

Developing Data Services Skills in Academic Libraries

Justin Fuhr
General Librarian
Elizabeth Dafoe Library
25 Chancellor's Circle
University of Manitoba
Winnipeg, Manitoba, Canada R3T 2N2
Tel. 204-290-0083
Justin.Fuhr@umanitoba.ca

ABSTRACT

Introduction: Research data services are increasingly offered by academic libraries. Librarians may need to upskill to provide sufficient services or to build capacity within their institution. This study measures the current level of data services skills of academic librarians and explores their preferred method of continuing education to support data services.

Methods: A twenty-two question online survey was circulated. Respondents self-assessed their data skills in four categories: general data services, programming languages and software, library instruction, and soft skills. Likert-scale questions were also used to measure how important various professional development initiatives were to respondents.

Results: There were 120 responses to the survey. There is statistical significance between percentage of time providing data services and higher levels of technical data services skill.

Conclusion: Academic libraries that plan to or currently offer technical data services should hire dedicated data librarians. In addition, training should be provided for librarians who provide any level of data services, with respondents preferring workshops, communities of practice, and mentorship with peers.

1. INTRODUCTION

Library and information studies (LIS) has long emphasized continuing education as the field moves quickly and regularly adopts new services. Librarians may need to acquire skills to offer new services to their users. Professional development is especially relevant in new areas of academic librarianship where librarians may not already have the skill sets required to adequately provide services.

One new area of academic librarianship is research data services, or simply, data services. Tenopir, et al. define data services as “services that address the full data lifecycle, including the data management plan, digital curation (selection, preservation, maintenance, and archiving), and metadata creation and conversion.”¹ An academic librarian requires a varied skill set to provide data services, a skill set which may include data mining, metadata knowledge, technical details of repository hardware and software, legal and policy knowledge, research consultations, among others.

Increasingly, digital research data is produced in research post-secondary institutions. As more and more digital data is produced, it is increasingly important to manage and organize data, so it remains accessible in the future. Studies have shown there has been growth of data services in academic libraries over the past decade.² Tenopir, et al. recognize some institutions have hired specialized data librarians to provide library data services, however other librarians have found themselves reassigned to this role, with variable amounts of time spent providing data services. In these cases, it is sensible to provide librarians with training to enhance existing skill sets.

Using an online survey of academic librarians, this study measures the current level of data services skills of academic librarians. Additionally, this paper explores the preferred method of professional education in support of academic librarians providing data services.

2. LITERATURE REVIEW

2.1 Data Services as an Emerging Area of Academic Librarianship

Managing research data has come to the forefront of academic library services as more research data are created digitally and post-secondary institutions adopt and accept new ideas and trends like the open science movement. Academic librarians have positioned themselves to offer these services at the same time as data services are adopted by many post-secondary institutions. Researchers have described how academic librarians are integral to providing research data services.^{3,4} Pryor and Donnelly also recognized early on the role librarians could play in data services.⁵ They describe the 2008 Research Data Management Forum where attendees envisioned four data-related roles: data librarians alongside data creators, data scientists, and data managers.

LIS researchers have made a convincing case for academic librarians to provide research data services, and many academic libraries have demonstrated that librarians can confidently provide data services.^{6,7,8,9,10} There is a recognition that academic librarians, who have long provided services to access and preserve research outputs, are a natural fit to manage inputs such as data.

As research data services picked up momentum in academic libraries, Tenopir, Birch, and Allard's seminal Association of College and Research Libraries' (ACRL) survey was published.¹¹ The survey invited library directors of ACRL institutions to respond and provide a baseline of what data services were offered in the early 2010s, and what services would be offered in the future. While a very small number of institutions were offering data services at the time, the authors found a quarter to a third planned to in the near future. Tenopir, et al. also found the majority (71.1%) of data services providers were subject-area liaison librarians, compared to dedicated data librarians (5.8%).

Just a year later, Tenopir, Sandusky, Allard, and Birch published results of a survey sent to librarians working at Association of Research Libraries (ARL) member libraries.¹² While they found almost 75% of libraries do not offer data services, the majority of survey participants responded that they believe they have the skill set to support such services moving forward.

The growing prevalence of academic library data services is shown in large empirical studies and institutional analyses, undertaken to determine the extent of academic library data services in the

early 2010s,^{13,14} and in a follow-up analysis in Tenopir, Sandusky, Allard, & Birch, using the same data from Tenopir, et al.'s study in 2012.¹⁵ In their 2014 survey, Tenopir, et al. found close to 83% of North American academic libraries plan to offer data services by 2016. However, only 31 of 99 libraries offered data services training to library staff.

Subsequent studies have shown continued research data services offered in academic libraries.^{16,17,18} In particular, the ACRL white paper by Tenopir, et al. finds data services growing, and acts as a follow-up study to the 2012 study by Tenopir, Birch, and Allard. The authors find 44.1% of respondent libraries are not providing any research data services. This is down from 60% of libraries not providing any data services in 2012, meaning a considerably higher percentage of libraries are offering these services when compared with the previous survey. Tenopir, et al. divide data services into two categories: informational and consultative services (e.g. consulting with faculty about data management plans), and technical and hands-on services (e.g. managing data repositories). Like the results from their 2012 survey, in 2019 libraries more commonly offer informational/consultative data services compared to technical/hands-on services. This is important to keep in mind as we explore training required for academic librarians to provide data services.

2.2 Challenges in Providing Data Services

With research data services becoming more common, there remains significant challenges to overcome. As with any newly emerged service, it is inevitable that there is some period of transition. Authors have noted lack of skill and confidence^{19,20,21,22}; financial limitations such as equipment costs^{23,24}; inadequate staffing^{25,26}; and little to no institutional support.²⁷

Among these challenges, ensuring academic librarians are trained and possess an appropriate skill set is a particularly significant challenge. Research has been undertaken to determine what competencies are needed to provide full, mature academic library data services. With actionable competencies, training can be provided to bridge skill gaps that exist. Federer has completed work to develop data services competencies, which aids librarians who wish to develop their skill set in this area.²⁸ Along with a list of competencies, Federer found two clusters of survey participants emerge: *subject specialists* (specialize in a specific subject and focus on a smaller number of tasks) and *data generalists* (broader range of tasks who work broadly across disciplines).

Federer, Foster, Glusker, Henderson, Read, and Zhao review the Medical Library Association's Data Services Competency framework, a framework which prepares librarians to provide data services.²⁹ The competency framework includes five performance indicators: applies principles of data literacy; establishes and advances data services; supports research data best practices across the data lifecycle; applies knowledge of research methods, research ethics and rigor, and open science practices; and provides training and consultation for data-related topics.

