CHAPTER 21

Data Librarianship:
A Day in the Life—Science Edition

Danianne Mizzy and Michele Hayslett

AFTER SURVEYING AND interviewing a cross-section of librarians supporting scientific work, we conclude that there is no one path to data librarianship: some start in the sciences (whether physical or biomedical) and find librarianship later; some start in librarianship and find the sciences later. This agrees with Kellam's observation that "Data librarians … come from a variety of disciplines and by diverse routes."[1] At least among our sampling, this area of practice has many commonalities with social science data support, but with more of an emphasis on data management and data sharing than on data discoverability and reuse. The growth in the importance of data for all areas of scholarship makes this topic broadly relevant. Advice and insights from current practitioners will offer a useful guide, especially for those considering this area as a profession.

Methodology

The authors followed the approach taken by Kellam of posing informal interview questions about the respondents "experiences becoming and being data librarians."[2] The interview questions were adapted from Kellam's questions but were also based on the authors' professional experience and were fine-tuned with help from a survey expert.

The authors drew upon their professional contacts to compile a list of librarians and informationists who provide data services for scientific, health, and biomedical researchers and invited them to participate. Using snowball sampling, we invited people from the original list to suggest others with whom we should speak.† Thirty respondents completed a brief online survey, and twenty-seven

† Appendix 21.A provides more detail on respondents' locations and affiliations as well as the survey questionnaire.

of them agreed to participate in a personal interview conducted by telephone or Skype.†

Survey Results

Education

Respondents possessed a wide range of educational backgrounds, some with no formal library training, some with no or minimal science training. Most respondents (22/30, or 73%) had a bachelor's degree from a science field; twenty-seven percent (8/30) had degrees in humanities or social science fields or did not specify a field. Twenty-three percent (7/30) had no science-related degrees. Seventeen percent (5/30) had doctoral degrees, all of which were in science fields. All but three had master's degrees in library science or information science. Two had advanced certificates focused in data or data curation; two others had graduate certificates in biomedical or health informatics.

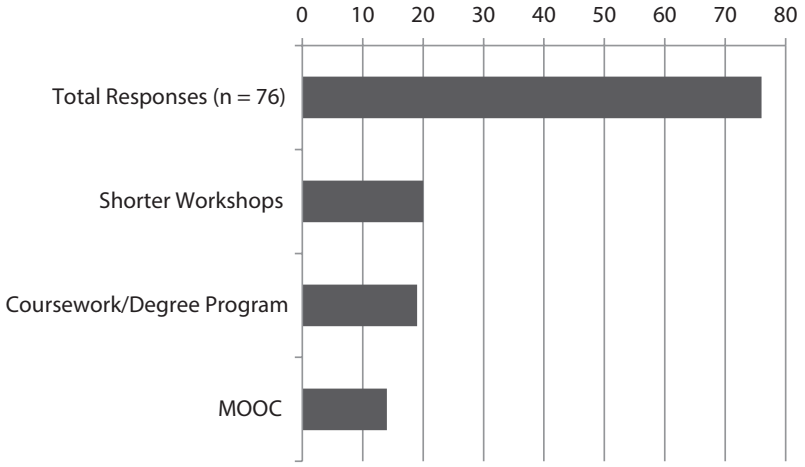
Data-Related Training

When we asked what other data-related training, coursework, and programs of study they had taken, respondents provided 76 answers, describing an enormous variety of programs they employed to gain additional knowledge, including both degree programs and many less formal workshops and courses (See Figure 21.1 for frequency of training by type). Twenty-six percent of the respondents (20/76) indicated use of workshops as a means of gaining data management training, but almost as many, twenty-five percent (19/76), learned the material within degree programs. Four of those who mentioned week-long in-person workshops (counted as part of the shorter workshops category) attended the *Curating and Managing Research Data for Re-Use* workshop usually offered in the summer by the Inter-university Consortium for Political and Social Research (ICPSR). Two people mentioned multi-month workshops without specifying which ones and three mentioned the e-Science Institute co-sponsored by the Association of Research Libraries (ARL) and the Digital Library Federation (DLF). Two more cited a two-week long immersion experience, the DigCCurr Institute, hosted by the School of Information and Library Science at the University of North Carolina.

---

† The data are available at http://dx.doi.org/10.17615/C6RP41.

