2-28-2014

Data Management at the University of Iowa: A University Libraries Report on Campus Research Data Needs

Shawn Averkamp
University of Iowa

Xiaomei Gu
University of Iowa

Ben Rogers
University of Iowa

DOI: https://doi.org/10.17077/t5j5-ry3i

DATA MANAGEMENT AT THE UNIVERSITY OF IOWA

A University Libraries Report on Campus Research Data Needs

Submitted by Shawn Averkamp, Xiaomei Gu, and Ben Rogers
University Libraries and ITS—Research Services

February 28, 2014

Executive Summary

This data management report was commissioned by the University of Iowa Libraries with the intention of performing a survey of the campus landscape and identifying gaps in data management services. The first stage of data collection consisted of a survey conducted during summer 2012 to which 784 responses were received. The second phase of data collection consisted of approximately 40 in-depth interviews with individuals from the campus and were completed during summer 2013. The individuals engaged during the data collection phase spanned a diverse set of campus programs but should not be considered comprehensive. Information Technology Services was invited to participate in the interview process and has also contributed to this report.

Key findings focus on data management and related areas but in some cases broader feedback is included because of the frequency with which the feedback was received.

Data Management Planning – The status of data management on campus is highly varied both across and within disciplines. Awareness of data management has been increasing as NSF has begun to require data management plans and NIH is enforcing data sharing plans. Many participants required to write plans indicated that they receive help for data management planning from colleagues or support services at the department or college level. The initial survey indicated a need for better support in this area but follow up interviews did not lend strong support to that conclusion. However, recent experiences indicate that some level of service would be beneficial. Requirements in this area are still evolving and at present appear driven by work by the White House Office of Science and Technology Policy (OSTP) on a mandate to require data sharing at large federal funding agencies.

Data Storage – A lack of campus data storage solutions that meet the needs of the research community was a common theme. The primary concerns focused on low capacity, high cost, and difficulty in collaborating using campus storage solutions. A lack of campus solutions for synchronous file sharing (i.e. Dropbox) and backups were frequently raised. It was also noted that a significant portion of those interviewed indicated that they wished to retain their data indefinitely.[1]

Data Organization and Analysis – Data organization and analysis practices are strongly dependent upon research discipline. Common themes in this area indicated a need for early career preparation for those working with data, inaccessibility of desired software tools, and a need for support personnel with skills beyond desktop support.

Data Publishing and Dissemination – Levels of data publishing and sharing varied widely across disciplines. Most individuals indicated a desire and willingness to share data with colleagues, but opinions were more varied regarding sharing with the general public. The greatest concerns are related to the burden of preparing data for sharing and dissemination, risks of sharing (compliance and intellectual property), lack of standards in some areas, and the possibility for misinterpretation of data made public.

Sensitive Data and Compliance – The majority of participants indicated working with sensitive or confidential data, usually human subjects data, and many found navigating compliance processes challenging, particularly with respect to sharing information with collaborators.

Researchers in arts and humanities raised concerns about challenges finding digital humanities-related support services. Participants reported issues in identifying appropriate software solutions for digitally publishing scholarship and data, coordinating and funding digital project development, and securing and maintaining reliable, long-term web hosting--tasks not traditionally associated with arts and humanities

[1] Data collection was completed prior to the launch of the ITS Research Data Storage Service in Fall 2013.

research or culture. While lack of grant funding for support services was cited as a challenge by participants across all disciplines, the absence of an established and effective research support structure in some disciplines may prevent research success, even when funding is available. Further exploration into existing domain-specific support services on campus, such as the College of Nursing's Office for Nursing Research and Scholarship, could provide insight into how to build out models for support in disciplines where none exist.Findings indicate that there are opportunities for improving services and processes. The main challenges point to a need for project or discipline-specific services, appropriately trained support personnel, and funding models that align with services or current research expectations.

Table of Contents

Introduction

As research becomes more data-intensive and collaborative, the University of Iowa is seeing a growing demand for skills and resources to deal with capturing, organizing, storing, publishing, archiving, and sharing data (i.e. data management). Funding agencies, such as National Science Foundation (NSF), National Institute of Health (NIH), and National Endowment for the Humanities (NEH), are increasingly requiring that researchers include data management plans in new grant proposals, and many now require that research data be archived and shared in digital repositories.

Currently, the Digital Research and Publishing unit in the University of Iowa Libraries offers dataset archiving (50GB or less) in its institutional repository and data management plan consultation services. In addition, ITS - Research Services and several other research centers on campus, such as the Clinical Trials Statistical and Data Management Center (CTSDMC), the University of Iowa Center for Health Effects of Environmental Contamination (CHEEC), the Institute for Clinical and Translational Science (ICTS), and the Iowa Social Science Research Center (ISRC), also offer data management services within certain research disciplines. The University of Iowa Libraries is interested in assessing the needs of researchers across campus for managing their data, in order to inform the coordination of services in the Libraries and across campus to support effective data stewardship throughout the research lifecycle.

The University of Iowa Libraries charged Shawn Averkamp, Data Services Librarian, and Xiaomei Gu, Clinical Education Librarian, to conduct a needs assessment using a mixed-mode survey methodology involving web-based questionnaires and face-to-face interviewing. The study protocol was approved by the Institutional Review Boards (IRB) of the University of Iowa. An online survey was conducted first, and results were used to guide the design of focus group and one-on-one interviews with researchers across campus. Averkamp and Gu partnered with Ben Rogers, Director of ITS Research Services, to conduct the interviews. This report summarizes the findings from approximately 40 semi-structured interviews, and survey results are incorporated in the findings where relevant to help support or contextualize interview subjects' responses.

Survey

Using an online questionnaire, this survey aimed to provide starting points to understand University of Iowa researchers' needs and concerns regarding data management. The All Faculty & Staff e-mail list was used to send a survey link to the target survey population, faculty and staff members at the University of Iowa who are directly involved in research (the exact number of target population is unknown). The survey was open from June 26, 2012 to September 18, 2012. A total of 784 responses, both partial and complete, to the survey, were received. Because participants were self-selecting and the exact target population is unknown, this total should not be interpreted as a representative sample of the University research community.

The 13-question survey included questions on types of data, sensitive or confidential data, funding agencies, data management plan requirement, data storage or sharing, and where researchers get assistance and guidance for data management and organization. The survey results illustrate a range of needs across the University's disciplinary spectrum as well as a view into current services, both internal and external, that researchers consult to meet those needs. Needs and services that surfaced in the survey results were then used to guide the design of the interview questions. The online survey and results can be found in Appendix 1. A full report on the survey results can be viewed at: http://ir.uiowa.edu/lib_pubs/152/.

Interviews

INTERVIEW PARTICIPANTS

To select interview participants, survey participants were asked about their willingness to be contacted to answer further questions about data management practices and needs. 183 survey respondents provided their email addresses for this purpose, and interview subjects were first drawn from this pool. In an effort to achieve representation across all Colleges, library liaisons and academic deans from underrepresented Colleges were invited to recommend possible interview participants. Upon agreeing to participate in the study, participants were emailed a copy of the interview questions and an IRB-required letter of consent. Participants were invited to bring colleagues or research assistants who could help them provide a full picture of research activity. In counting participants, these colleagues and assistants were counted with the subject as one "research perspective." Approximately 50 people were interviewed representing 41 "research perspectives." For simplicity, throughout this report these "perspectives" will be referred to in the singular as "interviewees," "participants," or "researchers."

