



Full-Length Paper

Assessing Data Management Support Needs of Bioengineering and Biomedical Research Faculty

Christie A. Wiley and Margaret H. Burnette

University of Illinois at Urbana-Champaign, Champaign, IL, USA

Abstract

Objectives: This study explores data management knowledge, attitudes, and practices of bioengineering and biomedical researchers in the context of National Institutes of Health (NIH) funded research projects. Specifically, this study seeks to answer the following questions:

1. What is the nature of biomedical and bioengineering research on the Illinois campus and what kinds of data are being generated?
2. To what degree are biomedical and bioengineering researchers aware of best practices for data management and what are the actual data management behaviors?
3. What aspects of data management present the greatest challenges and frustrations?
4. To what degree are biomedical and bioengineering researchers aware of data sharing opportunities and data repositories, and what are their attitudes towards data sharing?
5. To what degree are researchers aware of campus services and support for data management planning, data sharing, and data deposit, and what is the level of interest in instruction in these areas?

Correspondence: Margaret H. Burnette: phburn@illinois.edu

Keywords: data management, research data, librarians

Rights and Permissions: Copyright Wiley and Burnette © 2019

Disclosures: The authors report no conflict of interest.



All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Abstract Continued

Methods: Librarians at the University of Illinois at Urbana-Champaign campus conducted semi-structured interviews with bioengineering and biomedical researchers to explore researchers' knowledge of data management best practices, awareness of library campus services, data management behavior, and challenges managing research data. The topics covered during the interviews were current research projects, data types, format, description, campus repository usage, data-sharing, awareness of library campus services, data reuse, and anticipated impact of health on public and challenges (interview questions are provided in the Appendix).

Results: This study revealed the majority of researchers explore broad research topics, various file storage solutions, generate numerous amounts of data, and adhere to differing discipline-specific practices. Researchers expressed both familiarity and unfamiliarity with DMPTool. Roughly half of the researchers interviewed reported having documented protocols for file names, file backup and file storage. Findings also suggest that there is ambiguity about what it means to share research data, and confusion about terminology such as "repository" and "data deposit". Many researchers equate publication to data sharing.

Conclusions: The interviews reveal significant data literacy gaps that present opportunities for library instruction in the areas of file organization, project workflow and documentation, metadata standards, and data deposit options. The interviews also provide invaluable insight into biomedical and bioengineering research in general and contribute to the authors' understanding of the challenges facing the researchers we strive to support.

Introduction

Science, technology, engineering and medical (STEM) research produces large volumes of data creating increased requirements for active data management, curation, and stewardship. Some benefits of managing research data are that it increases opportunities for accessibility, discoverability, preservation, transparency, and reproducibility. As early as 2011, funding agency mandates and the open access movement created the initial need for comprehensive support. Researchers are encouraged or required to share research data along with their research outputs (Chen & Wu 2017).

Academic libraries have taken this opportunity to re-engineer their services to support and accommodate faculty and institutional research data management requirements and responsibilities (Lyon 2012; Webster 2012). Ogier et al. (2018) states research data management (RDM) services in the US and UK have grown. Numerous academic libraries have and are assessing researchers' perspectives (Carlson 2012; Diekema et al. 2014; Perrier et al. 2017; Peters & Dryden 2011). These studies reveal different attitudes, challenges and expectancies among researchers for managing and sharing data.

In a previous study, the first author interviewed engineering and science faculty about their current research projects and data management practices. Respondents indicated their current research varied from discipline specific to interdisciplinary in the area of health research. This led to the present project that explores bioengineering and biomedical faculty data management attitudes, practices, and perspectives in the context of National Institutes of Health (NIH) funded projects. Specifically, this study seeks to answer the following questions:

1. What is the nature of biomedical and bioengineering research on the Illinois campus and what kinds of data are being generated?
2. To what degree are biomedical and bioengineering researchers aware of best practices for data management and what are the actual data management behaviors?
3. What aspects of data management present the greatest challenges and frustrations?
4. To what degree are biomedical and bioengineering researchers aware of data sharing opportunities and data repositories, and what are their attitudes towards data sharing?
5. To what degree are researchers aware of campus services and support for data management planning, data sharing, and data deposit, and what is the level of interest in instruction in these areas?

Answers to these questions can help librarians, subject specialists, and research data services staff gain insights into common challenges and frustrations related to data management overall. Librarians need to adopt a more holistic approach to research data management support services that closely align with the research and data lifecycles and provide researchers with knowledge and resources starting at the beginning of a research project. A better understanding of issues and challenges will help build stronger relationships, enhance services, and reveal areas needing improvement.

Literature Review

The body of literature related to library support for research data management has increased significantly in recent years and reveals a wide array of opportunities and challenges in the areas of programming, instruction, and infrastructure. For librarians, supporting research data often begins with an exploration of local researchers' aptitude, attitudes, and practices to identify gaps and assess needs.

In their comprehensive scoping review of research data management practices in academia, Perrier et al. (2017) characterize the data management literature as both qualitative and quantitative, and skewed towards data access, including sharing and preservation. They note a gap in research that looks at data management interventions occurring early in the research cycle. In her "call to action" directed at data librarians, Martin (2016) proposes a data management framework with the user situated at the center and surrounded by "buckets" that represent areas of potential action for librarians. These buckets, which include data management practices, data archiving and preservation, data literacy, and data services, provide the framework for this literature review.