2.3 Bridging the Research Data Services Skill Gap

From the literature emerge useful solutions to bridge research data services skill gaps. Attebury suggests "activities that include the characteristics identified by participants as contributing to meaningful or transformational activities. These include sustained and interactive activities with opportunities for reflection."³⁰ However, given the time and cost investment in such learning

activities, these may be restrictive. Two training initiatives, (1) LIS graduate school programs and courses, and (2) workshops, are reviewed below.

2.3.1 LIS Graduate School Programs

Some researchers suggest reforming LIS graduate school programs to provide additional opportunities for research data services training. Lyon suggests two ways LIS programs can do this: define core competencies for research data services to add to the current curriculum, and analyze LIS program applicants for background in STEM fields.³¹ Lyon echoes this sentiment in later work from 2015 and Lyon and Brenner suggest immersive experiences and practicums for graduate students interested in research services.^{32,33}

In addition to her work in developing data services competencies, Federer suggests MLIS graduate students should be afforded the opportunity to take relevant data-related courses throughout their programs.³⁴ “Library schools and professional organizations should...stay up to date on trends,” Federer states, “in this rapidly evolving field to ensure that their curricula and continuing education programs are suitable to prepare information professionals to take on new data librarian roles.”³⁵ Heidorn also recommends data services training should take place in LIS graduate programs. He lists the University of Illinois, University of North Carolina, and the University of Arizona among the post-secondary institutions that offer this training.³⁶

Ma, Stahl, and Knotts’ excellent scoping review of an updated curriculum for the current health information professional includes data management among the nine roles that health sciences librarians engage in.³⁷

Though data-related courses are not offered at every ALA-accredited institution, librarians involved with the Research Data Management Librarian Academy created a document listing 163 data-related courses available throughout North American, Asia Pacific, and European post-secondary institutions offering LIS graduate programs.³⁸

2.3.2. Workshops and Courses

Workshops can be useful to have a group of library workers receive a large amount of training in a relatively short period of time. One example is a 2008 full-day workshop offered by Michael Witt (Purdue University) and Melissa Cragin (University of Illinois at Urbana-Champaign) on institutional data repositories.³⁹ Southall and Scutt describe a two-part workshop offered for library staff at the University of Oxford.⁴⁰ The workshop is intended to train library staff in the general principles of contemporary data services.

In 2016, a course was developed at the University of California, Berkeley Libraries to train their library staff in not only general data principles, but putting them in context into different subject areas and delivered by subject liaison librarians.⁴¹ The authors conclude the course was very successful, suggesting that other academic libraries would do well to offer something similar, depending on data services capacity, finances, and time.

Federer and Qin review a 1.5-day workshop offered by the National Library of Medicine (NLM) in 2019, entitled “Developing the Librarian Workforce for Data Science and Open Science”.⁴² Fifteen librarians and faculty attended, with a variety of expertise in both subjects, with the goal of developing training for a library workforce in both data services and open science.

In addition to the NLM workshop, the NLM, with support from the National Network of Libraries of Medicine Training Office, offers the online course, “Research Data Management On Demand,” consisting of four standalone classes.⁴³ There is also the Research Data Management Librarian Academy (RDLMA), a collaboration between Elsevier, Harvard Medical School, Harvard Library, Simmons University, Boston University, Brown University, Massachusetts College of Pharmacy and Health Sciences University, Northeastern University, and Tufts University.⁴⁴

Though competency frameworks and training initiatives for data services are coming into view as data services matures, the literature does not include self-assessment of current data services skills on which to benchmark the success or failure of current and future training opportunities.

3. AIMS

This paper measures the current level of self-described data services skills of academic librarians. Taking academic librarian’s current level of skill, this paper then explores the preferred method of professional education to support academic librarians providing data services.

The results of this research study will be of interest to academic librarians providing data services, academic library administrators, institutions increasing their data services capacity, and graduate students considering data librarianship as a potential career path.

4. METHODS

4.1 Study Concepts and Scope

For this study, library data services are defined as library services to manage researcher data, whether that is informational or technical in nature.⁴⁵ Academic librarians that are employed at a university or college library at the level of “librarian.” The study is directed towards, but not limited to, academic librarians providing data services. This study was open internationally to respondents, but respondents came from one of four geographic regions: Canada, the United States, the United Kingdom, and Australia.

4.2 Data Collection and Analysis

A twenty-two question survey was developed and administered in LibWizard, the survey module of LibApps software. A survey was chosen as the data collection tool as the researcher wanted to reach as many respondents as possible using standardized questions. The survey is intended to be a census, by surveying a large population of academic librarians from Canada, the United States, and abroad.

The survey consists of Likert-scale, multiple choice, and short answer questions, in three parts. The first part of the survey collected demographic information. Multiple choice and short answer questions to collect information such as what type of library respondents work for, their location, current job title, and what percentage of time is spent doing data-related work, among others. In the second part, Likert-scale questions were used to measure the respondent’s self-assessment of specific data services competencies. Lisa Federer’s article⁴⁶ was consulted to develop some of the competencies and develop the four overarching categories: general data services, programming languages and software, library instruction, and soft skills. Within each of the four overarching

categories are associated skill sets, such as data mining and data curation in the general data services category, and oral communication and management and leadership within soft skills.

In the third and final part, Likert-scale questions were also used to measure how important different professional development initiatives were to respondents. Examples include self-directed learning, mentorships, workshops, webinars, and conferences. Short answer questions were used to gather additional information about the respondent's self-assessment of data skills and professional development initiatives.

The study has been approved by the University of Manitoba's Research Ethics Board. After completing the survey, respondents were invited to enter a draw for a \$50 (CDN) Amazon gift certificate by submitting their email address.

The survey was circulated on listservs (DataLibs, CANLIB-Data, Code4Lib, CdnLIS-L, univers@IFLA, ScholComm (ALA), Research Metrics list (ALA), dss-rdm_dg, MEDLIB-L, CANMEDLIB) and social media (Twitter, LinkedIn) over a period of two months beginning February 20, 2020. The survey data was analyzed using regression analysis in R. A model was run for each of the four skill set categories, with the following variables: geographic regions (Canada, U.S., U.K., and Australia), percentage of time spent providing data services, time spent in their current role, and time spent in the LIS field.