INTERVIEW DESIGN

Participants were invited to interview in focus groups or as individuals, depending on their preference. Scheduling constraints often required individual interviews, regardless of the participant's preference. Interviews were kept to an hour and discussion was semi-structured, based on a prepared list of questions around four major areas of data management: data management planning, data storage, data organization, and data publishing and dissemination. The interview questions were informed by the results of the survey and were intended to spark conversation about areas of common concern in data management. In order to get a better look at specific challenges in data management, participants were not expected to answer every question but were instead encouraged to elaborate on areas of particular interest or concern to their research process. A full list of interview questions can be found in Appendix 2.

Each interview focused more heavily on some aspects of data management than others, so participant comments were recorded, coded, and analyzed qualitatively rather than quantitatively. Comments were first categorized by area of data management (data management planning, data storage, data organization, and data publishing and dissemination) then coded by the function of the comment (describing existing practices, recounting gaps or challenges, demonstrating strengths or opportunities, and offering suggestions or criticisms). Comments were then coded again to identify themes that emerged within each subcategory. This report organizes findings by area of data management and the challenges and opportunities present in each area. "Challenges" inventories the range of issues encountered by participants, while "Opportunities" includes existing solutions and participants' suggestions for addressing challenges.

Findings should not be interpreted as quantitatively representative of any demographic group identified in this study. The purpose of the study was to inventory data management practices and challenges across a sampling of researchers on campus and to surface exemplary services already in place with the goal of improving data management services at the University of Iowa. To protect the identity of participants, College affiliations have been withheld unless deemed relevant to the issue discussed. As this is an administrative report, however, exemplary services mentioned by participants may be identified by name in order to highlight opportunities for growth or collaboration.

Findings

Participant Demographic Information

SURVEY

The survey drew 784 respondents from the pool of all faculty and staff at the University. Though the survey introduction invited responses from a target population of "faculty and staff members at the University of Iowa who are directly involved in research," there is no way to calculate what percentage of respondents actually fell within that target. Survey participants were asked to identify their primary research area (Question 1), their position—faculty researcher, P&S researcher, or other (Question 3), and their department or unit (Question 12). Respondents were largely from the health and biosciences, and just over half of all respondents claimed to be faculty or P&S researchers. A full breakdown of demographic distributions can be found in Appendix 1.

INTERVIEWS

Willing participants were found in almost every College (Figure 1). The total of participants from College of Liberal Arts and Sciences far exceeded all other Colleges as attempts were made to represent the wide diversity of disciplines within that College.

College	Participants
College of Liberal Arts and Sciences	15
Carver College of Medicine	10
College of Engineering	3
College of Public Health	2
College of Pharmacy	2
College of Nursing	2
Graduate College	2
Tippie College of Business	1
College of Dentistry	1
College of Education	1
Other	2
TOTAL	**41**

Figure 1. Distribution of interview subjects across Colleges and other University-affiliated organizations.

Though primary College affiliation may sometimes determine a researcher's access to specialized software or technical support, the shared characteristics of each discipline seemed to be a stronger indicator of specific researcher needs than College affiliation. Therefore, interview participants were also categorized by the primary nature of their research, their funding sources, research personnel, and types of data generated into four broad research disciplines—Health and Biosciences, Education and Social Sciences, Arts and Humanities, and STEM sciences (physical, engineering, and mathematical sciences) (Figure 2). Throughout the report, these broad research disciplines are highlighted where potentially relevant.

Discipline	Characteristics	Participants
Health and Biosciences	NIH, NSF funding; research teams, labs; big data—image, genomic, numeric; sensitive data; data sharing requirements	22
Education and Social Sciences	Small grants or unfunded; student assistants; numeric, survey, textual, video data; some sensitive data	7
Arts and Humanities	Small grants or unfunded; solo research or student assistants; textual, video, audio data	6
STEM Sciences	NSF, NASA, DOD, EPA funding; research teams; small to big data—image, numeric, specimen, GIS data; data management plans	6
TOTAL		**41**

Figure 2. Distribution of interview subjects across broad research disciplines.

Types of Data

The survey defined "data" as "any recorded material necessary to validate your research findings. This is often numeric, tabular data, but it can also be textual data, images, audio/video files, or even artifacts" (Question 2), and the results showed that researchers generate data in many types and formats, most working with two or more (Figure 3). Every data type listed in the survey results was also discussed in greater depth at some point during the interviews, revealing a wide range of discipline-specific data types, including (but not limited to): flow cytometry data, gene sequencing data, microscopy data, biomarkers, air and water quality data, patient data, ethnographic interviews, audio linguistic data, CT scans, satellite imagery, and more.

Answer		Response
Tabular or relational data (e.g. spreadsheet, database)		404
Textual data (e.g. text files, Word documents, PDF)		391
Images		276
Survey / questionnaire data		258
Artifacts, samples, and/or specimens		188
Video		130
Computer programs or code		128
Audio		102
Genomic data		96
Geospatial data		76
Other. please specify:		42

Figure 3. Responses to Survey Question 5, "What kinds of data do you collect, generate or use in the course of your research? Check all that apply." 558 of 784 participants responded.

Data Management Planning

Data management plans (DMPs) are supplementary grant application documents that outline how a researcher will ensure data is secure and well-organized during the research process and provide access to that data after the research is complete. Data management plans are currently required for National Science Foundation (NSF) and National Endowment for Humanities Office of Digital Humanities (NEH ODH) grant funding, and it is anticipated that other federal granting agencies will follow suit. The National Institutes of Health (NIH) requires a similar document called a data sharing plan. Participants were asked if they write data management or data sharing plans and where they find assistance. Interviewees were additionally asked to discuss what challenges they have experienced in data management planning.

A substantial number of survey respondents (186 of 551) reported writing DMPs (Figure 4), though an additional 99 responded that they don't know if they are required to write DMPs (Question 7). This large number may be in part due to confusion over the meaning of "data management plan," as many NIH grant applicants are required to complete data sharing plans. Both groups of respondents were prompted with a question regarding assistance with DMPs (Question 8).

Answer		Response
No		266
Yes		186
I don't know		99
Total		551

Figure 4. Responses to Survey Question 7, "Some funding agencies require a data management plan as part of the grant proposal. Are you required to write data management plans?"

In the interviews, 18 participants discussed writing some sort of data management or data sharing plan to receive research funding, mostly NSF data management plans or NIH data sharing plans. In addition, four interviewees report assisting others in writing DMPs. Two more interviewees are not required to write plans but do engage in some sort of data management planning at the outset of a research project. 13 interviewees do not write DMPs or engage in data management planning.

Challenges

Seven interviewees commented on challenges in data management planning. Two noted a lack of guidance or clear purpose in DMP documentation. Another wished that funding agencies would include more explanation about the purpose of the DMP and what parts of the plan researchers will actually be held accountable for in the future. Three researchers expressed concerns about infrastructure and cost structure in planning for the future, particularly in regard to new expectations for data sharing and publication beyond the life of the grant. Two researchers identified IRB as a bottleneck in efficient data management planning. From the survey responses to the question on data management planning assistance (Question 8), 76 participants answered that they do not receive assistance, but would like help.

Opportunities

When asked where they find assistance in writing DMPs, four interviewees report finding help from colleagues, often senior faculty. Three interviewees receive college, departmental, or lab support. College of Nursing was commended for its Office of Nursing Research & Scholarship[2], which provides

[2] http://www.nursing.uiowa.edu/research/office-of-nursing-research

staff assistance and DMP templates to researchers. A researcher in the College of Engineering credited good administrative support in the lab for making their grant proposals more competitive. This researcher helped the lab develop templates for each research area represented and reports that this system has worked well for them. From the survey responses to the question on data management planning assistance (Question 8), seven respondents selected "other" and specified colleagues, mentors, research staff, and collaborators. 109 respondents said they receive help from their department or college, which could be interpreted to include colleagues, collaborators, and staff.

