2018

NN/LM NER e-Science Program, Researcher Support Committee White Paper on Supporting Researcher Needs

Sarah Oelker
Mount Holyoke College, soelker@mtholyoke.edu

Thea P. Atwood
University of Massachusetts Amherst, tpatwood@umass.edu

Andrew Creamer
Brown University, andrew_creamer@brown.edu

Margaret Henderson
San Diego State University, margaret.henderson@sdsu.edu

Kristin Lee
Tufts University, kristin.lee@tufts.edu

See next page for additional authors

NN/LM NER e-Science Program, Researcher Support Committee White Paper on Supporting Researcher Needs

Thea Atwood
Andrew Creamer
Margaret Henderson
Kristin Lee
Lora Leligdon
Sarah Oelker (Committee Chair)

May 4, 2018

Executive Summary

As more libraries become involved with Research Data Management (RDM) services, and the new strategic plan of the National Library of Medicine has a data focus, it is important to consider the researchers who create or collect data. There are many surveys and reviews in the literature about how researchers are managing their research data, as well as many case studies and surveys on how libraries are working to provide data management services. But anecdotal evidence and some surveys show that researchers aren't always going to the library for data support. What can be done to change these perceptions and insure that libraries and librarians are valued partners in research data management?

The Committee reviewed the literature and online information to learn more about what researchers actually want help with for data management, as well as what they probably need help with, although they don't realize it. Whether surveys were conducted by librarians or other researchers, the top things researchers want are storage, including help with the various formats of data that need to be combined, help with analytics and other computational needs, and help with sharing. Many surveys show a need for metadata help, although researchers aren't asking for it. This suggests that libraries developing data services need to provide, or facilitate access to, data storage and data analysis. They also need to learn about funder sharing policies and required and/or reliable outlets for sharing all types of data.

Another survey result is the lack of awareness many researchers have about library data services. Some researchers do not think libraries are capable of storing data or helping with research, but others just aren't aware that the service exists. Librarians setting up data services will need to make sure that outreach and marketing are part of the planning process from the beginning.

NN/LM NER e-Science Program, Researcher Support Committee White Paper on Supporting Researcher Needs

Introduction

> "The e-Science Portal for Researchers Committee will work in collaboration to create and implement a central web portal for New England (and outside) focused on researchers' needs. Members will identify the region's need for such a portal, communicating with experts, and identifying relevant resources for a web hub, similar to that of the existing e-Science Portal for Librarians."

There have been many changes in the goals of this committee since the initial charge was given to the group. When the e-Science Portal was no longer situated at the New England Region (NER), the ability to create a separate identity for the researcher portal was lost. Affiliation with librarian training materials, no matter how good, was not considered ideal for a researcher portal, so the focus of the committee changed.

Our initial work on the portal involved research into the literature on the research data management needs and wants of researchers, in order to develop a portal they would use. This white paper collects these needs and wants from a researcher perspective, and makes suggestions for librarians on how to best reach out to researchers. The committee recognizes that many librarians are already embedded in various institutional departments and schools, assisting with teaching and literature searching, and these relationships are important when trying to introduce support for research data management services.

While understanding researcher needs is important to providing good research data services, it is crucial that librarians have the support of library administration when starting research data management services. Some suggestions for librarian education and general information on the development of research data management (hereafter RDM) services will be provided, along with some examples from the authors.

Background

The National Network of Libraries of Medicine New England Region (NN/LM NER) recently hosted its 10th annual e-Science Symposium, "Libraries in Data Science: Past and Future." Academic and health sciences libraries in New England have developed an impressive number of new research data services and new positions aimed at supporting researchers with managing, sharing, and preserving the data that they collect and create. Looking back on the topics of the papers and posters presented at

the symposia since 2009, one sees an evolution that mirrors the changing landscape of biomedical research becoming more digital, more interdisciplinary, and funders and publishers requiring and expecting more accountability for researchers' management and sharing of publicly funded data. The symposium expanded and shifted focus over the past decade. Initial presentations on bioinformatics, clinical and translational science, and data sharing by NIH-funded researchers led to discussion of research data management as the NSF began requiring a DMP in 2011, just before the third symposium. Compliance with federal public access policies, particularly the Obama White House OSTP memo (Holdren 2013), has given way to recent concentrations on data science, coinciding with NIH appointing its first Associate Director for Data Science, who keynoted the 2015 symposium.

The term "e-science" has not survived the test of time. Its detractors have made valid arguments that its origins were too focused on "big data", ignoring the "long tail" of scientific research, as well as pointing out the fact that the term was in and of itself superfluous, ignoring that research of all stripes has become digital, making the 'e' unnecessary. However, the term data science has proven to be a much wider tent under which fit researchers with big and small data sets. Librarians need to become fluent with the terms used by researchers and start using data science or data management to describe the work they are doing. Under the umbrella of data science, biomedical researchers can collaborate with those in other STEM fields as well as those in the social sciences and even humanities. Where e-science once concerned itself with surmounting challenges posed by the deluge of data and increasingly high performance computation of data, data science concerns itself more with what insights can be extracted from data a researcher has, no matter the quantity or methods, which drives the innovation in developing new tools for data extraction, wrangling, and analysis as well as cyberinfrastructure.

Data science entails the tools, resources, and skills that can help researchers to have data savvy in carrying out their daily work. The National Library of Medicine (NLM) describes the field of data science as "broad in scope, encompassing approaches for the generation, characterization, management, storage, analysis, visualization, integration and use of large, heterogeneous data sets that have relevance to health and biomedicine" (National Library of Medicine 2017). In the same ways that biomedical libraries over the last decade have pivoted to embrace tools, resources, and new services and personnel to support bioinformatics, clinical and translational science, and research data management and sharing, biomedical libraries must now explore how to adapt to help researchers respond to the current trends in data science that are affecting how biomedical research is conducted.

One librarian who was ahead of the curve in introducing librarians in New England to the concept of data science was Christopher Erdmann, formerly at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. In 2012 he planned a course “Data Scientist Training for Librarians” (DST4L, Erdmann 2013) . The inaugural group of DST4L librarians met in early 2013 to learn and practice using tools for extracting, wrangling, analyzing, and visualizing data. While some libraries in New England had been up to that point supporting research data management, statistical software, and Geographic Information Systems (GIS), DST4L opened a new door to the types of data science tools and skills that researchers were interested in using and learning, such as R and Python, among others, and incorporating into their daily work. In addition to DST4L, many librarians in New England have sought similar skills and experience by participating in Software Carpentry, Data Carpentry, and Library Carpentry events, professional development workshops and courses such as North Carolina State University Libraries Data and Visualization Institute, and online courses.

