AC 2011-326: CONNECTING WITH DATA: FIRST STEPS TOWARD AN EMERGING AREA OF LIBRARY SERVICE

Megan Sapp Nelson, Purdue University, West Lafayette

Megan Sapp Nelson is Associate Professor of Library Sciences at Purdue University Siegesmund Engineering Library. She is liaison to the Schools of Civil and Electrical Engineering and related College of Technology disciplines. You can contact her at msn@purdue.edu.

Connecting with data: First steps toward an emerging area of library service

Introduction

The library profession is constantly changing in response to technology and our users' needs. Recently, academic libraries have confronted the profound impact of cyber-infrastructure, which results in born-digital data sets that are rapidly growing in number, size, and complexity. These data sets are new challenges for librarians.[1]This paper describes how one librarian began to engage data producers and collections.

The author found that this new service required brushing up on old skills and acquiring new skills as well. Identifying and assessing needs for data services is an extension of liaison roles as previously understood. Moreover, the data services often require framing for a non-library audience (i.e. explaining how domain knowledge of the library sciences applies to the given situation.) Selecting appropriate tools for the data interviews and partnerships require a familiar process of learning tools and applying them. Forming partnerships with faculty and other librarians is an essential tool of librarianship as a whole, and is particularly important to data engagement. On the other hand, creating infrastructure or collaborating to improve infrastructure takes a systems librarian approach into the realm of cyber-infrastructure. Finally, library science domain knowledge is used and new skills are acquired throughout the entire process.

A wakeup call for connecting to data

As a liaison to civil engineering, the author's first introduction to data curation and management came via traffic sensor data. The collection included sensor data for vehicles traveling an interstate bypass loop around a nearby city, as well as pictures of the vehicles passing over the sensors. The sensor data collection was large and was funded by a state agency. Therefore, the researcher wanted to provide access to the data set digitally for traffic engineers within the state. As the civil engineering liaison, the author knew that research similar to this collection was being performed within the department, but was unprepared to determine if this particular data set was worthy of addition to the library collection. This interaction served as a wakeup call. The author didn't have enough knowledge to be able to make a collection accession decision. If the library collects it, what is the best way to present the data for later reuse? Who is the target audience for this data if it is collected? Is this data set going to be relevant in 15 years? In 50 years? Fortunately, in the case of this data set, the data research scientists employed by the university libraries were able to help the disciplinary faculty member to create a test case and get the data online. The author began using lifelong learning skills to pick up needed tools in data management. The process is ongoing, but the following resources and tools have proven helpful or enlightening.

Literature of digital curation

The task of picking up new skills while carrying on with existing duties is familiar to all librarians. The rate of change in librarianship over the last twenty years has been enormous and librarians have responded to those changes by adding skills to their repertoire. Data stewardship is the next format to require this process of skills acquisition.

Gold presents a list of action items for librarians who are interested in establishing a role for themselves in the management and curation of data. She particularly emphasizes reading data science literature, participating in data curation conferences, and if possible undertaking advanced training in data science and information. [2] All three of these methods have been helpful to the author. They present multiple ways to learn additional information and skills for application with research collaborations. Each of these is addressed in more detail below, with suggestions for particularly helpful resources available to those entering the field of data curation.

Using the literature

The literature of data curation is growing rapidly and is broadly indexed across the disciplines of information science, computer science, and library science, as well as literature from the viewpoints of specific disciplines (i.e. biological sciences, medical sciences, and physics).This can make it difficult to identify appropriately scoped, useful resources. For this reason, determining how to focus limited professional reading time is difficult. However, there are several good overview resources available to librarians and researchers to quickly grasp the major issues of this new discipline.

A webliography of digital data curation resources, written by Westra et al, [3] is primarily consists of organizational reports, and as such, gives an excellent broad overview of current issues in the entire field of digital data curation. It was written to provide easy reference to the seminal reports in the past decade that have shaped the current practices of digital curation. It also includes references to listservs, standards, software and open-source journals.

The Digital Curation Centre is in the process of creating a Curation Reference Manual.[4] At this time, twelve chapters are written, peer reviewed and published. Another ten are currently being written. As a reference manual, the whole work can be read as an introduction to the field, but individual chapters stand alone. This makes this work very valuable for learning about topics as the need arises. Not all chapters will be needed in all situations. The chapters are intended to support curation activities regardless of discipline. Therefore, if the information needed is specific to a disciplinary context (i.e. metadata standards for civil engineering or repositories for visualization software) this work is not an authoritative source.