Three interviewees received assistance from other campus units, including the Office of the Vice President for Research (OVPR), University Libraries' Digital Research & Publishing unit (DRP), and the Bioinformatics Division of the Iowa Institute for Human Genetics. One interviewee in the STEM sciences found sufficient external support from the funding agency, NASA, which provides templates and boilerplate language for grant applicants. From the survey responses, roughly 20 respondents each found assistance from ITS, Institute for Clinical and Translational Science (ICTS), and Division of Sponsored Programs. Seven listed University Libraries as a source of help, and 17 respondents listed specific research centers or institutes, including IIHR—Hydroscience & Engineering, Clinical Trials Statistical and Data Management Center (CTSDMC), Iowa Social Science Research Center (ISRC), Preventive Intervention Center, and Health Effectiveness Research Center (HERCe).

One interviewee suggested that more DMP examples and templates would be helpful, as did a survey respondent who recounted an experience seeking such assistance from Division of Sponsored Programs. Another interviewee suggested that the IRB could improve the online application by either embedding or linking out to supporting resources at the point of need in the workflow. For instance, a link to DMP examples or best practices could be included in the IRB form where such a plan is required.

Data Storage

Data storage includes all methods researchers used to store, access, analyze, back up, and collaborate on data. Most participants in both the surveys and the interviews listed more than one method of data storage. Survey participants were asked how they store or share their data (Question 10). Survey results, illustrated in Figure 5, include respondents' methods of storing, sharing data for collaboration, and sharing data publicly.[3] Survey participants were not asked further questions about data publishing or dissemination.

[3] The term "sharing" was initially used in the survey questions, but the survey responses revealed that people understand this term to mean both sharing for purposes of collaborating on research and sharing of data as a means to disseminate, publish, or meet funding requirements for making data available after research. This term was clarified in the subsequent interviews.

Answer	Response
On a shared drive or server on campus	388
On my computer	375
On external media (e.g., external hard drive, memory stick, CD)	280
On paper (e.g. lab notebooks, printouts)	257
With a commercial online storage service (e.g. Dropbox, Google Docs)	94
Other. Please specify:	37
In a disciplinary repository (e.g. Dryad, ICPSR) or institutional repository (e.g. Iowa Research Online). Please specify:	34
I don't know	6

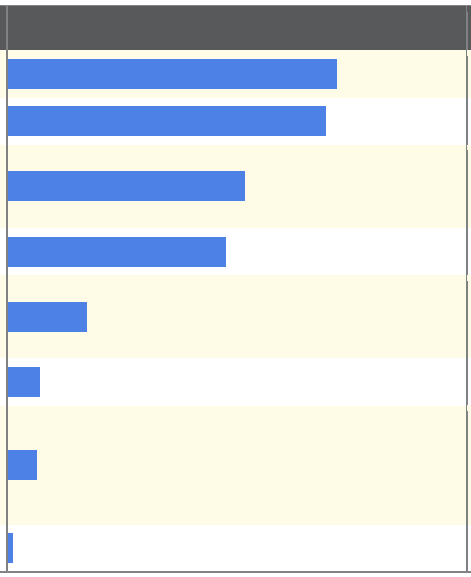


Figure 5. Responses to Survey Question 10, "Where do you store or share your data? Check all that apply." 541 participants responded.

In the interviews, participants were asked to address these methods of "sharing" data separately from discussions on storage, backup, and collaboration. Data publishing and dissemination is discussed later in this report in the section, "Data Publishing, Sharing, and Dissemination." Interviewees were asked to describe their current storage practices and needs, where they find assistance with storage and backups, and what challenges they have experienced in storing and accessing their data, backing up their data, and enabling access for research collaborators.[4] 23 interviewees use departmental or college servers, purchased servers, or the University-provided H or R drive. 14 interviewees store or back up data on external media, such as portable hard drives. This use was usually cited as a means to store data long-term or to conserve space on networked drives. 18 use Dropbox, Google Drive or other third-party cloud services for data storage or collaboration. Four use University-licensed cloud services, such as Qualtrics, REDCap or Xythos. Seven discussed storing data on their desktop computers. Four use other ITS or University services. The Helium HPC cluster was also mentioned as a place where a researcher stored their data for computational analysis. Eight interviewees discussed recording or storing data in paper lab notebooks for some stages of the research lifecycle.

STORAGE SIZE LIMITS

14 survey respondents and nine interview participants spoke directly to the issue of storage space constraints. While researchers frequently supplement network file storage with other means, such as external hard drives or third-party cloud storage services, attitudes expressed in interviews indicated that these means were not preferable, usually because of issues related to ease of access and security.

Challenges

As evidenced above by researchers' primary choice for data storage—networked storage—this concern reflects a common desire for synchronous access to their data. One survey respondent said that "[i]nstitutional servers have had inadequate memory for my data storage needs, which forced me to use external hard drives. I maintain these hard drives with dual backups in a fire proof safe in the lab." From the interviews, a researcher in the STEM sciences cited the H drive's 10GB limit as a "major constraint." A researcher in the arts and humanities was enrolled in ITS Research Services' Research Data Storage

[4] Five interviewees did not discuss this question, usually because the researcher did not directly manage the data and did not know where it was stored.

Service pilot program but had already exhausted the allotted free 3TB of storage and was storing excess data on naked hard drives. Another researcher in the art and humanities described a storage solution consisting of five laptops, two external hard drives, H drive, and at least three cloud services. The researcher explained that this cobbled-together, distributed approach only exacerbated the problem, as it was often necessary to duplicate data across systems for remote access and analysis.

Three interview participants cited cost as an obstacle to securing enough storage space. An interviewee from the health and biosciences expressed concern that grant funding is decreasing as network connectivity and equipment costs are increasing and feels the University should shoulder some of the cost of bandwidth, since they bring in substantial grant funding, and it is often difficult to build infrastructure costs into larger program grants.

Opportunities

In Fall of 2013, after this study was conducted, ITS Research Services moved the Research Data Storage Service[5] from pilot to production status. This service provides all faculty with 3TB network file storage at no cost with the option of purchasing additional space at the rate of $0.60 per GB per year. It is anticipated that this new service may alleviate many researchers' storage constraints.

SYNCHRONOUS STORAGE

Seamless access to research data was frequently discussed as an important factor in research success. Many participants pointed to the lack of a University-provided solution as a major obstacle to effective data management.

Challenges

Many of those interviewed and surveyed expressed the need for more distributed, synchronous storage, both for their own ease of access and for research collaboration with colleagues, assistants, and students. As some participants noted, distributed storage can be difficult to manage and can put additional constraints on research. Many noted the importance of remote access to desktop file systems and departmental servers with some reported challenges related to operating systems and network connections. While a handful of participants reported using University-licensed services, such as Xythos and REDCap, many discussed their use of third-party services like Google Drive and Dropbox.

It is worthwhile to note the use of and attitudes toward Dropbox and other third-party cloud storage. 18 interviewees discussed the use of Dropbox for storing, backing up, or sharing data with collaborators, but the comfort level with third-party cloud storage use varied widely. Most of these participants admitted using Dropbox for its convenience, but one third expressed concerns about security, confidentiality, and copyright. Three made sure to note that they do not use Dropbox for sensitive data, and one raised questions about copyright and ownership of data stored with the service. 12 participants did not express any concerns about the service, in fact, one stated he uses third-party cloud services exclusively because of privacy concerns with University-hosted storage solutions. Four of the six concerned Dropbox-users' research areas are in education and social sciences, where human subject data collection is common, but this concern may positively reflect an awareness amongst researchers of the risks involved with third-party cloud services. Conversely, researchers who do not manage sensitive data may not see any risk in using third-party cloud services.