In 2015, UMass Amherst created a Center for Data Science. Boston, Brown, Harvard, and Tufts Universities each launched campus-wide data science initiatives in 2017. Numerous institutions have created data science degree programs, such as Dartmouth’s Master’s in Health Data Science. In 2016, the NIH included data science in its strategic plan:

> Data science also holds tremendous potential, not only for enhancing the efficiency of the conduct of science, but also for increasing the impact of fundamental science, along with many other areas of biomedical research. To this end, NIH will serve as a focal point for catalyzing this historic research opportunity. . .

The following year the NLM issued "Request for Information (RFI): Next-Generation Data Science Challenges in Health and Biomedicine" (National Library of Medicine 2017), to

> seek community input on new data science research initiatives that could address key challenges currently faced by researchers, clinicians, administrators, and others, in all areas of biomedical, social/behavioral and health-related research.

In October of that year, the Association of College and Research Libraries (ACRL) responded to the NLM’s RFI (Association of College & Research Libraries 2017). Key to

its response was the inclusion of librarians and articulation of their role in meeting these challenges:

> Academic librarians are on the forefront of teaching about, facilitating access to, and preserving information and data across extensive resources. As foundational educational partners for many who will go into healthcare professions, we are ideal partners for workforce development in data management for health and biomedical professionals.

In 2016, the University of Pittsburgh and NCSU received funding from IMLS to “convene a group of experts from inside and outside the library community to articulate a vision and roadmap for data science in libraries.” The resulting “Data Science in Libraries Project” held a two day meeting to address the tension between the need for librarians to meaningfully engage the tools and techniques of data science but the challenge of what they called the librarians’ “data science skills gap”: “While practicing mid-career librarians are learning some data science skills, it is through ad-hoc, uncoordinated continuing education programs” (Burton et al. 2016). Before librarians know which skills to prioritize, we must have more input from our researcher communities on what their needs are in order to able to align goals with needs. In this report we will present several library case studies to characterize researchers needs regarding the skills, tools, and resources that are being sought and utilized. We hope that these examples will help librarians in New England to develop services and hire positions to help support the data science needs of their researcher communities.

Existing surveys of researchers' needs

There have been efforts to assess researchers' needs in data services, both within librarianship and outside of it. These efforts have come from a number of disciplines and contexts. In 2016 Kjellberg et al. brought together participants at ASIST 2016 to discuss needs for creating and using research data with both a panel and participant discussions. Work at Microsoft's external research organization brought together contributors from a variety of disciplines and types of work to comment on what will be needed to provide the infrastructure, training, and tools for research involving large data sets and new types of collaboration (Hey, Tansley, and Tolle 2009). Colorado State University librarians conducted five focus groups with thirty-one faculty, research scientists, and research associates. The groups explored: (1) The nature of data sets that these researchers create or maintain; (2) How participants manage their data; (3) Needs for support that the participants identify in relation to sharing, curating, and preserving their data; and (4) The feasibility of adapting the Purdue University Libraries’

Data Curation Profiles Toolkit interview protocol for use in focus groups with researchers (McLure et al. 2014). The authors report their review of related literature, themes that emerged from analysis of the focus groups, and implications for related library services. Barone et al (2017) surveyed over seven hundred NSF principal investigators in the Biology directorate, and noted that the respondents said the most pressing unmet needs are training in data integration, data management, and scaling analyses for high performance computing (HPC), acknowledging that data science skills will be required to build a deeper understanding of life (Barone, Williams, and Micklos 2017). This portends a growing data knowledge gap in biology and challenges institutions and funding agencies to redouble their support for computational training in biology.

There have been many library studies looking at researchers and data. Akers and Doty have characterized differences in the needs of faculty based on their rank, and on their disciplinary homes, both of which are valuable information (2012). They found that while faculty rank doesn't seem to make differences in how faculty are storing data, for instance, but that it does make differences in stated reasons for reluctance to share data. They note that an important detail is that earlier career scientists want to learn more about ways to share with limitations that might be protective to the researchers' careers, whereas for older faculty the barriers may be more about lack of time. In their paper on disciplinary differences, they found significant differences in four major disciplinary domains of research on questions ranging from familiarity with funder requirements, to whether they share data, to familiarity with metadata (Akers and Doty 2013).

The same study on disciplinary differences discusses which groups of faculty are more or less likely to be familiar with funder requirements, and that a majority of researchers were not familiar with such things at the time of this research. Whitmire also addressed this (Whitmire, Boock, and Sutton 2015). Meanwhile, a number of papers have addressed who is more likely to share their data, where they share it, and how much importance different populations of researchers place on sharing their data (Akers and Doty 2013; Akers and Doty 2012; Buys and Shaw 2015; L. M. Federer et al. 2015; Tenopir et al. 2011; Tenopir et al. 2015). And finally, Federer's 2016 paper addressed the training needs of a population of researchers, namely in biomedicine, which begins to map out where to put our efforts in this area (Federer, Lu, and Joubert 2016).

Goben and Griffin (2017) reviewed 40 library studies and found that storage, sharing, and issues that revolve around long term access to data are the top concerns of researchers. Data management plan assistance, security/privacy concerns, data organization, and deciding on the party responsible for data management were also

issues for many researchers. Goben and Griffin recommend the reuse or recombination of surveys, rather than recreation or newly creating tools, in order to facilitate future studies of aggregate data needs or opportunities.

While this sort of surveying has been very useful, surveys of researcher needs have not standardized the questions that they ask, which frustrates attempts to do a meta-analysis of the results of these projects. Perrier, et al. (2017) conducted a scoping review on research data management in academic institutions. After reviewing 301 articles, they found there was an issue of data quality, including a deficiency in standardized or validated data collection tools, and a lack of transparency in reporting that makes comparison of results difficult.

At this point, the librarianship literature on RDM articulates many ideas about researcher needs. We can begin to categorize those needs, in terms of subject area as well as in terms of disciplinary differences in needs. While there remain gaps in the characterization of needs, both in terms of the needs of researchers in different sizes and shapes of institutions and in areas where technological advance runs well ahead of description of some of the resulting needs, the existing characterizations are helpful to those seeking to serve researchers at their own institutions.

