Chapter 3

People Who Need Data People: A Case Study in Building Cross-Campus Data Support

Cameron Cook  
*University of Wisconsin-Madison, USA*

Heather Shimon  
*University of Wisconsin-Madison, USA*

Sarah Stevens  
*University of Wisconsin-Madison, USA*

Trisha L. Adamus  
*University of Wisconsin-Madison, USA*

Clare Michaud  
*University of Wisconsin-Madison, USA*

**ABSTRACT**

UW-Madison has a thriving community of support around research data management and sharing, computational skills, and research reproducibility which are all building blocks for supporting campus data science needs. Support is provided through many campus organizations such as the Center for High Throughput Computing, the Research Cyberinfrastructure Initiative, Research Data Services, the Data Science Hub, and more. However, informal and formal interviews with researchers at UW-Madison have shown that as the campus invests in technological solutions and infrastructure to address common research bottlenecks, researchers are less limited by access to data and computing resources, but rather the knowledge to find, understand, and use those tools effectively. Specifically, researchers need "data people," or people who can help them address these knowledge-based needs, who can fill in gaps in research support, who can advocate for their needs across campus silos, and who can build broad networks of relationships across campus to facilitate a better support experience.

DOI: 10.4018/978-1-7998-9702-6.ch003
INTRODUCTION

The University of Wisconsin–Madison (UW–Madison) has a thriving community of support around research data management and sharing, computational skills, and research reproducibility, which are all building blocks for supporting campus data science needs. Support is provided through many campus organizations such as the Center for High Throughput Computing (https://chtc.cs.wisc.edu), the Research Cyberinfrastructure Initiative (https://researchci.it.wisc.edu), Research Data Services (RDS) (https://researchdata.wisc.edu), the Data Science Hub (https://datascience.wisc.edu/hub), and more. However, informal and formal interviews with researchers at UW–Madison have shown that as the campus invests in technological solutions and infrastructure to address common research bottlenecks, researchers are limited not by access to data and computing resources but rather the knowledge to be able to find, understand, and use those tools effectively (Bloom et al., 2018; Cook et al., 2021). Specifically, researchers need “data people,” or people who can help them address these knowledge-based needs, who can fill in gaps in research support, who can advocate for their needs across campus silos, and who can build broad networks of relationships across campus to facilitate a better support experience.

This case study focuses on the work of librarians with RDS and the facilitators in the Data Science Hub to fill these needs. RDS is an interdisciplinary group made up of librarians, information technology staff, and research staff, who are committed to advancing research data management and sharing best practices on the UW–Madison campus. The Data Science Hub, with a staff of three facilitators, carries out a mission of community engagement and learning opportunities around data science and computing for campus researchers through a variety of services.

This chapter begins with a framing and history of the current landscape of research data and data science support on the UW–Madison campus, including a history of research IT support, the development of both groups, and how the units have worked together and leveraged each other’s expertise to provide vital programs and resources at a large, decentralized institution. Important to this discussion is the value of interpersonal skills such as communication, active listening, and creativity that have been found critical in these support roles. The “The Need for Data People” section explores the library literature pulling from surveys, job descriptions, and essays of data services in libraries showcasing how interpersonal skills are crucial in addressing researchers’ data needs.

The authors believe that the efforts outlined in the chapter illustrate that human infrastructure is necessary alongside technical infrastructures to develop robust and sustainable data science support within libraries and on academic campuses in general. Support services that provide people with interconnected infrastructure are vital in the growth of any new area of research, especially as it becomes increasingly data-intensive and requires coding skills (Baker, 2017).

BACKGROUND

The development of RDS, the Data Science Hub, and the UW–Madison Carpentries Community—which are the focus of this case study—are some of the units and efforts that were formed by and have contributed to the current system. Research-focused information and technical support services at UW–Madison are still often provided by small units that are managed at departmental or school levels, meaning that many researchers struggle to identify where to go for help or may end up purchasing support, tools, or services through their own channels. With the onslaught of data growth along with increasing interest
in computational methods over the last decade, gaps in data management and computing support for research became increasingly evident. Given the often slow-to-change environments of research-intensive universities, these gaps were filled through a variety of grassroots efforts that have since been formalized. The authors want to recognize that there is a constellation of other data and computing support services that also formed through grassroots efforts and are inextricable partners in the success of the work the units all do; these include the Advanced Computing Initiative, the Center for High-Throughput Computing, the Research Cyberinfrastructure Initiative, among others.