Figure 21.1. Most Commonly Cited Types of Data-related Training



Twenty-five percent (19/76) of responses cited degree programs, and about 8% (6/76) noted certificate programs (although not all were semester- or degree-length certificate programs). One person conducted an independent study, and one took a semester-long online course. The next most frequently cited source of training was MOOCs (Massive Open Online Courses), with 17% (13/76) listing that source. Of those, seven people attended Johns Hopkins' *Data Scientist Toolbox* class via Coursera.‡

While the authors did not ask how long respondents had been practicing, these results seem to suggest a divide between professionals who entered this profession before funding agencies began requiring data management plans, and those who attended graduate school after curricula caught up with developments in the field. The good news for people considering this specialty is that there are plenty of learning opportunities both in and out of graduate school. Likewise, respondents described diverse paths to their current positions so there is no one set of education and experience that qualifies individuals for this profession. The survey also did not explore motivation for choosing one type of program over another, but perhaps some found it easier to complete a MOOC in addition to a full-time job or wanted the formal credential of a certificate program.

‡ For more detail on the programs respondents cited, see Appendix C—Table 1 on the companion website (https://databrarianship.wordpress.com/).

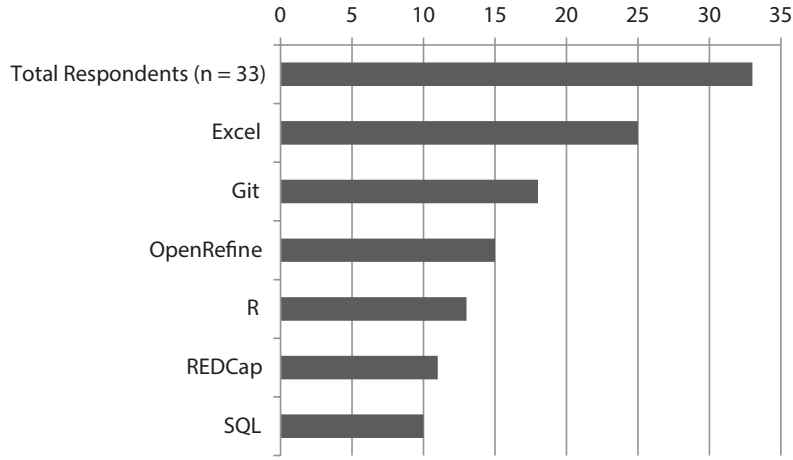
Other Learning Opportunities

For this question, we asked respondents to list specific types of resources they found helpful in their work: conferences/meetings, professional associations, blogs, listservs and other. As with the question about training, coursework, and programs of study, respondents took advantage of an enormous variety of other learning opportunities (each respondent was able to list as many as he or she wished). Conferences were the most popular option, with 57% (17/30) attending the Research Data Access and Preservation (RDAP) Summit, a conference organized by the Association for Information Science and Technology (ASIS&T). Forty-three percent (13/30) attended the International Digital Curation Conference (IDCC) and 23% (7/30) attended the annual conference of the International Association for Social Science Information and Technology (IASSIST). Listservs were the next most popular resource. Forty percent (12/30) of respondents cited the RDAP listserv, and 33% (10/30) the UK-based research-dataman listserv (short for data management). Twenty-three percent (7/30) of respondents participated in the ARL/DLF-sponsored eScience Community blog. Seventeen percent (5/30) cited the data management community on Twitter.†

Tools

Figure 21.2 presents the tools mentioned most frequently. Excel was the most commonly cited tool with 83% (25/30) of respondents mentioning it. Git was the next most frequently cited tool with 57% (17/30) of respondents, followed closely by OpenRefine and R with 50% (15/30) and 43% (13/30) respectively.