The recently published "Shifting to Data Savvy: The Future of Data Science In Libraries" (Burton et al. 2018) further fleshes out the concept of a "skills gap" that the authors defined in their earlier Data Services in Libraries Project cited previously, and lays out a framework for structures, stakeholders, service and skills necessary to implement support for data science in libraries, by librarians. It also describes the responsibility of managers to learn enough about the skills required of their reports to successfully re-organize workloads, encourage appropriate professional development, and provide other support necessary for library staff to be successful in implementing such programs.

There are a number of areas in which we would like to see further characterization of researcher needs: data collection and storage, data documentation and dissemination, education on and use of specific tools, and the support of both specific hardwares and softwares in use in data management. Most of these articles are also conducted at R1 universities, and more characterization of needs in smaller institutions, or more subject-focused institutions, would also help in meeting these researchers' needs. Much of the current research is survey focused, and methodologies like ethnographic study of research data management, as is being pioneered at the University of Washington ("Data Science Ethnography"), would be welcome additions to the literature.

Research aside, faculty also need to consider data literacy or data information literacy when teaching classes in data intensive subjects. There are already concerns that librarians shouldn't be teaching information literacy (Bivens-Tatum 2009), so it can be difficult to get faculty to consider librarians for data literacy instruction. And yet, DIL studies (Carlson and Johnston 2015) have shown that librarians can have a positive impact on student learning of data management. And teaching collaborations have shown the value of libraries in data literacy (Murillo and Jones 2017).

Supporting your initiatives

In supporting researchers' RDM needs, it is important to make sure that each library assess what skills, administrative backing, and infrastructure they have to be able to provide the necessary support. Providing services for RDM in your context could look very different from those offered at other institutions, as the options can come from the personal interest of a particular person, a need from a prominent group on campus, a grant that funds a specific project, or legacy services that have grown and flourished. It is also important to understand the depth of the scale of the service offerings, because sometimes it can be as simple as having a webpage. Determining the services that will have the most impact for your community is a far more important goal than trying to match all of the services being offered by peers, whether they are true peers or aspirational ones.

One way to support new initiatives is to find regional or national collaborations that can be built upon. These collaborations can come in a variety of shapes and sizes, from national and grant funded to small and with support from a number of institutions. An example of the former is the Data Curation Network, a joint project begun by six institutions to evaluate and create a common model for data curation services (Johnson, et al 2016). The website for the Data Curation Network states that the project has since expanded to nine institutions total, with new grant funding to implement the model that they created in the first phase (Data Curation Network 2018). This model involves hiring a coordinator for the Digital Curation Network, who will manage the requests for data curation services, make sure it gets to the correct curator at one of the 9 member institutions, and then track the work to completion (Data Curation Network 2018a). By leveraging the expertise at nine different institutions, the members of the Data Curation Network will be able to provide value-added curation services to their members in a variety of disciplines while also further deepening their local expertise.

The second example of a collaborative effort to provide services is the regional New England Software Carpentry Library Consortium ("NESCLiC"). NESCLiC was created in 2017 with seven member schools across four states to provide an opportunity for fifteen people to get Software Carpentry Instructor Training. The cohort of instructors has come together to participate in Software Carpentry as learners to get the full experience in addition to becoming Certified Instructors. Members have been offering workshops using the Carpentry materials and have been asked to teach Library Carpentry in the Boston area. The network of data service providers that has grown from this shared experience has created lines of communication that have allowed us to expand opportunities for data instruction, both on our individual campuses and for others in the area.

No matter what you are doing on your campus to support RDM, make sure that you are meeting the expectations that are being set by your marketing and outreach activities. This may mean slowly building services instead of trying to start everything at once, but this will have the added benefit of allowing time for reflection and discussion with the key stakeholders. It is also important not to try and create services just to compete with similar offerings that your researchers are finding. It is the personal touches and additional value that a librarian can add to the research endeavor that will bring more people to a service, not the number of other kinds of things you can do.

Researchers unaware of what library can do for them in the areas of data and research support.

Faculty and researcher awareness of library services, other than books and journals, has been a problem for a long time. In 1973, Nelson found that the average faculty member was aware of less than half of the reference services available in their library. Forty years later, a survey at two Georgia universities found similar awareness amongst faculty. (Slutskaya et al. 2013) In a March 2018 tweet, Lisa Federer, at the NIH Library, wrote “I heard a researcher give a talk today and he said libraries desperately need to update their services to address digital curation and other data-intensive research needs. He said librarians were still "holding on to buggy whips" and hadn't moved past printing presses.” (Federer 2018) If basic library services, such as LibGuides, are not known by faculty, they probably don’t know about scholarly communication and data services offered by libraries.

As noted in the literature review, there can also be a disconnect between what faculty/researchers need for data management, and what they actually want. Usually, researchers say they want storage, backup, and software for their research data, but

the underlying need for best practices and preservation, or an understanding of funder mandates, is not noted because they don't realize these needs, or cannot articulate them clearly enough to get assistance. Libraries often cater to such underlying needs, especially given that upfront needs, such as storage, backup, and software, have material costs which are typically outside of the library budget. Researchers and faculty usually don't consider the library when looking for data help, so outreach for all new library services is essential if libraries are to prove they have a place in the research lifecycle.

When developing outreach, librarians first need to check their vocabulary. Even the term data has different meanings when talking with researchers in different disciplines. Terms such as e-science or e-research may have meaning to researchers in some areas, but most will not recognize the terms. Terms like storage, preservation, and backup can also be problematic. And metadata can be collected at many levels, so a researcher may know they have metadata for a particular dataset, for example a series of images that include the metadata about the microscope settings and magnification, but not realize that there can be metadata at a higher level, such as the analysis of measurements in those images with R.

Researchers' needs as researchers, versus researchers' needs as educators

For librarians to be taken seriously as research collaborators, especially when offering help with data management and analysis, skills are needed to back up the offer. Extensive and usable knowledge on statistical tests, storage, preservation, or backup should not only be on the best solutions available, but also fit with the necessary rules and mandates. Additionally, working knowledge on DMPs, including discipline best practices, stated requirements, and reasonable researcher compliance, is needed prior to offering assistance.

