyn, T.P., Resnick, T. & Camina, S.K. (2012). Translational researchers' perceptions of data management practices and data curation needs: Findings from a focus group in an academic health sciences library. *Journal of Web Librarianship*, 6, 274-287. [doi:10.1080/19322909.2012.730375](https://doi.org/10.1080/19322909.2012.730375)
- Borgman, C.L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63, 1059-1078. [doi:10.1002/asi.22634](https://doi.org/10.1002/asi.22634)
- Borgman, C. L., Wallis, J. C. & Enyedy, N. (2006). Building digital libraries for scientific data: An exploratory study of data practices in habitat ecology. In J. Gonzalo, C. Thanos, M. F. Verdejo & R.C. Carrasco (Eds.), *Lecture Notes in Computer Science: Vol. 4172. Research and Advanced Technology for Digital Libraries* (pp. 170–183). [doi:10.1007/11863878_15](https://doi.org/10.1007/11863878_15)
- Briney, K. (2012). Lab notebooks as data management. *Special Libraries Association Winter Virtual Conference*. Retrieved from <http://www.slideshare.net/kbriney/lab-notebooks-sla-talk>
- Chavan, V. & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, 12, S2. [doi:10.1186/1471-2105-12-S15-S2](https://doi.org/10.1186/1471-2105-12-S15-S2)
- Connaway, L.S. & Dickey, T.J. (2009). Towards a profile of the researcher of today: common themes identified in an analysis of JISC Virtual Research Environment and Digital Repository Projects. Retrieved from <http://repository.jisc.ac.uk/418/>
- Cragin, M.H., Palmer, C.L., Carlson, J.R. & Witt, M.W. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A*, 368, 4023-4038. [doi:10.1098/rsta.2010.0165](https://doi.org/10.1098/rsta.2010.0165)

- DeWolf, V.A. (2002). Issues in accessing and sharing confidential survey and social science data. *Data Science Journal*, 2, 66-74. [doi:10.2481/dsj.2.66](https://doi.org/10.2481/dsj.2.66)
- Digital Curation Centre. (2010). Data dimensions: disciplinary differences in research data sharing, reuse and long term viability. Retrieved from <http://www.dcc.ac.uk/sites/default/files/documents/publications/SCARP-Synthesis.pdf>
- Doucette, L. & Fyfe, B. (2013). Drowning in research data: Addressing data management literacy of graduate students. *Association of College and Research Libraries Conference*. Retrieved from http://www.ala.org/acrl/sites/ala.org/acrl/files/content/conferences/confsandpreconfs/2013/papers/DoucetteFyfe_Drowning.pdf
- Freyman, J.B., Kirby, J.S., Perry, J.H., Clunie, D.A. & Jaffe, C.C. (2012). Image data sharing for biomedical research: Meeting HIPAA requirements for de-identification. *Journal of Digital Imaging*, 25, 14-24. [doi:10.1007/s10278-011-9422-x](https://doi.org/10.1007/s10278-011-9422-x)
- Gu, X., Averkamp, S., Walton, L., Saylor, N. & Soderdahl, P. (2012). Report on the University of Iowa Libraries' Data Management Needs Survey. *University of Iowa Libraries*. Retrieved from http://blog.lib.umn.edu/lmcguire/hslm/Data_Management_at_UIowa_SurveyReport_20121121.pdf
- Gutmann, M.P., Abrahamson, M., Adams, M.O., Altman, M., Arms, C., Bollen, K., Carlson, M., Crabtree, J., Donakowski, D., King, G., Lyle, J., Maynard, M., Pienta, A., Rockwell, R., Timms-Ferrara, L. & Young, C.H. (2009). From preserving the past to preserving the future: The Data-PASS project and the challenges of preserving digital social science data. *Library Trends*, 57, 315-337. [doi:10.1353/lib.0.0039](https://doi.org/10.1353/lib.0.0039)
- Heidorn, P.B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, 51, 662-672. [doi:10.1080/01930826.2011.601269](https://doi.org/10.1080/01930826.2011.601269)
- Jahnke, L.M. & Asher, A. (2012). The problem of data: Data management and curation practices among university researchers. *Council on Library and Information Resources*. Retrieved from <http://www.clir.org/pubs/reports/pub154/problem-of-data>
- Karasti, H., Baker, K.S. & Halkola, E. (2006). Enriching the notion of data curation in e-science: Data managing and information infrastructuring in the long term ecological research (LTER) network. *Computer Supported Cooperative Work*, 15, 213-358. [doi:10.1007/s10606-006-9023-2](https://doi.org/10.1007/s10606-006-9023-2)
- King, G. (2011). Ensuring the data-rich future of the social sciences. *Science*, 331, 719-721. [doi:10.1126/science.1197872](https://doi.org/10.1126/science.1197872)
- Monastersky, R. (2013). Publishing frontiers: The library reboot. *Nature*, 495, 430-432. [doi:10.1038/495430a](https://doi.org/10.1038/495430a)

