

Developing the Librarian Workforce for Data Science and Open Science

by

Lisa Federer, PhD, MLIS
Sarah Clarke, MLIS
Maryam Zaringhalam, PhD
Michael Huerta, PhD

Office of Strategic Initiatives
National Library of Medicine
National Institutes of Health

Table of Contents

Executive Summary	3
1. Introduction	5
2. Workshop Findings	6
2.1 The Landscape of Data Science and Open Science.....	6
2.2 Data Science and Open Science Services in Libraries	7
2.3 Skills for Librarians	8
2.3.1 Data Skills.....	9
2.3.2 Computational Skills	10
2.3.3 Research and Subject Matter Knowledge	11
2.3.4 Traditional Library Skills.....	12
2.3.5 Skills for Developing Programs and Services.....	13
2.3.6 Interpersonal Skills	15
2.3.7 Skills for Lifelong Learning.....	16
2.3.8 Finding the Right Skillset	18
2.4 Next Steps for Data Science and Open Science Librarianship	18
2.4.1 Expanding Data Science and Open Science Expertise in Librarianship	18
2.4.2 Training the Next Generation of Information Professionals for Data Science and Open Science	19
2.4.3 The Role of Experiential Learning	20
2.4.4 Implications for Library Leadership.....	21
3. Conclusions	22
4. Sources.....	23
Appendix A: Workshop Participants.....	25
Appendix B: Workshop Agenda	25
Appendix C: Existing Courses	28

Executive Summary

This report summarizes the discussions and major themes that emerged from a 1.5-day workshop on Developing the Librarian Data Science and Open Science Workforce held April 15-16, 2019 at the National Library of Medicine (NLM). The workshop aimed to identify the set of skills that librarians will need to advance work in data science and open science (DS/OS). Workshop participants included practitioners who provide DS/OS services, as well as faculty from schools of library and information science who are advancing DS/OS in academic settings. Broad topics of discussion included:

1. **The landscape of data science and open science**, with a focus on organizational settings, varied researcher needs, and how librarians communicate services.
2. **Challenges and opportunities for librarians**, including librarian communication to stakeholders, evaluation of program and service impact, and integration with DS/OS teams.
3. **Core skills for librarians for data science and open science**, to determine which skills are already taught in library school, what skills are needed, and existing resources for training. These skills, which can be divided into seven broad areas, include:
 - Data Skills
 - Data management plans and data workflows
 - Data and metadata standards and curation
 - Data sharing and reuse
 - Data citation
 - Data policy and governance
 - Computational Skills
 - Computational literacy
 - Database design
 - Familiarity with relevant coding languages, such as R and Python
 - Machine learning and data or text mining
 - Data visualization
 - Research and Subject Matter Knowledge
 - General understanding of the relevant science or subject matter
 - Research design and workflows
 - Statistics
 - Responsible conduct of research and research ethics
 - Methods for reproducibility
 - Traditional Library Skills
 - Consultation and reference
 - Literature searching
 - Scholarly communication
 - Bibliometrics
 - Training and instruction
 - Assessment and evaluation
 - Skills for Developing Programs and Services
 - Interview and assessment skills to understand institutional needs
 - Scoping and planning for sustainability
 - Willingness to embrace failure

- Communication and marketing skills
 - Interpersonal Skills
 - Team science skills
 - Entrepreneurship
 - Advocacy skills
 - Community building
 - Skills for Lifelong Learning
 - Flexibility and adaptability
 - “Anthropological” mindset
 - Logic and problem-solving
 - Design thinking
 - Computational thinking
4. **The future of library data science and open science services**, to address the sustainability of research products, librarian roles in promoting research reproducibility, and facilitating data science and open science in a quickly-changing technological and policy climate.
 5. **Preparing the current and future librarian workforce**, by identifying areas where training does not yet exist and considering how schools of library and information science and other stakeholders can address these gaps.

The findings presented here have implications for a range of stakeholders in libraries. First, this report provides a roadmap for librarians interested in gaining new skills and expertise to engage with work in DS/OS. This report offers information to library directors and other administrators who wish to develop their staff and advocate for their library’s role in providing DS/OS services. Finally, the set of skills discussed as core to DS/OS can serve as a guide for schools of library and information science, providers of continuing education, professional organizations, and other groups involved in developing and delivering training for current and future librarians and information professionals.

1. Introduction

More open research paradigms, in which the products and processes are shared broadly and in (re)usable ways, offer significant opportunities to enhance scientific insight, address reproducibility, and increase the return on investment in research. As these and other opportunities become widely recognized, and as data management plans (DMPs) are increasingly required of researchers, research data management (RDM) has become a focus of attention in scientific communities. Accordingly, many academic and research libraries now offer RDM services, with particular emphasis on data management planning services and training on RDM principles and practice. As early as 2014, nearly half of academic research libraries were offering RDM services, with an additional third planning to do so within two years.¹

With increased and better-managed research data comes the opportunity to use these datasets on their own or in combination, giving rise to the set of methodologies and inquiries that constitute data science. At the same time, more researchers have started to adopt open science practices, enabling broader sharing of research outputs including not only the final research paper, but also research data, software and code, peer reviews, and more. While effective RDM is essential to enabling successful data science and open science (DS/OS), many other skills are needed to effectively conduct (and support) data science such as programming, statistics, and data visualization. Many libraries have been successful in developing such services², thanks to motivated librarians with specialized expertise, but this type of knowledge is not widespread in the field, nor are such skills regularly taught in library schools of library and information science. Similarly, few opportunities exist for librarians to develop the needed skills and expertise to support data science.

Given the rapidly evolving nature of library support for DS/OS, researchers have attempted to define the role of the data librarian, including the types of services such an information professional might offer, the skills needed to offer them, and the best ways to develop those skills.³ Various library-related organizations and funders have also undertaken efforts (or are planning activities) to elucidate the role of librarians in supporting DS/OS and determine how best to prepare librarians to provide such services. The Institute of Museum and Library Services held a workshop in November 2017 to help strengthen the library school curriculum that included a session on “21st Century Skills, Expertise, and Competencies,” within which data science-related skills were mentioned.⁴ In early 2019, the Medical Library Association

¹ The Carpentries, “Assessment and Impact.”

² Hosburgh, “Developing a Bioinformatics Program and Supporting Infrastructure in a Biomedical Library”; Ogier and Stamper, “Data Visualization as a Library Service: Embedding Visualization Services in the Library Research Lifecycle”; Courneya and Mayo, “High-Performance Computing Service for Bioinformatics and Data Science”; Surkis et al., “Data Day to Day: Building a Community of Expertise to Address Data Skills Gaps in an Academic Medical Center.”; Read and LaPolla, “A New Hat for Librarians: Providing REDCap Support to Establish the Library as a Central Data Hub.”

³ Tenopir et al., “Research Data Management Services in Academic Research Libraries and Perceptions of Librarians.”

⁴ Federer, “Defining Data Librarianship: A Survey of Competencies, Skills, and Training”; Cope and Baker, “Library Carpentry: Software Skills Training for Library Professionals”; Tenopir et al., “Research

convened a committee to define the “performance indicators” or skills that would comprise a Data Services competency. The question of how to re-envision library education to develop the 21st century library professional also emerged as a topic at the Online Computer Library Center (OCLC) ResearchWorks event, which gathered experts to help shape OCLC’s research agenda.⁵

The National Library of Medicine’s (NLM) Office of Strategic Initiatives (OSI) held a workshop in April 2019 to explore in-depth the question of how librarians can become more deeply involved in DS/OS and identify the skills needed to do so successfully. The workshop convened library practitioners with experience working on DS/OS issues from a range of scientific disciplines, as well as faculty from schools of library and information science (see Appendix A for the full roster of participants). Over the course of the day-and-a-half workshop, participants engaged in conversation and facilitated brainstorming activities to explore directions for developing the library workforce for DS/OS (see Appendix B for the workshop agenda). The unique contribution of this workshop is that it surfaced a set of specific skills that can be used to develop and cultivate a library workforce that is prepared to meet the challenges of DS/OS not only today, but for the foreseeable future.