Figure 21.2. Most Commonly Cited Useful Tools

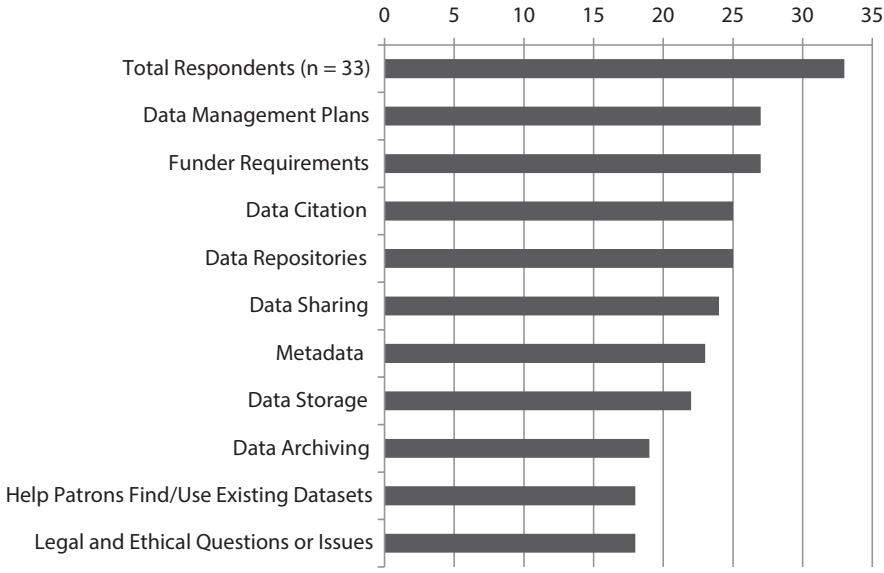


† For a list of these resources, see Appendix C on the companion website (https://databrarianship.wordpress.com/).

Tasks

The survey listed 25 different tasks. The authors gleaned the tasks from job descriptions and data services positions advertised through listservs to which they subscribe and from various library web sites. Respondents were invited to check the ones they performed. Figure 21.3 presents the most commonly selected tasks. As it shows, nearly all respondents assist with writing data management plans, understanding funder requirements, formatting data citations and finding repositories in which to preserve data.

Figure 21.3. Most Commonly Cited Tasks



The fact that all 25 tasks listed on the survey had responses speaks to the wide variety of interpretations of what it means to be a data librarian. Each institution has different needs and, while some tasks are common, each position has its own unique mix of job elements. Just as there is no one path to become a data librarian, job seekers in this area will find many different expectations in position announcements. As will be seen in the next section, overall the results of the survey are borne out in the responses from the interviews.

Interview Responses

In addition to a survey, we conducted interviews with several data librarians who support science disciplines. We have organized this portion of the chapter by the

interview questions which included several open-ended queries about the training and experiences needed to be a scientific data librarian.†

1. ***How does your educational background relate to the data responsibilities in your current position?***

The degree of correspondence between formal training and current data responsibilities varied greatly. Some respondents undertook their graduate library work before research data management entered the library science curriculum. Some received statistics, research methods, or data management training as part of their undergraduate or graduate work in non-library- science disciplines. Others sought out data training as an elective, a concentration within their library degree, or a within a certificate program. Many respondents shared the opinion that a science background is very useful but not a prerequisite. For example, Katie Houk offered these words of encouragement for non-scientists considering this area of practice:

> Librarianship is its own domain of specialized knowledge and skills. You don't necessarily need to have a higher degree in science, or currently be doing scientific research yourself, to bring data or information management best practices to the table and help scientists better deal with these issues.

2. ***Please describe what you had to learn on-the-job and how you went about it.***

Several respondents answered, only half-jokingly, "Everything!" meaning everything from the vocabulary to the software to the various stakeholders involved. Multiple respondents stated, "This was not part of my MLIS program." They had to learn the local or disciplinary context before they could apply what they had been taught, including the politics around data in their setting. What researchers really do with their data was eye-opening, and the lack of documentation about that data was only one of many ways that actual and ideal practices, as taught in a graduate curriculum, may diverge. Some respondents with scientific training observed that prior experience with research and managing data, while providing an excellent foundation, does not make you a data specialist.

Many had to discover just what a data librarian does, how reference services for data differ from other kinds of reference interactions, and how scientists use data. Metadata, programming, preservation, and data citation practices were the skills most frequently mentioned as something respondents needed to learn or learn more about on-the-job. Almost all respondents stressed the importance of getting the lay of the land, both within the library and at the institutional level. For

---

† Respondents reviewed this material and gave consent to include their excerpted answers.

example, those entering a biomedical setting might be interested in Alisa Surkis' early experience:

> I learned about differences in the research process and how data are viewed between basic and clinical researchers. On a clinical research project, beyond the PI (Principal Investigator) everybody is more an employee of the project. In a basic science lab you have a lot of essentially independent or semi-independent researchers who are all doing their own experiments. Post-docs and graduate students cycle through the lab, leaving after a few years. So those create completely different sets of problems.

