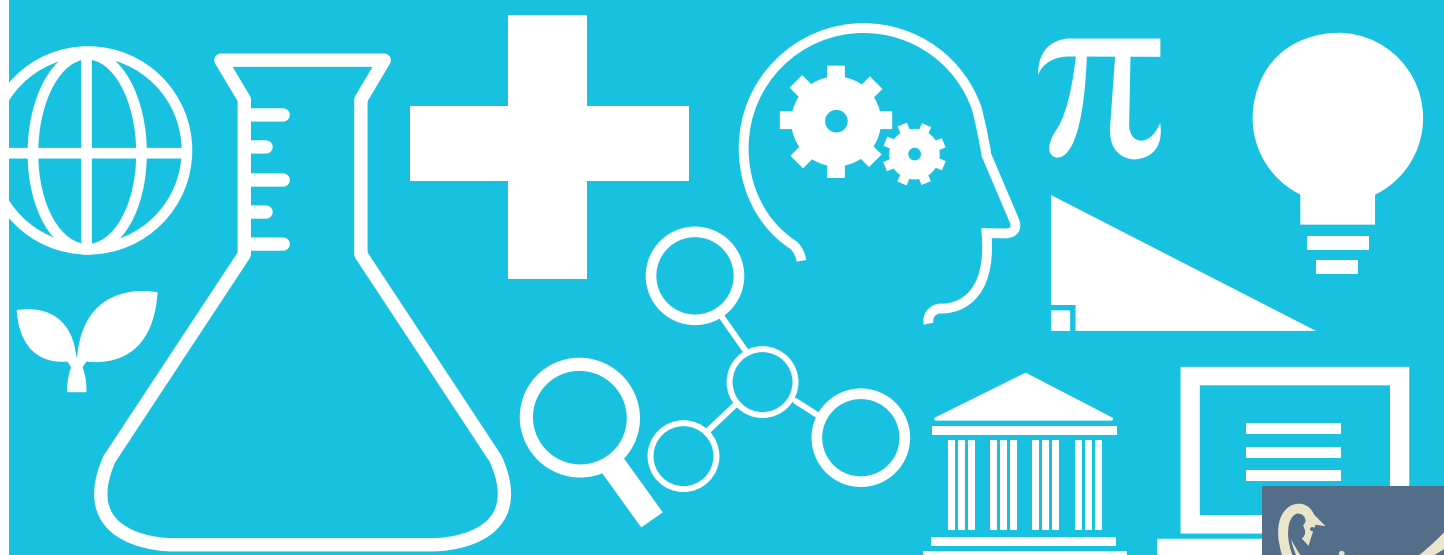


ISSUE BRIEF

Research Data Management Roles for Libraries

October 22, 2015
Neil Rambo



ITHAKA S+R



Ithaka S+R is a strategic consulting and research service provided by ITHAKA, a not-for-profit organization dedicated to helping the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways. Ithaka S+R focuses on the transformation of scholarship and teaching in an online environment, with the goal of identifying the critical issues facing our community and acting as a catalyst for change. JSTOR, a research and learning platform, and Portico, a digital preservation service, are also part of ITHAKA.

Copyright 2015 ITHAKA. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of the license, please see <http://creativecommons.org/licenses/by-nc/4.0/>.

ITHAKA is interested in disseminating this brief as widely as possible. Please contact us with any questions about using the brief: research@ithaka.org.

Background: the emerging role of data management in research libraries

I first became aware of research data management as a frontier area of expertise for libraries and librarians almost 10 years ago. Tony Hey was one of the first to popularize the term ‘e-science’ and the idea that librarians had a role to play in managing research data.¹ This call might have stirred little interest at another time. But at least two things were happening around then that might have caused this to stand out and catch interest: 1) libraries were in the midst of redefining their roles and place in the digital scholarly communication ecosystem; and, 2) the ‘data deluge’ made possible by the ubiquity and power of computation and networks was beginning to overwhelm traditional methods of data storage and management. It was becoming alarmingly clear that a new approach was needed to grapple with the burgeoning need.

This convergence of need and opportunity was not lost on the research library community. Not long after becoming interested in research data management as a role for libraries and librarians, I had the privilege of working with the Association of Research Libraries (ARL) in that community’s early efforts to explore the potential of this area.