Opportunities

Some participants pointed to commercial third-party cloud storage services, such as Google Drive or Dropbox, as models for University-managed storage services. One survey respondent suggested, "A

[5] http://its.uiowa.edu/support/article/100964

program that closely mimics Google Drive would be good, or just purchasing a site license for campus use of Google Drive would be fine." While many participants report using third-party cloud services, they often expressed a preference for comparable University-provided services if ever offered. One interviewee in education and social sciences reported using Dropbox but would much prefer to use a University service that would allow storing IRB-protected data. At the very least, some thought the University could offer guidance and education to researchers and students on using third-party services.

COLLABORATION

Providing collaborators with access to research data was also an issue for researchers. In addition to methods of storage, interviewees were asked about their methods of sharing access to data with collaborators and student assistants.

Challenges

For some researchers, the access controls provided for departmental servers suffice for collaboration with immediate members of their research team, but many researchers found challenges sharing data with colleagues in other departments and outside the University, and with student collaborators and assistants. After experiencing difficulties accessing collaborators' departmental network drives in the College of Medicine, one interviewee in the College of Engineering described having to create a separate silo for sharing research data.

Of the 13 participants who discussed working around these challenges, seven reported using Dropbox or other third-party cloud services to share data, two used the University-licensed Xythos, and one used University-licensed REDCap. Four participants shared data with collaborators over email, three provided collaborators with external media, and two transferred data via FTP. Three participants made use of University guest accounts to give collaborators direct access to data on their servers.

Opportunities

Though one researcher found the process of creating University guest accounts tedious, two others found it to be a good solution for sharing data with external collaborators.

Researchers' sentiments about using third-party cloud services for collaboration echoed comments and concerns raised in the previous section on synchronous storage. Many were highly satisfied with the functionality provided by Dropbox and Google Drive, but some had reservations about using these services with students. Researchers appreciated the flexible and self-service access controls provided by these services but expressed a desire for a University solution that would match this functionality.

DATA CURATION, BACKUPS, AND RETENTION

Interviewees were asked about needs and requirements related to data backups and retention. Discussion on data storage practices reflected a variety of practices and attitudes towards access to data both during the research process and after research has been completed.

Backing up data came up in discussion with 17 interviewees. 15 of those interviewed said they back up their data, while two specifically indicated they do not. 12 participants discussed data retention periods. Three interviewees store data for the minimum period required by policy or common disciplinary practice. Five expressed a desire to save data forever, while four said they save data indefinitely. One researcher in health and biosciences saves data in case questions arise after the research is complete or new analytical techniques become available.

Challenges

Storage limits were a primary challenge to backing up or retaining data. One researcher in the STEM sciences said they back up only data that would be difficult to reproduce. Another researcher

complained of too much data to back up, claiming they couldn't find a good solution from ITS and didn't have time to learn good data curation practices on their own. A lack of good recordkeeping and data organization practices also complicated data curation, as another researcher said it is hard to know what to keep or discard when people in the lab leave.

Three interviewees noted data obsolescence or degradation as obstacles to good data management. These researchers raised concerns about backward compatibility of software (ex. SAS) and implications of changing data formats on longitudinal studies. One researcher thought that NIH does not do enough to address long-term data issues, such as format obsolescence. Another researcher claimed to have had problems with data degradation but did not indicate if this happened on University storage systems.

Opportunities

Issues in backing up active data and retaining data long-term were closely related to challenges in data organization, storage limits, and synchronous access, so pursuing opportunities in those areas may also alleviate challenges in data curation. The researcher who reported difficulty curating their lab's data suggested a simple, email-like interface to file systems to help weed files more efficiently. Two interviewees suggested that more IT support with storage-related issues, such as guidance on using backup software, could help in implementing better data curation practices.

Data Organization and Analysis

Data organization includes all of the interactions involved in creating, using, analyzing, cleaning, storing and sharing data. Question 9 of the survey asked respondents to identify where they receive guidance for data organization tasks such as creating metadata, file naming, file versioning, and data storage. Interviewees were asked to describe their research process as it relates to collecting, generating, recording, analyzing, and storing data, noting who is involved in these stages, what software is used, and what challenges they have experienced. Challenges fell mostly into three broad areas: software-related issues, training and education, and support services.

SOFTWARE

10 interviewees discussed software-related issues, and 10 survey respondents submitted comments related to software training and costs.

Challenges

In the interviews, cost and University or departmental licensing was the most common problem for researchers—either software was too costly for researchers or their preferred software was not the option licensed by the University or their department. One participant also noted the difficulty in dedicating staff time to a data management system migration while minimizing data analysis services interruptions for clients. Several participants expressed that electronic lab notebooks would help data organization in their labs, citing the ability to provide a better audit trail and to improve accuracy, reproducibility, and verifiability.

Software-related comments submitted through the survey more frequently addressed a lack of software training. One respondent indicated a need for more humanities-focused software training.

Opportunities

Five interviewees and seven survey respondents expressed interest in software training. One interviewee praised the video tutorials available through the University-licensed Lynda.com[6], though

[6] http://lynda.uiowa.edu/

this sole mention of the service in both the survey responses and interviews may point to a need for better promotion across campus.

Two interviewees suggested a broader selection of University-licensed software. One of these participants would like to see a program to help get humanities faculty the software they need, saying that departments often won't fund it and it is often too expensive to purchase on their own. REDCap received positive reviews from four interviewees and two survey respondents. REDCap users cited its user-friendly interface and HIPAA compliance as the primary benefits of the system.

TRAINING

The need for training and education came up in the survey comments and interview discussions both directly, in suggestions for training opportunities on campus, and indirectly, in descriptions of data organization issues.

Challenges

Eight interviewees described challenges related to tracking and organizing data and files. Almost all of these accounts were attributed to student turnover and lack of knowledge of best practices. One researcher noted that the biggest challenge is getting students to use the data and software protocols in place for metadata creation and file naming. Another researcher in the arts and humanities relied on students to create and maintain a database but lacked the technical expertise to guide those students in their progress. While most solo researchers seemed comfortable with their data organization practices, faculty with research teams or assistants expressed more dissatisfaction, which may reflect the consequences of a common division of labor that puts the responsibility of data collection and organization on less-skilled students and assistants.

Opportunities

While a few participants in both the survey and the interviews expressed interest in data management training for themselves, most who discussed training expressed a need to educate students or others members of their research groups. Five interviewees reported offering or requiring training for students. Two of those offer formal coursework and two others work with students one-on-one at the start of their appointments. Another requires incoming students to take a programming class in computer science, which he believes will teach them the importance of file naming and version control. Three of these researchers also noted having a strong background in data organization either through previous careers in database development or mentoring and training early in their careers. Other suggestions from interviewees for data management training included lab orientations, courses for students, and interdisciplinary opportunities to learn from researchers in other fields.

SUPPORT STAFF AND SERVICES

Participant comments from both the survey and the interviews indicated the need for additional support services in data analysis, database design and maintenance, and project or software-specific assistance.

Challenges

Three interviewees and two survey respondents cited general issues in receiving assistance from IT beyond desktop support—high turnover of IT staff and lack of Mac support were two examples of factors that researchers perceive having an impact on effective data organization. Two survey respondents and three interviewees were interested in support for database design. Even though there are units on campus that provide this service, participants were either unaware of them or found the cost to be prohibitive.