2. Workshop Findings

This report provides an overview of the workshop findings and is divided into four sections. Section 2.1 details participants’ discussion of the current landscape of DS/OS, which helps frame the conversation and identify some of the major challenges that researchers face in these areas and that librarians can potentially address. Section 2.2 describes how participants are addressing and supporting DS/OS at their institutions, highlighting examples of existing DS/OS services. Section 2.3 outlines a set of core skills for librarians working in DS/OS, as well as discussing how these skills can be applied in various library settings. Finally, section 2.4 discusses the implications of the workshop findings for library practice and training.

2.1 The Landscape of Data Science and Open Science

Understanding how librarians can support DS/OS requires first having insight into the current landscape of DS/OS methodologies and practices, as well as how researchers are using them and the challenges that they face in adapting to more data-intensive and open research ecosystems. Accordingly, the workshop opened with a discussion of participants’ experiences and understanding of the current DS/OS landscape.

An initial, foundational challenge to addressing DS/OS is the problem of definition; the term “data science” does not have a single, agreed-upon definition, even among different groups

Data Services in Academic Libraries: Data Intensive Roles for the Future?”; Cox et al., “Developments in Research Data Management in Academic Libraries: Towards an Understanding of Research Data Service Maturity.”

⁵ Institute of Museum and Library Services, “Positioning Library and Information Science Graduate Programs for 21st Century Practice.”

within the NIH. Data science comprises a range of activities, skills, and methodologies, so definitions of data science and the specifics of its application to scientific problems may differ from one discipline to another, or even one research team to another. Generally speaking, data science typically involves computational approaches to research that enable discovering meaning and patterns in data, grounded in subject matter expertise and sound statistical approaches. Similarly, open science is not a single practice, but a continuum of activities that contribute to enhancing the transparency of and access to scientific information across the research life cycle, including open access to literature, open data and code, and open peer review. The fluidity of these terms means that librarians who work in these areas should ensure that they reach common ground with their users on terminology.

Just as definitions and understandings of DS/OS may differ, the specific needs of an individual user will differ not only from one institution to another, but even within institutions. For example, research suggests early-career researchers may be more open to sharing than their later-career colleagues,⁶ and therefore may be more likely to have an interest in using data and open science services, a finding that was borne out by the experiences of workshop participants and supported by assessment findings.⁷ Workshop participants also pointed to differences in the needs of large, well-funded research groups, which may have adequate resources of their own, compared to smaller research groups with less funding, which could significantly benefit from DS/OS library services.

2.2 Data Science and Open Science Services in Libraries

Given such an expansive and varied set of DS/OS needs across institutions, disciplines, and individual researchers, the opportunities for libraries to provide DS/OS services are ample. During and prior to the workshop, participants shared various DS/OS initiatives and services happening at their libraries. Each of these services is grounded in the participants' understanding of their institutions' unique needs, informed by information-gathering through needs assessments, surveys, user feedback, and other inputs. These services also draw on the participant's skillset, complemented as necessary by partners from within or outside the institution. For example, some participants who were not experienced with programming recruited knowledgeable partners within their institutions to teach coding classes or sought volunteers from The Carpentries, a non-profit organization that provides training on coding and computational skills designed especially for researchers.⁸ Services and infrastructure currently being offered by workshop participants include:

⁶ OCLC Research, "ResearchWorks."

⁷ Campbell, Micheli-Campbell, and Udyawer, "Early Career Researchers Embrace Data Sharing"; Abele-Brehm et al., "Attitudes Toward Open Science and Public Data Sharing: A Survey among Members of the German Psychological Society."

⁸ The Carpentries, "The Carpentries"; Williams and Teal, "A Vision for Collaborative Training Infrastructure for Bioinformatics."

- A “data science core” – a physical space within the library that has designated computers with data-specific tools like R, Python, SPSS, and SAS, and a library staff member available to assist patrons using these tools.
- A reproducibility workshop series that covers experimental design, registering reports, varying peer-review formats, fostering an open lab culture, and data, code, and protocol sharing.
- Electronic lab notebook training, including best practices for formatting, content and user management, mapping research workflows, features and functionality, and an exploration of different notebook tools.
- Data literacy and awareness training modules focused on data management principles, DMPs, open science, research rigor and reproducibility, and computational work.
- Embedded data services which allow for a librarian to work on specific projects within a lab and participate in lab meetings.
- A data discovery index that links multiple institutions to increase discoverability and encourage data sharing.
- Extended RDM services focusing on the adoption and use of standards, data quality measures, data sharing processes, data compliance, and data and software curation.
- Bibliometrics and research impact assessments performed by a librarian, including database searching, data wrangling, data analysis, data visualization, and report writing.
- Evidence synthesis services to support systematic reviews, scoping reviews, narrative reviews, rapid reviews, clinical evidence summaries, and knowledge synthesis reports.

These services and infrastructure meet a range of DS/OS needs and fill important gaps in the participants’ institutions. Understanding how workshop participants have situated themselves within the landscape of DS/OS at their institution helped lay the groundwork for the discussion that formed the heart of the workshop: defining core skills for librarians working in DS/OS.

2.3 Skills for Librarians

Given the widely varied needs of different user groups across institutions, workshop participants identified a range of skills considered core to librarianship in support of DS/OS. Since most workshop participants work in academic research institutions with an emphasis on biomedical research, the discussion primarily focused on the sciences, but these skills would also be relevant for librarians working in digital humanities and other fields of data-driven or computationally-intensive research. These core skills can be divided into seven broad categories, to be described in further depth in this section:

- data skills,
- computational skills,
- research and subject matter knowledge,
- traditional library skills,
- skills for developing programs and services,
- interpersonal skills, and
- skills for lifelong learning.

This list should not be considered a prescriptive set of skills; one need not have *all* of these skills to be successful. Further, few institutions will likely require all of the expertise and skills listed here. When considering the skills to look for in a new hire, or to provide training on for existing staff to support DS/OS, libraries should critically and thoughtfully evaluate their institution's unique needs and consider how new library DS/OS services will be situated within the broader institutional context.

It should also be noted that some of the skills described here are not unique to DS/OS librarianship. For example, liaison librarians likely have some research and subject matter knowledge, and many librarians already embrace lifelong learning. This workshop surfaced the substantive and technical skills and knowledge that are *most essential* to DS/OS, as well as interpersonal skills that make a librarian likely to succeed in DS/OS work in the context of her or his institution. Therefore, this report includes both skills that are specific to DS/OS work and those that might be common to other types of librarianship to provide a holistic view of the DS/OS librarian role.