One of the products of that work was a report and recommendations from the ARL Joint Task Force on Library Support for e-Science.² Eight years on, I’m pleased to find that the report has held up reasonably well. It does so because it took an expansive view of the environment of scientific research and the effects of ‘e-science’ on it. It also steered clear of a proscriptive approach to what research libraries should do. Instead, it laid out a broad approach designed to enable ARL and its members to remain responsive to the needs of the scientific research community. Out of that, research data management is perhaps the most specific and concrete instance that has continued to develop.

Over several years now, many libraries supporting science researchers have developed data management services. An informational and consultative approach requires a modest investment of resources. This includes developing user guides on the topic, references to source material for more information, advising on funding agency requirements for developing a data management plan, and advising on publisher requirements for data sharing. Another approach is to engage more directly in several

¹ Hey, Tony and Hey, Jessie. “e-Science and Its Implications for the Library Community.” *Library Hi Tech*, Vol.24, No. 4 (2006) p.515-528.

² *Agenda for Developing E-Science in Research Libraries Final Report and Recommendations*. Prepared by the Joint Task Force on Library Support for E-Science. Association of Research Libraries. November, 2007.

aspects of managing data. This requires a more significant investment in staff and resources, and, fundamentally, in attempting to engage with the basic science research community. These direct services may include promoting best practices in data management, identifying appropriate metadata standards, managing metadata (description and organization of data), advising on file organization and naming, data citation, and data sharing and access. Carol Tenopir characterizes these approaches as informational and technical.³

I will describe the development of research data management services at my own institution and library and focus on that for the remainder of this piece. In the process, I will describe the environment that helped shape this development and why we made the choices we did.

Case study: developing research data management services

Environment of NYUHSL

The Health Sciences Library (HSL) at New York University is both an academic department in the School of Medicine and a shared service across the NYU Langone Medical Center (medical school, hospitals, residency programs, clinical network, and faculty group practices) and the university as a whole, including health professions education programs and several colleges. The HSL coordinates closely with the university library on the main campus on shared services (e.g., the library catalog) and on similar services that we provide to our various user communities (e.g., research data management). At the time of writing, the HSL has about 18 librarians and about the same number of staff. It has a dedicated IT support unit that is both part of the HSL and part of the medical center central IT organization.

Other key characteristics of the environment the HSL operates in include the NYU-HHC (NYC Health and Hospitals Corporation) Clinical and Translational Science Institute (CTSI), one of many such centers across the county funded by NIH, which is a locus of interest in fostering progressive clinical data management. Also, the medical center, like many academic medical centers, is developing an increasing appreciation of the need for active data governance and, from that, data management enabling higher quality and more usable data across the missions of patient care, research, and education. Over the

³ Tenopir, Carol, Robert J. Sandusky, Suzie Allard, and Ben Birch. "Research Data Management Services in Academic Research Libraries and Perceptions of Librarians." *Library & Information Science Research*. Vol. 36, No. 2 (2014) p. 84–90. [doi:10.1016/j.lisr.2013.11.003](https://doi.org/10.1016/j.lisr.2013.11.003)

last few years, the HSL has become a recognized player in this data management environment. How did this come about?

I arrived at NYU in late 2010. Up to that point, the medical library had operated more or less independently of the university library with no formal connection. My arrival coincided with the beginning of a more collaborative relationship between the two organizations. As medical library director, I report to the Dean of Medicine with a dotted line reporting relationship to the Dean of Libraries. Early on in this new relationship, Carol Mandel, Dean of the Division of Libraries, and I identified research data management as a strategic area ripe for development and deep collaboration. We jumped at an opportunity then on the horizon to participate in the inaugural offering of the ARL eScience Institute.⁴ I led the NYU team that included representation from Washington Square (main campus) and the medical center.

Two elements of our participation in the ARL institute were key. One was a series of interviews several team members conducted across NYU to gauge the state of data management. We asked researchers, administrators, and IT professionals what services were offered and what they thought about the role of the library in support of research data management. The other, and final, Institute exercise was to conduct a SWOT analysis of the institution.

...most stakeholders did not think of the library as providing
eScience, or data management, services

We learned from these exercises that most stakeholders did not think of the library as providing eScience, or data management, services. We also learned that not many services were offered in this area. Where there were services, it was largely piecemeal and limited. We concluded from these investigations that there was great opportunity and also significant barriers to overcome, especially regarding the perceptions of the library.