Data Management Practices

Several key studies seek to characterize researcher data management knowledge, attitudes, and practices (Federer et al. 2015; Federer et al. 2016; Hickson et al. 2016; McLure et al. 2014; Perrier et al. 2017; Tenopir et al. 2015; Whitmire et al. 2015). Common themes reveal multiple challenges for researchers and tremendous variation in size, scope, processes, and formats of data objects across disciplines.

Researchers generate many different types of data, even within a single study. Data are generated in both tangible and digital formats, including lab notebooks, spreadsheets, text, images, geospatial, and other software-specific proprietary files (Carlson 2012). The volume of research data is expanding seemingly exponentially, and file sizes vary widely from kilobytes to petabytes. Genomic and imaging data consist of huge files that are difficult to transfer for sharing and deposit (Kim 2013).

Reznik-Zellen et al. (2015) report that file storage is a significant challenge for both faculty and graduate student researchers. Many researchers rely on personal computers and external hard drives to store data. Others use departmental servers or cloud storage. Dropbox and USB drives are popular because they are convenient and easy to use (Hickson et al. 2016; Whitemire, Boock & Sutton 2015). For backup, researchers report using external hard drives, while others assume that backup happens automatically (Hickson et al. 2016).

Data Archives and Preservation

Increasingly, funder mandates require grantees to deposit research data into a public repository. More and more publishers are adopting policies that require deposit of publication data in open access repositories (Kim 2013; MacMillan 2014).

Beyond these mandates, the propensity to share data or not is often embedded in the disciplinary culture (Kim 2013; MacMillan 2014). Federer et al. (2015) found that basic and

clinical sciences researchers willingly share data with peers and colleagues but do not upload data for widespread dissemination. MacMillan (2014) suggests that “small science” research data can be particularly problematic because it is generated in formats not easily shareable. Some “small science” researchers don’t see how their data would be of interest to anyone else (Hickson et al. 2016; MacMillan 2014; Tenopir et al. 2015). Keil (2014), himself a biomedical researcher, admits that many scientists feel they already share data within the publication, and sharing the raw data is never even a consideration. Clinical research presents enormous privacy and security obstacles for data sharing. Participant consent often covers the duration of the study but not after (Federer et al. 2015).

Several studies focus on data sharing specifically (Fecher, Friesike & Hebing 2015; Federer et al. 2015; Kim 2013; MacMillan 2014; Sayogo & Pardo 2013). Motivations for sharing data include collaboration, advancing research in a specific area, open science, peer review, supporting colleagues, and potential benefit for promotion and tenure (Federer 2015; Keil 2014; Kim 2013; MacMillan 2014; McClure et al. 2014; Poole 2015; Tenopir et al. 2015).

Scientists can be stingy with data for a variety of reasons (Keil 2014; Poole 2015; Sayogo and Pardo 2013; Tenopir et al. 2015). Logistical barriers include lack of infrastructure and support, the time and effort needed to prepare data for sharing, and lack of knowledge about where and how to deposit, and the related licensing and user agreements and metadata requirements (MacMillan 2014; Poole 2015; Tenopir et al. 2015).

With the increasing emphasis on open science, re-use and reproducibility can serve to motivate or impede. Issues related to research integrity and reuse include data deposit, discovery, permissions, and documentation (Perrier et al. 2017; Poole 2015). Data deposit may open researchers to unwelcome scrutiny or fault finding (Sayogo & Pardo 2013). Once data is released, the researcher has little control over what happens to that it leading to fears that others may steal ideas or data, won’t use the data responsibly, or will be over critical of research methods and findings (Fecher, Friesike & Hebing 2015; Federer et al. 2015; Hickson et al. 2016; MacMillan 2014; Sayogo & Pardo 2013; Tenopir et al. 2015). Vogeli et al. (2006) found that scientists were less willing to share data with those who have industry relationships for fear the data might be used for commercial purposes. Others may not give up their data if there is potential for future monetization or for patents (MacMillan 2014). Recognition, acknowledgement, and attribution of data ownership are also critically important for sharing (Sayogo & Pardo 2013; Tenopir et al. 2015).

Data Literacy

Data literacy among researchers remains relatively low (Federer et al. 2016). In 2014 librarians at the NIH surveyed internal research personnel about data management skills to assess training needs and to better understand data “attitudes, experience, and knowledge” (Federer et al. 2015, 3). They found that most researchers lack baseline data literacy skills, even though they felt these skills were relevant. Many researchers are only vaguely familiar with the notion of proactive planning for managing research data or the concept of a Data Management Plan (DMP). Researchers who have done a DMP often feel they lack sufficient knowledge of things like metadata standards and other issues related to data management (Hickson et al. 2016; McClure et al. 2014). Federer, Lu & Joubert (2016) found that NIH researchers rated their own DMP expertise lower than any other data literacy skill.

Regarding data sharing specifically, researchers have the will but lack the skills to do so (Federer et al. 2015; Kim 2013; Sayogo & Pardo 2013). Considerations for long-term access are not normally part of the researcher's skill set (Federer et al. 2015; Kim 2013; MacMillan 2014). Metadata in particular is problematic. McLure et al. (2014) found that researchers were not typically aware of metadata standards for their discipline and MacMillan reports that "members of the scientific community remain unaware of the value of metadata" (2014, 543).