Whether serving as the first person in this job at an institution, or just joining an established service at a new institution, respondents advised undertaking at least a basic environmental scan. This will help librarians find out who the stakeholders for data management are on campus and what capabilities already exist at the institution. Surkis advises, "Spend time talking to people and gaining awareness of institutional issues before deciding what data management services to offer, rather than basing it on preconceived ideas or what other libraries are doing."

3. ***What is the one thing you wish you had known about being a science data librarian when you started your position?***

Five major themes emerged for this question, which we will frame as completing the stem sentence, "I wish I had known …"

a. ***…that the sciences are so diverse***

> Amy Hodge: As soon as we started talking to people about the repository, they began saying things like "I have a thousand files" or "I have 30 TB of data" or "I have protected information" or "I want my deposits to be automatically updated from my GitHub repository." There is always something new or extra that people want and sometimes providing those services is expensive, hard, time consuming or all of the above.
>
> Heather Coates: You can't really talk about scientific data management practices in any reasonable way. They just don't exist in a consistent way across all scientific research. When we have these high-level conversations about these issues, we tend to gloss over those disciplinary or method specific differences. Don't expect a one solution fits all kind of service.

b. …that disciplinary attitudes toward data sharing vary so widely

> Jenny Muilenburg: The widely variable cultures of data sharing are a particular challenge. Some disciplines have been sharing their data for years. As an example, the GIS community has a culture of sharing their data, sharing certain files so you don't have to recreate them. In other fields that culture doesn't exist yet. You run into different "personalities" in the fields and in labs regarding data sharing and data management. You'll quickly realize that you may have to come up with a customized solution for individual researchers or labs because of their culture.
>
> Dana Bouquin: Everyone wants to use everyone else's data but no one wants to share their data. It's a horrible oxymoron. There is so much cultural push back in medicine to the open data idea. I think it's mostly a literacy issue. Especially in medicine people don't necessarily realize that they can maintain their rights to their data, that they can tightly control what people can do with it if they want and that it's actually a good thing if you share—if you share it you get credit for it and people have to cite you.

c. …to focus on tools and new processes that people will really use

Sally Gore cautioned about investing a lot of time in developing tools that researchers say they want: "Trying to build something that people will actually use to meet their needs is a hard thing to do. If the resource or the tool isn't something that's easily adopted and integrated into an existing workflow, it's not useful. Or used."

d. …how researchers view data

Many respondents wished they had a better understanding of how scientists deal with data in the research process. Some expressed that it would have been helpful to have more exposure to some of the underlying science. Others described a time-scale difference, with librarians focusing on the whole data lifecycle and researchers caring about the day-to-day management of data coming in from all the different researchers and projects in their labs.

e. …the importance of institutional and political awareness

Several indicated the importance of administration and stakeholder support for library involvement in data management as a key factor for success. Some who were new to libraries commented on adjusting to how bureaucratic and political a librarian's position can end up being on a day-to-day basis. Surkis stressed that "at least as important as any subject knowledge is institutional awareness so that you can see where the opportunities are. Making connections and talking to people throughout the institution is probably the most important thing you can do."

4. What is the most challenging aspect of supporting scientific data services?

Respondents identified many of the same challenges outlined in the Kellam chapter including "the difficulty of determining which levels of service to offer, in addition to finding the resources with which to support those levels [and] … finding the cutoff point."[3] One respondent expressed it as negotiating the balance between serving the needs of the entire campus versus meeting the complex needs of any one particular group. Infrastructure challenges identified include scalability of services, storage, and sustainable funding models. Lynn Yarmey strongly endorsed "collaborative models for open, distributed, community-based networks" as the way forward. Many respondents echoed Carly Strasser's view of the biggest challenge:

> The biggest challenge facing libraries in this space is a total lack of understanding on the part of researchers about the role of the library. There is a perception that libraries are about books and help desks. I've spoken to tenured professors at major universities who had no idea that their library was involved in anything that had to do with data. Researchers won't show up at the library's doorstep—libraries have to advocate for their role. Librarians should aggressively collaborate with other campus stakeholders, like the vice chancellors of research or the grants offices.