We decided at this point that it made sense to pursue this opportunity on separate tracks: the university library would focus on the main campus and the medical library would focus on the medical center. We agreed to keep each other informed along the way and work together when that was practical. The infrastructures, needs, and pressures

⁴ See ARL E-Science Institute, <http://www.arl.org/focus-areas/e-research/e-science-institute#.ViQrRKKZ4QQ>.

inherent in each environment were different enough that treating it as one problem would prevent precise targeting of services and hinder progress.

And so we set forth at the Health Sciences Library to develop services and solve research data management problems. But which ones? We weren't sure how to start. We were sure that we wanted to start by addressing the needs of researchers and let our own agenda develop in response. The interviews gave us some baseline information on the needs of researchers in this area but we needed a more nuanced understanding to be confident that we knew how to proceed and commit to offering services.

Teaching

Researchers

At the time, our nascent data services team consisted of two librarians. One had previous experience as a basic science researcher and programmer, and the other had experience in programming, database development, and digital archives [Alisa Surkis and Karen Hanson]. In 2012, they jointly developed a standalone introductory data management class for postdoctoral fellows (postdocs). The Postdoctoral Office saw the need for the very practical nature of the class in the lab management course series they offered and that allowed us an easy “in” without having to convince skeptical researchers that the library had anything to offer them about managing their own data. The class was built around core best practices from trusted sources and included real stories of poor data management leading to dire consequences such as retraction and loss of data.⁵ The librarian instructors also developed a short video demonstrating some of the issues that can arise in this area, *Data Sharing and Management Snafu in 3 Short Acts*.⁶ One of the successes that grew out of this class, the video since took on a life of its own. It is viewable on YouTube [<https://www.youtube.com/watch?v=N2zK3sAtr-4>] and has been used in presentations and classes around the world.

The class was well received. Work soon began to revise it to include recent changes in publisher requirements about data sharing and to favor researcher-centric terminology rather than library-centric as much as possible (e.g., using data description and documentation rather than metadata). Because of the particularly disruptive

⁵ See Christine Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet* (Cambridge, MA: MIT Press, 2007). See also Salo, Dorothea. “Innkeeper at the Roach Motel.” *Library Trends*. Vol. 36. No. 2 (2008) p. 98-123. DOI: [10.1353/lib.0.0031](https://doi.org/10.1353/lib.0.0031).

⁶ Hanson, Karen, Alisa Surkis, and Karen Yacobucci. “Data Sharing and Management Snafu in Three Short Acts.” *YouTube*. December 19, 2013. <https://www.youtube.com/watch?v=N2zK3sAtr-4>.

intervention of Hurricane Sandy on our library and the medical center, and also because of a change in staff (one of our data services librarians left and a new one arrived) we were not able to offer the data management class to postdocs again until 2014. The 2014 class was also rated highly. Subsequent revisions have continued to incorporate more relevant examples of items getting attention in the research community; for example, experimental reproducibility and introducing discipline-specific repositories and the opportunities they present for increased collaboration within specific communities.

Although the audience for this class had been limited to postdocs, word spread that the content was extremely useful and not something that was taught anywhere else in the curriculum. The research dean suggested that all biosciences graduate students should benefit from this. This is an opportunity to move away from the standalone, one off library classes to incorporation in the required curriculum. At the time of writing, we are developing a proposal for a one credit, 12-week course, focusing on data management, to be debuted in spring 2016.

Librarians

Our data services team was invited by the Medical Library Association Continuing Education Committee to propose a data management workshop to be presented at the annual meeting. Based in large part on the postdoc class, this was first presented at the 2014 annual meeting and was again offered in 2015.

Part of the motivation for teaching this subject to medical librarians was that, while the interest of the community was evident, seemingly little was actually being done to provide data management services. As our team includes librarians with significant biomedical research experience, they were in a position to impart some of this knowledge to others who did not have the advantage of this experience.

In fall 2013 we were fortunate to entice a librarian Fellow from the National Library of Medicine (NLM) to spend the 2nd year of his fellowship working with us at NYU [Kevin Read]. He spent the first year of his fellowship onsite at NLM/NIH where he worked on the NIH Big Data to Knowledge (BD2K) initiative, particularly the Data Discovery Index.⁷ He was able to continue to work on portions of these projects once at NYU and to leverage his expertise in these areas to work we were just beginning.