5. RESULTS

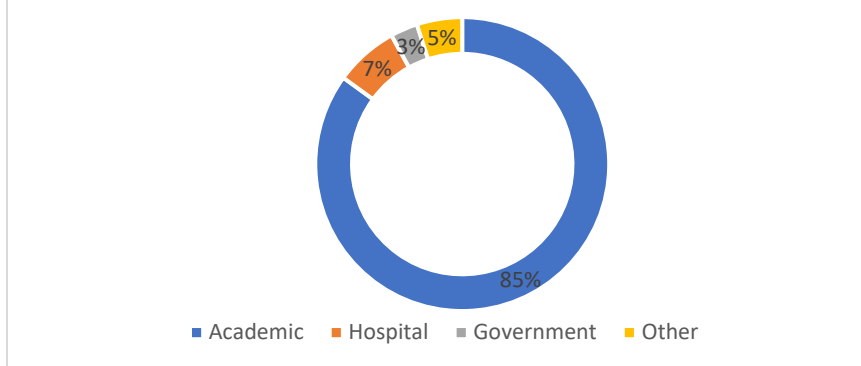
5.1 Respondent Demographics

In total, 120 respondents participated in the survey. Not all respondents filled out every question on the survey. As the survey was circulated over listservs and social media, the response rate is unknown.

On average, the respondents (n=120) reported that they have five years of experience in their current role. The mean years of experience for respondents (n=120) in the information studies field is twelve. Respondents (n=106) are from one of four geographic regions: 47.5% of respondents are from the US, 33.3% are from Canada, 5% are from the United Kingdom, and 2.5% are from Australia. 11.7% (n=14) of respondents have an unknown country of origin.

Out of 120 respondents, 85 percent work for an academic library, 7 percent work for a hospital library, 3 percent work for a government library, with the remaining 5 percent of respondents working for scientific research organizations, technology companies, and specialized libraries.

Figure 1. Type of library where survey respondents work



There is a large range of job titles for survey respondents. A sample of the respondent job titles includes: Data Services Librarian, Research/Education Librarian, Research Services Librarian, Software Developer and Librarian, Scholarly Communication Librarian, Data Curation Librarian, Health Sciences Librarian, Engineering Librarian, Research Repository Advisor, Web Librarian, among others.

In terms of the amount of time an academic librarian spends providing library data services (n=117), 52 percent spent between 0 and 24 percent of their time providing data services, 9 percent spent between 25 and 49 percent, 18 percent spent between 50 and 74 percent, and 21 percent spent between 75 and 100 percent.

5.2 Skill Sets

The survey had respondents rank twenty-five skill sets organized into four overarching categories: general data services, programming languages and software, library instruction, and soft skills. Each skill set has a corresponding Likert-scale, ranging from: no level of competence (0), low level of competence (1), average level of competence (2), moderately high level of confidence (3), and high level of confidence (4) (coded from 0 to 4 for data analysis). Respondents could also answer “not applicable (n/a)” for any skill set.

Many of the stronger skill sets came from the soft skills and library instruction categories, while many of the weaker skill sets came from the general data services and programming languages and software categories. The full listing of mean scores for individual skill sets is available in Appendix B.

Table 1: Mean score of strongest skill sets (n=120)

Skill set	Category	Mean score (n=120)
Written communication	Soft Skills	3.53
One-on-one session	Library Instruction	3.43
Working well with others	Soft Skills	3.4
Interpersonal customer service	Soft Skills	3.37
Library instructional sessions	Library Instruction	3.31
Oral communication	Soft Skills	3.29

In-class instructional sessions	Library Instruction	3.24
Project management	Soft Skills	2.78

As mentioned above, Tenopir, et al. divide data services into two groups: technical and informational/consultative services.⁴⁷ Throughout the 2010s, more academic libraries were offering informational/consultative services compared to technical.⁴⁸ Librarians may have higher skill in informational and consultative services since these services are more frequently offered by academic libraries and these services are more aligned with traditional library services such as library instruction. This survey's library instruction category is second only to the soft skills category in terms of average self-assessed scores (see Appendix B).

Table 2. Mean score of weakest skill sets (n=120)

Skill set	Category	Mean score (n=120)
Programming languages (JavaScript, R, Python, C++, etc.)	Programming Languages and Software	0.76
GIS software (e.g. ArcGIS)	Programming Languages and Software	0.77
Data mining	General Data Services Skills	0.94
Statistical software (e.g. SAS, SPSS, RStudio)	Programming Languages and Software	1.02
Text editors (e.g. Jupyter)	Programming Languages and Software	1.05
Data analysis software (e.g. OpenRefine)	Programming Languages and Software	1.15
Web development and maintenance	Programming Languages and Software	1.58
Markup language (e.g. HTML, XML)	Programming Languages and Software	1.78

5.2.1 Percentage of Time Correlates to Technical Data Services Skill

Table 3: Mean score of four main skill set categories, by time spent providing library data services

Time spent providing library data services (percent)	n (total n=117)	General Data Services Skills	Programming Languages and Software	Library Instruction	Soft Skills
0-24	61	0.89	0.60	2.06	2.44
25-49	11	1.94	1.63	2.11	2.5
50-74	21	1.88	1.08	2.89	3.17
75-100	24	2.22	1.5	2.72	3.44

When respondents (n=117) are divided into four groups by how much time they spend providing library data services, again there are stronger skills across all groups in library instruction and soft skills compared to the technical categories of general data services skills and programming

languages and software. However, there is a substantial increase in the self-assessed score for general data services skills in the group spending the most time providing data services (75-100 per cent of time spent providing library data services) compared to the other three groups.

Table 4. P-values for percentage of time spent providing data services and self-assessed score of the four skill set categories

Skill Set Category	P-value	Estimate
General Data Services	0.0000551	0.05617
Programming Languages and Software	0.00000532	0.08182
Library Instruction	0.2578	-0.01598
Soft Skills	0.8439	-0.002174

During data analysis, it was found the variable of percentage of time spent providing data services is statistically significant in regression models for the two technical skill sets: general data services, and programming languages and software categories. The more time a librarian spends providing data services, the higher their self-assessed score is likely to be for the two technical skill set categories.

5.2.2 Geographic Regions

When the respondents are divided into the four geographic regions (n=106), once again there is higher skill in library instruction and soft skills compared to general data services and programming languages and software.

Table 5: Mean score of four main skill set categories, by geographic region

Geographic Region	n (total n=106)	General Data Services Skills	Programming Languages and Software	Library Instruction	Soft Skills
Canada	40	1.73	1.14	2.82	3.02
United States	57	1.81	1.25	2.88	3.18
United Kingdom	6	1.79	0.29	2.58	3.36
Australia	3	2.67	1.21	3.39	3.83

Canadian librarians rank similar to the US, however the US has higher average ratings overall between the two countries (average of 5.42% comparatively across the four categories).