In order to properly grow data services, library administrators and managers need to allow librarians the time to learn about RDM, if they are using existing personnel, as well as time to learn about institutional resources, create relationships, and develop outreach to those working with data. In order to have the flexibility to help with grants or data disasters, librarians need to have open schedules. Making sure time isn't filled with classes or desk hours, or that an appropriate backup is available, will allow for the agility needed to respond to researchers in a timely fashion and eventually expand services.

The field of suitable tools is constantly shifting and discipline-specific

Researchers must grapple with the constant flow of new data tools being created and released to the researching public. There is no single source for learning of these new tools, and staying abreast of these developments requires engagement in communities doing this work, whether with in person information sharing, online groups, or a combination. While an individual researcher may hear of the tools that are most popular in their discipline, or those in which others in their discipline had a hand, they are less likely to be aware of tools that were not designed with their specific discipline in mind but which may still be useful. Librarians bear a responsibility for staying aware of new tools across a range of disciplines, and for connecting researchers with suitable tools for their work.

Organizations must stay informed of new tools and shifts in focus, as well. For example, The Coalition for Networked Information (CNI) provides resources for strategic planning for institutions and opportunities for networking, in the interpersonal sense, with others who are using networked information technology for scholarly projects ("CNI: Coalition for Networked Information"). While CNI has existed since 1990, it has new relevance as data tools proliferate, as a resource for information professionals, and the organization has expanded well beyond its original mission to provide high speed networking, in the technological sense, for academic institutions.

How is a librarian to interact with, for example, an organization like the Institute for Social Science Research at University of Massachusetts, Amherst ("Institute for Social Science Research | UMass Amherst"). This group is organized outside both the library and information technology organizations of the university, but with strong ties to both and with a mission to promote the highest quality social science research. It is the sort of organization that both librarians and researchers can look to for help identifying tools outside one's own expertise, and with resolving issues of methodology that might arise in the course of an interdisciplinary project.

There are also systems for finding and choosing between data science tools. The Data One software catalog is one such system ("Software Tools Catalog"), and another is the DCC Tools and Services listing ("Tools & Services"). Services like this do require some foreknowledge of the terminology for particular types of research. Some of this terminology can be acquired by interviewing the researcher seeking assistance, but a certain baseline from the librarian is also likely to be required, in order to interpret the results of this sort of finding aid. The balance of baseline knowledge and information acquired while consulting with an individual researcher may vary from librarian to librarian, as well. Librarians also feel pressure to jam existing data into existing systems

to feel relevant, as libraries, when sometimes there absolutely should be major paradigm shifts in where and how we store that data.

Sometimes the answer just is that the scene is changing more quickly than a catalogue of tools can keep pace with. The shifting nature of these things is difficult for library culture because libraries want to give cut-and-dried answers. In various areas, libraries have experimented with more just-in-time or patron-driven models for services, and Goldstein and Oelker posited that particularly for small liberal arts colleges, just-in-time provision of services for faculty researchers may be a key to growing services just fast enough (Goldstein and Oelker 2011). Once you are meeting just-in-time needs, to have sustainable services requires being willing to review services regularly, and to iterate services or even discontinue some older services to make room for new ones.

Resources for librarian professional self-development to meet researcher needs

Self-development is, by its nature, an individual process. Meeting researcher data needs is often a moving target, and as such, your own framework and plan for professional development will require flexibility. A well-configured framework will provide a method for librarians to build their skills, can help them achieve measurable outcomes, and should appropriately adjust to changes in the needs of organizations and their stakeholders. Many good resources for planning such a self-development framework exist. Goben and Sapp Nelson's 2016 work with the Association of College & Research Libraries on an RDM workshop yielded a variety of great resources to choose from to build a meaningful development path. More recently, Goben and Sapp Nelson (2018) have further unpacked their model, to help potential users understand the rationales behind each part, is valuable for understanding what to focus on, how to get institutional buy-in, and how to meaningfully and observably assess growth. Burton et al. (2018) have also released, in their "Shifting to Data Savvy" white paper, useful visualizations and frameworks that may be used alone or mixed with tools from other models, in building an individual's professional development plan.

These frameworks will assist librarians in assessing the structures, stakeholders, and needs around them, which is a necessary first step for building a baseline of understanding of what skills might be useful to acquire. Choosing one or two new skills to develop at a time is an excellent way to set achievable goals. Create a plan that outlines how you will learn and demonstrate these skills. Including management early in this process, to balance these goals with other workload, find opportunities to practice new skills, and set goals that are observable and achievable and meet whatever institutional assessment practices are in use in a particular library.

After a librarian has gained some skills in RDM and some buy-in from their institution, they should be able to execute a plan for self-development, assess that plan periodically, and iterate their approach. Librarians are wise to be generous with themselves both in terms of the time it will take to expand their skillsets and time for reflecting on what they have learned. Work with a peer, a mentor, or supervisor to get the benefit of self-reflection, whether informally over lunch or formally as part of a performance review process. Assessment might include writing more granular outcomes for yourself, such as SMART goals or other assessment techniques. To complete the cycle of inquiring, building skills, and assessment, it is important to periodically reassess structures and stakeholders in the institution, and then adjust self-development as necessary to keep up with changing needs.

Conclusion

Since 2009 health and science librarians in New England, with much support from NER, have made significant advancements in establishing initiatives within their libraries and institutions aimed at providing services to support the management, retention, preservation, and sharing of their researcher communities' digital data. It is now common to find in NER academic libraries a team of librarians or an individual librarian that is tasked with supporting researchers' navigation of funders' public access policies, with the writing of data management and sharing plans, and with providing of infrastructure and assistance with carrying out these plans. Similarly, NER health and science librarians have been actively engaged in helping researchers to meet the relatively new requirements from publishers to retain digital research data underlying published results and to provide a citation or statement for their availability. Lastly, local initiatives such as The New England Software Carpentry Library Consortium (NESCLiC) are helping librarians in the region to gain data science skills that could help bring them into new collaborations with researchers on parts of the research lifecycle where librarians have not had traditional roles.