⁷ See Data Science at NIH. "Big Data to Knowledge (BD2K)." <https://datascience.nih.gov/bd2k>

NLM Informationist Administrative Supplements

In 2012, NLM announced the first round of informationist administrative supplement awards.⁸ These awards were designed to fund a librarian (informationist) to work with a research team to provide information and data management support. The award is a supplement to an active NIH research grant. The funding mechanism requires a librarian to collaborate with the principal investigator of a grant to develop a compelling proposal. We saw this as an opportunity to gain deep knowledge and experience with research support in general and research data management in particular. Even though the funding is modest (up to \$25,000 per year for two years) the legitimacy conferred by NIH support is significant in the academic medical center context.

It has been described elsewhere how we choose particular research projects to collaborate with⁹ and some of the initial findings of this work.¹⁰ We decided to develop two proposals and, to our surprise, both were funded. One involved providing information process management support in bioinformatics and turned out to be fairly straightforward to accomplish. The other was a much more in-depth data management project that turned out to hold difficulties that were not anticipated by the library team or the principal investigator. The state of the existing data was worse and more intractable than had been understood and the eventual level of effort needed to work through to a partial solution was greater than expected. This project did provide a cautionary tale of scalability of a service (e.g., how much database development work, for example, is the library willing and able to commit to?) and the potential for problems to emerge after the initial commitment. It did provide a positive example of the library working through issues of commitment and negotiating a way forward, and providing substantial data management service to a research team.

A second round of informationist administrative supplements was announced in 2014. We again saw this as a valuable opportunity. We again selected two research projects,

⁸ See NLM Administrative Supplements for Informationist Services in NIH-funded research projects, <https://www.nlm.nih.gov/ep/InfoSplmnts.html>.

⁹ Williams, Jeff D., and Neil H. Rambo. "An extensible and successful method of identifying collaborators for National Library of Medicine informationist projects." *Journal of the Medical Library Association*. Vol. 103, No. 3 (2015) p.145-147. DOI: [10.3163/1536-5050.103.3.008](https://doi.org/10.3163/1536-5050.103.3.008)

¹⁰ Hanson, Karen, Theodora A. Bakker, Mario A. Svirsky, Arlene C. Neuman, and Neil Rambo. "Informationist Role: Clinical Data Management in Auditory Research." *Journal of eScience Librarianship*. Vol. 2, No. 1 (2013). DOI: [10.7191/jeslib.2013.1030](https://doi.org/10.7191/jeslib.2013.1030). Surkis, Alisa, Aileen McCrillis, Richard McGowan, Jeffrey Williams, Brian L. Schmidt, Markus Hardt, and Neil Rambo. "Informationist Support for a Study of the Role of Proteases and Peptides in Cancer Pain." *Journal of eScience Librarianship*. Vol. 2, No. 1 (2013). DOI:[10.7191/jeslib.2013.1029](https://doi.org/10.7191/jeslib.2013.1029)

using the same methodology as before, and again, we were surprised when both projects were funded. Both projects are still in progress at the time of writing. One project involves managing screening data collected from multiple community settings, including developing a data model and data dictionary and a web-based input tool. Once again, we committed to doing more than was practical. That, combined with personnel changes in the early phases of the project, forced us to renegotiate some of the parameters of the work to be done. We were able to bring the focus of our task back to data management and away from a more complex IT project.

The other project is with a neurophysiology laboratory that collects a variety of data types and is often responding to requests from other researchers for data. The aims we committed to working on with this group of researchers were varied but essentially involved improving multiple workflow and data organization issues, including lack of metadata uniformity, data stored in multiple locations and formats, and lack of identifier standardization. All of this makes for a tangle of interwoven issues, but all of it is squarely in the realm of data management, and, therefore, ultimately manageable.

For a host of reasons beyond the scope of this piece, our reach in two of the four projects described here was overly ambitious. In each case, we had to regroup and reassess what we could commit to and deliver. It forced us to be clear about distinguishing between data management problems and IT development projects. It also gave us experience in testing what scale of issues we could take on and reasonably affect within the limitations of the project period, as well as the overriding issue of sustainability. A more experienced project management team would most likely not have required this difficult testing period with a few false starts and restarts. Nevertheless, our team is now tested and confident in what we can do and what makes sense for us to commit to. Our interventions are benefiting the progress of research on several fronts.