Looking at specific skill sets, the US has significantly higher scores in data analysis software (1.36) and data visualization (2.04), when compared to Canada (0.83 and 1.59, respectively). Academic librarians from the US have significantly higher (>5% higher of total score) skill sets in data mining (1.02), programming languages (0.93), text editors (1.19), and project management (2.79), when compared to Canada (data mining (0.79), programming languages (0.68), text editors (0.88), project management (2.55)).

It is difficult to conclude much from the other two geographic region, the United Kingdom and Australia, due to small sample sizes. After regression analysis of the data, no statistical significance

was found between being from any of the four geographic regions and skill levels in any of the categories.

5.2.3 Early Career Librarians and Experienced Librarians

Comparison between early career academic librarians (≤ 5 years) and the three other groups of experienced academic librarians (> 5 years) sees early career librarians with the lowest mean score in library instruction and soft skills, and the second lowest mean score in general data services and programming languages and software.

Table 6: Mean score of four main skill set categories, by years of experience

Years of experience in the library field	n (total n=120)	General Data Services Skills	Programming Languages and Software	Library Instruction	Soft Skills
0-5	37	1.71	1.15	2.52	2.91
6-10	26	2.25	1.42	3.16	3.39
11-19	28	1.74	3.48	3.5	3.64
20-35	29	1.58	1.06	2.77	3.3

Comparison of specific skill sets found experienced academic librarians have higher averages in every skill set except data curation, programming languages, statistical software, GIS software, data analysis software, and text editors, where early career librarians have higher average scores. Early career librarians with higher scores in specific technical skill sets could be attributed to people in other disciplines being hired for academic librarian positions due to demand for technical skill sets in emerging areas. Analysis of participants with advanced degrees (e.g. MA, MS, Ph.D) without an MLIS shows early career academic library workers with a much higher percentage. Six out of thirty-seven (16.22%) of early career academic library workers who filled out the survey have advanced degrees without a MLIS, compared to seven out of eighty-three (8.43%) for experienced librarians. This indicates people working at academic libraries from other disciplines, possibly due to new demands for skills.

Data analysis found no statistical significance between the number of years spent in the LIS field and skill level in any category.

5.3 Professional Development Initiatives

Respondents ranked twelve professional development initiatives according to how important they felt they were in data services training. Each initiative had a corresponding Likert-scale question ranging from: not important, low importance, average importance, moderate importance, and very important (coded 0 to 4 in data analysis). Respondents could also answer “not applicable (n/a)” for any initiative.

Table 7: Mean score of importance of professional development initiatives (n=120) (0=not important; 4=very important)

Professional development initiative	Mean score
Learning by doing (trial and error)	3.38
Self-directed learning	3.33
Workshop or bootcamp	3.19
Communities of Practice	3.04
Mentorship with peer (library staff)	3.02
Online courses	2.98
Attending conferences	2.83
Mentorship with researcher or other faculty (non-library staff)	2.69
Webinar	2.60
Courses during LIS graduate school	2.56
Job shadowing	2.11
Fellowship (e.g. early or mid-career fellowship)	2.1

Table 8: Mean score of importance of professional development initiatives, by time spent providing library data services

Time spent providing library data services (percent)	n (total n=117)	Most important (mean score in parentheses)	Least important (mean score in parentheses)
0-24	61	Mentorship with peer (library staff) (3.33)	Fellowship (e.g. early or mid-career fellowship) (2)
		Mentorship with researcher or other faculty (non-library staff) (3.33)	Courses during LIS graduate school (2)
		Workshop or bootcamp (3.33)	
25-49	11	Self-directed learning (4)	Fellowship (e.g. early or mid-career fellowship) (2)
		Learning by doing (trial and error) (4)	Job shadowing (2)
			Courses during LIS graduate school (2)
50-74	21	Attending conferences (3.67)	Fellowship (e.g. early or mid-career fellowship) (1.67)
		Online courses (3.67)	Job shadowing (1.67)
		Workshop or bootcamp (3.67)	
75-100	24	Workshop or bootcamp (3.67)	Webinar (2.33)

		Learning by doing (trial and error) (3.67)	
		Communities of Practice (3.67)	

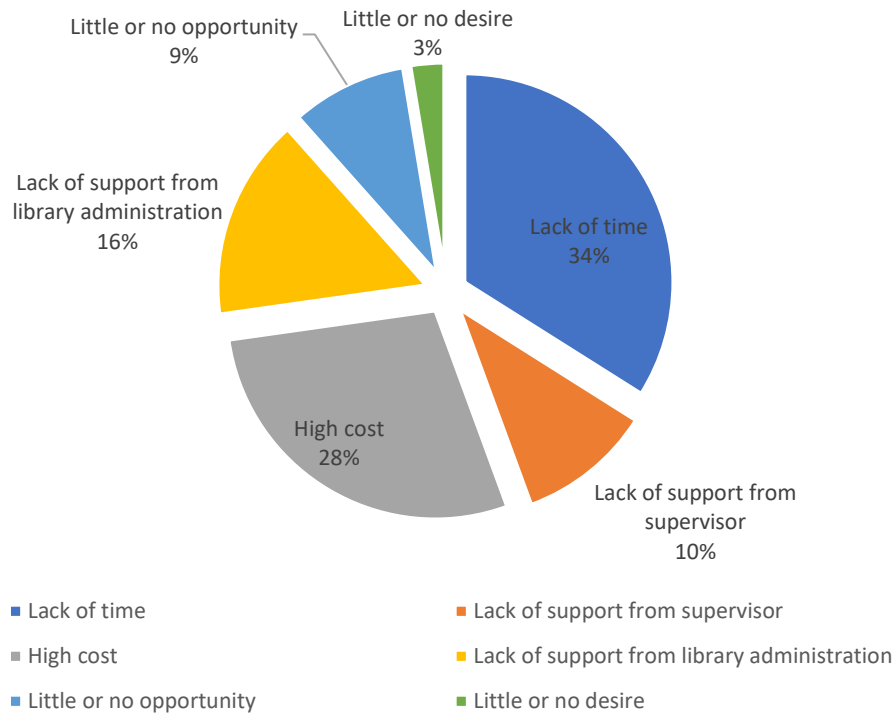
The two highest-rated professional development initiatives are independent: “learning by doing (trial and error)” and “self-directed learning.” While there is little cost to both of these independent learning initiatives, there is considerable investment in time and a need for support from supervisors. Indeed, both are noted by respondents as impediments to their professional development. Though researchers suggest one possible solution for bridging the data services skills gap is taking courses during LIS graduate school, it ranks tenth out of the twelve initiatives as being important for respondents.⁴⁹ One possible reason for the relative unimportance is due to not taking data services-related courses during the respondent’s time in graduate school, or not having the opportunity due to these courses not being offered until recently,

Respondents indicated impediments to their professional development, including lack of time, high cost, and lack of support from library administration and their supervisor.