Three interviewees cited challenges with data analysis or conversion. One of these researchers expressed the need for more assistance with the "last mile" of data analysis. Graduate student work is

often used to convert data to necessary standards but the researcher would prefer to pay for more reliable services. Related to this challenge, three additional interviewees noted a lack of available skillsets necessary to perform data-intensive tasks, particularly with student assistants.

Four interviewees stressed the need for more specialized support staff and services. One of these researchers in the arts and humanities found limited research support from the Digital Studio for Public Arts and Humanities (The Studio) and would like to see more specialized staff to assist with database design and other technical issues. A researcher in health and biosciences believes that other institutions have invested more in lab notebooks and data entry solutions and would like to see a dedicated center for clinical trials support services.

Opportunities

Some participants credited dedicated data management staff as major sources of assistance. The College of Nursing's Office for Nursing Research and Scholarship and College of Public Health's Clinical Trials Statistical and Data Management Center (CTSDMC)[7] were both named as campus models for data management support. Researchers who had access to a full-time data manager on staff, often through grand funding, reported satisfaction with their data management.

It should be noted that support service needs were often framed as a shortfall in IT support at various levels—departmental, college, central ITS, and HCIS—which may reflect where researchers expect to find these services. Efforts to improve support services may need to consider managing these expectations or implementing a consultation or clearinghouse service to help researchers better articulate technical needs and then direct them to the most appropriate services.

Data Publishing, Sharing, and Dissemination

Some funding agencies are now requiring researchers make available or "publish" their data after the research is complete, through a disciplinary or institutional repository, the researchers' website or, in some cases, on request. Many journals are also now asking authors to include supplementary raw data in associated repositories, such as Dryad[8] and Dataverse,[9] as part of the publication process. Participants were asked about requirements to publish or disseminate their data and how they publish and disseminate. Some participants discussed attitudes towards sharing post-research data with the public and other researchers as well as challenges they face in publishing or disseminating data.

Requirements to publish data was discussed with 28 interviewees.[10] 11 said they are required to share data publicly by funders or journals. Seven of these were from health and biosciences, three from STEM sciences, and one from education and social sciences. Of the 17 who are not required to publish data, 11 interviewees, from across all broad disciplinary categories, expressed interest in sharing data publicly.

Participants who did not show interest in publicly sharing data cite a number of reasons. One researcher in health and biosciences expressed concerns about being scooped. Another researcher in health and biosciences thinks that journals require too much supplementary data that they don't even review, which puts more of a burden on resources. One researcher reports sharing data through publication and conference presentations, but doesn't see data preparation as worth the effort. Data in this researcher's field, psychology, is often specific to a research question, and it is not seen as reusable.

[7] http://www.ctsdmc.org/
[8] http://datadryad.org/
[9] http://thedata.org/
[10] This question of requirements was not applicable to all interviewees—some were in service or support roles and do not publish--but all participants discussed practices and attitudes for other questions in this area.

Two interviewees said that public data sharing is common within their disciplines. One researcher said that journals in political science require authors to submit supporting data and are now turning to data repositories to host this data rather than manage it themselves. The researcher also noted that standards for data sharing and replication are starting to emerge within the discipline. Another researcher commented that open and limited-access data repositories within socio-linguistics are becoming increasingly common and hopes to one day digitize and make data openly accessible. Five participants said data sharing is not common practice in their fields. One said that in the field of chemistry, researchers will often just pull data off of charts from published research by hand, rather than contact the authors for raw data.

Researchers disseminate data in a variety of ways. Disciplinary repositories and journals or journal repositories were the most frequently mentioned. Five interviewees reported sharing data publicly on personal or lab websites. At least nine interviewees share data upon request. Three interviewees, all in health and biosciences, share data through restricted-access repositories. One researcher reported publishing data through a University repository, the Iowa Digital Library.[11]

DATA PREPARATION

Many researchers expressed concerns about how their data would be interpreted if presented outside of the context of a research publication. Data must often be cleaned, de-identified, contextualized, and reformatted in order for researchers to feel comfortable making it public.

Challenges

Of the 17 interviewees who discussed challenges in data dissemination and publishing, 13 found data preparation to be a bottleneck or deterrent to sharing their data. Six of these researchers pointed to the time and cost necessary to prepare data for public consumption. Two researchers noted the difficulty in finding the resources to prepare data for sharing once the grant is over and funding is gone. Another researcher thought that there should be more assistance at the University level to support the effort needed to prepare and standardize data. Three researchers noted variance in standards as a challenge in data preparation. One of these said that genomic research is still so new that standards are just starting to emerge. Another researcher in health and biosciences explained that standards differ between industry and academic use, and that preparing data in formats for both purposes can be time-consuming. Three more researchers expressed concern about methods and lack of guidance for de-identifying sensitive data. One worried that even given best efforts to anonymize data, subjects may still be identifiable.

Opportunities

Researchers who did not have difficulties with data preparation were usually depositing to an established, discipline-specific repository, like the National Center for Biotechnology Information (NCBI) GenBank, or managing their data within a data management platform, such as REDCap or EPIC. This may suggest that well-established standards, practices, tools, and repositories within a field may provide a supportive foundation for post-research data publishing.

REGULATIONS AND RESTRICTIONS

Researchers may encounter challenges sharing data because of restrictions related to regulations on human subjects, intellectual property issues, and security.

[11] http://digital.lib.uiowa.edu

Challenges

Six interviewees discussed how regulations and restrictions are a deterrent to disseminating and publishing data. HIPAA restrictions and current practices for gaining patient consent limit what data researchers can publicly share. One researcher identified this as the biggest challenge for medical imaging and research on campus right now. Another researcher in the health and biosciences showed interest in sharing data but said external sponsors may limit data sharing and reuse. Access to data from research funded by certain sponsors is often limited after the research is completed, which can make it difficult to field inquiries after the research is published.

Opportunities

Obtaining patient consent was named by six interviewees as an obstacle to sharing research data. Most saw great benefit in opening certain research data for reuse but were doubtful that the necessary regulatory changes would occur. One researcher suggested that the University's Office of the Vice President for Research could play a larger role in helping implement processes for gaining default or lifelong consent for all patients.

OUTLETS

Participants in both the survey and interviews identified a number of disciplinary, institutional, and journal repositories. A full list is available in Appendix 3. All but two of the 11 researchers who said they were required by funders or journals to publish their data used established disciplinary repositories associated with their federal funding agency or a journal repository associated with the journal.

Challenges

Researchers required to publish or share data did not find as much difficulty in locating an outlet for their data as researchers without the financial and repository support of a funding agency. Two interviewees in the arts and humanities discussed difficulties in identifying and implementing appropriate sharing mechanisms for digital humanities data online. Because digital humanities projects are often technologically one-off websites, they frequently require project-specific technical expertise and sustainable web hosting solutions.

It is important to note that the interviews were conducted shortly after the announcement of the White House Office of Science and Technology Policy (OSTP) mandate, which will require federal agencies to promote data deposition into public repositories.[12] Though researchers do not yet report immediate challenges in finding appropriate repositories for data publication, this may change in the coming years.

Opportunities

Five interviewees offered opinions and suggestions regarding University support for publishing and archiving data. All but one were either researchers or service providers who assist other campus researchers with data management and storage. One interviewee, who participated in early campus discussions on institutional repositories, would like to see a data repository on campus but expects that policy development around content, data migration, and access controls would be complex. Another interviewee suggested a more holistic repository that would encompass the full research lifecycle—documentation, data, and publication. Another suggested the University or the Libraries could help connect data with those who can reuse it, help write data dictionaries, and prepare the data to be more accessible.