- Muñoz, T. & Renear, A. (2011). Issues in humanities data curation. *Humanities Data Curation Summit*. Retrieved from <http://cirssweb.lis.illinois.edu/paloalto/whitepaper/>
- National Science Foundation. (2013). *Chapter II: Proposal Preparation Instructions*. Retrieved from http://www.nsf.gov/pubs/gpg/nsf04_23/2.jsp
- Olendorf, R. & Koch, S. (2013). Beyond the low hanging fruit: Data services and archiving at the University of New Mexico. *Journal of Digital Information, 13*. Retrieved from <http://journals.tdl.org/jodi/index.php/jodi/article/view/5878/5882>
- Palmer, C.L. & Cragin, M.H. (2008). Scholarship and disciplinary practices. *Annual Review of Information Science and Technology, 42*, 163-212. [doi:10.1002/aris.2008.1440420112](https://doi.org/10.1002/aris.2008.1440420112)
- Parsons, T., Grimshaw, S. & Williamson, L. (2013). *Research Data Management Survey*. University of Nottingham. Retrieved from http://eprints.nottingham.ac.uk/1893/1/ADMIRE_Survey_Results_and_Analysis_2013.pdf
- Piwowar, H.A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS ONE, 6*. E18657. [doi:10.1371/journal.pone.0018657](https://doi.org/10.1371/journal.pone.0018657)
- Pryor, G. (2009). Multi-scale data sharing in the life sciences. *International Journal of Digital Curation, 4*(3), 71-82. [doi:10.2218/ijdc.v4i3.115](https://doi.org/10.2218/ijdc.v4i3.115)
- Raggett, M. (2012). DICE final report. *Digital Communication Enhancement Project at London School of Economics and Political Science*. Retrieved from <http://lusedice.files.wordpress.com/2012/07/dice-final-report-v-1-2.pdf>
- Reznik-Zellen, R.C., Adamick, J. & McGinty, S. (2012). Tiers of research data support services. *Journal of eScience Librarianship, 1*(3). [doi:10.7191/jeslib.2012.1002](https://doi.org/10.7191/jeslib.2012.1002)
- Scaramozzino, J.M., Ramirez, M.L. & McGaughy, K.J. (2012). A study of faculty data curation behaviors and attitudes at a teaching-centered university. *College & Research Libraries, 73*, 349-365. Retrieved from <http://crl.acrl.org/content/73/4/349.full.pdf+html>
- Soehner, C., Steeves, C. & Ward, J. (2010). E-Science and data support services: A study of ARL member institutions. *Association of Research Libraries*. Retrieved from <http://www.arl.org/storage/documents/publications/escience-report-2010.pdf>
- Starr, J., Willett, P., Federer, L., Horning, C. & Bergstrom, M.L. (2012). A collaborative framework for data management services: The experience of the University of California. *Journal of eScience Librarianship, 1*(2). [doi:10.7191/jeslib.2012.1014](https://doi.org/10.7191/jeslib.2012.1014)