Data Services

Research data services (RDS) are at the forefront of many academic libraries. RDS can encompass data literacy education, data management support services, and/or data infrastructure. Services can potentially touch all points of the data lifecycle, including data management planning, locating and using existing data, data analytic tools and methods, and data dissemination and deposit (Surkis & Read 2015).

For Martin (2015), data literacy encompasses instruction, workshops, online guides, and tools that promote best practices for data management. Federer (2016) suggests aligning data literacy with the research and data life cycles to provide a logical guide for program planning efforts. Programming often includes workshops and tutorials, and informal outreach during consultations or other instructional sessions. One-on-one consultations tailored to discipline-specific or literacy-specific levels may be more effective than generic workshops (Federer et al. 2016; Johnson & Knuth 2016).

Reference librarians and subject specialists are well positioned to participate in research data support programming—they have intimate disciplinary knowledge and understand the research realm of the discipline, including core publications, resources, topics and terminology, research methods and materials, and research outputs (Federer 2016). From the perspective of one biomedical researcher, "who is better suited with skills for organizing, linking and preserving data than librarians?" (Keil 2014, 234).

Whitmire, Boock and Sutton (2015) suggest that librarians can help researchers keep up with data sharing and preservation requirements and help to identify local or remote options for sharing and archiving. MacMillan suggest librarians "develop and provide lists of data repositories for faculty" and "make sure institutional repositories are findable" and help users "work through data-deposit processes" (2014, 548).

Kim calls upon librarians to "assume a more proactive role in support of digital scholarship" (2013, 503). In order to do this, librarians need to better understand researchers' data literacy levels, and what they need in the context of specific disciplines.

Efficient and effective research data management is fundamental to quality research. Best practices address issues of: file management, including file structure, versioning, security, storage and backup; data documentation and metadata; and data sharing and preservation. Data sharing has been on the radar of researchers and librarians for decades. This study seeks to further that understanding in the context of biomedical and bioengineering research.

Methods

The investigators conducted in-person, semi-structured interviews using a set of 18 questions that were adapted from a questionnaire used by Wiley and Mischo (2016) that explored data management practices of atmospheric and engineering faculty. The authors revised the questions for this study to account for bioengineering and biomedicine focus which have different data sharing demands.

A search in the NIH Report website (<https://report.nih.gov>) for University of Illinois Urbana-Champaign, Champaign, Illinois active projects from 2013 – 2016 yielded 143 grant projects. The authors exported project data to a spreadsheet, sorted by PI and matched names to departments on the UIUC campus. Seventy-six of the 98 PIs on the list were affiliated with either biomedical or bioengineering projects. The authors divided the cohort based on primary affiliations and assigned them to the bioengineering group (those with any affiliation with bioengineering) or biomedical (other departments such as molecular and cellular biology). The authors sent personalized emails requesting interviews to 36 bioengineering and 40 biomedical faculty. A total of 16 researchers agreed to be interviewed. The low response may have been due in part to the fact that the study was conducted after the end of the spring semester and some faculty leave town as soon as classes are finished.

The authors each interviewed the group associated with their liaison area. Topics covered during interviews included descriptions of research projects, research collaborations, scope of data collected, data management plans, data documentation, data storage, data sharing activity and intentions, and knowledge of library research support services. The authors interviewed participants for an average of 45 minutes in the researcher's campus office or lab and each investigator transcribed interviews verbatim into a Word document.

The authors manually coded their respective interview transcripts and then identified themes related to a pre-established framework of overarching focus areas:

- Data management attitudes and practices;
- Major challenges related to data management;
- Characteristics of research data;
- Attitudes and practices related to data sharing;
- Level of awareness of campus support for research support;
- Level of interest in instruction on data management strategies and best practices.

Initial thematic codes were recorded for each question along with interesting or significant quotes were also collected. Once the initial thematic coding was complete, the second author conducted an additional pass of all transcripts to verify alignment of themes with focus areas and to quantify data as appropriate.

Results

Biomedical and Bioengineering Research at Illinois

Sixteen individuals participated in this study, including two assistant professors, two associate professors, and twelve full professors. Many interviewees hold multiple departmental affiliations. Primary affiliations represented the following departments within the College of Engineering: Electrical and Computer Engineering (1); Computer Science (1); Chemical and Biomolecular Engineering (1); and Bioengineering (1). The remaining 12 were from the College of Liberal Arts and Sciences: School of Chemical Sciences (4); and School of Molecular and Cellular Biology (8).

All interviewees were PIs on at least one NIH grants representing both basic sciences and clinical research. Table 1 outlines funding types and provider agencies. PIs reported employing from one to 15 personnel in their respective labs, with an average of some combination of five to seven postdocs and/or graduate students. Three labs also employed non-research support personnel.

Table 1: NIH Institutes and Funding Types

Funding Institute	Funding Type							Total
	R01	R33	R37	P01	P30	U01	U54	
NIGMS	4		1	2			2	11
NIAMS	1							1
NIDCD	1							1
NIHLBI	3							3
NIMH						1		1
NIDA					3			3
NIBIB	1							2
NIAIA		1						1
NINDS	1							1
NCI	1							1
NIDDK	1							1

The projects represent very diverse research areas including biochemistry, cell biology, biomaterials, electrophysiology, metabolic disease, nano-medicine, stem cell research, and therapeutics. Most projects generate multiple types of data as outlined in Table 2.