The expertise that librarians are developing and bringing to the table include: knowledge of funders' data management and sharing requirements; data management and organization best practices; resources for locating and tools for applying metadata standards for data, such as minimal information (mins) and reporting standards; resources for locating and using data and code sharing repositories; curation and documentation of data sets deposited into repositories; consulting on copyright and licenses for data sets; consulting on publisher policies regarding the citation and retention of data; the promotion of open standards; use of electronic laboratory

notebooks; use of tools and resources for rigor and reproducibility; and use of persistent identifiers, among others.

So if NER librarians are providing data services for researchers, are these services aligned with researchers' needs? Pinfield et al. (2014) found in their study on data service providers and users that we as providers need a clear understanding of where our users' demands exist and where there are needs for services, even if these are not explicitly shared with us. Our awareness of users' needs and demands, they conclude, "will shape ongoing RDM activity." Yet, their major finding was that participants saw the challenge was not in creating services for researchers but in *persuading* researchers to recognize the importance of data management and to seek our help.

So then what does the data say about these perceived needs? Mainly that over the last decade they have not changed much. In 2011, Tenopir et al. found among the leading reasons for researchers to not share data were insufficient time and lack of funding, copyright issues, and lack of knowledge about metadata, funder requirements and repositories. This year, nearly a decade later, Stuart et al. (2018) asked researchers about their challenges to sharing their data. The main challenges identified by respondents were organizing data in a presentable and useful way and a lack of knowledge about funder requirements, copyright and licensing, and repositories, and the lack of time to deposit data and the costs to share data. The question that we now need to start asking is if they had the time and the funds, then what would they still need to do to make their data available? The answer is a lot. But these activities are too difficult to do at the end of a study.

Over the last decade librarians have developed repositories and data management consultation services for curating and depositing data and invested a lot in their own education about data management and training in data science skills as well as invested in offering data management education to their researcher communities. However, researchers are still reporting the same challenges. Our takeaway from this is that this is evidence that they are either unaware of us or they need to be persuaded. It is unreasonable to expect a librarian to take a researcher's data set from a completed study and go back and re-perform hours of experiments and in order to collect the missing documentation and go back and comment and document their code and then clean the data set, metadata and code files in order to get them presentable and in shape for sharing. But the researcher and librarian can connect upstream to, from the very outset of a project, make a plan to capture the documentation necessary at point of data collection/creation and coding necessary for reuse, repository deposit, and publication, and organize and collocate files, and appraise and select for long-term

preservation the files necessary for the validation of results and those having enduring value.

While we feel the promotion of a data savvy librarian and data science skills should continue to be among the NER's foci, we should not lose sight that the literature is telling us the main researcher need is not a technical one, but a human one. We must do better to promote our services, perform more and prioritize outreach, and be confident in our role and value to the researcher and the research process. Going back to clean up a data set is too difficult; there is never going to be enough time and it will be expensive to re-run experiments again to electronically capture details such as the settings of the instrument and characteristics of the sample at time of collection. Researchers need us upstream to help get advice on how to create and collect metadata and integrate organizational best practices necessary to get their data, metadata, and code presentable for sharing as well as downstream to help them with selecting a license and a repository, and obtaining a DOI for the citation of their data set, metadata, and code in their publication.

So what happens if we and our services continue to be invisible and researchers unpersuaded for another decade? The for-profit publisher Springer-Nature, which sponsored the Stuart et al. 2018 study, rolled out a new Research Data Support service around the same time of the study's publication (Springer Nature 2018). For $340.00, they will "organize files into a logical structure and collections"; add keywords; perform checks for spelling and human subject identifiers; they will issue the data set a DOI; and they will deposit the data set in a repository. Basically they will charge authors hundreds of dollars for downstream services that many of the authors' libraries do for free. But what this service does not do is address the researcher needs for a partner at the outset of a study. This service cannot make experimental metadata and code documentation appear that were never collected in the first place or go back to re-label files and re-order directories to make collaboration more efficient.

The for-profit publishers monetizing the research lifecycle is concerning. Yet their investment in and charging money for these downstream services should embolden us and our confidence that the expertise and services that we have developed over the last decade are indeed needed and have value. We should not be ceding this territory; the for-profit publishers' presence and potential successes in this realm will be aided by the continued lack of visibility of our expertise and services and persuasion in this space. Our recommendation as a committee is for the continued investment by NN/LM NER in the development of librarians' data science expertise and services while prioritizing the investment in the outreach, promotion, and visibility of librarians' expertise and services.

Moreover, we recommend targeting programming aimed at the building of confidence and development of soft skills necessary to engage with and promote ourselves and our services to the researcher communities outside of our libraries.

Bibliography

Akers, Katherine G. and Jennifer Doty. 2012. "Differences among Faculty Ranks in Views on Research Data Management." *IASSIST Quarterly* 36 (2): 16-20. http://ejournals.library.ualberta.ca/custom/pdfs/IQarticles/iqvol36_2_doty.pdf.

———. 2013. "Disciplinary Differences in Faculty Research Data Management Practices and Perspectives." *International Journal of Digital Curation, Vol 8, Iss 2, Pp 5-26 (2013)* (2): 5. doi:10.2218/ijdc.v8i2.263.

Association of College & Research Libraries. "ACRL Comments to NLM on Data Science Challenges in Health & Biomedicine." American Library Association. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/ACRL%20response%20on%20next%20gen%20data%20science%20FINAL.pdf.

Association of College and Research Libraries, Megan Sapp Nelson and Abigail Goben. 2016. “Data Engagement Opportunities Scaffold.” Association of College and Research Libraries. https://acrl.libguides.com/scholcomm/toolkit/RDMWorkshop.

Barone, Lindsay, Jason Williams, and David Micklos. 2017. "Unmet Needs for Analyzing Biological Big Data: A Survey of 704 NSF Principal Investigators." *PLoS Computational Biology* 13 (10): 1-8. https://doi.org/10.1371/journal.pcbi.1005755

Bivens-Tatum, Wayne. "The Victim of Library Instruction | Academic Librarian.", https://blogs.princeton.edu/librarian/2009/10/the_victim_of_library_instruction/.