Sarah Higgins of the Digital Curation Centre produced a model of the lifecycle of data which is very helpful for conceptualizing how data changes and is shaped by the process of creation and

analysis. The DCC model is not considered definitive but it is referenced frequently in the literature and has been altered to meet specific situational needs[5]. Most library liaisons can rattle off the publication lifecycle as presented in the Encyclopedia of Library and Information Science. [6] It takes that kind of familiarity with the data lifecycle to work with researchers to troubleshoot data management and curation issues.

The data lifecycle also provides an easy visual model to understand the concept of "upstreaming". Librarians have generally been downstream archivists and consumers of information. However, good data management begins before the research project with a plan. Involvement in data requires a librarian's subject expertise to move upstream in the research project, just as our involvement in research is moving upstream in the research/publication cycle.[1]

Familiarity with the needs and common practices of engineering research

The most useful reading time spent for truly understanding the reality and scope of data issues for researchers is reading reports by LIS researchers who interviewed scientists and engineers about their data management practices and problems. It is invaluable to be introduced to the lifecycle of specific data sets through the eyes of the researcher, as they reflect on how they produce the data, how they manage the data, and their own expectation for the data after their immediate project is completed. This type of reading gives the liaison an opportunity to become familiar with the issues that come up in the process of interacting with department faculty about data. More importantly, it also gives an opportunity to see the language of data curation in a specific discipline in use. The data curation practices of few sub-disciplines are represented in the literature at this point, but a number of libraries around the country are initiating interviews with researchers and the depth of the literature grows accordingly. With the introduction of new tools for capturing the researchers' experience, the literature will likely expand dramatically in the next few years.

It is also important to stay up to data not only on the literature of data curation, but also on changes in e-science that are impacting liaison disciplines. This implies not only keeping up to date on what has just been published by a faculty member within a disciplinary department, but also tracking emerging areas within the discipline, the predicted next steps in a research thread being carried out by a professor, and staying up to date on relevant data standards for your discipline. [8]

A number of libraries have performed extensive interviews with research faculty to determine the current state of data management at their institution, as well as predict needed services. A comprehensive survey of the data practices and needs of academic departments was performed by University of Minnesota in 2007. Departments across University of Minnesota, Twin Cities were interviewed regarding scholarly work and communication, with a special focus on data

management. The final report is full of quotes from professors working in a variety of disciplines talking about their perception of their own data management and sharing.[9]

Cornell University librarians conducted an extensive environmental scan of existing data resources and cyber-infrastructure initiatives at their institution. They also examined data production of a select group of Cornell departments, examining data production for each and noting large scale data collections housed within the departments. They also noted significant relationships with existing external data repositories such as ICPSR and Protein Data Bank. Cornell identified prominent repositories in disciplines while simultaneously looking for opportunities for creation of data services for data that does not have an obvious home elsewhere. Recommendations were made for structuring library resources to support data and providing education in data curation. [10]

The Data Curation Profiles Toolkit, the web presence of an IMLS funded project of the University of Illinois Urbana-Champaign and Purdue University, provides access to completed Data Curation Profiles (DCP), reports of interviews conducted with researchers at both institutions in a number of STEM disciplines. The DCP profiles provide a good feel for what a liaison may encounter when interacting with researchers. The current profiles provide excellent background on disciplinary subfields. The repository is intended to grow as the DCP tool is used at other institutions nationwide. UCSD has already contributed the first outside DCP, and more are anticipated. [11]

In 2009, MIT Libraries participated in the HMI? How Much Information? Project hosted by the Global Information Industry Center at UCSD. In the process, MIT librarians interviewed 29 researchers from six disciplines at MIT regarding data generation, growth, retention, and sharing. [12] While the information gathered was specific to MIT, the report provides an overview of data accumulation across multiple engineering disciplines and gives particularly interesting insight into data sharing preferences. This report provides the most detail in terms of rate of data accumulation and data file size.

An understanding of the workflow and practices of the discipline is important to fully understand how science and engineering research is changing. Tenopir and King's "The Communication Patterns of Engineers"[7] provides a helpful baseline understanding of the general practices of engineers in the lab regarding not only data, but analyzing and publishing that data as well. It also offers a perspective on problems that were already in existence in engineering communications prior to the growth of e-science. These problems are growing or morphing due to the heightened pressure on communications around data caused by external expectations for data management.