2.3.1 Data Skills

- Data management plans and data workflows
- Data and metadata standards and curation
- Data sharing and reuse
- Data citation
- Data policy and governance

As would be expected, various skills for working with data figured prominently in the workshop discussion. These skills span the data life cycle, from initial research planning to disposition of data at the end of a research project. While this workshop aimed to consider roles beyond consultation on DMPs, such expertise is foundational to DS/OS, as is an understanding of best practices for data workflows. DS/OS work also requires an understanding of the data curation skills that enable logical organization of data, including awareness of relevant metadata standards and their application and use. Relevant standards will differ depending on the subject matter of the research the librarian supports; for example, librarians working with clinical researchers might be familiar with clinical common data elements,⁹ whereas those working with genomic or basic science researchers might be familiar with the family of Minimum Information for Biological and Biomedical Investigations (MIBBI) standards.¹⁰

As funders and journals increasingly adopt policies that require researchers to share their data, librarians supporting DS/OS can play a key role in both data sharing and reuse of shared data. Working with researchers who are sharing their data can involve a number of different skills. Researchers often need advice on where or how to share their data, such as guidance on

⁹ Sheehan et al., "Improving the Value of Clinical Research through the Use of Common Data Elements."

¹⁰ Taylor et al., "Promoting Coherent Minimum Reporting Guidelines for Biological and Biomedical Investigations: The MIBBI Project."

selecting an appropriate data repository. In some cases, researchers need assistance with areas related to data sharing governance, such as preparing a data sharing agreement or selecting an appropriate license for reuse. Data documentation and quality assurance are also important to ensuring that shared data are both accurate and FAIR (findable, accessible, interoperable, and reusable).¹¹ Conversely, researchers who are interested in reusing existing data may also need assistance in locating a relevant repository and a dataset that can answer their research question, as well as potentially negotiating access to datasets that are not publicly available and understanding the terms of data sharing agreements. All of these activities involve skills and knowledge of data practices that librarians interested in DS/OS should consider.

More generally, librarians working in DS/OS should ensure they stay up-to-date on policy and practice, particularly since the DS/OS landscape is rapidly evolving. Given the proliferation of funder and journal policies, and the potential consequences for non-compliance, workshop participants considered their awareness of data-related policies and their ability to interpret them within their specific institutional setting a valuable service to researchers. In addition, practices in conducting and communicating DS/OS research are quickly changing, with new platforms, tools, standards, and scholarly communication norms frequently emerging. For example, the scientific and publishing communities are still working toward standardizing practices for citing datasets in scholarly articles. Researchers are not likely to closely follow such developments, so DS/OS librarians can be an important source of information on new practices.

2.3.2 Computational Skills

- Computational literacy
- Database design
- Familiarity with relevant coding languages, such as R and Python
- Machine learning and data or text mining
- Data visualization

Both data science and open science are highly computational, in many cases drawing on cutting-edge technology, so librarians working in the DS/OS environment need a level of computational literacy and comfort with new technologies. The specific computational skills that are most relevant differ across institutions and disciplines, but at a minimum, DS/OS librarians should be computer-savvy and adept at learning and applying new technical skills. In particular, DS/OS librarians should be familiar with basic tools for interacting with different types of data, such as Excel or OpenRefine for spreadsheets. Workshop participants also mentioned database design as a relevant skillset that enables complex methods of organizing and interacting with data and instills a deeper understanding of data structures.

For some librarians working in DS/OS, knowledge of one or more common coding languages such as R or Python may be useful. However, workshop participants emphasized that they did

¹¹ Wilkinson et al., “The FAIR Guiding Principles for Scientific Data Management and Stewardship.”

not consider it necessary for librarians to know how to code, even for those who specifically specialize in DS/OS. In fact, most of the participants did not know how to code, nor did they perceive lack of coding ability to be a barrier to their success in the field. Still, those participants who did have knowledge of coding considered it a useful skill that enabled them to engage with researchers on DS/OS work more deeply. Data science and open science librarians should at least be aware of the coding languages most frequently used in the research areas they support and be able to point users to relevant campus resources or online materials for learning to code. Similarly, workshop participants considered it useful (but not necessary) to have a conceptual understanding of specific data science techniques widely used in the research fields they support, such as machine learning, data mining, and text mining.

In addition to computational techniques for organizing and interacting with data, workshop participants also mentioned data visualization as an important skill. Data visualizations can range from simple Excel charts to interactive visualizations with tools like Tableau to highly customized, novel chart forms created from scratch using an appropriate coding language. As with other computational skills, the specific tools and depth of knowledge needed for success are highly dependent on the institution and the discipline with which the librarian works. Though we classify data visualization as being computational in nature in this report, mastery of this skill also involves additional non-computational knowledge. Creating a successful visualization requires not only being able to use software, but also knowing how to visually convey meaning from raw data, selecting the most appropriate chart form to tell a given dataset's story, and drawing on design best practices, such as selection of color-blind friendly color schemes.

The level and specific types of computational skills each participant identified having themselves seemed to vary the most of any of the broad skill sets. Computational skills that one participant considered essential to her work might be completely unknown to another, and the depth of expertise in a given topic or tool differed considerably among workshop participants. The variety of skills and knowledge in this group of experts reflects the broad range of services that DS/OS support can entail and further underscores the importance of understanding the institution's unique needs to find the right fit in terms of a librarian's computational expertise.

2.3.3 Research and Subject Matter Knowledge

- General understanding of the relevant science or subject matter
- Research design and workflows
- Statistics
- Responsible conduct of research and research ethics
- Methods for reproducibility

As with other areas of librarianship that involve working closely with users on their research issues, DS/OS librarians can benefit from understanding the basics of the science or discipline. For many workshop participants, this knowledge was gleaned on the job in the course of working with users and learning about their needs and challenges. Others had formal education in the discipline in which they worked, including some participants who held doctorate degrees

in a scientific field. Some of these participants felt that their degree enabled different types of engagement because users perceived them as peers who were conversant in the language of the discipline. However, the general consensus was that, while formal scientific education could be helpful, it is not a prerequisite for success in working in DS/OS.

While deep experience with the science may not be necessary, workshop participants did see value in understanding of the conduct of research more broadly. In fact, some participants felt that understanding the general end-product of research and the processes that lead to it was more important than having deep knowledge in any one area, since the specifics of how researchers conduct their work differs from one discipline to another, and even from one group to another in the same discipline. Having a familiarity with research design and typical research workflows provides background that enables librarians working in DS/OS to make more meaningful recommendations to their users, as well as understand some of the pain points that users may encounter during the process of conducting research. A basic understanding of statistical methods can be helpful for the same reason, especially since proper collection, management, and organization of data can have significant implications for the statistical methods that can be used and the types of conclusions that can be reasonably drawn from the research.

Workshop participants also considered it important for librarians working in DS/OS to be familiar with research ethics and the responsible conduct of research. Such knowledge is especially relevant in the context of current concerns about research reproducibility. Librarians who work with users on DS/OS issues are well positioned to provide guidance on best practices for enhancing reproducibility. In fact, many such practices fall under the umbrella of open science, including study pre-registration and data and code sharing. Other skills and services described elsewhere in this report as falling within the purview of DS/OS librarianship also contribute to improving reproducibility, such as metadata and documentation, data curation, and use of coding languages.

2.3.4 Traditional Library Skills

- Consultation and reference
- Literature searching
- Scholarly communication
- Bibliometrics
- Training and instruction
- Assessment and evaluation

Traditional library skills taught in graduate-level library science courses lend themselves well to the responsibilities of DS/OS librarians. Workshop participants identified various skills that form the basis of most library services and can be utilized across all disciplines served: consultation and reference, literature searching, scholarly communication, training and instruction, and assessment and evaluation.