Table 2: Types of Data Generated in Grant Funded Projects

Date Type Reported	Number of Projects Reporting Data Type
Assays	1
Graphs	1
Microscopy	1
Tissue samples	1
Lab Notebook ¹	2
Tables	2
Transcription	2
Video	2
Sequencing	3
Spectroscopy	3
Spreadsheets	3
Text files	5
Image files	9

Several questions were designed to assess specific data management knowledge and behavior.

When asked about data management plan (DMP) requirements specifically, five of sixteen researchers indicated that their funding exceeded the \$500,000 threshold and required a data management plan (Table 3). The DMPTool provides a customizable template to help researchers develop DMPs and Illinois offers an institutional version of the DMPTool. However, only one of the five PIs required to submit a DMP had used the DMPTool. Two others were familiar with the tool but had not used it for current projects and one respondent reported using the DMPTool for a previous grant submission. Of the twelve PIs who were not familiar with the DMPTool, four indicated interest in learning more about it.

Research Data File Management

Roughly half of the researchers interviewed reported having documented protocols for file names, file backup and file storage. Documentation tended to correlate with larger clinical research projects.

1 Multiple PIs reported using Lab Notebooks in their research, but only two mentioned lab notebooks as a type of data that was generated.

Table 3: DMP requirements and PI familiarity with DMPTool

Data Management Planning (DMP)	Yes	No	n
Were you required to submit a DMP?	6	10	16
Are you familiar with DMPTool?	4	12	16
Have you used the DMPTool?	2	14	16

Bioengineering researchers were associated more with human subject research and were more likely to have established file naming conventions (i.e. project file names, instrument, and date). Researchers in the biomedical group conduct smaller-scale basic science research and were more likely to rely on their graduate students and postdocs to manage their own data, including things like file organization, file naming, and backup (Table 4).

Table 4: Current Practices for File Management

Established Protocols	Documented	Verbal Agreement	Up to Individuals	n
File Backup	7		9	16
File Storage	6	2	8	16
File Naming Convention	6	3	7	16

Respondents reported using a variety of file storage solutions including local computers, cloud based storage, external hard drives, departmental servers, and tape backup. Most researchers understand the importance of backup and rely on daily or weekly automated computer back-ups, and/or back up on an external hard drive. One PI stressed that data backup is “the most important thing...not just backed up duplicate—backed up copies stored at three different locations so that if the whole building burns down, I’m not worried.” Some rely on local department computers or servers for data storage, assuming that their respective department, unit, or campus IT backs up systems automatically.

Small-scale researchers tend to be less concerned with backup and reported little in the way of tracking to ensure that redundancy happens. Even as they recognize the importance of backup, the routine is not always in place. “We’ve been burned on this before—having a hard disk crash—I need to get more backup going myself—but the type of data that is being lost under those circumstances is—oh now I have to do that experiment again and we generate that data again.”

When asked about file access control and tracking protocols, clinical researchers were more likely to report using network tools, sharing access by directory and project to all those working on the project. Other access options include password protected computers, or using the cloud for information in the public domain.

For many researchers, the lab notebook remains the gold standard for documentation of research methods, observations and findings. “In the old days—this shows how long I have been in this business—it was really simple—everybody’s notebook [was data] and that’s it—you could find everything in the physical notebook.” Even computer-generated data analyses are printed and added to the notebook. And while there are now options for online lab notebooks, the iterative nature of basic science research does not lend itself to computer documentation. Researchers adopt their own “shorthand” for recording what happens and everything is documented in the lab notebook. The “shorthand” is often not codified in any way. It may be easily interpreted by the others in the field, but would be meaningless to those without a deep understanding of the area of scientific inquiry. According to one researcher, “even I have trouble picking up the notebook of someone in my lab and make sense of what is there.” All PIs in this study reported saving lab notebooks indefinitely—there was not a single report of a lab notebook being discarded, even for research spanning decades. “I have every notebook from every student that has ever been in my lab.” No interviewee reported backing up notebooks (photocopies or digital scans) and most saw no need to do so.

For others, the interviews did seem to increase awareness of data management practices and generated reflection on current practice. Some respondents indicated that they could or should do more in the way of standardizing file management. One researcher noted, “one can imagine if we had something searchable in place alongside standardization for file naming that we could probably be saving time and money because we wouldn’t have to repeat experiments.” Another participant stated, “Yes, I can—in talking to you—imagine something that is more systematic—file naming and storage—primarily file naming because we might save time in the end.”

The cost of data management, in both time and money, creates additional barriers. “In the funding environment that we find ourselves in, is that where I want to spend my money; that’s what I am doing on a daily basis—risk analysis/cost/benefit analysis.” The return on investment for documenting project processes is not sufficient for researchers who feel that project or data management just takes too much time away from research time.