Formal training

Formal training programs for professional librarians are needed. It is difficult to learn additional skills while on the job, particularly if the institution does not have experts in data curation and is

starting a data service from scratch. A few of these training opportunities exist. They require registration, housing and travel expenses, so require more material support. However, they make up for additional cost in time and money by providing benefits that outlast the few days on campus.

University of North Carolina-Charlotte's School of Information and Library Science hosts a week-long intensive training program and two day workshops for professionals currently working with data. These workshops feature individuals all over the world who are experts in data management. The curriculum is underpinned by the six facet matrix of core competencies, which focuses not only on skills but also on values and processes required to create a data management workflow across the data lifecycle.[13]The DigCCur institute focuses broadly on skills needed in digital curation, rather than focusing on particular disciplinary-related tools and trends. This includes practical sessions featuring the technical side of data curation, as well as discussion about the process of data curation and the values of data curation. The topics for the symposium reflect the matrix created and taught through the DigCCur curriculum. [14]

The University of Illinois has been hosting Summer Institutes on Data Curation for the last several years. In 2010, the topic was Earth Sciences Data.[15] The focus of the institute alternates by year between humanities, physical sciences, and life sciences. The workshop is three days in length, covering topics that include the cyber-infrastructure needed to manage data, the standards that (potentially) govern e-data curation, and, particularly, the existing projects that have been created to manage data sets. The focus on specific disciplines means that the content is built on the issues and needs of targeted user groups.

This type of focused training on digital curation has several advantages. The individuals attending the session become a natural cohort of similar focus and interest. With the assistance of communication tools such as listservs, wikis, and blogs, these groups exchange ideas, encourage each other, and expose new and interesting work to the rest of the cohort. Additionally, while the amount of information presented is at times overwhelming, it also gives a good overview to the current best practices in a brief amount of time. Generally, it is difficult to spend extended periods of focused independent study while in the library, simultaneously working on other projects.

Conferences

The primary conference in the field of digital curation for librarians is the International Digital Curation Conference. This gathering of digital curators allows curation professionals from different countries with different digital curation approaches to exchange ideas and research results. It is held annually and is primarily held in the UK with occasional host sites in the US. [16] University of North Carolina hosts the DigCCur conference annually. This conference is intended to allow digital curation practitioners to explore current practice and research. [14] There are a number of other technology-oriented conferences on topics concerning data curation. IEEE

hosts the International Conference on E-Science, which is focused on large multidisciplinary research group. IEEE also hosts the Conference on Mass Storage Systems and Technologies, which focuses on the technology necessary to store large data sets. ACM hosts the SIGCOMM Conference on Data Communication, which looks at the networking technology necessary to transfer and share large data sets. These are all venues for learning the newest research and techniques in the management of large data sets.

Cohorts

Gabridge suggests that those individuals who are interested in data curation at a given institution should come together to form a cohort.[17] This was reiterated by the Cornell Library Data working group in their report.[10] The opportunity to share knowledge, experience, and techniques is quite helpful. If a librarian is embarking on this new field of library work as an individual without colleagues who are experienced data curators, there are several listservs available that provide an electronic cohort to share experiences and ask questions. North Carolina State University hosts the numeric data listserv.[18] For a list of a variety of data-related listservs, see Westra et al. [3] University of North Carolina – Charlotte's School of Information and Library Sciences hosts DigitalCurationExchange.org[19]. Those who use the Data Curation Profiles can share experiences and questions via datacurationprofiles.org.[11]

Connecting with data

Educating oneself is important, but at some point, connecting with data requires interacting with researchers. Gold noted that individual librarians "who are actively working to increase their own data literacy and awareness, and equipping themselves to provide educational and consultative services related to data management and curation" are engaging faculty and students, as well as other information professionals, in research and instruction. [20] Outreach has long been part of the librarian's toolkit. It is particularly useful in the context of data curation and e-science. [8]

Researchers generally do not realize that library and information science skills may alleviate practical problems in their own laboratory. It does not occur to researchers to come to a librarian for help with managing data. In the author's experience, data projects often come about due to a situation of being in the right place, at the right time. Actively engaging departments frequently opens doors to discuss issues of data management. Communication of the value of the skills that librarians have and/or are developing is imperative.

"Liaisons know their disciplinary communities, their information practices, and understand a community's data needs and can help shape the services that will meet them". [17] This message does not always come from the liaison. Senior management of the library can be very effective messengers about services offered and skill sets or collaborations in existence.[21] The Office of Research for a college or university is also a likely venue for encouraging researchers to use library data services. Building relationships with the research administration of the college or

university allows messages with great internal authority (that of the Office of Research) to be broadly disseminated.