While many of the areas that workshop participants provide support for extend beyond what some might traditionally consider the purview of the library, most of these services still exist within the structure of what could be considered a typical academic library setting. As a result, many of the ways services are delivered reflect the ways that information professionals have long provided services. Many participants deliver support via consultations with individuals or groups, similar to traditional reference services. Training and instruction were also described as a major service area for many workshop participants. With interest in and need for DS/OS skills becoming increasingly widespread, delivering classes can be an effective way of maximizing the number of users who can be reached.

Data science and open science services are also situated within the broader scientific ecosystem and research life cycle. Many workshop participants draw on their library background to provide additional services to the researchers and teams they work with, such as literature searching and consulting on scholarly communication topics, such as licensing and copyright. Some of this same understanding of licensing can be parlayed into working with data, as described above, by either helping researchers select licenses for data they share or understand license terms for data they wish to reuse. Data science and open science librarians' adeptness with data also positions them to work effectively in bibliometric analyses, which are typically data-intensive and draw on many of the same tools that researchers are using in their own work. However, it should be recognized that *interpretation* of bibliometric analyses also requires specialized knowledge beyond simply being able to analyze data to ensure that bibliometric methods and metrics are responsibly applied.¹²

Finally, as with any library service, DS/OS services should be carefully evaluated and assessed to ensure that they are meeting their aims and successfully reaching their intended audience. As will be discussed, librarians and library leaders should be aware that not all services will ultimately succeed or may not be the right fit for an institution at a given time. Being willing to embrace failure necessarily involves knowing how to define success. Thoughtful evaluation and assessment of DS/OS services can also be invaluable in further tailoring services going forward. The successes (and failures) of a DS/OS service provide insight into the institution's and users' needs.

2.3.5 Skills for Developing Programs and Services

- Interview and assessment skills to understand institutional needs
- Scoping and planning for sustainability
- Willingness to embrace failure
- Communication and marketing skills

At present, DS/OS services are relatively new to libraries; many of the workshop participants had originated, developed, and eventually expanded the programs at their institutions. Many libraries will be building entirely new DS/OS services from the ground up, but even those with mature DS/OS services will likely need to evolve over time, given the rapidly changing nature of

¹² Hicks et al., "Bibliometrics: The Leiden Manifesto for Research Metrics."

the DS/OS landscape. Therefore, skills for developing programs and services are relevant both to librarians just beginning in DS/OS work as well as those who have established expertise in DS/OS. Developing successful programs requires a significant amount of leadership, including developing a vision for the library's DS/OS services and communicating that vision to library and campus leadership, users, and other stakeholders. Librarians who have DS/OS expertise can also provide leadership within their library by serving as mentors for their colleagues and providing guidance for liaison librarians on how to engage with their users on DS/OS issues.

As has been previously discussed in this report, identifying institutional needs and determining how the library can address them are crucial skills for success in providing DS/OS services. This process may involve formal assessment of the unique institutional setting, such as conducting a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis or other needs assessment. Such assessment can be useful to identify not only what services are needed, but also what services already exist within the institution. For example, those responsible for campus information technology or individual departments may offer training or services that meet some DS/OS needs. Being familiar with all relevant services at the institution helps prevent duplication of efforts and enables DS/OS librarians to make referrals to other services when applicable. For example, a librarian may not have adequate experience to address a user's question on statistical methods for data science but could point the user to a statistical consulting service at the institution.

Getting input from potential users can also be helpful, whether in the form of structured information collection, such as a survey, or more informal conversations. This type of user input is not only helpful to determine the services that are most relevant, but also to demonstrate need to leadership and gain buy in for developing new services. Eliciting input from users about their DS/OS needs can also serve as a form of marketing for the library's DS/OS services; a user may not have thought to seek help from the library for such needs, but being part of a survey or interview can increase their awareness that the library is working in the DS/OS space. On the other hand, participants warned against "analysis paralysis," or overthinking the planning and assessment phase to the extent that no action is ever taken.

While identifying user needs is important for developing DS/OS services, also crucial is consideration of what the library can realistically provide. When developing new services, appropriate scoping and planning for sustainability are essential. Such planning requires consideration of the budget, physical space, and staffing that will be required. Staffing is especially important for ensuring the success of a new service. Will the library hire a new librarian to provide DS/OS services, or train and reassign an existing staff member to such services? If the latter, the library should consider how portions of that librarian's work can be handed over to others in the institution to ensure adequate staff time for new services.

Another essential skill for launching a successful DS/OS service is understanding how to communicate and market it. Even more than with traditional library services, marketing is important in the context of DS/OS services given that many users are not aware that librarians have DS/OS expertise and therefore may not think to look to the library for such services.

Forming partnerships and building relationships with key stakeholders, such as administrators and offices dealing with research broadly, can be a useful approach to marketing, allowing the library to leverage existing communication channels to raise awareness.

Developing a new service sometimes involves accepting failure. Even with the necessary background assessment and highly qualified DS/OS librarians, sometimes a service is just not the right fit for an institution. Some of the participants spoke of the concept of “failing faster” - that is, being willing to experiment with new ideas, but recognizing when an approach isn’t working and either iterating on or discontinuing the service. On the other hand, participants also urged caution in abandoning ideas before they have had the chance to succeed. For example, one participant spoke of offering a new DS/OS training course at the library. Though only two people came for the first class, the course eventually became so successful that it was integrated into one academic department’s formal instruction for graduate students. Developing programs that will go on to be successful in the long term requires understanding when a service should be discontinued and when it just needs to be adjusted or better marketed to the right audience.

2.3.6 Interpersonal Skills

- Team science skills
- Entrepreneurship
- Advocacy skills
- Community building

Workshop participants also identified a number of interpersonal skills that, while not specific skills as such, may help lead to success in DS/OS librarianship. As has been discussed, many of these skills are desirable for librarianship in general, and they are specifically highlighted here because of their particular importance for DS/OS librarianship.

It is often noted that data science is team science - given the broad range of requisite skills and expertise, from subject matter knowledge to statistics to programming, much data science is conducted by teams of individuals who each contribute their own specialized knowledge and disciplinary background to the problem at hand. Librarians wishing to engage with researchers on DS/OS issues can play an important role in such teams. Doing so requires having skills for working effectively in a team, including excellent skills in communication and conflict management, as well as being able to learn and adapt to the language and culture of the team. Librarians should consider how best to position themselves within a team to maximize their contribution by exploring how to match their expertise with gaps in the team’s skills.

Getting a foot in the door to work with a team is just one aspect in which librarians working in DS/OS need to take significant initiative, creating opportunities rather than waiting for users to come to the library. Workshop participants spoke of the importance of taking an entrepreneurial approach to developing services. Such an approach involves not only developing services to meet an identified need, but also knowing how to “sell” those services by being able to

communicate the value of the service and its potential positive impact to users. Collecting and sharing success stories from users who have tangibly benefited from the library's DS/OS services can be one compelling way of doing so.

In addition to marketing specific DS/OS services, librarians must also be able to advocate for the library's role in DS/OS more broadly. It has been the experience of workshop participants that users often think of libraries as a provider of books and journals and not as a source for DS/OS services. Building awareness of the library as a partner for DS/OS requires persistence in communicating the librarian's expertise in and availability for DS/OS projects. Workshop participants also suggested that library directors and leaders should take an active role in advocating for the library's role in DS/OS at a high level in the institution.