- Steinhart, G., Chen, E., Arguillas, F., Dietrich, D. & Kramer, S. (2012). Prepared to plan? A snapshot of research readiness to address data management planning requirements. *Journal of eScience Librarianship*, 1(2).
[doi:10.7191/jeslib.2012.1008](https://doi.org/10.7191/jeslib.2012.1008)
- Svensson, P. (2011). From optical fiber to conceptual cyberinfrastructure. *Digital Humanities Quarterly*, 5(1). Retrieved from
<http://www.digitalhumanities.org/dhq/vol/5/1/000090/000090.html>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M. & Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PLoS ONE*, 6. E21101. [doi:10.1371/journal.pone.0021101](https://doi.org/10.1371/journal.pone.0021101)
- Tenopir, C., Birch, B. & Allard, S. (2012). Academic libraries and research data services: Current practices and plans for the future. *Association of College and Research Libraries*. Retrieved from http://www.ala.org/acrl/sites/ala.org/acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf
- Wells Parham, S., Bodnar, J. & Fuchs, S. (2010). Supporting tomorrow's research: Assessing faculty data curation needs at Georgia Tech. *College & Research Libraries News*, 73, 10-13. Retrieved from <http://crln.acrl.org/content/73/1/10.full>
- Westra, B. (2010). Data services for the sciences: A needs assessment. *Ariadne*, 64. Retrieved from <http://www.ariadne.ac.uk/issue64/westra>
- Whyte, A. & Pryor, G. (2011). Open science in practice: Research perspectives and participation. *International Journal of Digital Curation*, 6(1), 199-213.
[doi:10.2218/ijdc.v6i1.182](https://doi.org/10.2218/ijdc.v6i1.182)
- Williams, S.C. (2013). Data practices in the crop sciences: A review of selected faculty publications. *Journal of Agricultural & Food Information*, 13, 308-325.
[doi:10.1080/10496505.2012.717846](https://doi.org/10.1080/10496505.2012.717846)
- Williford, C. & Henry, C. (2012). One culture: Computationally intensive research in the humanities and social sciences. *Council on Library and Information Resources*. Retrieved from <http://www.clir.org/pubs/reports/pub151>
- Wilson, J.A.J., Jeffreys, P., Patrick, M., Rumsey, S. & Jefferies, N. (2013). University of Oxford research data management survey 2012: The results. *University of Oxford*. Retrieved from <http://blogs.it.ox.ac.uk/damaro/2013/01/03/university-of-oxford-research-data-management-survey-2012-the-results/>
- Witt, M., Carlson, J., Brandt, D.S. & Cragin, M.H. (2009). Constructing data curation profiles. *International Journal of Digital Curation*, 4(3), 93-101.
[doi:10.2218/ijdc.v4i3.117](https://doi.org/10.2218/ijdc.v4i3.117)

Appendix

Table 1. Survey questions, responses, and results of statistical analysis. (n.a.: not analyzed).

Question	Possible Responses	Statistical Analysis
1. Do you conduct research that generates some type of data (e.g., spreadsheets, text, images, videos, audio files, instrument files, photographs, physical samples/specimens, etc.)?	Yes	n.a.
	No	
2. Approximately how much digital research data are you currently storing?	Megabyte range	$\chi^2 (9, n = 329) = 55.82$
	Gigabyte range	$p < 0.001$
	Terabyte range	
	Don't know	
3. What is your current method of data storage or back-up?	Desktop/laptop computers	$\chi^2 (3, n = 330) = 16.74,$ $p = 0.001$
	External hard drives (including USB drives)	$\chi^2 (3, n = 330) = 13.02$ $p = 0.005$
	Department/university server	$\chi^2 (3, n = 330) = 52.06$ $p < 0.001$
	Hard drive of instrument used to collect data	$\chi^2 (3, n = 330) = 22.13$ $p < 0.001$
	Internet-based storage (Dropbox, GoogleDocs, etc.)	$\chi^2 (3, n = 330) = 17.09$ $p = 0.001$
	Lab books/field notes/other printed or handwritten materials	$\chi^2 (3, n = 330) = 17.17$ $p = 0.001$
	CDs/DVDs/tapes	$\chi^2 (3, n = 330) = 5.40$ $p = 0.145$
	Other (please specify)	$\chi^2 (3, n = 330) = 6.51$ $p = 0.089$
4. How familiar are you with the requirement for data management and/or sharing plans as components of many funding applications (e.g., NSF, NIH, NEH)?	Not at all familiar	$\chi^2 (6, n = 329) = 58.74$
	Somewhat familiar	$p < 0.001$
	Very familiar	

Question	Possible Responses	Statistical Analysis
5. Do you currently share your data with people outside of your research group?	Yes	$\chi^2 (3, n = 328) = 16.48$ $p = 0.001$
	No	
5a. If yes, how do you share your data with others?	E-mail upon request	$\chi^2 (3, n = 325) = 18.23$ $p < 0.001$
	Supplementary material linked to journal article (e.g., PLoS journals, Dryad)	$\chi^2 (3, n = 325) = 17.17$ $p = 0.001$
	Data repository/databank (e.g., GenBank, ICPSR)	$\chi^2 (3, n = 325) = 18.84$ $p < 0.001$
	Department/university website	$\chi^2 (3, n = 325) = 8.10$ $p = 0.044$
	Personal website	$\chi^2 (3, n = 325) = 15.05$ $p = 0.002$
	Other (please specify)	$\chi^2 (3, n = 325) = 7.74$ $p = 0.052$
6. With whom are you willing to share your data?	Researchers (PIs, students, staff) working on the project	$\chi^2 (3, n = 315) = 21.28$ $p < 0.001$
	Researchers outside of the project	$\chi^2 (3, n = 315) = 18.75$ $p < 0.001$
	Project funders	$\chi^2 (3, n = 315) = 5.16$ $p = 0.160$
	Instructors interested in using data as a teaching tool	$\chi^2 (3, n = 315) = 13.84$ $p = 0.003$
	General public	$\chi^2 (3, n = 315) = 26.62$ $p < 0.001$
	No one	$\chi^2 (3, n = 315) = 3.81$ $p = 0.283$
	Other (please specify)	$\chi^2 (3, n = 315) = 2.79$ $p = 0.425$
7. What reasons might prevent you from sharing data with people outside of your research group?	Data contain personal/sensitive information	$\chi^2 (3, n = 315) = 18.34$ $p < 0.001$
	Might not get credit (e.g., citation, acknowledgement, authorship)	$\chi^2 (3, n = 315) = 8.89$ $p = 0.031$
	Data might be misinterpreted/misused	$\chi^2 (3, n = 315) = 2.19$ $p = 0.535$