Frequent collaborators with the library from a liaison department may not be the ideal audience for first forays into data curation. Researchers with active research programs involving the collection of digital data and the management of multiple graduate students in the research laboratory are likely to have encountered the problems of digital curation. If a frequent library supporter meets this description, he or she may have a data management project that they are interested in addressing.

As Carlson and Garritano note, "the classic reference interview, used to identify the information needs of patrons, still has an important place when dealing with researchers that produce and work with large amounts of data."[8] The subject of the interview may present a range of topics across the data lifecycle or focus on a narrow subset of issues directly related to data management, data information literacy, or data preservation. The approach may well depend upon time and resources available for this new service at both the liaison and library level.

The initial conversations will certainly require that the liaison primarily collect and synthesize information regarding the current practices and issues of the researcher. Regardless of the type of tool or approach chosen by the librarian, using a digital voice recorder is an effective way to make the most of the little time with the researcher. The librarian can focus on what is being said, and not on writing or capturing information. It also allows the librarian to return to the interview transaction and to create follow up questions or clarify issues. [11]

There are multiple ways to approach the conversation with the disciplinary faculty member. The first is to address the NSF data management requirement. The second is to approach the subject of the data management practices of graduate students in the laboratory. The third is to use the Data Curation Profiles as a complete package to look at the lifecycle of a specific data set. The fourth is to perform a data audit for the entire department or academic unit. Each of the methods has its positive and negative attributes, which may influence which type of interview is best for a given department or researcher. Each of these tools simultaneously allow the librarian to get to know the researcher's individual scientific process and help the researcher to learn the breadth and depth of the issues involved in curating scientific data. They also provide openings for the development of relationships, both with the individual and with the department as a whole.

In the wake of the NSF data management requirement, speaking to a department faculty meeting about data management plan requirements and the contribution information science can make to those plans may be all the advertisement that is needed to begin data liaison work. This interview requires little initial preparation beyond familiarizing oneself with the NSF data management mandate, the Engineering Directorate guidance documents, and existing tools. [22] While the NSF mandate is light on details of what is expected, the Engineering Directorate does provide some sense of the expectations, and allows comparison of the specifications of the tools available to

determine which best fits the situation. Other directorates are now creating their own guidance documents as well. Many universities are creating specific tools to suit the local situation, so there may be proscribed tools available for a local institution.

The librarian may need to make a case for ways that library and information science can be helpful to a researcher. As a reference for oneself and preparation for interaction with researchers around data management plans, identifying and becoming familiar with metadata standards and disciplinary repositories is useful. This is a challenge because no one standard has been adopted for the engineering disciplines, nor has a single disciplinary repository taken hold for engineering. A derivative of ISO 10303 (Standard for the Exchange of Product Model Data) called STEP has been developed to govern archiving of product models and is increasingly being used by industry. LOTAR (Long Term Archiving) is an OAIS application that has specifically been designed to archive STEP data packages.[24] These are standards worth investigating. Familiarity with metadata standards commonly used in the discipline or under development for a specific type of data is a help for faculty members who are vaguely aware of their existence but who have yet to implement them. A useful tool for assisting researchers to think through the development of metadata is the use of Competency Questions. [23]

Repositories housing data produced by faculty from the target discipline are also a fruitful area of discussion because an existing repository with respected scientists and engineers as participants provides authority for discussion with the faculty. It allows librarians to give suggestions for the appropriate information source or type in a given situation, just as we would recommend a book or journal article. Knowledge of relevant repositories also helps the subject liaison to have a better understanding of the culture of data management for that specific discipline.

The Digital Curation Centre in the UK recently unveiled the DMP online[25], a web application that provides a number of prompts for facets of data management and curation that should be taken into consideration while developing a project (and while submitting a grant to a funding agency). The DMP online is currently in the process of adding the NSF requirements, so that it will be an appropriate tool to recommend to US-based researchers. The Digital Curation Centre also has an abbreviated Checklist for a Data Management Plan that is a nice tool for the easy generation of questions for a data management plan interview. [26,27]

The Data Curation Profile can be used to help a researcher think through potential issues and needs for a project data management plan as well. The following modules contain relevant questions:

Module 2 – Overview of the research

Module 3 – Data kinds and stages

Module 5 – Organization and description of data

Module 7 – Sharing and access

Module 8 – Discovery

Module 12 – Data management

Module 14 – Personnel

The other modules will provide more detail, but are not necessarily needed to meet the criteria of the NSF Data Management Plan.