Finally, workshop participants spoke of the importance of community building skills to help develop communities of practice around DS/OS at the institution. Given that data science and open science are relatively new approaches to research, such communities of practice may not yet have coalesced. As librarians become more engaged in DS/OS and build an awareness of others who are involved in DS/OS at the institution, they have the opportunity to help create connections to develop communities. In fact, some workshop participants described themselves as connectors within their institution, facilitating introductions between groups or individuals who would not otherwise have met. Becoming familiar with others interested in DS/OS can also help librarians cultivate a network of campus experts who can help the library's DS/OS services succeed. For example, workshop participants described having success in enlisting campus experts, particularly graduate students, to teach classes in areas where the librarian lacked expertise, such as coding in R or Python. Such arrangements can be mutually beneficial, giving graduate students an opportunity to gain teaching experience while also supporting data science at the institution.

2.3.7 Skills for Lifelong Learning

- Flexibility and adaptability
- "Anthropological" mindset
- Logic and problem-solving
- Design thinking
- Computational thinking

The skills discussed in this report thus far provide a snapshot of DS/OS librarianship today, but with the quickly evolving nature of data science and open science, and with new technologies and practices constantly emerging, roles for librarians in DS/OS will likewise shift over time. As a result, discussion of DS/OS skills should involve consideration of not only the skills that are needed today, but also what skills will likely be needed in the future. Accordingly, workshop participants suggested that librarians who are flexible and easily adapt to change may be well suited to working in DS/OS roles. An openness to lifelong learning could be considered a key characteristic for many types of librarianship, and is especially so for librarians interested in supporting DS/OS.

Lifelong learning may take the form of continuing education or formal study, but participants also pointed to the need for an “anthropological” mindset that allows a DS/OS librarian to observe and adapt to the cultures and norms of different groups with whom they work. Researchers from different fields may approach problems and work with data in very different ways, so the successful DS/OS librarian must learn to pick up on these distinctions, both to gain credibility with their users and to ensure that they are providing the most appropriate guidance and resources. Workshop participants described the value of being able to speak with researchers in their own “language,” understanding that the terms that are commonly used in conversations among librarians may be unfamiliar to researchers. Spending time working with and talking to researchers, as well as attending lectures and meetings, are strategies that workshop participants found useful in learning the language of the user communities they serve.

Workshop participants also described the importance of being able to think logically as well as creatively to effectively problem-solve in complex situations. Successful problem-solving in DS/OS involves not only an understanding of users’ DS/OS needs and challenges, but also an ability to think outside the box to apply available resources to the problem. Design thinking – the practice of applying user-centered design principles – was mentioned as a useful approach to problem solving in the DS/OS space. Figure 1 outlines the five phases of approaching a problem using design thinking.

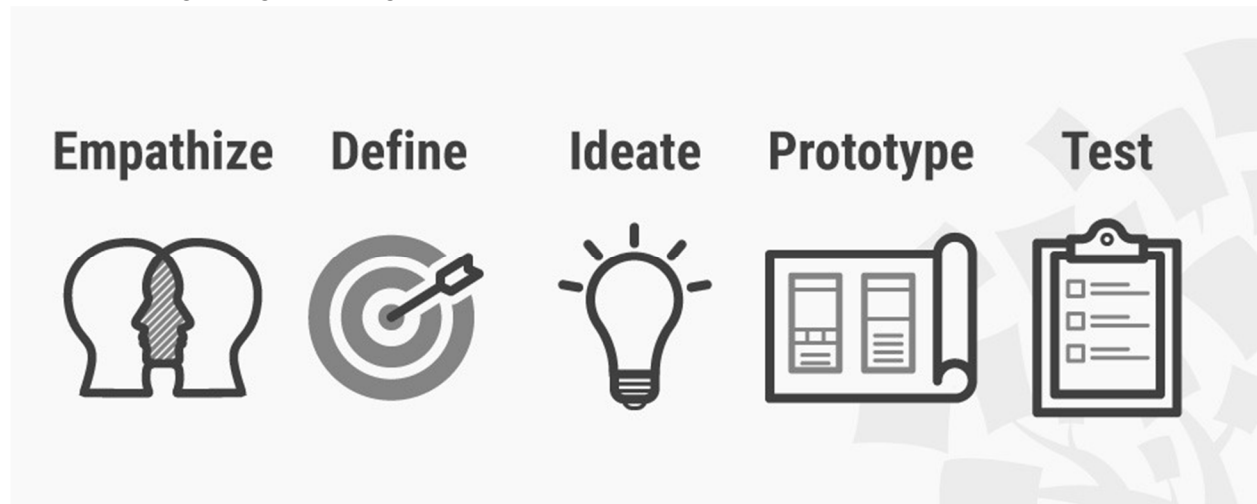


Figure 1. Design thinking steps. Author/Copyright holder: Teo Yu Siang and Interaction Design Foundation. Copyright terms and license: [CC BY-NC-SA 3.0](https://creativecommons.org/licenses/by-nc-sa/3.0/)

Computational thinking can also be an important approach to problem solving for librarians working in DS/OS. Computational thinking involves understanding how to break a problem apart into workflows or processes that a computer could execute and collecting and organizing the needed data. In fact, workshop participants suggested that a capacity for computational thinking could actually be *more* relevant to DS/OS work than in-depth knowledge of a specific programming language. The particular programming languages in use will vary between disciplines and over time, but an understanding of information processes and computational approaches for solving problems is broadly applicable.

2.3.8 Finding the Right Skillset

As this report has emphasized, the list of core skills described above is intended to serve as a guide to the types of expertise that contribute to success in DS/OS librarianship. An individual librarian need not have mastered *every* skill from this list to be successful in DS/OS work. Indeed, even among the workshop participants, who are recognized for their expertise and contributions to DS/OS work, few, if any, would have every single skill from this list. More important is mastering the skills that are most relevant to the individual's institution.

The depth and breadth of any one librarian's set of skills may also be guided by whether their role is as a generalist or specialist DS/OS librarian. A generalist DS/OS librarian might be expected to work across many different disciplines, potentially including not only biomedical science, but other sciences and the humanities. Such a generalist would likely be best served by becoming conversant in the basics of RDM and DS/OS principles that are applicable regardless of discipline. For example, best practices for managing and preserving data are fairly generalizable across disciplines. On the other hand, some librarians work in settings that require them to focus more deeply on a particular discipline. Librarians working in biomedical libraries, for instance, would be better able to serve their users by gaining a deep understanding of the issues that are unique to their audience, such as NIH policy, standards for biomedical data, and the repositories most commonly used in biomedical research. A librarian may focus even more deeply to serve a particular subfield of research within a broader discipline. For example, a DS/OS librarian supporting bioinformatics could benefit from deeper knowledge of the unique needs of this community of researchers, such as understanding the programming languages and types of analyses that are typical of bioinformatics. In institutions where there is significant need for DS/OS services, such as R1 universities with extensive research activities, it may even be reasonable to have multiple specialist DS/OS librarians who each focus on their own unique area.

2.4 Next Steps for Data Science and Open Science Librarianship

This report has explored the landscape of data science and open science, as well as highlighted a few services that libraries have begun to explore. From these examples and the experiences of the workshop participants, along with existing studies of DS/OS librarian activities, a picture emerged of the set of skills and competencies that define DS/OS librarianship. In this section, the report concludes by outlining potential next steps for advancing DS/OS librarianship, including considerations for various stakeholders.

2.4.1 Expanding Data Science and Open Science Expertise in Librarianship

At present, much of the data-related training content developed for librarians and information professionals who want to continue their education or reskill focuses on research data management skills (Appendix C includes a set of representative courses). Most of these courses are broad in scope and designed to provide beginner information on topics like data life cycles; data curation and documentation; data standards, taxonomies, and ontologies; data

security, storage, and preservation; data sharing, publishing, and re-use; data management plan creation and implementation; and data policies and ethics. Such courses typically do not delve into the more technologically-intensive and advanced topics identified as core skills within this report.