Table 9. Impediments to professional development initiatives (respondents could select multiple answers)

Impediment	Responses (n=118)
Lack of time	91
High cost	76
Lack of support from library administration	42
Lack of support from supervisor	28
Little or no opportunity	24
Little or no desire	7

Figure 2. Impediments to professional development initiatives (n=118, respondents could select multiple answers)



5.3.1 Qualitative comments on professional development

Respondents also expanded on impediments to their professional development with text-based responses, including lack of support from their institution and library administration, lack of support from their peers, and lack of a defined role or projects:

Lack of support from institution and library administration

“[I]f the [institution] do not recognize librarian as a legitimate partner in this area. Several colleagues don't want to develop there [professional] skills in the direction of data services. It is a heavy inertial constraint. Also, all of our librarian managers are not enough literate with what implies the manipulation of data. Any coding seems complicated for them and they don't...want for the moment to [consider specialization] of some librarian. They want all librarian to move towards data. So the boat turns very slowy and may miss all the race"..."

“Institutional indifference until having an open data policy is linked to some larger strategic initiative like the Research Excellence Framework [available here: <https://www.ref.ac.uk/>]. Then there is a flurry of activity to 'evidence' good practice.”

“My library administration is so utterly lacking in vision that I can't implement most of what I take the time to learn. My current manager would be happy if I did what I did 20 years ago (which of course is not what my liaison departments want).”

“Where I still feel lacking is in taking the initiative to lead work at my institution. I lean heavily on leadership in my libraries, which tends to only really turn their attention to support of data

services when the topic has caught their attention in other channels that library deans, etc. tend to follow.”

“Lack of support from colleagues/resentment of colleagues who do not understand specialty.”

Lack of support and hostility

“Sustainability -- if I start something now, there is no guarantee that there will be ongoing support to continue or expand on that service.”

“Hostile attitude from peers to devoting time to PD.”

Lack of defined role or projects

“It's hard to identify training wheels" projects -- real world projects where I can *start* to apply new skills I'm developing. The path from novice to expert is basically a big map with 'here be dragons" and no more details.”

“It's difficult to know for sure what is the task for us (librarians) versus Data Scientist and even Data [steward] ... What is for us to learn (what is for Librarians) and what is to let to others...?”

“Working with data services can often be included as part of a role, rather than being a role itself (my previous post was an example where I did that along with Open Access)...So time, and importance of the work to managers...can make learning quite difficult.”

“I feel that sometimes the barrier to professional development in data services for librarians is the lack of clear 'end' location...For data services, so many librarians and information professionals are participating in disparate but related 'data things' that the typical librarian does not see a 'compass' to point the direction. The scariness of choosing one's own direction can be a barrier here, with the constant 'imposter syndrome' talk in library-land, but also the hesitation to act without clear precedent...It is often the internal (or external) quibbling of priorities and clear path for 'what is a library in the 21st century' that inhibit this area of growth for librarians and our ilk.”

“The challenge is that data librarians are not all doing the same thing and so there is no one key skill set.”

These quotations show support is needed for academic librarians who provide data services. It is clear from the survey data that practitioners rely on financially feasible but time intensive options like self-directed learning, but should also be provided with opportunities like workshops and mentorship with library staff.

6. DISCUSSION

6.1 Dedicated data librarians

It is encouraging to see academic librarians who are providing data services consistently have higher data services skills. This corresponds to Tenopir, et al.'s claim that “growth in current performance of [research data services (RDS)] by librarians can be expected to follow growth in current availability of RDS by libraries.”⁵⁰ As more data services are offered, and offered at a higher capacity, librarians providing these services become more skilled and adept.

Why are dedicated data librarians skilled in providing technical data services? Two reasons immediately jump out. Firstly, if an academic library has a vacancy for a position where the successful candidate will be spending seventy-five to one hundred percent of their time performing data services, it is likely the successful candidate will have previous technical data services experience. One example would be hiring developers, other IT staff, or otherwise people from other disciplines than librarianship, who already have significant technical skill sets. The survey data trends this way for newly hired librarians, as reported above. The survey data shows 16.22% of early career respondents have a graduate degree without an MLIS. However, this assumes a higher percentage of data librarians have recently been hired, which is more likely since data services is an emerging area of academic librarianship. Another example could be a data librarian who spent time, whether that was the majority of their time or otherwise, previously providing data services. Librarians with previous data services are more likely to be successful candidates for data librarian vacancies.

Secondly, dedicated data librarians, due to their specialized role, are more likely to have more opportunity (e.g. financial, time) to spend on training to bridge specific skill gaps. Additionally, data librarians may have more familiarity and confidence with technical data services due to more time spent providing the services, including time devoted to working with data tools common in technical data services. There may also be an expectation to improve your data skills in a dedicated data librarian position.

The more a librarian practices data services, the more skilled and confident they will be. This shows the value for academic libraries to hire dedicated data librarians, when those libraries are currently offering, or planning to offer, data services. These are positions where data librarians perform data services between seventy-five and one hundred percent of their time. Dedicated data librarians not only provide data-related informational and consultative support, but actively develop and promote data services initiatives in their library systems, consisting of services supporting all stages of the data lifecycle, from creation to discovery, to preservation and destruction.

There are numerous benefits to having a dedicated data librarian. Dedicated data librarians have higher technical skills, whether due to having a variety of technical experience or taking additional training. Academic libraries will increase their capacity to provide a higher level of data services to their users. Dedicated data librarians have the time to devote to concentrating solely on data services, to provide users with current services and for planning and coordinating new service offerings. This is important since, as Tenopir, et al. found academic libraries are looking to increase their data services capacity beyond informational and consultative data services.⁵¹ Dedicated data librarians can increase current services, to develop increased services for the future, depending on user need and demand, library and institutional priorities, financial considerations, among others.

Academic libraries have capacity to provide data services, including technical services. Since dedicated data librarians have higher data services skill sets, academic libraries are encouraged to create and retain these positions. Dedicated data librarians can focus on developing data services initiatives and programs and they have a higher capacity to do so when compared to librarians who have other duties and support other roles.