Challenges

Multiple respondents cited dealing with legacy data when students leave the university as a major challenge. Specific data files are difficult to track down once someone has left or graduated. As one PI noted, “I just feel every time a student leaves the lab I don’t have a complete grasp of exactly where everything is; I tell them to deposit, I quite often have to go back to the student who is no longer in the lab and ask them for information; and I do feel like I’m not doing a good enough job...” Even when data are readily available, issues may arise. For example, another researcher shared that one postdoc had recorded everything in French.

Large-scale researchers revealed challenges in the areas of file organization, centralized resources, size of data, and making data open and accessible to the general public (Table 5). Data storage in general is an issue, particularly with large imaging projects. “We have not really encountered any major problems. But I would say maintaining the large video files size of video microscopy that leads to an explosion of data volume. We compress video files to a smaller size. We also make decisions about data (i.e. if we believe the data is faulty we delete it).” Labs often resort to purchasing external hard drives, or they rely on departmental servers

to store large data sets.

Two faculty reported having no challenges with their data management.

Table 5: Data Management Challenges

Data Management Challenge	Number of PIs Reporting
Finding Data when Students Leave or Graduate	6
Storage Solutions for Large Data Sets	6
File Management Challenges	5
Sharing Mandate Compliance	4
Sharing Data	2
Loss of Data	1

Data Sharing

When asked if they receive requests for their data, most of the researchers with large-scale projects indicated they receive data requests via email from colleagues. And in some cases, they get requests for pre-prints or more detailed data.

Small-scale basic science researchers tend to think there is minimal interest in their research data beyond others in their field who are studying similar things (Table 6). One respondent said, "I can't imagine anyone being interested in raw data like I just showed you." Another PI mentioned that data might be requested for the purposes of reproducing the data, but they had never had anyone request data to date. Several respondents said they were required to share data related to NSF funded projects.

Table 6: Category of Those who Might be Interested in your Research Data

Who Might be Interested in Your Data?	Number of Mentions
Researchers	9
Others in my field	8
Policy Makers	3
Industry	3
Practitioners	2
No One	1

The majority of researchers report sharing data in journal and conference publications. They also share data on websites, conference presentations, and with the media. In general, small-scale basic science researchers are willing to share their data, but they just don't see that it is of interest outside of their narrow respective fields. Large-scale and clinical researchers tend to be more proactive about disseminating research findings: "If you make a statement in the paper that data has to be in the paper or supplementary files. Anything that is published has to be backed up with accessible hard data" (Table 7).

Many researchers indicated that their data is shared in the publication and raw data would not be of interest, even if it could be interpreted by someone else which is problematic. "We are required to publish—no data—it's just not that kind of research; the data is really literally the paper—like this is what we saw," and, "...certainly we publish our data, in the article." The burden of preparing raw data for sharing is not one most researchers want to take on. As one PI noted, "someone on the outside needs additional information before they understand raw data...it is a lot of work to make data understandable." Another PI commented "I would be flattered frankly if someone wanted to look at my data, but it is not a huge priority because if I were to spend a lot of time doing it—to make a bunch of our data available—I imagine that it would take a fair amount of time." When asked if it was important for their data to be "findable" one researcher said it just wasn't that important. "I haven't really thought about it—it sounds like chaos to me." Another participant stated the data does not lend itself to searchability. Other responses included, "it's a non-issue in our field," and, "there is no reason to see unpublished data." Most respondents have never deposited their data into a repository.

Regarding potential reuse of their data, basic science researchers don't have a clear idea as to who might ultimately be interested. They generate a lot of information that doesn't directly relate to what they are looking for. "We take a lot of images of cells and there is a lot of information there that we don't necessarily see—or we do see but we are actively choosing to ignore because it doesn't directly relate to the project that we are studying." One PI brought up the problem with depositing data with publishers. If someone wants to get to that data, they need a subscription to the journal and that creates an additional barrier.

Table 7: Current Practices for Disseminating Research Findings

How do you Plan to Disseminate Your Findings?	Number of Mentions
Publication	15
Conference	5
Website	3
Poster	1
ResearchGate	1

Awareness of Campus-Level Research Data Support

Lastly, researchers were asked about their awareness of campus services and support for data management planning, data sharing and data deposit. As noted above, small-scale

researchers generally were not aware of the campus DMPTool and some expressed interest in learning more, as with the comment “Does that [DMPTool] extend to NSF? ...so this is worth it already,” Most researchers were somewhat familiar with IDEALS, the library’s repository for publications, but many had not used it. Some PIs had heard about the campus Research Data Service (RDS) but again, few had contacted RDS or knew what specific services were available.

We also asked if they would be interested in data management instruction for their research groups. Overall, faculty doing large-scale research were more aware that the University library offers data management and preservation services. Two PI responded as follows: “Yes, we know that. But we are computer scientists and currently there are so many commercial tools, so we rely on relational databases and the cloud.” Three respondents were not aware of data management and preservation services.

Three researchers indicated interest in obtaining data literacy instruction for their graduate students and project team. Another PI noted that most of their graduate students have good knowledge of data management because they were trained systematically. Another faculty member indicated they might consider partnering with the University library for archiving purposes. Universally, small-scale researchers were not as aware of library support for research data or other research support; however, several expressed an interest in learning more. This was particularly true for data management issues such as file organization, file naming, and data storage solutions. On PI asked, “Do you have specific recommendations [for file management]—yeah, I am interested.”