**Research Data Services**

RDS was formed circa 2009 as a pilot collaboration by a coalition of invested parties from the Libraries, the Division of Information Technology (DoIT), the Office of the Chief Information Officer, the iSchool (formerly School of Library and Information Studies), and the Office of the Vice Chancellor for Research and Graduate Education (formerly the Graduate School). The original mission of RDS was to “assist UW–Madison researchers in their efforts to preserve and maintain their data” (Research Data Services, 2011b), and these efforts were aligned with campus IT having identified digital data curation as a strategic initiative at that time (Research Data Services, 2011a). This pilot offered a few services, including consultation on data management, assistance with technologies, and referrals to campus resources (Research Data Services, 2011b).

By the mid-2010s, RDS still offered these original services but expanded its scope to include on-demand training and a new focus on community building, largely through events such as lunch-and-learn events, in which invited researchers shared personal experiences, projects, or ideas for improving and championing data management and sharing. Invited speakers ranged across all major research domains and also included representatives from campus IT services such as Cybersecurity and Identity and Access Management. Past presentation titles included: “Geographic Data and Confidentiality in Health Research,” “Data Pipelines and Computational Methods for the Social Sciences,” “Recovering the London Stage Information Bank (1970–1978): Data Preservation Lessons from an Early Humanities Computing Project,” and “Identity Federation for Research Data Sharing.” As federal funding agencies released plans in response to public access requirements around 2015, RDS continued to develop its services in line with the impacts of these plans. Further support included assistance with creation, review, and feedback on data management plans through DMPTool (https://dmptool.org), tailored training and education, expanded online resources, and support for compliance with public access requirements, including support for depositing data into MINDS@UW (https://minds.wisconsin.edu) the UW–Madison institutional repository, and to the Dryad (https://datadryad.org) data repository through an institutional membership. Over the years, RDS has continued to seek partnerships with several campus units to pursue special projects. Examples of such projects include: piloting data management education in the undergraduate classroom with an honors biology program; an undergraduate-focused program of evening dinner events on specialized data management topics; partnering with Research Cyberinfrastructure Initiative (RCI) to use Globus with the institutional repository to provide public access to bigger datasets; and partnering with RCI and Ebling Library for the Health Sciences staff to launch a UW–Madison version of the Data Storage Finder (https://storage.researchdata.wisc.edu) tool for researchers based on the module originally built by Cornell (https://finder.research.cornell.edu).

Since its inception, RDS transitioned through many iterations of staffing and sponsor support. Currently, RDS is a library-led program that is staffed by a dedicated chair position housed in the Libraries,
a portion of a data and digital scholarship-focused librarian position, and then many disciplinary and functional experts from all across the UW–Madison campus who provide portions of time that are either directly allotted or volunteered. These experts come from partnerships with the already mentioned RCI and Ebling Library within the School of Medicine and Public Health, along with the Science and Engineering Libraries, the Map Library, Memorial Library, the Data Science Hub, the Division of Information Technology, and others. RCI, within the Division of Information Technology, has been a critical supporting partner for RDS. They have provided not only consultation support and time given from their specialists but also infrastructure support by hosting the RDS website and the Data Storage Finder, as well as supporting communications and marketing needs.

This coalition of invested partners and experts helps to bring a breadth of expertise and unique lenses to user questions. In turn, this helps to create a more holistic support pipeline for data management and sharing for the campus, as well as maintaining a broad network of relationships for referrals. This wide network and their dedication to working together to support researchers across the gaps in campus services and communications remains the greatest strength of RDS at UW–Madison.