Question	Possible Responses	Statistical Analysis
	Data require secure/restricted access	$\chi^2 (3, n = 315) = 43.69$ $p < 0.001$
	Requires too much time/effort	$\chi^2 (3, n = 315) = 6.67$ $p = 0.083$
	Data of little value to others	$\chi^2 (3, n = 315) = 7.64$ $p = 0.054$
	Commercialization/ patent concerns	$\chi^2 (3, n = 315) = 20.71$ $p < 0.001$
	Not licensed to share data	$\chi^2 (3, n = 315) = 2.46$ $p = 0.483$
	Other (please specify)	$\chi^2 (3, n = 315) = 3.21$ $p = 0.361$
8. Do you currently deposit some or all of your data in a data repository/databank (e.g., GenBank, ICPSR)?	Yes No	$\chi^2 (3, n = 315) = 33.61$ $p < 0.001$
8a. If yes, which data repository/databank do you use?	(please specify)	n.a.
8b. If no, how interested are you in starting to deposit your data in a data repository/databank?	Not at all interested Somewhat interested Very interested	$\chi^2 (6, n = 262) = 18.53$ $p = 0.005$
9. How familiar are you with documenting and/or creating data for your data (i.e., so the contents of datasets can be understood by others and/or computer-readable)?	Not at all familiar Somewhat familiar Very familiar	$\chi^2 (6, n = 299) = 4.79$ $p = 0.571$
10. What services would you use if they were offered by Emory University?	Workshops on data management practices for faculty Assistance preparing data management plans or data sharing plans for grant applications	$\chi^2 (3, n = 302) = 13.01$ $p = 0.005$ $\chi^2 (3, n = 302) = 3.23$ $p = 0.358$

Question	Possible Responses	Statistical Analysis
	Assistance with confidentiality/privacy/legal/intellectual property issues associated with data preservation and/or sharing	$\chi^2 (3, n = 302) = 13.49$ $p = 0.004$
	Personalized consultation on data management practices (for specific labs or groups)	$\chi^2 (3, n = 302) = 3.98$ $p = 0.263$
	Institutional repository/databank for preservation and/or sharing of research data	$\chi^2 (3, n = 302) = 3.88$ $p = 0.274$
	Assistance with documenting data for sharing with others or saving for future use (i.e., metadata creation)	$\chi^2 (3, n = 302) = 4.58$ $p = 0.205$
	Workshops on data management practices for students/technicians/administrative assistants/postdocs	$\chi^2 (3, n = 302) = 4.00$ $p = 0.262$
	Digitization of print or other types of physical records	$\chi^2 (3, n = 302) = 21.98$ $p < 0.001$
	Assistance with identifying/using appropriate data repositories/databanks	$\chi^2 (3, n = 302) = 17.13$ $p = 0.001$
	Data citation services (e.g., assignment of permanent digital object identifiers (DOIs))	$\chi^2 (3, n = 302) = 3.03$ $p = 0.388$
	Other (please specify)	$\chi^2 (3, n = 302) = 2.70$ $p = 0.441$

Table 2. Representativeness of samples by college and professional school.

College	Survey respondents	Statistical analysis
Candler School of Theology	$n = 6$ out of 98	$Z = 0.07, p = 0.94$ (representative)
Goizueta Business School	$n = 8$ out of 106	$Z = 0.66, p = 0.51$ (representative)
Oxford College	$n = 8$ out of 97	$Z = 0.90, p = 0.37$ (representative)
Nell Hodgson Woodruff School of Nursing	$n = 8$ out of 210	$Z = 0.83, p = 0.41$ (representative)
College of Arts and Sciences	$n = 105$ out of 975	$Z = 6.50, p < 0.001$ (over-representative)
School of Law	$n = 3$ out of 225	$Z = 2.87, p = 0.004$ (under-representative)
School of Medicine	$n = 163$ out of 3312	$Z = 3.43, p < 0.001$ (under-representative)