Discussion

This study revealed the majority of researchers explore broad research topics, various file storage solutions, generate numerous amounts of data, and adhere to differing discipline specific practices. This confirms results found in similar studies (Akers & Doty 2013; Carlson, & Stowell-Bracke 2013; Wiley & Mischo 2016). Researchers expressed both familiarity and unfamiliarity with DMPTool. Twenty-five percent of the respondents indicated interested in learning more about it. Future work should explore the interest and creating services to meet this need.

Researchers repeatedly discussed the volume of data they are producing and listed it as a challenge as well. Margolis et al. 2014 suggests improved infrastructure and tools could possibly help with the volume of data biomedical research produces. Future work can explore both local and remote infrastructure solutions and tools to facilitate data storage and file transfer.

Research data management is ensuring that research data is properly documented, organized, stored, archived, and curated so that it is available for access, use, and reuse. Project documentation is highly variable and subject to the culture set by the PI. Publishing norms can also dictate the level of documentation. For example, it is expected that instruments used in research will be reported by make and model. Settings are sometimes recorded by associated software programs and researchers were not concerned if different machines were used in the same experiment. We found that PIs with larger labs and major grants were likely to be more diligent about documenting this level of project metadata.

There were some clear distinctions in data management practices between basic research, which tends to be smaller in scale, and clinical research. How do the differences between basic science and clinical research impact the application of data management? The Association of American Medical Colleges (AAMC) defines basic science research as laboratory studies involving cell cultures, animal studies or physiological experiments. The National Institute of Health, defines clinical research as a range of studies and trials in human subjects that can be categorized as patient-oriented, epidemiologic, behavioral and outcomes research and health services research. Basic sciences researchers tend to do smaller studies that are more iterative in nature.

Research involving human subjects brings the challenges of informed consent, participant recruitment and protection of confidentiality (Johnson et al.... 2016). The majority of researchers in this study share data in research publications, upon request and on websites. One respondent indicated a challenge with sharing research data in a repository. Data privacy and security are critical and data must be de-identified before it can be shared. For example, research focusing on enhancing breast cancer imaging would benefit others researchers studying this, but the researcher was unsure if a repository would be best suited for the results of this study. Sharing clinical data continues to raise questions about what data needs to be shared, and whether ethical or legal regulations permit it.

For small-scale researchers, the iterative and exploratory nature of basic science research was a reason several gave for not formalizing/codifying information regarding how things should be done or documented. One researcher noted “nowadays a lot of data is electronic and that creates a big problem to be honest—to manage—to keep the so called raw data—many journals now require raw data.” But without knowledge of the notation shorthand, or the context of experiments, raw data is not generally useful.

Johnson et al. 2016 assessed data management needs in clinical research from the perspectives of what researchers believe they want. The results of this study determined individual researchers exhibited significant diversity of perspectives that did not correlate by role or site. Basic science and clinical research have different areas of focus and in some cases they work together. For example, a researcher designing drug vaccines will at some point impact the direction of clinical research. Both areas benefit from applying research data management practices to their research.

Overall there is little difference in the application of data management for basic science, clinical research and human subjects. For librarians, future research can explore these issues and suggest solutions to these issues. The reality is these are common in many disciplines and with many researchers. Despite the challenges of sharing data researchers are sharing, data is often not visualized or in the uniformity that library communities or even the discipline themselves perceives. Sharing and managing research data has to be important for more reasons than funding requirement and scholarly publications policies.

Researchers in this study were asked how they planned to share and disseminate the findings from research. The majority of the respondents indicated they plan to share data within journal publications. Is this considered data sharing? Read et al. (2017) suggests sharing research data is essential for research transparency in scholarly literature. *Public Library of Science (PLOS One)* is requiring authors to share data if they want to publish in there journal.

Whereas, *New England Journal of Medicine* has released less stringent deadlines regarding sharing data. The *Journal of Medical Library Association (JMLA)* has not enacted policies related to sharing research data². This suggests ambiguity in regards to sharing data as well as disappointing open data advocates interested in seeing biomedical research (Oransky 2017).

While researchers are generally aware of research support services available to them, they still may not make seek them or apply the information gained from them. Awareness of services and resources does not have a direct correlation to application. The results in this study confirm some similar findings in previous studies (Perrier et al. 2017). Yet this study adds to the literature in the field and by identifying new opportunities to provide services and build relationships with faculty and graduate students. More specifically, UIUC's new school is in full operation as of 2018, and this provides librarians and, data managers with the ability to expand research data management support services to health and medical researchers. In addition, half of the participants in this study are interested in having instruction about best practices for managing research data.

There is also significant ambiguity about terms like "deposit" and "repository." Researchers consider publisher sites as repositories for example, while others equate publisher sites with research data storage. One PI summarized it this way: "Its fundamental research so therefore it flies way underneath the radar screen and people don't have any interest in it until—oh my gosh—this is important for what we are studying."

Federer et al. 2015 indicates researchers are more likely to share data with another researcher than to have deposited research data in a repository. These results are confirmed in the research data behaviors of engineering and atmospheric researchers as well as the usage of campus repositories (Wiley & Mischo 2016; Wiley 2015).