A few opportunities do exist for librarians interested in gaining DS/OS skills. One example is Library Carpentry, a series of workshops developed to teach software and data skills to people working in library- and information-related roles.¹³ As of this writing, the Medical Library Association is also developing a continuing education specialization for a data services competency, which will include many of the core skills included in this report. However, further work is needed to develop a range of courses and other training resources tailored to enhancing the skills of information professionals who already have experience and education in libraries.

In many cases, it is not necessary to reinvent the wheel by creating new course material from scratch. A wealth of massive open online courses (MOOCs) exist that address many of the skills covered in this report, but the sheer quantity and variety of MOOC offerings (as well as their variable quality) can make it challenging to find the most relevant course. A roadmap or course catalog that helps librarians locate worthwhile training resources on the skills and topics that will be most useful to them could be a valuable contribution to the profession. Another training approach that could help increase DS/OS expertise within libraries could involve looking to researchers who already have training and experience in DS/OS and providing training on the fundamentals of librarianship that will enable them to successfully apply their subject matter background to work in libraries.

Whether upskilling librarians or scientists, attention should be paid to delivering training resources in formats that will be accessible to professionals already working in the field. Making training resources useful for this audience may involve considering how to provide free or low-cost opportunities, offering asynchronous or self-paced courses that can be completed during one's free time, or pairing courses with meetings or other events that the audience of interest is already attending. Novel training methodologies may also be useful for such groups; participation in or observing codeathons or similar events can help librarians gain on-the-ground understanding of how computational researchers work, and librarians can also learn from each other through experiences like mentoring or shadowing an established DS/OS librarian.

2.4.2 Training the Next Generation of Information Professionals for Data Science and Open Science

In addition to focusing on those already working in libraries, developing the librarian workforce for DS/OS also necessarily involves considering how to prepare the next generation of information professionals to succeed in increasingly data-driven and open science ecosystems. Offering the range of courses necessary to provide a set of core skills may require hiring new faculty with the requisite skill sets or encouraging students to look outside of schools of library

¹³ Library Carpentry, "Library Carpentry"; Cope and Baker, "Library Carpentry: Software Skills Training for Library Professionals."

and information science by providing credit for courses in computer science, data science, and other related topics.

However, library and information science school faculty who participated in the workshop noted that although students expressed interest in data science, when related classes were offered, they frequently had to be canceled due to low enrollment numbers. This observation was not limited to one iSchool, but rather appeared to be a common experience for the iSchool faculty participating in the workshop. While workshop participants could not be sure of the exact reason for low enrollment in DS/OS-related courses, the discussion did touch on some potential steps that might help increase students' participation in such courses. First, a potential "rebranding" of such courses may help increase appeal; students who do not have significant math, science, or computer science backgrounds could consider data science courses intimidating. Nevertheless, given that many of the workshop participants came to DS/OS librarianship from non-scientific backgrounds, such experience may not be necessary, and students should know that data science courses could be right for them regardless of their previous area of study.

Students in schools of library and information science may also be dissuaded from pursuing data science courses if they do not see a clear career path ahead of them. Most schools of library and information science offer courses or even specific tracks that allow students to focus on a specialty like law librarianship, school or children's librarianship, public librarianship, or archives. The career path for students who pursue such specialties is relatively well-defined. By comparison, schools of library and information science do not typically have a track for students interested in DS/OS, so it may be challenging to determine how best to assemble a curriculum of study that will best prepare a student for such work. Further complicating the issue, given that DS/OS librarianship is itself an evolving field, it can be difficult for a student to even envision what that future work might look like. As DS/OS roles become more common in libraries, schools of library and information science may be able to help students better craft a course of study that will prepare them to pursue such work. This report may also contribute to that effort.

2.4.3 The Role of Experiential Learning

Workshop participants broadly agreed that hands-on, experiential learning opportunities are extremely beneficial for those interested in pursuing DS/OS librarianship. Given the complexity of the issues and practices involved in DS/OS work, coursework alone may not be enough to prepare librarians to take on DS/OS roles. Workshop participants emphasized the value of their own experiential learning opportunities in preparing them for such work. For some, this learning took the form of internships during their library school work, which are offered for credit at many schools of library and information science, but often are not required for completion of the master's degree, and in many cases are unpaid. The participants who had been trained in Canadian programs described benefitting from the Canadian "co-op" program that enabled them to pursue paid internships as part of their degree work. These types of learning opportunities helped prepare workshop participants to take on DS/OS work in ways that their coursework alone would not have, giving them insight into the types of real-world problems that researchers face in working with data.

Such hands-on opportunities may also be useful for librarians who are already in the field and wish to expand their skills. Several participants spoke of the valuable experiences and knowledge they gained through their work as informationists on NIH-funded grants. NLM Administrative Supplements for Informationist Services in NIH-funded Research Projects, which were awarded between 2012 and 2017, gave NIH-funded research teams the opportunity to add an information professional to their team.¹⁴ Besides providing value to the research teams who had the chance to work with an informationist, these supplements also benefited the informationists themselves by enabling them to gain deep experience working with researchers and providing a better understanding of their data- and open science- related challenges. Other mechanisms like fellowships within libraries or research labs could give working librarians additional opportunities to get hands-on, real-world experience that complements their library background with new DS/OS skills.

2.4.4 Implications for Library Leadership

In the experience of workshop participants, library directors and other library leaders play an important role in their library's success at launching and sustaining DS/OS services. Before even launching a service, directors must make important staffing decisions to ensure that they have the right people to fill their institution's unique needs. Such decisions require an understanding of what services already exist at an institution and where gaps could be meaningfully filled by services the library could provide. Once the service needs are determined, a clearer understanding of the skills needed to fulfill such roles will provide a guide for staffing decisions. Such decisions may include hiring new librarians with specific expertise, but as additional courses and training opportunities as described above are developed, it may become possible to deliver new services by providing training opportunities for existing staff (as well as adequate time off from other duties to complete training) and reassigning them as appropriate.

Many workshop participants had previously noticed job ads for DS/OS librarians that expected deep mastery of every one of the skills included in this report (or variations thereof), including fluency in one or more coding languages. Finding librarians with all of these skills – or “unicorns,” as some workshop participants put it – is not only difficult (since few librarians will have *all* skills), but is also probably unnecessary. Instead, it would make more sense to think about what specific services would be most valuable to the institution and look for librarians with the set of skills most likely to lead to success in developing such services. It may also be useful to think of data services as a team effort, rather than the purview of one single librarian. As discussed above, data science is often team-based, with a set of specialists each offering their own unique expertise to a shared effort. In the same way, libraries that need to provide a comprehensive and broad-ranging set of DS/OS services could think about composing a team of several specialist DS/OS librarians rather than expecting a single person to know (and be responsible for) the entire suite of services.

¹⁴ National Library of Medicine, “NLM Administrative Supplements for Informationist Services in NIH-Funded Research Projects (Admin Supp)”; National Library of Medicine, “Awards for NLM Administrative Supplements for Informationist Services in NIH-Funded Research Projects.”