[12] http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

Sensitive Data and Compliance

In many disciplines research data may include personally identifiable, culturally sensitive, confidential, or classified information that may require special treatment, such as de-identification or compliance with regulations.

382 of 562 survey respondents said they work with sensitive or confidential data (Figure 6). 30 interviewees have worked with or plan to work with sensitive or confidential data. Most of the interviewees reported working with personally identifiable patient or human subject data. One researcher works with classified data, and another records potentially culturally sensitive data.

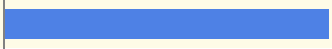
Answer		Response
Yes		382
No		170
I don't know		10
Total		562

Figure 6. Responses to Survey Question 4, "Do you work with sensitive or confidential data?"

Challenges

Issues related to sensitive data have been addressed throughout this report, but the unique challenge in managing this data warrants its own discussion.

Most interviewees' concerns were related to managing personally identifiable data. As noted in the section on Data Publishing, Sharing, and Dissemination, some participants who were interested in making de-identified research data open and available for reuse felt that regulations on patient consent greatly limited their ability to share data. On the other hand, three interviewees expressed concern about methods and lack of guidance for de-identifying sensitive data. One worried that even given best efforts to anonymize data, subjects may still be identifiable.

Compliance with IRB and other regulatory bodies was an obstacle for some researchers, particularly in data storage and providing access for collaborators. Two researchers commented on the limitations that IRB compliance put on their ability to leverage the functionality of third-party cloud services, such as SurveyMonkey and Dropbox. One suggested that IRB provide more guidance on how researchers can safely use these tools, while the other suggested the University provide IRB-compliant solutions that mimic Dropbox's seamless synchronous access. A researcher in the STEM sciences noted compliance with national security regulations, such as International Traffic in Arms Regulations (ITAR) and Federal Information Security Management Act (FISMA), as a challenge to storing data.

Opportunities

In addition to guidance on IRB-compliant software and tools, interviewees had other suggestions for improvements to IRB services. One participant would like to see IRB expand services to include more instruction on best practices and recommendations for other data management services on campus. Another would like to see the online application form enhanced with embedded guidance and links to relevant documentation, examples, and templates.

As noted earlier in the report, REDCap received praise from almost all participants who reported using the service, citing its user-friendly interface, fine-grained access controls, and regulatory compliance as major features.

Additional Findings

Research culture and funding models were not addressed directly in the interview questions, but enough concerns were voiced around these themes to warrant additional discussion in the report findings. Future service development may benefit from further investigations into these areas.

Research Culture

Culture plays a significant role in many aspects of data management as well as perspectives on existing and future services. The University of Iowa is a collection of subcultures, and during the interview process there were often conflicting opinions expressed. For example, four interviewees expressed concerns related to trust, reliability, and security with University storage services. Two of these said they did not trust H or shared drives as reliable storage, and one Dropbox user did not feel comfortable keeping data on University systems as it is subject to search at any time. On the other hand, several individuals expressed concern about storing their data on free cloud services due to security, compliance, or intellectual property concerns.

Two interviewees shared concerns about University administration's influence on research culture. One was looking for more infrastructure transparency and the other one felt that administration doesn't really understand the technical and data needs of humanities faculty, which they indicate could be a reflection of the lack of humanities representation on Research Council. This participant also felt that IRB could do a better job of understanding humanists' research needs.

Three other interviewees, all from Arts & Humanities, found frustration with the lack of collaboration on campus. One felt the tech culture on campus does not respect the perspectives of those with less experience, noting that tech conversations are usually one-way and create barriers to progress. This participant believes that researchers across campus have different views of goals and scope and have much to teach each other. Another noted a lack of support for interdisciplinary research, that "interdisciplinary" seems to be a buzzword, honored in name only. This researcher would like to see more opportunities for researchers to learn about each other's fields and points to the Obermann Center and The Studio as campus organizations that are doing a good job of this.

It was also clear that the research culture within participants' disciplines heavily influenced their perspectives on data management. In areas such as High Energy Physics, data format standards are well established while in other areas, such as Genomics, the formats are rapidly evolving, which in turn affects researchers' abilities and willingness to make data publicly available. Most repositories discussed were also disciplinary indicating that many fields are building national or international locations for their data interchange needs in preference over local solutions. One example of a successful repository that was cited by multiple participants was the Interuniversity Consortium for Political and Social Research (ICPSR). This however does not include all disciplines and becomes substantially more complicated in areas where research involves sensitive or confidential information.

One particularly striking culture change appears to be happening in the humanities. As digital scholarship increases in popularity in this area it is changing not only the types of projects and presentation format but also the how research is done. In traditional humanities scholarship, sole authorship has been expected and rewarded, and humanities research has been primarily an individual endeavor using widely accessible, low cost tools. Digital humanities projects typically require a team of collaborators with a breadth of skills related to information management and information technology To bridge this gap and successfully implement and maintain a digital humanities project, researchers may need to make a cultural shift and adopt a team approach.

Funding

A number of interviewees indicated that funding was a challenge in one form or another. Some of these participants indicated that some services, such as data storage, should be included as part of facilities and administrative costs. One participant from Health and Biosciences spoke about the difficulties in funding infrastructure on large program grants and that areas such as connectivity costs should be covered by the institution. Another individual, however, indicated that more indirect costs should be passed back to the researchers so that they could purchase the services that best meet their needs. Another participant stated that they had problems finding funding for the software they need. Multiple others expressed a desire for increased campus software site license coverage.

Interviewees indicated that further clarification from funding agencies is needed on data management and sharing expectations and that the agencies need to be willing to fund these initiatives. Two participants thought NSF and NIH policies were not clear on who should be paying for data retention after the end of a project, and another indicated that it was important to catch researchers before a project ended so that funding would be available to pay for long term data retention.

Although funding was cited as a challenge across disciplines, there are acute challenges to digital scholarship in the arts and humanities. At the core of this is a funding model and support services designed for traditional research that do not accommodate new requirements associated with digital scholarship in these areas. One interviewee talked in depth about the challenges of funding humanities projects given the one-off nature of the projects. The participant talked about the need for personnel who could assist with prototyping, project management, application development, maintenance, and technology consultation. Another participant indicated that they had received funding from The Studio to support graduate students to work on their project. Another participant indicated that they had received internal funding to hire research assistants but they are concerned about ongoing funding. As noted earlier in this report, however, some researchers found that student assistants alone could not help them meet their project and technology needs.

While lack of grant funding for support services was cited as a challenge by participants across all disciplines, the absence of an established and effective research support structure in some disciplines may prevent research success, even when funding is available. Further exploration into existing domain-specific support services on campus, such as the College of Nursing's Office for Nursing Research and Scholarship, could provide insight into how to build out models for support in disciplines where none exist.

Funding Note

This project has been funded in whole or in part with Federal funds from the National Library of Medicine, National Institutes of Health, Department of Health and Human Services, under Contract No. HHS-N-276-2011-00005-C with the University of Illinois at Chicago.

Appendix 1: Survey Instrument and Results

Email Message Sent to Faculty/Staff Listserv:

You are invited to participate in a brief survey being conducted by the University of Iowa Libraries.

As research becomes more data-intensive and collaborative, The University of Iowa is seeing a growing demand for skills and resources to deal with capturing, organizing, storing, publishing, archiving, and sharing data (i.e. data management). Funding agencies (including NSF, NIH, and NEH) are increasingly requiring that researchers include data management plans in new grant proposals, and many now require that research data be archived and shared in digital repositories.