Burton, Matt, Liz Lyon, Bonnie Tijerina, and Chris Erdmann. 2016. "The Data Scientist as the 21st Century Librarian?" https://www.imls.gov/sites/default/files/grants/re-43-16-0149-16/proposals/re-43-16-0149_university_of_pittsburgh.pdf

Burton, Matt, Liz Lyon, Chris Erdmann, and Bonnie Tijerina. 2018. "Shifting to Data Savvy: The Future of Data Science in Libraries." http://d-scholarship.pitt.edu/33891/1/Shifting%20to%20Data%20Savvy.pdf

Buys, Cunera M. and Pamela L. Shaw. 2015. "Data Management Practices Across an Institution: Survey and Report." *Journal of Librarianship & Scholarly Communication* 3 (2): 1-24. http://jlsc.ubiquitypress.com/articles/10.7710/2162-3309.1225/galley/87/download/.

Coalition for Networked Information. "CNI: Coalition for Networked Information." accessed May 1, 2018 https://www.cni.org/.

Data Curation Network. “Project Timeline.” accessed May 6, 2018 https://sites.google.com/site/datacurationnetwork/timeline.

Data Curation Network. “DCN Implementation 3-Pager.” accessed May 6, 2018a, https://docs.google.com/document/d/1pDHZtMj76R37bQVxCWc3st6g2Gcr8A1nzFz7ExtOrEU/edit

DataOne. "Software Tools Catalog." https://www.dataone.org/software_tools_catalog.

Digital Curation Centre. "Tools & Services." http://www.dcc.ac.uk/resources/external/tools-services.

Erdmann, Christopher. "Data Scientist Training for Librarians." accessed April 24, 2018, http://altbibl.io/dst4l/.

Federer, Lisa. "I Heard a Researcher Give a Talk Today....", last modified March 14,

https://twitter.com/lisafederer/status/973970613403611136.
Federer, Lisa M., Ya-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* (6). doi:10.1371/journal.pone.0129506.
Federer, Lisa M., Ya-Ling Lu, and Douglas J. Joubert. 2016. "Data Literacy Training Needs of Biomedical Researchers." *Journal of the Medical Library Association* 104 (1): 52-57. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4722643/.
Goben, Abigail, and Tina Griffin. 2017. "In Aggregate: Trends, Needs, and Opportunities from Faculty Research Data Management Surveys" [presentation]. IASSIST Annual Conference, May 23-26, 2017, Lawrence, KS.
Goben, Abigail, and Megan R. Sapp Nelson. 2018. "The Data Engagement Opportunities Scaffold: Development and Implementation." Journal of eScience Librarianship 7(2): e1128. https://doi.org/10.7191/jeslib.2018.1128
Goldstein, Sarah and Sarah Oelker. 2011. "Planning for Data Curation in the Small Liberal Arts College Environment." *Sci-Tech News* 65 (3). http://jdc.jefferson.edu/scitechnews/vol65/iss3/4
Hey, Anthony J. G., Stewart Tansley, and Kristin Michele Tolle. 2009. *The Fourth Paradigm : Data-Intensive Scientific Discovery* Redmond, Wash. : Microsoft Research.
Holdren, John P. 2013*. Increasing Access to the Results of Federally Funded Scientific Research.* https://obamawhitehouse.archives.gov/blog/2016/02/22/increasing-access-results-federally-funded-science.
Human-centered Data Science Lab. "Data Science Ethnography." https://depts.washington.edu/hdsl/research/data-science-ethnography/.
Johnston, Lisa R., Jake R. Carlson, Patricia Hswe, Cynthia Hudson-Vitale, Heidi Imker, Wendy Kozlowski, Robert K. Olendorf, and Claire Stewart. 2017. "Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services." Journal of eScience Librarianship 6(1): e1102. https://doi.org/10.7191/jeslib.2017.1102
Perrier, Laure, Erik Blondal, A Patricia 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). doi:10.1371/journal.pone.0178261.
McLure, Merinda, Allison V. Level, Catherine L. Cranston, Beth Oehlerts, and Mike Culbertson. 2014. "Data Curation: A Study of Researcher Practices and Needs." (2): 139. doi:10.1353/pla.2014.0009.
Murillo, Angela P. and Kyle M. L. Jones. 2017*. The Development of an Undergraduate Data Curriculum: A Model for Maximizing Curricular Partnerships and Opportunities*. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3091834.
National Institutes of Health. 2016. "NIH-Wide Strategic Plan Fiscal Years 2016-2020." https://www.nih.gov/sites/default/files/about-nih/strategic-plan-fy2016-2020-508.pdf
National Institutes of Health. 2017. "Request for Information (RFI): Next-Generation Data Science Challenges in Health and Biomedicine."

https://grants.nih.gov/grants/guide/notice-files/NOT-LM-17-006.html
Nelson, Jerold. 1973. *Faculty Awareness and Attitudes Toward Academic Library Reference Services: A Measure of Communication.* Chicago, Ill. : College and Research Libraries.
New England Software Carpentry Library Consortium. "NESCLiC" https://nesclic.github.io/home/
Pinfield, S., A. M. Cox, and J. Smith. 2014. "Research Data Management and Libraries: Relationships, Activities, Drivers and Influences." PLoS ONE 9(12): e114734. https://doi.org/10.1371/journal.pone.0114734
Slutskaya, Sofia A., Rebecca Rose, Anne A. Salter, and Laura Masce. 2013. "Assessing Faculty Awareness of Library Services in Two Georgia Undergraduate Institutions." *Georgia Library Quarterly* 50 (1): 23-30. https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1671&context=glq.
Springer Nature. 2018. "Pricing for Research Data Support." https://www.springernature.com/gp/authors/research-data-policy/pricing/15499842
Stuart, David, Grace Baynes, Iain Hrynaszkiewicz, Katie Allin, Dan Penny, Mithu Lucraft, and Mathias Astell. 2018. "Whitepaper: Practical challenges for researchers in data sharing." https://doi.org/10.6084/m9.figshare.5971387
Tenopir, C., S. Allard, K. Douglass, A. U. Aydinoglu, L. Wu, E. Read, M. Manoff, and M. Frame. 2011. "Data Sharing by Scientists: Practices and Perceptions." *Plos One* 6 (6). https://doi.org/10.1371/journal.pone.0021101.
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): 1-24. https://doi.org/10.1371/journal.pone.0134826.
University of Massachusetts, Amherst. "Institute for Social Science Research | UMass Amherst." https://www.umass.edu/issr/.
Whitmire, A. L., M. Boock, and S. 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): 382-397. https://ir.library.oregonstate.edu/concern/articles/ht24wm21w.