The grant data management plan approach creates avenues for further exploration and collaboration with faculty members. However, the initial consultations appeal only to those researchers who have short to medium term plans to submit an NSF (or NIH for specific disciplines) grant.

The second approach is to meet with a specific researcher about his or her graduate students' abilities to manage data. (Data management often falls to graduate students.) This approach is similar to information literacy consultations performed to address ARL standard information literacy performance indicators.[28,29] In this case, the performance indicators are identified as technical skills that the students are missing, such as data cleanup, backup and storage, or data analysis skills. The skills that are lacking in graduate students are taught through partnership with the advisor professor. Performance indicators may be a knowledge gap that the librarian can address, such as metadata, ontologies, or citation of data or a technical knowledge gap addressed by the advisor. This approach requires spending a moderate amount of time with the researcher, and depending upon the educational intervention developed, could result in multiple sessions with students, or the development of a traditional "one shot" workshop or class presentation. The educational intervention depends upon the collaboration between the researcher and librarian, and the performance indicators that need to be addressed.

The third approach, the Data Curation Profiles, provides the opportunity to explore a dataset with the data author using a library and information science perspective. The Data Curation Profiles Toolkit[11] provides a step by step interview process to discuss one single data set, and learn from the researcher the data lifecycle, current handling, and future intentions for sharing of and access to the data. The interview must be audio recorded. This interview will take at least two sessions of one hour with the faculty member, as well as the time to either transcribe or index the interview, and the time to analyze the interview and slot it into the Data Curation Profile template. The completed profile is a document that allows for educated decisions to be made regarding long term preservation and access. The interview may also raise awareness of digital curation issues that are relevant for that specific data set.

This approach is time consuming but opens up many doors for further collaboration, collection development of a specific dataset, and identifying institutional barriers to good data hygiene.

This tool is particularly appropriate for a researcher who has a project that is important in the life of the institution (large grant award or noteworthy outcomes of the project). It may also be suitable for interviewing a representative faculty member of a department to identify barriers to good data management within the department. The approach is also very effective at teaching the librarian about the data lifecycle or typical data practices.

Finally, the fourth approach is to conduct a broad overview of all data assets in the disciplinary department. The data audit framework proposed by Digital Curation Centre[30] requires the department liaison to have ready access to department servers. At the author's university this is not the case. Therefore, multiple interviews with individual researchers of a department serve as a substitute for access to files. These interviews cover the size and location of data sets, the data management and sharing plans for these assets, and current data management practices. The data audit identifies any departmental and university policies in place for the management of the identified data. It also names all stakeholders for the data.

A data audit is a large research initiative on the part of the librarian. If completed, the inventory will provide basic information with which to prioritize data curation services to the department, identify likely research labs for data information literacy educational interventions, and guide collection development decisions regarding specific data assets. The department can also use the data to create policies governing data management and to predict the growth of data and plan resource allocation accordingly. This is a very time consuming, broad ranging project that provides direct pay off to the librarian and is a service to the department.

The author's experience with data curation work

As mentioned above, the author's first experience came through an interview with a disciplinary faculty member. Initially, the role emphasized was that of subject liaison, not data curation. Data services started at our university with the hiring of a data research scientist. That staff member works with liaison librarians to provide service for the researchers. In the case of the interview with the civil engineering faculty member, the data research scientist played a primary role while the author was expected to contribute subject expertise. In reality, existing expertise did not provide adequate preparation for providing useful input in that interview. It proved to be a catalyst for learning more, however.

With the development of the Data Curation Profiles, the author possessed a tool that broke down the data curation process in an approachable and understandable way. The author received a grant to use the Data Curation Profile to help international research collaboration participants to communicate about their expectations for the outcomes of projects. Preparation required for this project substantially increased knowledge of the data lifecycle, primarily by familiarizing the author to the terminology of data creation.

The UIUC Summer Institute for Data Curation proved to very helpful. Individuals attending the institute were all interested in earth sciences data and the presentations were targeted to that

specific disciplinary data culture. Librarians with many years of experience managing data sets discussed the details of metadata standards, best practices for data management, observations of data use within the discipline, and the many real challenges of data management for the earth sciences. The Summer Institute participants also joined a listserv which provides a venue for asking questions about earth sciences data. The cohort created at the Institute has continued to interact at other conferences as well.