We are interested in hearing from faculty and staff researchers from across campus including the scientific, social sciences, and humanities communities to learn what kind of data is being generated and how researchers are finding help in managing it. Your participation will help us assess data management needs at the University of Iowa and will inform the coordination of services across campus to support effective data stewardship throughout the research life cycle.

The survey should take approximately 10 minutes to complete. Click on the following link to access the survey:

If you have any questions or comments, please contact Shawn Averkamp, Data Services Librarian, at shawn-averkamp@uiowa.edu or Xiaomei Gu, Clinical Education Librarian, at xiaomei-gu@uiowa.edu.

If you have questions about the rights of research subjects, please contact the Human Subjects Office, 105 Hardin Library for the Health Sciences, 600 Newton Rd, The University of Iowa, Iowa City, IA 52242-1098, (319) 335-6564, or e-mail irb@uiowa.edu.

/**********
Distribution of this message was approved by the Provost and the VP for Human Resources. Neither your name nor e-mail address was released to the sender. The policy and guidelines for the UI Mass Mail service, including information on how to filter messages, are available at: http://its.uiowa.edu/apps2/support/massmail.

**********/

Survey Introduction:

We invite you to participate in a research study being conducted by investigators from The University of Iowa Libraries. The information collected in this study will be used to assess data management needs at The University of Iowa and will inform the coordination of services across campus to support effective data stewardship throughout the research life cycle.

If you agree to participate, we would like you complete the 12-question survey about your primary research area and your collection of study data. You will also be asked to provide your primary department or unit. You are free to skip any questions that you prefer not to answer. It will take approximately ten minutes to complete the survey. If you would like to elaborate on any of your responses, please feel free to use the comment form at the end of the survey.

If you collect, generate, or use data in your research, you will be asked to provide your email address if you agree to be contacted about a follow-up interview procedure. You may participate in this part of the study without providing your email address.

You will be asked to provide information over the Internet. It is possible that your responses could be viewed by persons who have access to the computers hosting the web site or by unauthorized persons who gain access to the web site computers. We will use a secure web site and computers to collect and store the study information. We will separate your email address from your survey responses and assign your survey responses a study code number. The study code number will be linked to your email address only if you agree to participate in an in-depth interview regarding data management. The list linking your email address and your study identification code will be stored in a separate location that is accessible only to the researchers.

Taking part in this research study is completely voluntary. If you do not wish to participate in this study, close your web browser window without submitting your survey responses.

If you have any questions or comments, please contact Shawn Averkamp, Data Services Librarian, at shawn-averkamp@uiowa.edu or Xiaomei Gu, Clinical Education Librarian, at xiaomei-gu@uiowa.edu.

If you have questions about the rights of research subjects, please contact the Human Subjects Office, 105 Hardin Library for the Health Sciences, 600 Newton Rd, The University of Iowa, Iowa City, IA 52242-1098, (319) 335-6564, or e-mail irb@uiowa.edu.

Thank you very much for your consideration of this research study.

If you agree to participate in this study, click the NEXT button to go to the first survey question. If you wish to keep a copy of this information page, please save or print the page before going to the next page.

Question 1: Which one of the following best describes your primary research area?

- ❍ Interdisciplinary. Please specify: _____
- ❍ Arts and Humanities
- ❍ Business
- ❍ Computer and information science
- ❍ Education
- ❍ Engineering
- ❍ Health sciences
- ❍ Life sciences
- ❍ Law
- ❍ Physical sciences
- ❍ Psychology
- ❍ Social sciences
- ❍ Other. Please specify: _____
- ❍ Not applicable. I do not conduct any research.

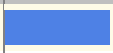
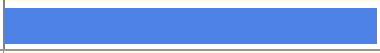
Answer		Response
Law		6
Business		12
Computer and information science		13
Education		14
Engineering		19
Psychology		19
Other. Please specify:		32
Physical sciences		32
Life sciences		42
Interdisciplinary. Please specify:		50
Social sciences		51
Arts and Humanities		75
Not applicable. I do not conduct any research.		174
Health sciences		241
Total		780

Question 2: "Data" is any recorded material necessary to validate your research findings. This is often numeric, tabular data, but it can also be textual data, images, audio/video files, or even artifacts. Do you collect, generate, or use data in your research?

- ❍ Yes
- ❍ No

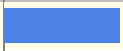
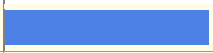
If No is selected, then skip to “What is your department or unit?”

Total Responses	735

Answer		Response	%
No		162	22%
Yes		573	78%
Total		735	100%

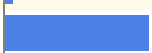
Question 3: Which one of the following best describes your position?

- ❍ Faculty researcher
- ❍ P & S researcher
- ❍ Other. Please specify: _____

Answer		Response
Other. Please specify:		136
P & S researcher		184
Faculty researcher		241
Total		561

Question 4: Do you work with sensitive or confidential data?

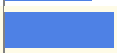
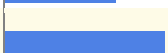
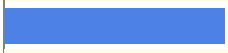
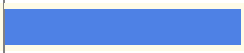
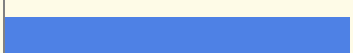
- ❍ Yes
- ❍ No
- ❍ I don't know

Answer		Response
I don't know		10
No		170
Yes		382
Total		562

Question 5: What kinds of data do you collect, generate, or use in the course of your research? Check all that apply.

- ❑ Artifacts, samples, and/or specimens
- ❑ Tabular or relational data (e.g. spreadsheet, database)
- ❑ Textual data (e.g. text files, Word documents, PDF)
- ❑ Geospatial data
- ❑ Images
- ❑ Audio
- ❑ Video
- ❑ Computer programs or code
- ❑ Genomic data
- ❑ Survey / questionnaire data
- ❑ Other. please specify: _____

Total Responses	558

Answer		Response
Other. please specify:		42
Geospatial data		76
Genomic data		96
Audio		102
Computer programs or code		128
Video		130
Artifacts, samples, and/or specimens		188
Survey / questionnaire data		258
Images		276
Textual data (e.g. text files, Word documents, PDF)		391
Tabular or relational data (e.g. spreadsheet, database)		404

Question 6: Where do you receive funding for your research? Check all that apply.

- ❑ Department of Defense (DoD)
- ❑ Department of Education (ED)
- ❑ Department of Energy (DOE)
- ❑ National Institutes of Health (NIH)
- ❑ National Science Foundation (NSF)
- ❑ National Endowment for the Humanities (NEH)
- ❑ National Aeronautics and Space Administration (NASA)
- ❑ Privately funded
- ❑ Not applicable. I don't receive funding for my research
- ❑ Other. Please specify: _____

Total Responses	547

Answer		Response
Department of Energy (DOE)		11
National Aeronautics and Space Administration (NASA)		12
Department of Education (ED)		13
National Endowment for the Humanities (NEH)		16
Department of Defense (DoD)		25
National Science Foundation (NSF)		82
Other. Please specify:		122
Privately funded		147
Not applicable. I don't receive funding for my research		160
National Institutes of Health (NIH)		229

Question 7: Some funding agencies require a data management plan as part of the grant proposal. Are you required to write data management plans?

- ❍ Yes
- ❍ No
- ❍ I don't know

If No is selected, then skip to "Where do you find guidance for data ..."

Answer		Response
I don't know		99
Yes		186
No		266
Total		551

Question 8: Where do you find assistance for writing data management plans? Check all that apply.