Appendix A: Brown University Case Study

The bulk of data-related services offered by the Brown University Library are housed in the Library's Center for Digital Scholarship (CDS). Core CDS staff include a scientific data management librarian, a social sciences data and GIS librarian, two digital humanities librarians, and a data visualization coordinator. The scientific data management librarian focuses on helping faculty and student researchers with writing data management and sharing plans, carrying out these plans by providing data curation services, such as documenting data and depositing data in a repository, and serving as the liaison for the Brown Digital Repository (BDR). The social sciences data and GIS librarian focuses on helping faculty and student researchers with quantitative and qualitative data analysis software, geographic information systems, and with depositing and access data in ICPSR. The digital humanities librarians partner with faculty on digital scholarship projects and the data visualization coordinator supports data visualization instruction and implementation across all disciplines in addition to managing the Library's data visualization wall and makerspaces. CDS offers several research data management-related workshops on writing data management and sharing plans, best practices for managing and sharing research data, and depositing data in the Library's data repository, using the Open Science Framework, and using the LabArchives@Brown electronic laboratory notebook (ELN). Other data-related workshops taught include using OpenRefine, Regular Expressions, tools for textual analysis, and using Gephi, Tableau, and Excel for data visualization in addition to GIS tools.

In 2017 Brown was awarded a $1.5 million grant to establish a data science research institute. The award is one of 12 nationwide Transdisciplinary Research in Principles of Data Science (TRIPODS) grants announced by the National Science Foundation. Brown's Institute will focus on the foundations of model-driven discovery from massive data. In the fall, Brown officially launched its Data Science Initiative (DSI) which will manage the institute. The launch involved inviting departments and units on campus that were willing to host a roundtable at an event targeted at faculty and students. The aim was to have a conversation with attendees of the launch who chose to come to sit at the roundtables and to listen to these researchers about their needs and how units could support them. The Library was invited to participate in this launch and hosted two roundtables, one focused on library data services and the other on digital humanities. The needs expressed by attendees of the Library's data services table can be categorized into two themes. The first theme was workshops. Attendees recognized that the Library was already heavily involved in holding workshops on research data management and various research skills and tools, and they felt that it should develop a

data science series to teach introductory-level workshops on tools relevant to data extraction, wrangling, analysis and visualization to student and faculty researchers. The second theme was support of their own teaching. Two faculty attendees expressed that they teach non-computer science courses, but their courses were designed to emphasize solving problems with computers and to engage with their subject matter, but not to dwell on the technical aspects of computing. Their course assignments required students to learn Python. The issue, they explained, was that students often needed a “module 0”-- that is extra help installing the correct version of Python and downloading the relevant libraries on their laptops using whichever operating system, and then basics of using command line and Jupyter notebooks to teach basics of creating simple programs. As more courses begin to incorporate Python or R, these faculty saw a role for the library in helping students who are non-computer science concentrators and who have less of a programming background with these module 0s. The last one was visualization support for grants. The Library manages the VIVO instance Researchers@Brown and therefore has access to publication data. Attendees were interested in exploring how they could extract these data and visualize citation networks, cross-campus and external collaborations, as well as areas of research.

In December the Library partnered with Brown’s Instructional Technology Group (ITG) to work with a Cognitive Sciences professor who was teaching a course on deep learning from the perspective of both machine learning and neuroscience. As highlighted at the DSI launch, this faculty member was interested in attracting students to the course who were interested in the subject matter and did not want students to be intimidated by the computational requirements, such as learning Python. The Library and ITG organized a 2-hour bootcamp for students to meet with programmer volunteers from the Library’s Digital Technology group and Computer and Information Services (CIS) to get hands-on help with installing the needed version of Python, solve any issues related to their operating system, set up their Jupyter notebooks, and work through basics of working with the command line and writing simple programs. In January, the Library and ITG will partner with a Physics professor to offer the same module 0 for his students. In response to these researchers needs, the Brown Library has joined the New England Software Carpentry Library Consortium (NESCLIC) to have two staff trained as carpentry trainers who could then help train other library staff and help build capacity. The challenges that CDS staff are facing include concerns that they have little work time to learn and practice data science skills that are in demand by researchers while juggling their other responsibilities as well as the scaling of in-person bootcamps if more faculty were to begin requesting these for their courses.

The Brown Library is just one of many units on campus who also are interested in supporting faculty and student researchers' data science skills. The Center for Computation and Visualization (CCV) offers a workshop on Unix and using the campus high performance computing system, Oscar. The Brown Center for Biomedical Informatics (BCBI) offers researchers workshops on RedCap, biostatistics and data analysis, and programming in R and Julia, as well as ones on natural language processing and machine learning for researchers that are interested in working with electronic health record data. The Data Science Practice Group is a new unit of campus IT comprised of data scientists who can be funded by a faculty grant to work on data science-related aspects of the project, such as visualization or mining large amounts of data. In addition, several general data science courses have been developed, including "What's the Big Deal About Data Science", "Data Fluency for All", and the less flashy "Introduction to Data Science" and "Applied Data Analysis", which means that the pool of data-savvy students that may be interested in a student work position in the Library's CDS is growing. Students who have completed one of these courses have taken student jobs in the Library to work on a variety of digital projects for faculty who do not have the grant resources to hire a data scientist from the data science practice group or not large enough of a project to warrant it, for example one student helped a historian create an online indigenous slavery database based on his corpus of data and another helped a public humanities professor create an online visualization of an 1853 international trade exhibition in New York's Crystal Palace based on digitized exhibition documents.

Appendix C: Assessing the needs of a research facility: a simplified DMP for Mount Holyoke College's Microscopy Facility(MF)

1. Types of data

a. What types of data will you be creating or capturing? (experimental measures, observational or qualitative, model simulation, existing)

Images and some video from microscopy observations:

- Digital monochrome and pseudo-colored Z-stack images of various materials.
- Materials likely to be imaged include fixed, whole *Drosophila* brain and brain sections, fluorescently dyed colloidal silica suspensions in microfluidic channels, live Jurkat cells, live fat body cells in *Drosophila* pupae, and fixed *Mus* brain slices.

b. How will you capture, create, and/or process the data? (Identify instruments, software, imaging, etc. used)

Electron Microscopes:

- Philips CM100 Transmission electron microscope (TEM) capable of acceleration voltages from 40 to 100 kV, using an AMT BiosprintM ActiVu 16megapixel digital camera for image acquisition, with film camera imaging an alternative option.
- FEI Quanta 200 Scanning electron microscope (SEM) with EDAX Genesis X-Ray microanalysis (EDS), capable of acceleration voltages between 200 V and 30 kV.