Respondents identified getting buy-in from researchers as a big hurdle. Even getting time to meet with them can be difficult. Surkis' advice is, "Be with or ahead of the curve on what they need to know. Be really careful not to talk to them from a library-centric place using terminology that is not meaningful to them. Try to understand their world, how they think about this stuff and communicate to them where they are." Heather Coates agreed about the challenge and had some additional advice:

> I learned in my first couple years that most of our faculty don't care about data management. It's an activity they have to do but what they care about are their priorities, the things that interest them or are rewarded. I try to find out what those priorities are—funding for a new project, meeting promotion and tenure criteria, etcetera. I tend to start with the incentives and work backwards.

Sarah Oelker identified another big challenge. "The nut we haven't cracked yet is, what structure will encourage them to be thinking about it in small doses routinely rather than in a rush at grant time? Educating graduate students will generate a generation of PIs who are used to thinking about it."

Respondents working with biomedical researchers remarked on some special issues with medical and health data that can best be addressed at the institutional level. Barrie Hayes identified the lack of consistency of data management practices across projects as the biggest challenge, along with the data being sensitive and requiring special handling. Dana Bouquin observed:

> At the institutional level, I'd like to see a shift—even though [biomedical] funders don't necessarily require data management plans, how can we make data curation tasks part of our IRB [Institutional Review Board] process? How can we integrate this into researchers' regular workflows as opposed to trying to just create impetus? How can I facilitate you producing documentation for the institution, not necessarily the federal funder?

5. ***If you could give a new science data librarian advice, what would be the first thing you would say?***

We have organized these into ten themes, and have combined answers from both the survey and the interviews in this section.

a. So what exactly is a data librarian?

More than one respondent commented that no one knows what a data librarian or databrarian is, so librarians should not focus on trying to build a brand. Instead they should focus on getting services in front of researchers and be prepared with a simple explanation of what a data librarian does. Chris Eaker's answer to this question is, "I do what a traditional librarian normally does for books, but I do it for data. I catalog, organize, describe, preserve and make [data] available for reuse." There is still a lot of confusion about the role of libraries in data management, and researchers do not necessarily see why they should talk to a librarian about data. Even those librarians with domain expertise and relevant degrees may encounter resistance or skepticism from researchers, and it may be that much harder for those who lack these credentials. Sharing specific examples and success stories can help to counter this attitude.

b. Be patient.

Many respondents shared that they were surprised at how long it takes to develop connections.

> Chris Eaker: The most surprising thing to me was how slow the process was of getting a data curation program off the ground and starting to accept datasets into our repository. It's going to take a long time and researchers may not always be cooperative about wanting to share and archive their research data. Be

willing to be patient and take things slowly. Try to explain the benefits to them as researchers and to science. Be willing to make concessions, such as making data publicly available two or three years from now.

Amy Hodge: How long the time horizon is for things that go on in libraries. In the business world that I was used to, things get done very quickly, because businesses aren't sure if they are going to be around next quarter. Libraries are planning to be here in one hundred years, so there is a distinctly slower—but more thoughtful—pace to the activities they undertake.

c. Get out of your office.

Get in touch with your liaison librarians and research deans. Find champions among researchers who can help promote your services to their colleagues.

Amy Hodge: I sense a tendency for some librarians to sit in their offices and expect that people will come to them. I recommend that you go to seminars and hear about the kind of research that is happening on your campus. Meet in researchers' offices or in their labs so you can get a better feel for the kind of work people are doing. See them in person in their space and in their environment.

d. Work from the bottom up.

Carly Strasser: Immerse yourself in the community of researchers you are trying to support. Ask the intro-level grad class professor if you can sit in on his class. You could also casually visit lab groups. Ask them about their work, and introduce yourself and why you're interested in what they do. Talk to graduate students about what they do with their data, and offer to help work through issues they might have. A lot of librarians aren't comfortable inserting themselves in the researcher community at their institution, but it's the only way to really understand what your constituency needs, and shows them you are committed and interested.

e. …and the top down.

Regina Raboin: A new science librarian needs to be or learn to become comfortable speaking with your library administration and the administrators at your school, such as the VP for research, provost for research, deans, and department chairs. You must read and learn about the institution's bigger picture, understand its critical mission and how your library's mission

> melds with it. You need to be able to outreach to the offices within research that are charged with moving forward the institution's research interests and mission. Become very familiar with the federal guidelines and agencies and how their policies are going to impact your university.

f. Get your hands dirty.