Data interviews

To inform the development of research data management services, we decided to interview a variety of researchers to better understand their practices, needs, and challenges regarding data management. As described in “Starting the Data Conversation,” we selected a sample of active researchers with current grants that included a mix of basic science and clinical researchers, and a mix of research experience.¹¹ Interview questions were developed based on previous studies evaluating

¹¹ Read, Kevin. “Starting the Data Conversation: Using Interviews to Inform Data Services at the NYU Health Sciences Library.” Presented at the National Library of Medicine Associate Fellowship Colloquium. National Library of Medicine. June 3, 2014, Bethesda, MD.

data-related challenges and needs of researchers. In approaching these interviews, we felt it important to maintain a steady focus on the researcher and take the library out of the equation as much as possible. We avoided library jargon and each interviewer familiarized her-/himself with the basics of the research conducted by the interviewee. We wanted to elicit information that would better enable us to design services to fit in the researcher's workflow, rather than the researcher attempting to understand or fit into ours.

We wanted to elicit information that would better enable us to design services to fit in the researcher's workflow, rather than the researcher attempting to understand or fit into ours.

In all, we interviewed 11 basic scientists and 19 clinical researchers. Although not a large number, we felt we had reached a saturation point because the same themes and responses began to be repeated. The primary finding from these interviews is that the practices and challenges differ significantly between basic science and clinical research.

Among basic science researchers there was generally a perceived lack of standards and procedures, leading to developing their own custom collection and storage methods and reinventing the wheel. There are problems with disparate types of data, e.g., image and numerical data stored in different places and analyzed using different tools. Post-docs and graduate students come and go and data and knowledge of methodologies used to collect and store data may go with them.

Clinical researchers often deal with variable quality of data resulting from multiple people involved in a study with inconsistent interpretations of variables (e.g., lb vs kg) and inconsistent data collection methods. They have difficulties transferring data from one format to another. Moving data from one statistical analysis platform to another can result in data degradation or loss.

The basic scientists interviewed were not keen on sharing their data broadly. Some of this may be unease at losing control over who accesses it and for what purpose. It also may reflect skepticism that someone not immersed in his or her research would understand it or be interested. Clinical researchers, on the other hand, presumably see a more direct connection between their work and benefits to human health and are

therefore more predisposed to sharing their data beyond those engaged in similar research.

Through the interviews we gained insights into the data management challenges of the medical center research community and a deeper awareness of issues around collecting, organizing, and sharing their research. What we learned from this process feeds directly into the development and planning of improved library data services. Specifically, that population health researchers are interested in and willing to share their data and find other datasets that they can use for their research provided a basis for the Library to build a data catalog to first address the needs of these likely and motivated users. Also, the difficulties that basic science researchers have in organizing data led to the development of a low bar lab organization tool that is being piloted at the time of writing. Each of these is described below.

Data catalog

The need for a catalog of datasets emerged from discussions with key members of the CTSI, which included the library liaison to the CTSI, our translational science librarian, who is a core member of our data services team. Population health researchers needed to know about large external datasets that are heavily used for their areas of study. Researchers in other areas seconded this need. There was a concern that individuals and departments may be paying license fees several times for the same access. This need led to the idea of designing and building a discovery tool that would describe datasets, their uses, and how to access and use them. Our data interviews that were being conducted in parallel also pointed to the need for a discovery tool. The interviews are described in more detail below.¹² The early phase of this catalog effort (2013) became the basis for a close collaboration between the library and the Department of Population Health.

The initial scope was to provide relatively standardized descriptions of commonly used external datasets, based in large part on the inventory created by the University of California San Francisco CTSI.¹³ The library team built a preliminary version of this and previewed this with population health researchers to get feedback on further development before making it available.

¹² Read, Kevin, Jessica Athens, Ian Lamb, Joey Nicholson, Sushan Chin, Junchuan Xu, Neil Rambo, and Alisa Surkis, "Promoting Data Reuse and Collaboration at an Academic Medical Center." *International Journal of Data Curation*. Vol. 10, No. 1 (2015) p. 260-267. DOI:10.2218/ijdc.v10i1.366

¹³ See UCSF Clinical and Translational Science Institute. <https://ctsi.ucsf.edu/>.