6.2 Data services training initiatives

Overall, academic librarians have average to moderate self-assessed scores across technical data services skill sets, and moderately high to high scores across library instruction and soft skills. While hiring dedicated data librarians can help, it is clear academic librarians need training opportunities in data services, especially technical skill sets, to improve their skill sets. Responses to the question of whether respondents felt proficient before they started in their current role resulted in answers such as “I never feel fully proficient, there's always more to learn and you never know what's going to be thrown at you” and more bluntly, “I do not feel proficient in data librarianship.”

The importance of professional development should not be understated. While there are high average scores for traditional academic librarian skills such as library instruction (2.86) and soft skills (3.18), technical skills in general data services (1.83) and programming languages and software (1.18) are low. Since technical skills are low, librarians need training to bridge this skill gap. It should be noted that depending on the specific data services offered, librarians may not need these skills. Individuals, library administration, and their institution will need to decide which skills their library staff need training in, based on current and future data services offered by their academic library.

Data services skills are overall similar across different geographic regions, which for this study includes Canada, the United States, the United Kingdom, and Australia. This indicates that where there is a need for more training, similar initiatives can be provided in these regions. Based on the survey data, independent learning, such as self-directed learning, should be emphasized, followed by workshops and bootcamps, communities of practice, and mentorship by other library staff.

While there is an emphasis in the literature on increasing data services training in LIS graduate school curriculums, during data analysis, there was no statistical significance found between years of experience and data services skills. The demographic most likely to have taken data-related graduate school courses, early career librarians, had lower average scores in every skill set except several specific technical skills: data curation, programming languages, statistical/GIS/data analysis software, and text editors. This could indicate lack of experience since they are still early in their career, or not receiving any or not enough training, and/or relevant training, throughout their program. There was also little recognition by respondents of the value of taking data-related courses during graduate school. This indicates the relative insignificance of these courses by respondents.

This does not mean graduate programs should scrap all current and future data-related courses. However graduate programs training future information professionals should assess their data-related course offerings. This could mean offering more data-related courses or shifting course content to different types of library data services. As well, different educational formats should be considered. For example, if an LIS program already offers data-related courses, they could offer work placements with data librarians. This could provide a different mode of training compared to for-credit courses, and students may find they also receive one of the most highly rated types of training: mentorship by other library staff.

Academic libraries should provide their staff with the specific skills they need to offer data services at a consistent level. For some academic libraries, this will be informational and

consultative services, or technical, and for others, a mixture of both. Training initiatives should be provided that give staff the skills to be confident in providing data services.

6.3 Range of data services

It is clear not all data librarians do the same thing. One survey response reads:

“[t]he challenge is that data librarians are not all doing the same thing and so there is no one key skill set. I've always thought of data librarians fitting into three different categories: (1) Acquisition librarians who focus on procuring, ingesting, and support access to data collections; (2) Analytical data librarians - the unicorns who do have program and analysis skills often gained from other jobs. They seem to focus on teaching and training basic skills to meet short term needs of researchers; and (3) Curation librarians who focus on end products like research output data, sharing, preservation etc. I don't think one person can effectively do all three at a high level.”

One only needs to view the survey respondent job titles to see the large range of different data services roles: Data Librarian, Research Services Librarian, Software Developer and Librarian, Data Curation Librarian, Research Data Management Analyst, STEM and Research Data Outreach Librarian, and so on. This is also seen in the differing scores of the various skill sets for data librarians. The survey data shows librarians are involved in different areas, and provide different levels, of data services. Federer aptly notes data librarians provide an array of diverse services, including data management, curation, preservation, visualization, and more. “[D]ata librarianship,” Federer writes, “may not be a single role but rather one that allows professionals to focus on areas related to their own interests or their users’ needs.”⁵²

Data librarians should focus their training on those data services an academic library currently provides, and will provide in the future, to increase and diversify their skill set in those areas. While one librarian who supports faculty and researcher data preservation may need training using preservation software, another librarian may need training in consultative data services, such as hosting library information sessions using the institutional and data repositories.

Academic librarians looking for positions providing data services should be mindful of what skills a particular data librarian position requires. As well, bear in mind academic libraries may hire outside the profession. Based on the survey data and summarized above, there is a higher percentage of early career librarians without an MLIS compared to librarians with six or more years of experience. While this could be explained by getting a MLIS while working as an academic librarian, it remains that academic libraries are filling at least some data-related positions outside the profession. Why is this the case? Academic librarians may not have the required skills for a position. Academic librarians should keep this in mind when looking at their long-term career goals and what training they take in data-related skill sets.

7. CONCLUSION

Data librarianship is an emerging field in academic librarianship, one which comprises a wide-range of services and different levels of expertise among practitioners. This study determined areas of strength and weakness of academic librarians providing data services. As well, the importance of different training opportunities to academic librarians is shown.

Statistically significant data shows academic librarians providing data services have higher scores in technical skill sets. Evidence is provided of the benefit for dedicated data librarians in academic libraries currently offering, or planning to offer, technical data services. It is this author's recommendation that academic libraries hire dedicated data librarians.

Academic librarians providing data services should be given the opportunity to take a mixture of independent learning and nonautonomous training initiatives, depending on their specific areas of expertise and what is available to them. As data services is an increasingly common academic library service, training needs to be provided to those academic librarians who are providing this service. Data librarians are diverse; there is no 'one-size fits all' model for the role data librarians play. The same is true for data services training initiatives; academic librarians should explore what data services their institution is currently offering or planning to offer and participate in training appropriate to their expertise.

Academic librarians have shown the value they provide to faculty, researchers, and students with data services. Data services has moved beyond 'emerging' and has emerged as a significant area for academic libraries.

7.1 Limitations

Limitations to this research include small samples sizes of two geographic regions in this study (UK, Australia), and the small number of regions surveyed overall (4). The survey was self-selecting, so there may be self-selecting bias. Also, survey participants were only presented with predefined skills to measure their competency in data services, as well as an English-language only survey. Potential survey participants was limited to the reach of listserv invitations and various social media posts.

7.2 Future Research

Future research could include a follow-up survey to update the skill set baseline of academic librarians, providing data services now or in the future. Future researchers could also track skill set assessment after training opportunities are provided to library staff. Future studies could also expand the scope of geographic regions, including those with small sample sizes in this study (UK, Australia) and increase the sample size for the survey overall.

Thought should be given to updating the measured skill sets which consider future data services. Adjusting the skill set categories based upon current uptake of library data services could provide more accurate data.

Acknowledgments

I acknowledge the generous financial support received from the Canadian Association of Research Libraries (CARL), through CARL's Research in Librarianship grant. I also acknowledge data analysis support, generously provided by Razvan Romanescu and Depeng Jiang at the George & Fay Yee Centre for Healthcare Innovation (Winnipeg, Manitoba, Canada). Thank you to my generous colleague, Nicole Askin, who provided helpful feedback when writing this article.