New practitioners need exposure to working with data, and familiarity with what researchers really do.

> Dana Bouquin: Make sure you have basic numeracy under your belt. With data you can't just talk about how to manage it, catalog it, [and] make sure it is clear and reproducible. You are going to have to talk to people about all the steps of their workflow: the analysis steps, the subsetting, the data clean up, etcetera. Get as much hands-on experience actually cleaning up and manipulating data as you can. Get your hands dirty, get a dirty dataset, and clean it up.

> Sally Gore: If you haven't done so already, do some research yourself. Find a topic of interest to you. Go through the process of collecting data, managing data, building your own dictionary, working through all of the issues and problems that come up. Even if you are already working in a library, do a research project in the library.

g. Find your community then grow it.

Respondents were practically unanimous in stressing the importance of finding a support network of people in similar positions at other universities. There is a very active group of data librarians on Twitter.† Yarmey reminds us to think of the data community broadly: "One of the fantastic things about this work is that most data scientists are generous, open and really interested in actual collaboration. Take advantage of that."

Respondents suggested engaging in cross-organization activities as much as possible such as participating on campus task forces or committees, faculty governance (if librarians hold faculty status), and research compliance training. The library data community is currently separate from other data communities, and is also internally divided (e.g., science v. social science and the data discovery v. data management areas). The more we can do to bridge these divides, the better off the data community and the research communities we are trying to support will be.

---

† Use the hashtag #datalibs to begin finding them as well as others listed on Sherry Lake's web page, https://sites.google.com/site/dmsocialnetworks/twitter.

h. Don't overextend yourself.

They also suggest that librarians need to have clear definitions of available services and realistic expectations of the capacity to fulfill obligations such as devoting time to specific research projects versus merely advising on available tools. Over-commitment is an issue especially for librarians with job responsibilities in addition to data management.

i. Don't be intimidated if you don't have a science background—it's very helpful but not a prerequisite for success.

> Amy Hodge: I think that if you haven't done scientific research it will be much harder to understand how it works and what the mindset is and how to relate to researchers. The practice of science is often quite different from the idealized theory of it. It's much more about personal interactions as opposed to technical knowledge than I had expected. I didn't realize I was going to spend so much time in one-on-one conversations with faculty and getting to know people. It turns out that the people skills are the most important part of my job.

j. Learn one statistical software package.

> Jenny Muilenberg: People in the sciences are using a wide variety of software, and there is no way that a science data librarian could be familiar with every kind that there is. Having said that, it's important that you have at least basic knowledge of one type of software, like R for statistical analysis or ArcGIS for GIS analysis.

6. What training would you advise them to take advantage of?

Respondents were divided on the benefit of taking basic science courses with a lab component. There was general enthusiasm for online and short course curricula for research data management. Some respondents also recommended Software Carpentry (https://software-carpentry.org/) and Data Carpentry (http://datacarpentry.github.io/) workshops, and said the value lay in developing a sense of the capacities of the software, meaning, not to worry about trying to become an expert after one workshop. Science boot camps were lauded for the opportunity to better understand the researcher perspective and were highlighted as especially useful for those without a science background.

Important competencies included appraisal, metadata creation, preservation, project management, communication and negotiation skills, statistical knowledge, data analytics, numeracy, and programming. One respondent observed that there is not as much training available specifically for biomedical librarians while another talked about taking advantage of bioinformatics cer-

tification. More biomedical-focused training is expected as an outcome of the recently awarded National Institutes of Health BD2K (Big Data to Knowledge) grants.[4] Multiple respondents stressed the importance of continuing education and we expect this to hold true for the foreseeable future. The data librarian community is a great resource to help you keep up with the ever-changing data landscape.

7. ***Please discuss the balance across your work between data, science, and librarianship.***

This question provoked a wide range of responses. On the practical level, responses partially correlated with the number of staff tasked with data support at their institution. One data librarian exclaimed "You mean the lack thereof?" while a member of a multi-person team talked about the distribution of these areas of responsibility across the team. Several data librarians said they did not do any traditional library work, and informationists had duties typical in their setting but unusual in academic libraries, such as systematic reviews, research evaluations, and support for collaboration and team science.

On the philosophic level, there was great diversity in the ways people view their jobs; some see these aspects as all interwoven with data as just another dimension of being a librarian while some saw them as quite separate.