Light Microscopes:

- Nikon TE2000 inverted microscope with Roper Coolsnap HQ camera and NIS Elements AR imaging software (fluorescence, phase, DIC modes).
- Nikon 50Ti upright microscopes (2) with PixeLink monochrome camera and Q-Cam w/ Micro-Manager imaging software (fluorescence, phase modes).
- Olympus BX41 upright microscope with PixeLink monochrome camera (fluorescence, phase modes).
- Olympus BX51 reflected light microscope with DP70 camera (brightfield, darkfield, transmitted light, reflected light modes).
- Nikon SMZ1500 stereo microscope with fluorescence illuminator and Spot camera (brightfield, darkfield, fluorescence modes).

Atomic Force Microscopes (AFM):

- Veeco Dimension.
- Innova.

Software in use:

- On Nikon microscopes, Q-Cam Micro-Manager imaging software on some Nikon microscopes.
- ImageJ/Fiji open source image processing software including Nikon ND Reader plugin.
- Camtasia screencasting software.

Computing equipment: 4-6 PCs connected to the instruments at any one time.

2. Contextual Details (Metadata) Needed to Make Data Meaningful to others

a. What file formats and naming conventions will you be using?

Formats:

- TIFF files.
- Nikon .nd2 image format files.
- Video from Camtasia software.

TIFF is widely readable in a variety of image software, and Nikon .nd2 files can be read with the open source imaging platform ImageJ/Fiji, which supports the plugin "Nikon ND Reader" that allows users to open .nd2 files using ImageJ/Fiji. ImageJ/Fiji plugin "Exif Reader" is also available to read Exchangeable Image File data associated with TIFF and .nd2 images. Both TIFF and .nd2 support multiple pages and the sorts of z-stack images common in microscopy.

Files are generally named with semester and date information, and there is an excel sheet in the lab notebook for logging the images taken. The lab does not currently maintain a database of metadata associated with these images and does not routinely convert proprietary .nd2 files to TIFF or other more widely used format. Both a separate metadata database and automated production of TIFF or other conversions for .nd2 file formats would be steps to consider for improving long term accessibility of MF files.

3. Storage, Backup and Security

a. Where and on what media will you store the data?

PCs have between 500GB and 1TB drives on each machine for storing images locally at time of observation. There is also a large external hard drive which the Director uses to back up microscopy data from individual machines periodically. Since early 2017 the computers have all been logged into a lab specific google account that replicates materials saved in the linked Google Drive space on each machine's hard drive to folders in a shared google drive account. Students are encouraged to save images they take to their own Google Drive spaces as they work.

b. What is your backup plan for the data?

The MF Director manually backs each individual computer's collected images tp to 1 external hard drive regularly. Director encourages students to save copies of all images they take in the course of classwork or research to their own campus Google Drive spaces. Since early 2017 the computers have all been logged into a lab specific google account that replicates materials saved in the linked Google Drive space on each machine's hard drive to folders in a shared google drive account, and instructions have been written instructing center users in saving data to these spaces as they take images.

c. How will you manage data security?

Only the MF Director and certain Library, Information, & Technology services personnel have the password for the MF google drive account, and machines and external hard drives are in the locked facility to which only research and coursework users have swipecard access. Students save their own copies of images they create to their own google drive spaces or other storage.

4. Provisions for Protection/Privacy

a. How are you addressing any ethical or privacy issues (IRB, anonymization of data)?

At this time no data is expected to fall under IRB provisions or require anonymization. Current digital imaging ethics will be followed to ensure accurate representation of data.

b. Who will own any copyright or intellectual property rights to the data?

Images are the property of the students, faculty and staff who took them. Data will only be shared with the permission from the PIs and co-PIs who use the Facility. Individuals are expected to notify the College's Sponsored Research Office of any Invention that the individual has made using funds or facilities provided or administered, in whole or in part, by the College. The Dean of Faculty and the Vice President for Finance and Administration will determine whether the Invention is within the scope of the College's patent policies.

5. Policies for re-use

a. What restrictions need to be placed on re-use of your data?

Permission to reuse images should be obtained from the PI of the research for which the image was taken, or from the student who took the image during the completion of coursework.

6. Policies for access and sharing

a. What is the process for gaining access to your data?

The MF director can assist researchers in getting in touch with the PIs on a project and with students who took images during the completion of coursework, if they would like to request permission for image reuse.

7. Plan for archiving and preservation of access

a. What is your long-term plan for preservation and maintenance of the data?

The MF Director and Library, Information, & Technology Services (LITS) staffers, particularly in the Digital Access and Preservation Services group, have begun working to develop both network-attached storage in the MF that can be mirrored to network-attached storage in a College data center, and have begun discussions for how some of this data might be stored in IDA, the College's institutional repository, for a long term archive.

Discussion:

A difficulty of writing a plan like this, even a simplified one, is deciding where to stop. Our purpose in compiling this simplified data management plan was to get an overall view of the environment in which data needs to be managed. When writing this, it became clear that it would, if we allowed it, turn into an extremely deep dive on the image metadata. So, because the goal was capturing the overall environment first, we made the decision to stop short of writing up all of the metadata that we could apply to microscopy images from the light microscopes. An initial, simplified plan like this leaves room to follow up by writing up a full metadata schema to capture all of the modes, optical strengths and other settings of the machines and attributes of the resultant images.

Stopping *before* the metadata deep dive stage on a data management plan also provides the librarian and researcher the chance to stop and look at existing standards that might be applied to the data. This is particularly useful in cases where the researcher is now to data curation, and in cases where the librarian is new to working with a particular type of data. In this case, resources such as the Open Microscopy

Environment format, and the Bio-Formats image reading and writing package, would be some examples of tools to investigate to improve metadata capturing and interoperability of the data in use.

Other reasons to do plans like this are to quickly diagnose urgent issues, such as a lack of "good enough" storage protocols, or to help facilities during staff changes. Librarians can intervene and suggest helpful tools or processes to adopt quickly, even if a deeper examination of the metadata has not yet been done. This sort of plan will also be useful during staff transitions, and to build trust in library RDM support across an institution.