The next phase of catalog development focused on expanding and improving the metadata and began to accommodate the idea of including internally generated datasets as well as external, something which had been in the background from the beginning. This was informed in part by findings from the data interviews that were being conducted in parallel. The interviews helped guide an overall strategy for subsequent catalog development. Although intended to be more than a population health data catalog, it made sense to develop the catalog from that area of strength and build outwards to related areas before expanding to more diverse and disparate areas of research. The strategy would be to focus on datasets, external as well as internal, of potential interest to the broad range of population health areas of interest. This meant that we would not focus on recruiting basic science researchers for their datasets at this stage because there would be little value to anyone of a few datasets that didn't relate to the rest of the catalog corpus in any meaningful way. There would be no synergy between these entries and the bulk of the content. Developing a viable body of content in the basic science area would, we think, come later.

Two areas worth noting in the development of the data catalog are decisions regarding metadata and local experts. As referred to above, one of the members of our data services team had experience developing metadata schemas for scientific research that greatly informed and influenced the data catalog. The schemas included in the data catalog were the NIH Common Metadata Elements,¹⁴ DataCite, Dryad, Nature's Scientific Data, and the W3C's Data Catalog Vocabulary. At several points, we favored pragmatism over purity in adjusting metadata applications that would maximize use over schematic consistency. Some metadata elements were of more use for external datasets and others for internal.

At several points, we favored pragmatism over purity in adjusting metadata applications that would maximize use over schematic consistency.

We decided to include local experts for each external dataset as an added value for potential users of a dataset. Local experts agree to be a contact for anyone wanting information on what the dataset includes and how to access and use it. Identifying local

¹⁴ Read, Kevin, Jessica Athens, Ian Lamb, Joey Nicholson, Sushan Chin, Junchuan Xu, Neil Rambo, and Alisa Surkis, "Promoting Data Reuse and Collaboration at an Academic Medical Center." *International Journal of Data Curation*. Vol. 10, No. 1 (2015) p. 260-267. [DOI:10.2218/ijdc.v10i1.366](https://doi.org/10.2218/ijdc.v10i1.366)

experts is labor intensive and will require ongoing maintenance but we think it is worth doing. It is an added resource for users, it offloads some of this role from the librarian, and it provides an element of buy-in by those researchers agreeing to be identified as local experts. The researcher generating the data will be the local expert for internal datasets unless otherwise noted.

At the time of writing, the data catalog is available through the Library website and is being viewed by a growing number of users across the institution. It includes more than 70 external datasets and internal datasets are beginning to be added. Outreach to invite additional datasets to include has begun.

Data catalog growth and expansion are likely to develop from collaborations with clinical research data services and with the university library. DataCore is the centralized clinical research data services unit for the medical center. DataCore is planning to use the data catalog as the discovery layer for analysis datasets generated from clinical research or drawn from our electronic health record (EHR). We are also discussing with data librarians at the main campus plans to ingest datasets in broadly related areas of research across NYU.

The data catalog is fundamentally an iterative effort, based in the library, involving multiple collaborations across research communities.

The data catalog is fundamentally an iterative effort, based in the library, involving multiple collaborations across research communities. We have invested a large effort in planning, developing metadata schemas, generating researcher interest and institutional support, and building it. We have been assiduous in making sure everything we do in developing the catalog is based on user input and is geared toward maximizing its utility. It will stand or fall on whether it delivers sufficient value to individual researchers and demonstrably facilitate research productivity and interdisciplinarity.

Lab organization tool

As mentioned above, a theme that emerged from the data interviews was many basic science researchers had minimal or inadequate methods for organizing data in their lab environments. This was a potential opportunity for the library but there were some daunting barriers: scalability and the likelihood of being met with considerable

resistance. If we were to provide a tool or service to address this need, it would have to be lightweight and require very little time and effort to implement and maintain. We felt there would be demonstrable value to a PI to have a central record of which researcher in a lab had done what experiment, when, and where the data was stored. With that as a start, another level would add basic metadata elements specific to the experiment and protocol.¹⁵

We developed a prototype tool using the REDCap platform¹⁶ and it is being piloted in two neuroscience labs. Neuroscience was chosen because one of our data management librarians (A. Surkis), who first developed the idea for the lab organization tool, has research experience in that field, and we are working closely with one of the labs on an informationist project (as described more fully above). We hope these pilots will answer many questions about the utility and suitability of the lab organization tool and how to promote its eventual uptake. We will be paying close attention to how specific the tool needs to be to a particular field of research and what the limits of its generalizability are.