Workshop participants also saw a role for library leadership in advocating for the library, as well as raising awareness of library DS/OS services at higher levels around the institution. DS/OS services are relatively new to libraries, and many researchers would not think to look to the library for help with their data- and open science-related issues. Library directors are uniquely situated to raise awareness of the library's involvement in DS/OS among other institution leaders. As has been discussed, establishing effective relationships with relevant institutional partners, such as IT, the Office of Sponsored Research or equivalent, and Clinical and Translational Science Awards (CTSA) programs, can be a useful approach to developing and promoting DS/OS services. Library leadership may be in a better position to broker such partnerships than the "on-the-ground" librarians providing DS/OS services.

Library leaders also have a role to play in creating an atmosphere that is conducive to the type of innovation that fosters DS/OS services. To be successful, DS/OS services must be effectively scoped and sustainable. Ensuring sustainability requires assigning staff appropriately and giving them adequate time to take on new responsibilities, which may mean that some other activities have to be given up or reassigned. It is also important to recognize that trying new things requires accepting that not all of them will work. Exploring new services, particularly in DS/OS, requires a willingness to embrace failure and to move on to new ideas. As previously discussed, it is important not only to define success and measure it via effective evaluation and assessment activities, but to act on those results and discontinue services when appropriate.

3. Conclusions

As this report has discussed, the landscape of data science and open science is both vast and quickly evolving. Given the range of researchers' needs, libraries have a wealth of opportunity to provide valuable services that can help advance their users' ability to excel in data science and open science. While specific technologies, policies, and practices may evolve over time, the set of core skills outlined in this report provide a foundation for librarians and information professionals to develop expertise that will enable them to provide innovative and beneficial DS/OS services to their institutions. Though some gaps and challenges remain, both new and current librarians have an increasingly broad range of training opportunities from which to draw to build these skills. This report provides a roadmap and a vision to build on existing training opportunities in order to develop a librarian workforce prepared to take on the data science and open science challenges not only of today, but of the future.

4. Sources

- Abele-Brehm, Andrea E., Mario Gollwitzer, Ulf Steinberg, and Felix D. Schönbrodt. "Attitudes Toward Open Science and Public Data Sharing: A Survey among Members of the German Psychological Society." *Social Psychology* 50 (2019): 252–60. <https://doi.org/10.1027/1864-9335/a000384>.
- Campbell, Hamish A., Mariana A. Micheli-Campbell, and Vinay Udyawer. "Early Career Researchers Embrace Data Sharing." *Trends in Ecology & Evolution* 34, no. 2 (February 1, 2019): 95–98. <https://doi.org/10.1016/J.TREE.2018.11.010>.
- Cope, Jez, and James Baker. "Library Carpentry: Software Skills Training for Library Professionals." *International Journal of Digital Curation* 12, no. 2 (2017): 1–8. <https://doi.org/10.2218/ijdc.v12i2.576>.
- Courneya, Jean-Paul, and Alexa Mayo. "High-Performance Computing Service for Bioinformatics and Data Science." *Journal of the Medical Library Association* 106, no. 4 (October 4, 2018). <https://doi.org/10.5195/JMLA.2018.512>.
- Cox, Andrew M., Mary Anne Kennan, Liz Lyon, and Stephen Pinfield. "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 1, 2017): 2182–2200. <https://doi.org/10.1002/asi.23781>.
- Federer, Lisa. "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>.
- Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. "Bibliometrics: The Leiden Manifesto for Research Metrics." *Nature* 520, no. 7548 (April 22, 2015): 429–31. <https://doi.org/10.1038/520429a>.
- Hosburgh, Nathan. "Developing a Bioinformatics Program and Supporting Infrastructure in a Biomedical Library." *Journal of eScience Librarianship* 7, no. 2 (March 16, 2018): e1129. <https://doi.org/10.7191/jeslib.2018.1129>.
- Institute of Museum and Library Services. "Positioning Library and Information Science Graduate Programs for 21st Century Practice," 2017.
- Library Carpentry. "Library Carpentry," 2019. <https://librarycarpentry.org/>.
- National Library of Medicine. "Awards for NLM Administrative Supplements for Informationist Services in NIH-Funded Research Projects." U.S. National Library of Medicine, 2019. <https://www.nlm.nih.gov/ep/InfoSplmnts.html#2017>.
- . "NLM Administrative Supplements for Informationist Services in NIH-Funded Research Projects (Admin Supp)." U.S. National Library of Medicine, 2017. <https://www.nlm.nih.gov/ep/AdminSupp.html>.
- OCLC Research. "ResearchWorks." OCLC, 2019. <https://www.oclc.org/research/researchworks.html>.
- Ogier, Andrea, and Michael Stamper. "Data Visualization as a Library Service: Embedding Visualization Services in the Library Research Lifecycle." *Journal of eScience Librarianship* 7, no. 1 (2018): e1126. <https://doi.org/10.7191/jeslib.2018.1126>.
- Read, Kevin, and Fred Willie Zametkin LaPolla. "A New Hat for Librarians: Providing REDCap Support to Establish the Library as a Central Data Hub." *Journal of the Medical Library*

- Association* 106, no. 1 (January 12, 2018): 120–26.
<https://doi.org/10.5195/JMLA.2018.327>.
- Sheehan, Jerry, Steven Hirschfeld, Erin Foster, Udi Ghitza, Kerry Goetz, Joanna Karpinski, Lisa Lang, et al. “Improving the Value of Clinical Research through the Use of Common Data Elements.” *Clinical Trials (London, England)* 13, no. 6 (2016): 671–76.
<https://doi.org/10.1177/1740774516653238>.
- Surkis, Alisa, Fred Willie Zametkin LaPolla, Nicole Contaxis, and Kevin B Read. “Data Day to Day: Building a Community of Expertise to Address Data Skills Gaps in an Academic Medical Center.” *Journal of the Medical Library Association : JMLA* 105, no. 2 (April 2017): 185–91. <https://doi.org/10.5195/jmla.2017.35>.
- Taylor, Chris F, Dawn Field, Susanna-Assunta Sansone, Jan Aerts, Rolf Apweiler, Michael Ashburner, Catherine A Ball, et al. “Promoting Coherent Minimum Reporting Guidelines for Biological and Biomedical Investigations: The MIBBI Project.” *Nature Biotechnology* 26, no. 8 (August 2008): 889–96. <https://doi.org/10.1038/nbt.1411>.
- Tenopir, Carol, Dane Hughes, Suzie Allard, Mike Frame, Ben Birch, Lynn Baird, Robert J. Sandusky, Madison Langseth, and Andrew Lundeen. “Research Data Services in Academic Libraries: Data Intensive Roles for the Future?” *Journal of eScience Librarianship* 4, no. 2 (2015): e1085. <https://doi.org/10.7191/jeslib.2015.1085>.
- Tenopir, Carol, Robert J. Sandusky, Suzie Allard, and Ben Birch. “Research Data Management Services in Academic Research Libraries and Perceptions of Librarians.” *Library & Information Science Research* 36, no. 2 (April 1, 2014): 84–90.
<https://doi.org/10.1016/J.LISR.2013.11.003>.
- The Carpentries. “Assessment and Impact,” 2019. <https://carpentries.org/assessment/>.
- . “The Carpentries,” 2019. <https://carpentries.org/>.
- Wilkinson, Mark D., Michel Dumontier, Ijsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (2016): 160018.
<https://doi.org/10.1038/sdata.2016.18>.
- Williams, Jason J., and Tracy K. Teal. “A Vision for Collaborative Training Infrastructure for Bioinformatics.” *Annals of the New York Academy of Sciences* 1387, no. 1 (January 2017): 54–60. <https://doi.org/10.1111/nyas.13207>.