Conclusions

Overall, biomedical and bioengineering researchers have established patterns for managing data, but these don't necessarily represent best practices for proactive management of data. Decisions regarding data file naming, storage and backup are often left up to individuals in the lab, particularly in basic science research labs. Basic science researchers are also less concerned with sharing or their data and view the publication as the main vehicle for disseminating results. Clinical researchers are more proactive about disseminating research findings but cite challenges to prepare data for sharing and deposit.

Many researchers are not fully aware of the array of services for supporting research data management. In this regard, the interviews served as a good vehicle to raise awareness and share information about available resources. Furthermore, our questions often prompted researchers to think about data in a way that they had not done previously.

The interviews revealed significant data literacy gaps that present opportunities for instruction in the areas of file organization, project workflow and documentation, and metadata standards

2 At the time of the writing and publishing of this article, The *Journal of the Medical Library Association's* data sharing policy wasn't in effect. This policy will go into effect on October 1, 2019.

and data deposit options. Interviewees expressed interest in improving their own data management practices, and almost half said that they might be interested in instruction in data management related issues.

For librarians, the interviews provide invaluable insight into biomedical and bioengineering research in general and contribute to the authors' understanding of challenges facing the researchers we strive to support. The number of participants who did not view sharing data as necessary is quite notable, especially the attitude that data are simply not of interest or use to others. If our data services are structured around data sharing, which is a natural tendency given library-led repositories and their ready extension to data archiving, then being relevant to such audiences will be particularly challenging. The interviews described here suggest that a greater focus on data management may be warranted, at least in the near term, for some communities. The dismissal of the need for data sharing also suggests that additional research into data re-use, or lack thereof, is critical. Without a better understanding of the realities of data reuse, we may struggle to build library services with the reach and impact initially intended and expected.

Supplemental Content

Appendix

An online supplement to this article can be found at <http://dx.doi.org/10.7191/jeslib.2019.1132> under "Additional Files".

References

- Carlson, Jake. 2012. "Demystifying the Data Interview." *Reference Services Review* 40(1): 7-23. <https://doi.org/10.1108/00907321211203603>
- Carlson, Jake and Marianne Stowell-Bracke. 2013. "Data Management and Sharing from the Perspective of Graduate Students: An Examination of the Culture and Practice at the Water Quality Field Station." *portal: Libraries and the Academy* 13(4): 343-361. <http://doi.org/10.1353/pla.2013.0034>
- Chen, Xiujuan and Ming Wu. 2017. "Survey on the needs for Chemistry research data management and sharing." *The Journal of Academic Librarianship* 43(4): 346-353. <https://doi.org/10.1016/j.acalib.2017.06.006>
- Diekema, Anne R., Andrew Wesolek, and Cheryl D. Walters. 2014. "The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories." *Journal of Academic Librarianship* 40(2-3): 322-331. <http://doi.org/10.1016/j.acalib.2014.04.010>
- Fecher, Benedikt, Sascha Friesike and Marcel Hebing. 2015. "What Drives Academic Data Sharing?" *PLoS ONE* 10(2): e0118053. <http://doi.org/10.1371/journal.pone.0118053>
- Federer, Lisa. 2016. "Research Data Management in the Age of Big Data: Roles and Opportunities for Librarians." *Information Services & Use* 36(1-2): 35-43. <http://doi.org/10.3233/ISU-160797>
- Federer, Lisa M., Ya-Ling Ly, and Douglas J. Joubert. 2016. "Data Literacy Training Needs of Biomedical Researchers." *Journal of the Medical Library Association* 104(1): 52-57. <http://doi.org/10.3163/1536-5050.104.1.008>
- Federer, Lisa M., Y-Ling Lu, Douglas J. Joubert, Judith Welsh, and Barbara Brandys. 2015. "Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff." *PLoS ONE* 10(6): e0129506. <http://doi.org/10.1371/journal.pone.0129506>