While at the Summer Institute, several librarians from our university brainstormed an article that presented learning objectives in data management skills targeted to graduate students. Upon return to our institution we recruited an additional writing partner and developed our article. The process of doing the research for the article proved to be the point of critical mass at which the author's understanding of data curation solidified. Preparation for the article included many hours reading transcripts of interviews with professors regarding their students' ability to perform data curation tasks. Reading the transcripts provided many examples of the types of tasks that had to be accomplished to successfully manage data. These examples, in tandem with the data curation profiles questions, brought the full picture of data curation to light.

The author then attended the International Digital Curation Conference in December 2010. This conference brought together data curation practitioners and researchers from all over the world. In a one day pre-conference workshop, the Digital Curation Centre staff from the JISC in the UK presented their model for eliciting data related needs from researchers. The workshop was specifically focused on the data management plan. The workshop provided handouts and tips for working with researchers to develop a data management plan.

Less than a month later, the author was asked to participate in a group of librarians who were trained to provide consultations regarding the NSF data management plan. Our role was to provide advice to NSF research grant writers with the goal of creating a stronger data management plan. Working directly with the grant writers was an enlightening experience. In some cases, the authors had too little information in their data management plan and required many questions and edits from the librarian. Frequently, the authors didn't discuss who was in charge of the management of data, what metadata standards would be used, the difference between data sets and types throughout the lifecycle, and either wanted to share all data regardless of usefulness, or were leery of sharing any of the data until the very end of the project.

On the other end of the spectrum, some authors took the example data management plan created for university wide dissemination too seriously and went so far as to predict exact size of each data set, and name servers where the data would be held. These authors required someone to help them step back from the project and understand the purpose of the data management plan in terms of access and reuse.

The author has used the data management plan as a way of capturing faculty interest in library services, and as a relevant service that meets the faculty members' immediate needs. This

includes presentations to entire departments, as well as within graduate student orientation classes, and to small faculty research meetings. The result is generally curiosity and cautious interest. However, a few faculty members have embraced the role of the libraries, are open and enthusiastic, and are prepared to participate with librarians in research projects to improve data information literacy.

Overall, the addition of data services has been a net positive experience. It has required substantial time to bring knowledge of data topics up to an effective level. Collaborators provided important external motivation to continue in the independent study of the data curation literature. The university libraries provided significant investment in skills development by sending the author to institutes and conferences. As time progresses, data projects comprise larger percentages of the author's work time. At the same time, the university libraries have relieved librarians of other traditional library tasks in favor of strategically important work like data curation. This has been both a relief in terms of practical day-to-day scheduling and a source of regret, as some of the tasks were ones that the author enjoyed (i.e. physical reference desk work.)

On the other hand, data work provides a ground for rapport with researchers that was not possible when talking about traditional library domain knowledge. The services offered are understood by researchers and requests are made for point of need service, which have greater value to the researchers. Data curation roles also provide opportunities for collaborative research with faculty members; the research done has great relevance to both disciplinary faculty members and librarians in this time of changing research practices.

Stepping out of the comfort zone: New roles and responsibilities

For data curation, best practices haven't been identified yet. Librarians and information scientists are starting to identify successful interventions for small portions of the data curation cycle, but there is no one model or intervention that will provide solid groundwork from which to start. This is primarily due to the fact that all research projects are different, have different needs, and have different expectations for data management imposed upon them.

Professors and staff with an active research agenda have been the primary audience of the services highlighted in this paper, but others within the university are also likely audiences. Graduate students, particularly in the STEM disciplines, must learn data management skills. They will never know their discipline without electronic data capture. However, graduate student education in data curation skills is patchy and frequently reliant on the advisor professor to teach those skills. For professors who are learning the skills themselves, it is difficult to teach all necessary skills across the data lifecycle. Additionally, not all graduate students work directly with an individual advisor in a laboratory. Fellowship students, who are expected to create their own research plan, may be missed by lab-based data management training. Add in the fact that graduate students come into a program with different strengths and weaknesses, and it quickly

becomes apparent that some sort of formalized data management assessment and instruction is needed.

Access to these students may come in the form of classroom based activities (a graduate seminar) or through individual professors who open their research groups for instruction. Research is needed on best practices for presenting these skills to students. At this point, a list of core competencies in data information literacy has been presented, but not yet validated.[28] Librarians and information scientists will continue to hone educational objectives for teaching data information literacy. Traditional one hour lectures will not be sufficient for presenting the broad range of skills that need to be transferred. More likely, successful data information literacy programs will demand embedding with research groups (student or faculty) over the life of a project and teaching skills at the point of need. (If this is the case, then the question of the scalability of data information literacy programs must be addressed as well.)