Funding

This work was financially supported by the Canadian Association of Research Libraries.

Notes

- ¹ Carol Tenopir, et al., "Academic Librarians and Research Data Services: Preparation and Attitudes," *IFLA Journal* 39, no. 1 (2013): 70, <https://doi.org/10.1177/0340035212473089>.
- ² Carol Tenopir, et al., "Research Data Services in Academic Libraries: Where are We Today?" Choice White Paper, no. 5 (2019), <https://www.choice360.org/research/research-data-services-in-academic-libraries-where-are-we-today/>.
- ³ Tony Hey and Jessie Hey, "e-Science and its Implications for the Library Community," *Library Hi Tech* 24, no. 4 (2006): 515-528, <https://doi.org/10.1108/07378830610715383>.
- ⁴ Elizabeth Lyon, "Open Science at Web-Scale: Optimising Participation and Predictive Potential," Joint Information Systems Committee (JISC), November 6, 2009, <http://www.jisc.ac.uk/media/documents/publications/research/2009/open-science-report-6nov09-final-sentojisc.pdf>.
- ⁵ Graham Pryor and Martin Donnelly, "Skilling Up to Do Data: Whose Role, Whose Responsibility, Whose Career?" *International Journal of Digital Curation* 4, no. 2 (October 15, 2009): 158-70. <https://doi.org/10.2218/ijdc.v4i2.105>.
- ⁶ Mark P. Newton, C. C. Miller, and Marianne Stowell Bracke, "Librarian Roles in Institutional Repository Data Set Collecting: Outcomes of a Research Library Task Force," *Collection Management* 36, no. 1 (2011): 53-67, <https://doi.org/10.1080/01462679.2011.530546>.
- ⁷ Sheila Corral, Mary Anne Keenan, and Waseem Afzal, "Bibliometrics And Research Data Management Services: Emerging Trends In Library Support For Research," *Library Trends* 61, no. 3 (2013): 636-674, <https://doi.org/10.1353/lib.2013.0005>.
- ⁸ Sheila Corral, "Roles and Responsibilities: Libraries, Librarians and Data," in *Managing Research Data*, ed. Graham (London: Facet, 2012), 105-133.

⁹ Liz Lyon, "The Informatics Transform: Re-Engineering Libraries for the Data Decade,"

International Journal of Digital Curation 7, no. 1 (March 11, 2012): 126–38,

<https://doi.org/10.2218/ijdc.v7i1.220>.

¹⁰ Lisa Federer, "Research Data Management in the Age of Big Data: Roles and Opportunities for

Librarians," *Information Services & Use* 36, no. 1-2 (2016), 35–43, [https://doi.org/10.3233/ISU-](https://doi.org/10.3233/ISU-160797)

[160797](https://doi.org/10.3233/ISU-160797).

¹¹ Carol Tenopir, Ben Birch, and Suzie Allard, "Academic Libraries and Research Data Services:

Current Practices and Plans for the Future," Association of College & Research Libraries (2012):

1–55,

http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch

[Allard.pdf](http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch).

¹² Carol Tenopir, et al., "Academic Librarians and Research Data Services: Preparation and

Attitudes."

¹³ Corrall, Kennan, and Afzal, "Bibliometrics and Research Data Management Services."

¹⁴ Andrew M. Cox and Stephen Pinfield, "Research Data Management and Libraries: Current

Activities and Future Priorities," *Journal of Librarianship and Information Science* 46, no. 4

(December 2014): 299–316, <https://doi.org/10.1177/0961000613492542>.

¹⁵ Carol Tenopir, et al., "Research Data Management Services in Academic Research Libraries and

Perceptions of Librarians," *Library & Information Science Research* 36, no. 2 (April 2014): 84–90,

<https://doi.org/10.1016/j.lisr.2013.11.003>.

¹⁶ Carol Tenopir, et al., "Research Data Services in Academic Libraries: Where are We Today?"

¹⁷ Andrew M. Cox, et al., "Developments in Research Data Management in Academic Libraries:

Towards an Understanding of Research Data Service Maturity." *Journal of the Association for*

Information Science and Technology 68, no. 9 (September 2017): 2182–2200,

<https://doi.org/10.1002/asi.23781>.

¹⁸ Alexandra Cooper, et al., “Institutional Research Data Management Services Capacity Survey: Executive Summary,” Portage Network, Research Intelligence Expert Group, January 2020,

[https://portagenetwork.ca/wp-](https://portagenetwork.ca/wp-content/uploads/2020/02/RIEGSurvey_execsummary_2020_02_12_EN.pdf)

[content/uploads/2020/02/RIEGSurvey_execsummary_2020_02_12_EN.pdf](https://portagenetwork.ca/wp-content/uploads/2020/02/RIEGSurvey_execsummary_2020_02_12_EN.pdf).

¹⁹ Tenopir, et al., “Academic Librarians and Research Data Services.”

²⁰ Tenopir, et al., “Research Data Management Services in Academic Research Libraries and Perceptions of Librarians.”

²¹ Corrall, Kennan, and Afzal, “Bibliometrics And Research Data Management Services.”

²² Bethany Latham, “Research Data Management: Defining Roles, Prioritizing Services, and Enumerating Challenges,” *The Journal of Academic Librarianship* 43, no. 3 (May 2017): 263–65,

<https://doi.org/10.1016/j.acalib.2017.04.004>.

²³ Andrew Creamer, et al., “An Assessment of Needed Competencies to Promote the Data Curation and Management Librarianship of Health Sciences and Science and Technology Librarians in New England,” *Journal of EScience Librarianship* 1, no. 1 (2012): 18–26,

<https://doi.org/10.7191/jeslib.2012.1006>.

²⁴ Andrew M. Cox, et al., “Maturing Research Data Services and the Transformation of Academic Libraries,” *Journal of Documentation* 75, no. 6 (2019): 1432–462, [https://doi.org/10.1108/JD-12-](https://doi.org/10.1108/JD-12-2018-0211)

[2018-0211](https://doi.org/10.1108/JD-12-2018-0211).

²⁵ Ibid.

²⁶ Tenopir, et al., “Research Data Services in Academic Libraries.”

²⁷ Creamer, et al., “An Assessment of Needed Competencies to Promote the Data Curation and Management Librarianship of Health Sciences and Science and Technology Librarians in New England.”