Appendix A: Workshop Participants

Suzie Allard, Ph.D.

Associate Dean for Research, School of
Information Sciences
University of Tennessee

John Borghi, Ph.D.

Data Services and Visualization Librarian
Stanford Lane Medical Library

Matt Burton, Ph.D.

Lecturer, School of Computing and Information
University of Pittsburgh

Patti Brennan, Ph.D., RN

Director
National Library of Medicine

Ariel Deardorff, MLIS

Data Services Librarian
University of California, San Francisco Library

Chris Erdmann, MLIS

Community and Development Director
Library Carpentry

Lisa Federer, Ph.D., MLIS

Data Science and Open Science Librarian
National Library of Medicine

Erin D. Foster, MSLS

Data Services Librarian
Ruth Lilly Medical Library
Indiana University School of Medicine

Julie Goldman, MLIS

Research Data Services Librarian
Francis A. Countway Library of Medicine
Harvard Medical School

Ursula Gorham, JD, Ph.D.

Co-Director, MLIS Program & Lecturer, College
of Information Studies
University of Maryland

Margaret Henderson, MLIS

Health Sciences Librarian
San Diego State University Library

Mike Huerta, Ph.D.

Director, Office of Strategic Initiatives &
Associate Director of NLM
National Library of Medicine

Elaine Martin, MSLS, DA

Director and Chief Executive Office
Francis A. Countway Library of Medicine
Harvard Medical School

Jian Qin, Ph.D.

Professor, School of Information Studies
Syracuse University

Neil Rambo, MLibr

Retired Library Director, NYU Langone Health
Adjunct Curator, NYU School of Medicine

Kevin Read, MLIS

Data Services Librarian and Lead, Data
Discovery
New York University Health Sciences Library

Amanda Whitmire, Ph.D.

Head Librarian and Bibliographer
Stanford Hopkins Marine Station

Shirley Zhao, MSLS, MS

Data Science Librarian
Spencer S. Eccles Health Sciences Library
University of Utah

Appendix B: Workshop Agenda

Workshop on Developing the Librarian Workforce for Data Science and Open Science

National Library of Medicine Board of Regents Room | April 15-16, 2019

Day One – April 15, 2019

9:00 – 9:30 Welcome & Overview

- Patti Brennan, NLM Director: NLM welcome and NIH/NLM data science and open science activities and vision
- Elaine Martin, workshop chair: Workshop welcome and workshop purpose statement (To describe how librarians can support data science and open science, and the skills they will need to do so.)

9:30 – 10:45 Group Discussion – The Data Science and Open Science Landscape

- What services and assistance do researchers need to facilitate data science and open science?
- What are the different settings in which researchers conduct data science and open science (e.g. academic, industry, etc) and how do researchers' needs vary?
- How can librarians communicate with researchers about their data science and open science needs and practices?

10:45 – 11:00 Break

11:00 – 12:00 Group Discussion – Challenges and Opportunities for Librarians

- How can librarians communicate our skills to administrators, researchers, and other stakeholders and form relationships with relevant partners?
- How can librarians measure or evaluate the impact of services on researchers' data science and open science practices?
- How can librarians build or become part of teams for data science and open science, and what specific roles can they fill?

12:00 – 1:00 Networking Lunch – Expertise and Collaboration Interests

1:00 – 2:00 Small Group Breakouts – Gaps in Library Services for Data Science and Open Science

- Brainstorming on how libraries can address gaps

2:00 – 2:45 Reporting back and discussion

2:45 – 3:00 Break

3:00 – 5:00 Group Discussion: Core Skills for Librarians for Data Science and Open Science

- How can librarians adapt the skills that are already taught in iSchools to meet researchers' new data science and open science needs?

- What skills are not part of current curricula that librarians will need to support data science and open science?
- What skills do data specialist librarians need, and what skills do all librarians need?

Day Two – April 16, 2019

9:00 – 10:30 Group Discussion: The Future of Library Data Science and Open Science Services

- How can libraries address sustainability of research products?
- What roles do librarians have in promoting reproducibility and research ethics?
- How can librarians facilitate data science and open science in a quickly-changing technological and policy climate?

10:30 – 10:45 Break

10:45 – 11:45 Group Discussion: Preparing the Current and Future Librarian Workforce

- What gaps exist between training and curricula that are already available and the core skills we have identified?
- How will iSchools need to modify or add to their curricula to prepare future data science and open science librarians?
- What opportunities exist (or should be developed) for practicing librarians to gain relevant skills?

11:45 – 12:00 Break

12:00 – 1:00 Working Lunch – Conclusions and Closing Thoughts

Appendix C: Existing Courses

The courses below provide training on topics related to research data management, data science, and open science, and may be useful for librarians interested in gaining some of the skills discussed in this report. This list was created by Sarah Clarke and Candace Norton, and is maintained as part of a larger project on research data management training for librarians at <http://doi.org/10.17605/OSF.IO/8SJR5>.

Course Name	Sponsor	Website
Best Practices for Biomedical Research Data Management	Harvard	https://www.canvas.net/browse/harvard-medical/courses/biomedical-research-data-mgmt
RDM 102: Beyond Research Data Management for Biomedical and Health Sciences Librarians	U.S. National Library of Medicine (NLM) and the National Network of Libraries of Medicine Training Office (NTO)	https://news.nlm.gov/nto/2018/11/28/applications-open-for-rdm-102-beyond-research-data-management-for-biomedical-and-health-sciences-librarians/
RDM 101: Biomedical and Health Research Data Management Training for Librarians	U.S. National Library of Medicine (NLM) and the National Network of Libraries of Medicine Training Office (NTO)	https://nlm.gov/classes/biomedical-and-health-research-data-management-librarians

Course Name	Sponsor	Website
Data Management Course	University of Minnesota (in partnership with Purdue University, Cornell University, and the University of Oregon)	https://sites.google.com/a/umn.edu/data-management-course_structures/home-1
Data Management Expert Guide	Consortium of European Social Science Data Archives (CESSDA)	https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide
Data Management for Clinical Research	Vanderbilt University	https://www.coursera.org/learn/clinical-data-management
Data Management Short Course for Scientists	ESIP Foundation (in partnership with NOAA and the Data Conservancy)	http://commons.esipfed.org/datanamagementshortcourse
Data Science and Visualization Institute	North Carolina State University and the Coalition for Networked Information	https://www.lib.ncsu.edu/data-science-and-visualization-institute
DataONE Education Modules	DataONE, US National Science Foundation	https://www.dataone.org/education-modules

Course Name	Sponsor	Website
Essentials 4 Data Support (formally Data Intelligence 4 Librarians)	Research Data Netherlands	https://datasupport.researchdata.nl/en/
Fundamentals of Research Data Management	University of Washington	https://canvas.uw.edu/courses/889213
Library Carpentry	Library Carpentry	https://librarycarpentry.org/
MANTRA; Research Data Management Training	University of Edinburgh	https://mantra.edina.ac.uk/
NNLM Research Data Management Webinar Series	National Network of Libraries of Medicine (NNLM)	https://nnlm.gov/classes/nnlm-research-data-management-webinar-series
RDMRose	University of Leeds, Sheffield and York, with the Information School at Sheffield	http://rdmrose.group.shef.ac.uk/
Research Data Management	Library Juice Academy	http://libraryjuiceacademy.com/082-data-management.php

Course Name	Sponsor	Website
Research Data Management and Sharing	University of North Carolina at Chapel Hill; University of Edinburgh	https://www.coursera.org/learn/data-management