Teaching faculty members these skills will likely come in informal settings. It will be important for each faculty member that only relevant skills are presented as required. In order for this to happen, librarians must identify which data skills are needed by a researcher. The Data Curation Profiles tool mentioned above is useful for this work. However, a tool that will assist a librarian to understand the depth of a faculty member's data management skills in a short interview has yet to be developed.

There may be an opportunity for peer teaching of faculty members. If librarians identify researchers who exhibit excellent data management skills, those researchers could serve as examples with personal experience that is immediately relevant and respected by the faculty of the liaison department. Those faculty members who excel in data management may be highlighted in data training during a faculty meeting, or the departmental liaison may work with those faculty members to prepare workshops or tools which the disciplinary faculty member then presents to his/her peers.

Gold indicates that the problems of data curation lend themselves to collaborative solutions from a variety of departments across campus [20] Expectations for data management differ by funding body, by university, by discipline, and by research field. As the ARL Workshop on Digital Data Stewardship notes, "Stakeholder groups have different expertise, outlooks, assumptions, and motivations about the use of data. Forging partnerships will require transcending and reconciling cultural differences." [31] These collaborations are dependent upon framing library science skills in the language of data curation and information science. "Translating" skills for technical audiences may require multiple attempts and a variety of descriptive information. However, the technical requirements of many aspects of data curation mean that librarians have the opportunity to serve in a mediator role between liaison and technology departments, gathering user requirements, providing expertise, and working with technical experts to generate solutions that will support and enhance data management within the liaison department.

These collaborations emphasize the librarian's role as a scholar of data management working in collaboration with scholars of other disciplines. Taking on this role of "data researcher" means fully embracing data as a major area of scholarship, and requires librarians to re-conceptualize their understanding of their own scholarly agenda and traditional work. Only when librarians are able to define our culture of scholarship of information (and data) to other disciplines will we be able to reconcile our differences with those of our disciplinary faculty members.

Conclusion

While data presents major challenges for a practicing librarian who is unable to devote large amounts of time to the development of new skill sets, it is an important emerging area of scholarship and a major area of need among liaison departments. There are several tools available to approach departments about data management and each tool has strengths and weaknesses. The implementation of those tools provides information to identify needs of individual researchers and students. No one solution will meet all of the needs of researchers or students. New learning objectives, educational interventions and pedagogy are needed to teach data curation skills. Librarians need to translate traditional library science skills into a data curation context in order to forge the partnerships necessary to have a successful data management program.

References

1 Gold, A. Cyberinfrastructure, Data, and Libraries, Part 1. *D-Lib Magazine* **13**, 9/10, doi:10.1045/september20september-gold-pt1 (2007).

2 Gold, A. Cyberinfrastructure, Data, and Libraries, Part 2. *D-Lib Magazine* **13**, 9/10, doi:10.1045/july20september-gold-pt2 (2007).

3 Westra, B., Ramirez, M., Parham, S. W. & Scaramozzino, J. M. in *Issues in Science & Technology Librarianship* no. 63 (Fall 2010). Available: http://www.istl.org/10-fall/internet2.html

4 *DCC Digital Curation Manual*. (Digital Curation Center, 2005) Available: http://www.dcc.ac.uk/resources/curation-reference-manual.

5 Constantopoulos, P. *et al.* DCC&U: An extended digital curation lifecycle model. *International Journal of Digital Curation* **4**, 34-45 (2008). Available: http://www.ijdc.net/index.php/ijdc/article/viewFile/100/75

6 Kent, A., Lancour, H., Nasri, W. Z. & Daily, J. E. *Encyclopedia of library and information science*. (M. Dekker, 1968).

7 Tenopir, C. & King, D. W. *Communication patterns of engineers*. (IEEE Press ; Wiley-Interscience, 2004).

8 Garritano, J. R. & Carlson, J. R. A Subject Librarian's Guide to Collaborating on e-Science Projects. *Issues in Science & Technology Librarianship*, no. 57 (Spring 2009). Available: http://www.istl.org/09-spring/refereed2.html

9 Marcus, C., Ball, S., Delserone, L., Hribar, A. & Loftus, W. Understanding Research Behaviors, Information Resources, and Service Needs of Scientists and Graduate

Students: A Study by the University of Minnesota Libraries (University of Minnesota Libraries, Minneapolis, MN, 2007). Available: http://lib.umn.edu/about/scieval/Sci%20Report%20Final.pdf