Here are a few of the different perspectives:

> Katie Houk: Librarianship is its own domain of specialized knowledge and skills. You don't necessarily need to have a higher degree in science, or currently be doing scientific research yourself, to bring data or information management best practices to the table and help scientists better deal with these issues.
>
> Heather Coates: This is just a different facet of librarianship. Data literacy can be provided using similar structures and services in place for information literacy. The services that I offer are still reference and instruction and research support and consultation. Only the content is different.
>
> Dana Bouquin: I don't think they are independent at all. Librarianship is about fostering the creation of new knowledge and creating communities. Data science is intensively collaborative, a team science. Getting people to better communicate about their work so they can better execute on their ideas is what librarianship is about. It's a continuum, not a balance.

8. ***Is there anything you would like to add that I haven't asked?***

There were a wide variety of responses to this catch-all question. Many respondents reflected on the challenge of adding data services with no additional or limited staffing. "Adding data to a job that is already chock full is not a strategy for success," "Library administration needs to figure out how to bring sufficient resources to bear and to balance the load," and "There has to be a commitment to data services at the institutional level" were three of the comments. One solution may be building a team. Philippa Broadley spoke about this approach at her university:

> Upskilling our liaison librarians to deal with research data management issues that are arising more and more frequently in their everyday jobs is an exciting opportunity. Our liaison librarians' willingness to embrace this new part of their job is fantastic and makes for a better support service for researchers, especially as they are the "face of the library."

Yarmey encourages a big-tent approach to these challenges:

> Think big in the sense that many people inside and outside of the library community are working on understanding and addressing data challenges. Doing this on your own isn't necessarily sustainable personally, professionally, or programmatically. With those two things in mind, how do we collaborate across communities and get this done together? Find something that you do well and enjoy doing and then contribute it to something bigger.

Observations from the Social Science Perspective

A few respondents touched on ways that data librarianship in the sciences compares to the same specialty in the social sciences. When asked whether they found the week-long workshop on data services hosted by ICPSR to be helpful despite its focus on the social sciences, one interviewee strongly agreed, stating that the principles of data management are pretty much the same for the social sciences versus sciences, and generally applicable regardless of discipline. Margaret Henderson noted:

> Data librarianship is different in the science and social sciences context. It shares a lot of similarities. It's not just [that science is] bigger and [has] more disciplines. The sciences, even broadly, want something different out of data services than the social sciences do. In the social sciences, people explore for

> data because there is a lot of it, but it can be a little difficult to find. They have an idea of the kind of data they want and they have to go out and look for data that matches their question. In the sciences there are few things that everyone uses. Any observational data that is big everybody knows about it and where to get it. Their focus is how you're generating your own data, where to put it. I wasn't prepared for the backend things that scientists wanted, like where to put things and how to comply with …funder mandates.

Henderson and author Michele Hayslett have both observed that scientists are generally more likely to create their own data from observations or measurement ("primary" research data, akin to primary source materials), whereas social scientists are more likely to be able to reuse data created by someone else (i.e., "secondary" data). ICPSR offers perhaps the longest-running training program for new data librarians† and, because ICPSR historically archives social science data, that program focuses on helping researchers obtain secondary data. But if a data librarian focuses only on secondary data, he or she will never connect with a substantial proportion of the researchers at their institution.

Nevertheless, the practices of assigning metadata and managing datasets are virtually identical in the sciences and the social sciences. Outreach to researchers, finding champions and networking on campus and with data colleagues are identical for the two domains. The day-to-day tasks and tools of data librarians are largely the same regardless of the disciplines on which they focus. The advice we would offer colleagues new to the specialty is also largely the same. As a result, data practitioners in the sciences and social sciences are moving closer and closer to one another professionally all the time.