-
- ²⁸ Lisa Federer, "Defining Data Librarianship: A Survey of Competencies, Skills, and Training," *Journal of the Medical Library Association* 106, no. 3 (2018): 294-303, <https://doi.org/10.5195/JMLA.2018.306>.
- ²⁹ Lisa Federer, et al., "The Medical Library Association Data Services Competency: A Framework for Data Science and Open Science Skills Development," *Journal of the Medical Library Association* 108, no. 2 (2020): 304–309, <https://doi.org/10.5195/jmla.2020.909>.
- ³⁰ Ramirose Ilene Attebury, "Professional Development: A Qualitative Study of High Impact Characteristics Affecting Meaningful and Transformational Learning," *The Journal of Academic Librarianship* 43, no. 3 (May 2017): 240, <https://doi.org/10.1016/j.acalib.2017.02.015>.
- ³¹ Liz Lyon, "The Informatics Transform: Re-Engineering Libraries for the Data Decade." *International Journal of Digital Curation* 7, no. 1 (2012): 126–38, <https://doi.org/10.2218/ijdc.v7i1.220>.
- ³² Liz Lyon and Aaron Brenner, "Bridging the Data Talent Gap: Positioning The iSchool as an Agent For Change," *International Journal of Digital Curation* 10, no. 1 (2015): 111-122, <http://doi.org/10.2218/ijdc.v10i1.349>.
- ³³ Liz Lyon, "Librarians in the Lab: Toward Radically Re-Engineering Data Curation Services at the Research Coalface," *New Review of Academic Librarianship* 22, no. 4 (October 2016): 391–409, <https://doi.org/10.1080/13614533.2016.1159969>.
- ³⁴ Lisa Federer, "Defining Data Librarianship: A Survey of Competencies, Skills, and Training." *Journal of the Medical Library Association* 106, no. 3 (2018). <https://doi.org/10.5195/JMLA.2018.306>.
- ³⁵ Federer, "Defining Data Librarianship," 301.
- ³⁶ P. Bryan Heidorn, "The Emerging Role of Libraries in Data Curation and E-Science," *Journal of Library Administration* 51, no. 7–8 (October 2011): 662–72, <https://doi.org/10.1080/01930826.2011.601269>.

-
- ³⁷ Jinxuan Ma, Lynne Stahl, and Erica Knotts, "Emerging Roles of Health Information Professionals for Library and Information Science Curriculum Development: A Scoping Review," *Journal of the Medical Library Association* 106, no. 4 (2018), <https://doi.org/10.5195/JMLA.2018.354>.
- ³⁸ "Training," Research Data Management Librarian Academy, updated January 31, 2020, <https://github.com/RDMLA/rdmla.github.io/blob/master/survey-documents/Training.xlsx>.
- ³⁹ Michael Witt and Melissa Cragin, "Introduction to Institutional Data Repositories Workshop," *Libraries Research Publications*, paper 83 (2008), http://docs.lib.purdue.edu/lib_research/83.
- ⁴⁰ John Southall and Catherine Scutt, "Training for Research Data Management at the Bodleian Libraries: National Contexts and Local Implementation for Researchers and Librarians," *New Review of Academic Librarianship* 23, no. 2-3 (July 3, 2017): 303-22, <https://doi.org/10.1080/13614533.2017.1318766>.
- ⁴¹ Jamie Wittenberg, Anna Sackmann, and Rick Jaffe, "Situating Expertise in Practice: Domain-Based Data Management Training for Liaison Librarians," *The Journal of Academic Librarianship* 44, no. 3 (May 2018): 323-29. <https://doi.org/10.1016/j.acalib.2018.04.004>
- ⁴² Lisa M. Federer and Jian Qin, "Beyond the Data Management Plan: Expanding Roles for Librarians in Data Science and Open Science," *Proceedings of the Association for Information Science and Technology* 56 (October 2019): 529-531, <https://doi.org/10.1002/pra2.82>.
- ⁴³ "Research Data Management Training - On Demand," National Network of Libraries of Medicine, updated 2021, <https://nnlm.gov/rdm-on-demand>.
- ⁴⁴ "Research Data Management Librarian Academy: Exploring and Providing Research Data Management Training for Librarians," Research Data Management Librarian Academy, updated 2021, <https://rdmla.github.io/>.
- ⁴⁵ Tenopir, Birch, and Allard, "Academic libraries and research data services."
- ⁴⁶ Federer, "Defining Data Librarianship."
- ⁴⁷ Tenopir, Birch, and Allard, "Academic libraries and research data services."

⁴⁸ Tenopir, et al, "Research Data Services in Academic Libraries."

⁴⁹ Heidorn, "The Emerging Role of Libraries in Data Curation and E-Science"; Lyon, "The Informatics Transform"; Lyon and Brenner, "Bridging the Data Talent Gap"; Federer, "Defining Data Librarianship."

⁵⁰ Tenopir, et al., "Research Data Management Services in Academic Research Libraries and Perceptions of Librarians," p. 88.

⁵¹ Tenopir, et al., "Research Data Services in Academic Libraries."

⁵² Federer, "Defining Data Librarianship," p. 299.

Appendix A. Survey

Description:

Available here: https://libguides.lib.umanitoba.ca/ld.php?content_id=35814630

Filename:

Appendix_A_-_Survey.pdf

Appendix B. Mean Scores for all Categories

Data Mining	0.94
Data Use and Analysis	1.95
Data Curation and Preservation	1.98
Data Visualization and/or Informatics	1.86
Data depositing and/or work with repositories	2.07
Policy and advisory skills (e.g. interpreting the upcoming Tri-Agency Research Data Management Policy For Consultation)	2.12
MEAN FOR GENERAL DATA SERVICES SKILLS	1.83
Programming languages (JavaScript R Python C++ etc.)	0.76
Markup languages (HTML XML)	1.78
Web development and maintenance	1.58
Statistical software (e.g. SAS, SPSS, RStudio)	1.02
GIS software (e.g. ArcGIS)	0.77
Data visualization software (e.g. Tableau)	1.25
Data analysis software (e.g. OpenRefine)	1.15
Text editors (e.g. Jupyter)	1.05
MEAN FOR PROGRAMMING LANGUAGES AND SOFTWARE	1.18
Library instructional sessions	3.31
In-class instructional sessions	3.24
One-on-one sessions	3.43
Development of online instruction tutorials or modules	2.62
Knowledge of blended and online learning theory	2.24
Experience with learning management systems	2.29
MEAN FOR LIBRARY INSTRUCTION	2.86
Oral communication	3.29
Written communication	3.53
Management and leadership	2.67
Interpersonal customer service	3.37
Working well with others	3.40
Project management	2.78
MEAN FOR SOFT SKILLS	3.18