Discussion

Our development of data management services has been ad hoc and opportunistic. We did not begin with a plan as to how this was going to develop. We started by getting out there and finding out what the needs are and where the opportunities may lie in the midst of researchers' needs, limited resources, and skepticism that we had much to offer. We combined that with what expertise we had and we took some calculated risks. We approached the risks with confidence based on careful and, in some cases, hard-won preparation, hard work on the part of all of our team, and some (who knows how much?) luck.

Fundamentally, our approach has been based on connection with the target community: researchers in an academic medical center. We developed that connection through a few years of establishing and growing liaison relationships with academic departments, centers, and programs. The informationist projects provided an opportunity for a deep, and sometimes uneasy, exposure to vagaries of research workflows as they are being conducted. We were able to conduct a series of fairly systematic interviews of many researchers about their data practices. Importantly, we did this as much as possible from the perspective of the researchers themselves and what we thought might constitute a reasonable return on the researcher's time to talk to us and allow our brief intervention.

¹⁵ Read, Kevin, and Alisa Surkis. "Building Data Management Services at an Academic Medical Center: An Entrepreneurial Approach," in *Data Management in Practice*. Lanham: Rowman & Littlefield Publishers (forthcoming).

¹⁶ See Research Electronic Data Capture (REDCap), <http://project-redcap.org/>.

It's such a simple idea but it is fundamental to our progress to this point: understand the researcher's needs, pressures, and environment and assess what you do with them and for them in that light. Is it going to solve a problem that matters to them? Are you going to save them time or make something they have to do easier? If not, don't do it.

...understand the researcher's needs, pressures, and environment and assess what you do with them and for them in that light. Is it going to solve a problem that matters to them? Are you going to save them time or make something they have to do easier? If not, don't do it.

If you gain a level of trust based on the hard work of outreach and understanding, then you can afford to demonstrate the knowledge you have (and they don't) about data standards, principles and techniques of organization, and data sharing.

The other side of the cost/benefit consideration – of making sure there is sufficient value to the researcher in what you are asking them to do – is making sure it makes sense and is of value to the library in the long run. Does the proposed service fit well in the library's portfolio or is it an awkward stretch? Is it more likely to be a one-off, or series of one-off instances, or is there a practical path toward sustainability? If sustainability is too high a bar to attain at this point, is a pilot project worth the knowledge and experience that could be gained from it, in which case sustainability doesn't have to be the only measure of success?

Although our approach to the development of data management services has not spooled out according to a plan, our initial foray has begun to coalesce into a set of services that have gained traction across the medical center. This, then, becomes part of the library's strategy and identity. It begins to make sense and look as if we planned it. At first, we were eager to take on a problem and, through service, try to run with it. Later, we began to realize that not every data management problem was something it made sense for the library to try to solve (and, many were problems we couldn't solve).

Once you have launched data management as a recognizable library service, you can begin to shape and focus that effort in such a way as to build and support the library's identity and value to the institution it serves. We think we are at that point with our data

consultation and education services, the data catalog, and, possibly, soon, with the lab organization tool. We may be able to add to this portfolio in time.

Concluding thoughts and future directions

I have described the development of data management services at one academic health sciences library. Although that development has been piecemeal and probably looked haphazard, we followed a path based on the best information we could get about local needs and opportunities. A question is increasingly being asked about which academic or research libraries should develop data management services? To start, those supporting a research mission should consider it. Beyond that, there are many factors that may influence this or indicate a more or less favorable environment for such development. Maybe another campus unit is already providing sufficient support. It wouldn't make sense just to duplicate that or to compete with an existing service. It might be an opportunity, though, to learn more about how this works and whether there are supplemental or other roles that the library could reasonably take on.

...in order to develop new, untried services, the library has to have access to sufficient resources (primarily in the form of people) to dedicate to this new effort. It also has to have a degree of flexibility and autonomy to take this on.