10 Cornell University Library Data Working Group. Digital Research Data Curation: Overview of Issues, Current Activities, and Opportunities for the Cornell University Library. 49 (Cornell University Library, Ithaka, NY, 2008). Available: http://www.library.carleton.ca/documents/researchdataseminar/DaWG_WP_final.pdf

11 Carlson, J., Brandt, D. S. & Cragin, M. *Data Curation Profiles Toolkit*, (2011). Available: http://www.datacurationprofiles.org

12 Madnick, S., Smith, M. & Clopeck, K. HMI? Case Studies: How Much Information? Case Studies on Scientific Research at MIT. (Global Information Industry Center at the School of International Relations and Pacific Studies; University of California, San Diego, San Diego, CA, 2009). Available: http://hmi.ucsd.edu/research_case_summary.php

13 Lee, C. *Matrix of Digital Curation Knowledge and Competencies (Overview)*, (2009). Available: http://www.ils.unc.edu/digccurr/digccurr-matrix.html

14 Tibbo, H. R., Lee, C. & Costello, K. *DigCCurr*. (2010). Available: http://ils.unc.edu/digccurr/

15 *GSLIS to host 2010 Summer Institute on Data Curation*, (2010). Available: http://www.lis.illinois.edu/articles/2010/05/gslis-host-2010-summer-institute-data-curation

16 Digital Curation Centre. *Conferences*, (2010). Available: http://www.dcc.ac.uk/events/conferences

17 Gabridge, T. The Last Mile: Liaison Roles in Curating Science and Engineering Research Data. *Research Library Issues*, **265,** 15-21, (2009).

18 Hayslett, M. Numeric Data Listserv. North Carolina State University, Raleigh, NC, 2011. Available: http://www.lib.ncsu.edu/data/listserv.html

19 *Digital Curation Exchange*, (2011). Available: http://www.digitalcurationexchange.org

20 Gold, A. *Data curation and libraries: Short-term developments, long-term prospects*. (2009). Available: http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1027&context=lib_dean

21 Brandt, D. S. Librarians as partners in e-research: Purdue University Libraries promote collaboration. *College & Research Libraries News* **68**, 365-396 (2007).

22 Engineering Directorate, National Science Foundation. *NSF Data Management Plan Requirements*, (2010). Available: http://www.nsf.gov/eng/general/dmp.jsp

23 Doyle, J., Viktor, H. & Paquet, E. Long-term digital preservation: preserving authenticity and usability of 3-D data. *International Journal on Digital Libraries* **10**, 33-47, (2009). doi:10.1007/s00799-009-0051-7

24 Lubell, J., Rachuri, S., Mani, M. & Subrahmanian, E. Sustaining Engineering Informatics: Toward Methods and Metrics for Digital Curation. *International Journal of Digital Curation* **3**, 59-73 (2008). Available: http://www.ijdc.net/index.php/ijdc/article/viewFile/87/58

25 Donnelly, M., Jones, S. & Pattendon-Fail, J. W. DMP Online: The Digital Curation Centre's web-based tool for creating, maintaining and exporting data management plans. *International Journal of Digital Curation* **5**, 187-192 (2010). Available: http://www.ijdc.net/index.php/ijdc/article/viewFile/155/218

26 Digital Curation Centre, DCC Checklist for a Data Management Plan v.3.0. Digital Curation Centre, Edinburgh, UK. (2010).
27 Digital Curation Centre. *DMP online*, (2011.) Available: http://www.dcc.ac.uk/dmponline
28 Carlson, J., Fosmire, M., Miller, C. C. & Sapp Nelson, M. Determining data information literacy needs: A study of students and research faculty. *Manuscript submitted for publication.* (2011).
29 Qin, J. & D'Ignazio, J. Lessons learned from a two-year experience in science data literacy education. *International Association of Scientific and Technological University Libraries,* 31st Annual Conference. West Lafayette, IN,(2010). Available: http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1009&context=iatul2010
30 Jones, S., Ball, A. & Ekmekcioglu, ς. The data audit framework: A first step in the data management challenge. *International Journal of Digital Curation* **3** (2008). Available: http://www.ijdc.net/index.php/ijdc/article/viewFile/91/62
31 Association of Research Libraries. *ARL: A Bimonthly Report* Vol. 249, 4-5. Association of Research Libraries, Washington, DC, (2006). Available: http://www.arl.org/bm~doc/arlbr249datasteward.pdf