- ❑ My department or college
- ❑ UIowa research center or institute. Please specify: _____
- ❑ Division of Sponsored Programs
- ❑ University Libraries
- ❑ Information Technology Services (ITS)
- ❑ Institute for Clinical and Translational Science (ICTS)
- ❑ I don’t receive assistance, but I would like help.
- ❑ I don’t need assistance writing data management plans
- ❑ Other. Please specify: _____

Total Responses	273

Answer		Response
University Libraries		7
UIowa research center or institute. Please specify:		17
Information Technology Services (ITS)		21
Institute for Clinical and Translational Science (ICTS)		23
Division of Sponsored Programs		26
Other. Please specify:		36
I don’t need assistance writing data management plans		53
I don’t receive assistance, but I would like help.		76
My department or college		109

Question 9: Where do you find guidance for data organization? (e.g., creating metadata, file naming, file versioning, data storage) Check all that apply.

- ❑ My department or college
- ❑ UIowa research center or institute. Please specify: _____
- ❑ Division of Sponsored Programs
- ❑ University Libraries
- ❑ Information Technology Services (ITS)
- ❑ Institute for Clinical and Translational Science (ICTS)
- ❑ I don't receive assistance, but I would like help.
- ❑ I don't need guidance in organizing my data.
- ❑ Other. Please specify: _____

Total Responses	535

Answer		Response
Division of Sponsored Programs		13
UIowa research center or institute. Please specify:		23
Institute for Clinical and Translational Science (ICTS)		28
University Libraries		34
Information Technology Services (ITS)		54
Other. Please specify:		57
I don't receive assistance, but I would like help.		141
I don't need guidance in organizing my data.		155
My department or college		192

Question 10: Where do you store and/or share your data? Check all that apply.

- ❑ On my computer
- ❑ On a shared drive or server on campus
- ❑ On external media (e.g., external hard drive, memory stick, CD)
- ❑ With a commercial online storage service (e.g. Dropbox, Google Docs)
- ❑ In a disciplinary repository (e.g. Dryad, ICPSR) or institutional repository (e.g. Iowa Research Online). Please specify: _____
- ❑ On paper (e.g. lab notebooks, printouts)
- ❑ I don't know
- ❑ Other. Please specify: _____

Total Responses	541

Answer		Response
I don't know		6
In a disciplinary repository (e.g. Dryad, ICPSR) or institutional repository (e.g. Iowa Research Online). Please specify:		34
Other. Please specify:		37
With a commercial online storage service (e.g. Dropbox, Google Docs)		94
On paper (e.g. lab notebooks, printouts)		257
On external media (e.g., external hard drive, memory stick, CD)		280
On my computer		375
On a shared drive or server on campus		388

Question 11: Are you willing to be contacted to answer further questions about your data management practices and needs?

- ❍ Yes. Please enter your email address. _____
- ❍ No

Total Responses	523

Question 12: What is your department or unit? List your primary affiliation if you have a dual appointment. (Optional)

Total Responses	477

Question 13: Please add any additional comments or concerns related to data management needs and services on campus. (Optional)

Total Responses	110

Appendix 2: Interview Questions

1. Data/Research overview
    - Tell us about your research. Who is involved? What are their roles? Where do you receive funding?
    - What kind of data do you collect, generate, or use?
        i. Size?
        ii. Formats?
        iii. Associated documentation (e.g., keys, metadata, notes, methodologies)
    - Do you work with confidential data? What is the nature of this data?
2. Data Management Plans
    - Are you required to write data management plans to receive funding?
    - If no, do you do any data management planning prior to your research?
    - Have you written a data management plan? (If no, has someone in your research group written a data management plan?)
        i. Where did you find assistance?
        ii. What were the challenges (if any) in writing it?
3. Storage
    - What kind of storage needs do you have?
        i. Temporary, during research
        ii. Long-term, post-research
        iii. How long do you anticipate storing data for your research?
    - Where are you currently storing data?
    - Are you backing up your data? How and where?
    - Who is helping you find/store data? Where do you look for information about data storage?
    - What challenges have you faced in storing your data?
    - If you collaborate with researchers outside your department or outside the University, have you experienced problems with access to research data?
4. Dissemination/Sharing
    - Are you required to publish or disseminate your data publicly?
    - Do you share (or are you interested in sharing) your data publicly? Privately?
    - How do you disseminate your data? (e.g., on my website, in a disciplinary repository, by email on request)
    - Do you apply any formal licenses to your data? What licenses do you use? Did anyone assist you with this?
    - What challenges have you faced in trying to disseminate or share your data?

- If you collaborate with researchers outside your department or outside the University, have you experienced problems in disseminating your data?

5. Data organization
    - Tell us, step by step, your process of collecting/generating data, recording data, analyzing data, storage.
        i. Who is involved in each stage?
        ii. What software/tools are used?
        iii. What challenges have you experienced in each stage?
    - What kind of assistance (training, software/tools, etc.) would help simplify these tasks? Where would you prefer to find this assistance?
6. Other concerns
    - What would you like to see the University of Iowa do to better support data management in your research?
    - Do you have any other issues related to data management that you would like to talk about?

Appendix 3: List of Selected Repositories

This appendix lists URLs of disciplinary and institutional repositories mentioned in the survey results and interview discussions.

- AphasiaBank http://talkbank.org/AphasiaBank/ - (note: mentioned but not yet in active use)
- Coordinated Data Analysis Web (CDAWeb) http://cdaweb.gsfc.nasa.gov/
- Child Language Data Exchange System (CHILDES) http://childes.psy.cmu.edu/
- ClinicalTrials.gov Clinicaltrials.gov
- The Dataverse Network http://thedata.org
- Dryad http://datadryad.org/
- EarthChem http://www.earthchem.org/
- Ensemble http://www.ensembl.org
- Electronic Laboratory Exchange Network (FDA eLEXNET) https://www.elexnet.com/elex/
- European Space Agency Cluster Active Archive (ESA CAA) http://caa.estec.esa.int/caa/home.xml
- The Interuniversity Consortium for Political and Social Research (ICPSR) www.icpsr.umich.edu/
- Iowa Digital Library (IDL) http://digital.lib.uiowa.edu
- Iowa Research Online http://ir.uiowa.edu/
- Morphbank http://www.morphbank.net/
- NASA CDAWeb http://cdaweb.gsfc.nasa.gov/
- NASA National Space Science Data Center (NSSDC) http://science.nasa.gov/planetary-science/planetary-science-data/national-space-science-data-center-nssdc/
- NGDC NOAA SPIDR (Space Physics Interactive Data Resource) http://spidr.ngdc.noaa.gov/spidr/
- National Cancer Institute (NCI) repositories
    - Biological Specimens http://www.cancer.gov/clinicaltrials/international/answers/biologicalspecimens
    - Cancer Genomics Hub https://cghub.ucsc.edu/
- National Center for Atmospheric Research (NCAR) http://ncar.ucar.edu/
- National Center for Biotechnology Information (NCBI)
    - GenBank http://www.ncbi.nlm.nih.gov/genbank/
    - Gene Expression Omnibus (GEO) http://www.ncbi.nlm.nih.gov/geo/
    - Genotypes and Phenotypes (dbGaP) http://www.ncbi.nlm.nih.gov/gap/
- PubChem http://pubchem.ncbi.nlm.nih.gov/
- OMNIWeb http://omniweb.gsfc.nasa.gov/ow.html
- The Research Data Archive http://rda.ucar.edu/
- Satellite Situation Center Web (SSCWeb) http://sscweb.gsfc.nasa.gov/
- Society for Assisted Reproductive Technology (SART)
    - Society website: http://www.sart.org/
    - Restricted access data repository: https://www.sartcorsonline.com/
- The Smithsonian http://www.si.edu/