Also, in order to develop new, untried services, the library has to have access to sufficient resources (primarily in the form of people) to dedicate to this new effort. It also has to have a degree of flexibility and autonomy to take this on. It's probably not going to work well if every new effort needs to show that success is, if not guaranteed, highly likely before being granted approval to proceed. Part of this autonomy includes the freedom to fail without the library being made to suffer catastrophic consequences. In developing a new, emerging service, failure is often the best way to figure out how to get it right.

In our case, we have found that the right mix of expertise in biomedical research methodologies and practice in information science has been an important factor in our success. Our experience has demonstrated to us that one without the other is not sufficient. This may play out differently in different situations. Personalities and how they interact also inevitably plays a role: e.g., balancing risk-taking with a thoughtful

analytic approach, being willing to make mistakes and continue to work to get it right; which may be until it works or is abandoned.

It is probably true that ready access to IT expertise and resources, whether within the library or through a close and collaborative working relationship with IT outside of the library, is necessary to develop services beyond the informational or consultative role. Having the services of a web developer readily at hand has been a great help to our data management team.

It is probably true that ready access to IT expertise and resources, whether within the library or through a close and collaborative working relationship with IT outside of the library, is necessary to develop services beyond the informational or consultative role.

I will speculate that an academic medical center and many health sciences libraries provide conducive environments for both finding a critical mass of needs and having a flexibility to respond to these needs. The data management needs side of the equation is biomedical and life sciences research; vast though this area of research is, it is not all of scientific inquiry. It is more contained than that of an entire university. Most public-facing academic medical and health sciences librarians are dealing with hundreds of medical and graduate biosciences students and thousands of faculty. Whereas, most of their (research intensive) university library counterparts are dealing with tens of thousands of undergraduate and thousands of graduate students, and perhaps hundreds of faculty. The dynamics, opportunities, and expectations of each are quite different. There may typically be more opportunities for deeper engagement in the more focused world, if you will, of the academic medical center.

Where are these services-in-progress likely to go from this early point of development? Although speculative, some of this is already being discussed. A logical path of progression can be easily imagined for each of the areas of development described earlier:

- Data management education will probably evolve from standalone offerings (“one offs”) to integration in required courses.

- Other institutions may adopt the software underlying our institutional data catalog. A spread of interoperable data catalogs may eventually lead to a federation of data catalogs across multiple institutions, possibly including NIH.
- The incorporation of data discovery tools in the conduct of research could provide an impetus for the development of data repositories and data curation. University libraries have been much more involved in repository development than health sciences libraries, where repository uptake has mostly sputtered. This could change that and provide an opportunity for meaningful health sciences and university library collaboration.
- The lab organization tool, now in pilot testing, could be an entry to electronic laboratory notebooks (ELNs). The lab organization tool is but a pale and static early prototype of a full-fledged ELN. But it is testing some of the same waters that an ELN will have to traverse: e.g., the burden it imposes not outweighing the benefit of using it. Early stage discussions of ELN use among the research community have begun at our institution, as is no doubt the case at most academic medical centers, and are probably well underway at many others. Widespread adoption of ELNs will open the door, at least in theory, to data sharing in advance of or separate from publication. This, in turn, will provide an entry to open science.

To return to the beginning: it's all about the researcher and the enormously difficult endeavor they are engaged in. How can libraries and librarians contribute to the advancement of the research mission? We have to start by raising researchers' expectations of what libraries and librarians can deliver. Part of that is easy: we already deliver more than they are aware of. The other part is more daunting. It requires taking a risk.

Acknowledgements

I am deeply indebted to our creative, talented, and bold data services team: Alisa Surkis, Kevin Read, and the most recent addition, Fred LaPolla. Ian Lamb has provided superb web development support, particularly for the data catalog project. Jeff Williams was instrumental in supporting the early informationist projects and more recently in nurturing the collaboration with our university library colleagues. I want to thank Karen Hanson who was an inspiration behind the *Data Sharing and Management Snafu in 3 Short Acts* video and who labored mightily through some of the more difficult informationist project entanglements.

Going back to the beginning of this journey, I am indebted to Betsy Wilson, dean of university libraries at the University of Washington, who provided me with the initial opportunity to follow this path and was always encouraging.