- Hickson, Susan, Kylie Ann Poulton, Maria Connor, Joanna Richardson, and Malcolm Wolski. 2016. "Modifying Researchers' Data Management Practices: A Behavioural Framework for Library Practitioners." *IFLA Journal* 42(4): 253-265. <http://doi.org/10.1177/0340035216673856>
- Johnson, Andrew M. and Shelley Knuth. 2016. "Data Management Plan Requirements for Campus Grant Competitions: Opportunities for Research Data Services Assessment and Outreach." *Journal of eScience Librarianship* 5(1): e1089. <http://dx.doi.org/10.7191/jeslib.2016.1089>
- Keil, Deborah E. 2014. "Research Data Needs from Academic Libraries: the Perspective of a Faculty Researcher." *Journal of Library Administration* 54: 233-240. <http://doi.org/10.1080/01930826.2014.915168>
- Kim, Jeonghyun. 2013. "Data Sharing and its Implications for Academic Libraries." *New Library World* 114(11/12): 494-506. <http://oid.org/10.1108/NLW-062013-0051>
- Lyon, Liz. 2012. "The Informatics Transform: Re-Engineering Libraries for the Data Decade." *International Journal of Digital Curation* 7(1): 126-138. <http://dx.doi.org/10.2218/ijdc.v7i1.220>
- MacMillan, Don. 2014. "Data Sharing and Discovery: What Librarians Need to Know." *Journal of Academic Librarianship* 40(5): 541-549. <http://doi.org/10.1016/j.acalib.2014.06.011>
- Martin, Elaine R. 2016. "The Role of Librarians in Data Science: A Call to Action." *Journal of eScience Librarianship* 4(2): e1092. <http://lx.doi.org/10.7191/jeslib.2015.1092>
- McLure, Merinda, Allison V. Level, Catherine L. Cranston, Beth Oehlerts, and Mike Culbertson. 2014. "Data Curation: A Study of Researcher Practices and Needs." *portal: Libraries and the Academy* 14(2): 139-164. <http://doi.org/10.1353/pla.2014.0009>
- Ogier, Andrea L., Anne M. Brown, Jonathan L. Petters, Amr E. Hilal, Nathaniel D. Porter. 2018. "Enhancing Collaboration across the research ecosystem: Using libraries as hubs for discipline-specific data experts." In *Practice and Experience in Advanced Research Computing Conference Series*. Pittsburgh, PA, USA, July 22-26, 2018. <https://doi.org/10.1145/3219104.3219126>
- Oransky, Marcus L. 2017. "New science data - sharing rules are two scoops of disappointment." *STAT News*, June 6, 2017. <https://www.statnews.com/2017/06/06/data-sharing-rules-disappoint>
- Perrier, Laure, Erik Blondal, Patricia A. Ayala, Dylanne Dearborn, Tim Kenny, David Lightfoot, Roger Reka, Mindy Thuna, Leanne Trimble and Heather MacDonald. 2017. "Research Data Management in Academic Institutions: A Scoping Review." *PLoS ONE* 12(5): e0178261. <http://doi.org/10.1371/journal.pone.0178261>
- Piowar, Heather A. and Wendy W. Chapman. 2008. "A Review of Journal Policies for Sharing Research Data." In *ELPUB 2008 Conference on Electronic Publishing*. Toronto, Ontario, Canada, June 25-27, 2008. <https://pdfs.semanticscholar.org/3577/8362c926f1f36ac1250b13415cc7214b3c28.pdf>
- Poole, Alex H. 2015. "How has your science data grown? Digital Curation and the Human Factor: A Critical Review." *Archives of Science* 15(2): 101-139. <http://dx.doi.org/10.1007/s10502-014-9236-y>
- Read, Kevin, B., Liz Amos, Lisa M. Federer, Ayaba Logan, T. Scott Plutchak, Katherine G. Akers. "Practicing what we preach: developing a data sharing policy for the Journal of the Medical Library Association." *Journal of Library Medical Association* 106(2): 155-158. <http://dx.doi.org/10.5195/jmla.2018.431>
- Reznik-Zellen, Rebecca, Lisa A. Palmer, and Nancy E. Harger. "A comparison of Student and Faculty Attitudes on the Management and Sharing of Research Data at UMMS." In *University of Massachusetts and New England Area Librarian e-Science Symposium*. Worcester, MA, USA, April 9, 2015. http://escholarship.umassmed.edu/escience_symposium/2015/posters/14
- Sayogo, Djoko Sigit, and Theresa A. Pardo. 2013. "Exploring the Determinants of Scientific Data Sharing: Understanding the Motivation to Publish Research Data." *Government Information Quarterly* 30(Supplement 1): S19-S31. <http://doi.org/10.1016/j.giq.2012.06.011>

Surkis, Alisa. 2015. "Research Data Management." *Journal of the Medical Library Association* 103(3): 154-156. <http://dx.doi.org/10.3163/1536-5050.103.3.011>

Tenopir, Carol, Elizabeth D. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. 2015. "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide." *PLoS ONE* 10(8): e0134826. <http://doi.org/10.1371/journal.pone.0134826>

Tenopir, Carol, Robert J. Sandusky, Suzie Allard, and Ben Birch. 2014. "Research Data Management Services in Academic Research Libraries and Perceptions of Librarians." *Library and Information Science Research* 36(2): 84-90. <http://doi.org/10.1016/j.lisr.2013.11.003>

Vogeli, Christine. Recai Yucel, Eran Bendavid, Lisa M. Jones, Melissa S. Anderson, Karen Seashore Louis, and Eric G. Campbell. 2006. "Data Withholding and the Next Generation of Scientists: Results of a National Survey." *Academic Medicine* 81(2): 128-136. https://journals.lww.com/academicmedicine/Fulltext/2006/02000/Data_Withholding_and_the_Next_Generation_of.5.aspx

Webster, Peter. 2012. "The Web-scale Discovery Environment and Changing Library Services and Processes." In *Planning and Implementing Resource Discovery Tools in Academic Libraries*, edited by Mary Pagliero Popp and Diane Dallis 646-661. Hershey, Pennsylvania: IGI Global. <http://dx.doi.org/10.4018/978-1-4666-1821-3>

Whitmire, Amanda L. Michael Boock and Shan C. Sutton. 2015. "Variability in Academic Research Data Management Practices: Implications for Data Services Development from a Faculty Survey." *Program: electronic library and information systems* 49(4): 382-407. <http://dx.doi.org/10.1108/PROG-02-2015-0017>

Wiley, Christie and William H. Mischo. 2016. "Data Management Practices and Perspectives of Atmospheric Scientists and Engineering Faculty" *Issues in Science and Technology Librarianship* 85(Fall 2016). <http://doi.org/10.5062/F43X84NJ>