Conclusion

We conceived of this chapter as the scientific counterpart to Kellam's focus on the social sciences. However, this distinction, at least within data librarianship, may be less and less relevant over time. There is a growing acknowledgment of the centrality of data to all disciplines in the digital age, regardless of the diversity of definitions of data. As such, data services provide a way for librarians and their libraries to advance their institution's goals and to prepare the next generation of scholars, be they scientists, social scientists or humanists. The main implication for administrators is the pressing need for additional resources to support this evolving and growing area of practice. Adding data as one more subject liaison area of responsibility is a fine start-up model but is neither scalable nor sustainable. The takeaway

† The summer program as of 2014 was called *Providing Social Science Data Services*.

for LIS students is the broad applicability of this skill set and that having a science background is extremely helpful, but lacking one need not be an impediment to becoming an effective provider of scientific data services. There is a welcoming and supportive community of data librarians to help nurture new practitioners.‡

1. Lynda M. Kellam, "Data librarianship: a day in the life," in *Numeric Data Services and Sources for the General Reference Librarian*, ed. Lynda Kellam and Katharin Peter (Oxford: Chandos Publishing, 2011), 151.
2. Ibid., 151.
3. Ibid., 154–5.
4. "BD2K Home Page: Data Science at NIH," last modified August 15, 2015, accessed August 2015, https://datascience.nih.gov/bd2k.

‡ Kellam's "Data librarianship: A day in the life" inspired this chapter. Kellam interviewed data librarians supporting the social sciences through statistical and numeric data services. In addition, the authors thank Matthew Leavitt, Erin Moore and Amanda Tickner for their assistance. Finally, we thank the participants for taking the time to share their insights, their passion and their eloquence.

Appendix 21.A. Interview Subjects

Note: Libraries outside of institutions of higher education and outside of the U.S. do not have Carnegie Classifications.

Name	Title	Level	Institution
Daina Bouquin	Data & Metadata Services Librarian	Special Focus Institutions—Medical schools and medical centers	Weill Cornell Medical College of Cornell University
Kristin Briney	Data Services Librarian	Research Universities (high research activity)	University of Wisconsin–Milwaukee
Philippa Broadley	Research Data Librarian	N/A	Queensland University of Technology
Heather Coates	Digital Scholarship & Data Management Librarian	Research Universities (very high research activity)	Indiana University-Purdue University at Indianapolis
Andrew Creamer	Scientific Data Management Librarian	Research Universities (very high research activity)	Brown University
Kiyomi Deards	Assistant Professor	Research Universities (very high research activity)	University of Nebraska-Lincoln
Chris Eaker	Data Curation Librarian	Research Universities (very high research activity)	University of Tennessee
Vessela Ensberg	Data Curation Analyst	Research Universities (very high research activity)	University of California, Los Angeles
Chris Erdmann	Head Librarian	Research Universities (very high research activity)	Harvard University
Jane Frazier	Data Librarian	N/A	Australian National Data Service
Sally Gore	Research Evaluation Analyst, UMass Center for Clinical and Translational Science	Special Focus Institutions—Medical schools and medical centers	Lamar Soutter Library, University of Massachusetts
Barrie Hayes	Bioinformatics and Translational Science Librarian	Research Universities (high research activity)	University of North Carolina at Chapel Hill

Margaret Henderson	Director, Research Data Management	Research Universities (very high research activity)	Virginia Commonwealth University
Amy Hodge	Science Data Librarian	Research Universities (very high research activity)	Stanford University
Katie Houk	Health & Life Sciences Librarian	Research Universities (very high research activity)	San Diego State University
Michelle Hudson	Science and Social Science Data Librarian	Research Universities (very high research activity)	Yale University
Sherry Lake	Data Specialist	Research Universities (very high research activity)	University of Virginia
Jenny Muilenburg	Data Services Curriculum and Communications Librarian	Research Universities (very high research activity)	University of Washington Libraries
Sarah Oelker	Science Librarian	Baccalaureate Colleges—Arts & Sciences	Mount Holyoke College
Regina Raboin	Associate Director for Library Education and Research	Special Focus Institutions—Medical schools and medical centers	University of Massachusetts Medical School
Yasmeen Shorish	Physical & Life Sciences Librarian	Master's Colleges and Universities (larger programs)	James Madison University
Carly Strasser	Program Officer; Formerly Research Data Specialist	N/A	Gordon and Betty Moore Foundation; California Digital Library
Alisa Surkis	Translational Science Librarian	Research Universities (very high research activity)	New York University, Health Sciences Library
Brian Westra	Lorry I. Lokey Science Data Services	Research Universities (very high research activity)	University of Oregon
Amanda Whitmire	Data Management Specialist	Research Universities (very high research activity)	Oregon State University
Stephanie Wright	Data Services Coordinator	Research Universities (very high research activity)	University of Washington Libraries
Lynn Yarmey	Lead Data Curator	N/A	National Center for Atmospheric Research