The Data Science Hub

In 2018 the Data Science Hub started as a new service to support researchers around data science and research computing at UW–Madison. The Hub was created in the Wisconsin Institute for Discovery (WID), the university’s interdisciplinary research institute, which originally identified the need for a coordinated campus strategy around data science. Initial funding came from a short-term campus grant for new collaborative research projects. Momentum for the development of the Hub came from a community of professors, researchers, and research support staff that recognized the need for more centralized support around applying data science to research in many subject areas. The effort started with the WID hosting a town hall meeting to get feedback from the campus stakeholders about data science needs. At the town hall, a variety of researchers on campus presented short talks about their needs and followed up with an open discussion. The interest and variety of perspectives shared at the town hall supported the idea that the need for data science support was not specific to only a few subject areas but rather a campus-wide research need. Using town-hall meeting feedback, a small group of professors and staff from several campus departments put together an initial proposal for the Data Science Hub, led by individuals from the Wisconsin Institute for Discovery, the Department of Computer Sciences, the Department of Statistics, the Department of Horticulture, and the Center for High-Throughput Computing. The development of the Hub catalyzed several new efforts to support the growing field of data science, including a research institute and degree programs.

The Hub has worked to build community and support researchers when using data science in their research through workshops, consultations, projects, and other events. As outlined in the following section, the Hub coordinates the Carpentries workshops at UW–Madison, which was a key entry point for collaboration between RDS and the Hub. Other activities provided in collaboration with RDS are described later in this chapter. The Hub also works to support communities of practice that allow researchers to share their experiences and learn from one another. An example of one such community is the Machine Learning Community of Practice which hosts a regular event called ML+X, where researchers can share and discuss best practices and ethical considerations for applying machine learning to their work. Many of these activities are coordinated by the Data Science Hub and then enacted by a campus community of individuals supporting and using data science.
Building the Carpentries Community

The Carpentries are a “global community teaching foundational computational and data science skills to researchers in academia, industry and government” and provide free online lessons, as well as additional instructional support to member organizations (Carpentries, n.d.a, footer). Software Carpentry has a long history of development at UW–Madison. A community of primarily physics researchers, in conjunction with support from the Center for High Throughput Computing (CHTC), started running Software Carpentries workshops as early as 2013. Initially, many of these efforts were part of the Advanced Computing Initiative (ACI, https://aci.wisc.edu), a short-term funding initiative that was developed because faculty and researchers recognized the need for more support around computing. Facilitators from the CHTC were instrumental in organizing the workshops on a regular basis and nucleating the development of an instructional community.

The initial community of instructors comprised facilitators from CHTC, campus research staff, graduate students, and faculty. RDS began attending and then assisting with local Carpentries efforts in early 2015, with many campus librarians also joining the community through a connection with RDS or by attending a workshop. The Libraries also co-hosted an early version of Library Carpentry with the local iSchool. Since then, multiple RDS consultants and campus librarians have become certified instructors. The development of the instructional community has included many individuals who attended a workshop, followed up by helping with the workshops, and then eventually began teaching at workshops.

Prior to the development of Data Carpentry, ACI hosted three Software Carpentry workshops each year. In 2016, after the success of the first Data Carpentry workshop, ACI expanded to hosting three Software Carpentry and three Data Carpentry Ecology workshops. In 2018, the funding for the ACI ended, and its services were distributed to a variety of other campus partners. The newly created Data Science Hub carried forward the Carpentries workshops and community organization role.

When the Data Science Hub became the workshop organizer and host, the Hub expanded the number and types of Carpentries workshops to ten or more workshops each year, including Software Carpentry, many different Data Carpentry workshops, Carpentries Instructor Training, and other half-day workshops using established and newly developed Carpentries lessons. The Hub also started monthly instructor development meetings for the instructional community to discuss their teaching practices and develop connections within the group. At present, the UW–Madison Carpentries instructional community includes approximately seventy individuals with about forty active instructors/helpers each year. Members of the instructional community include staff (mainly librarians and research staff), graduate students, and faculty with a variety of research backgrounds, including the sciences, such as physics and biology, and the humanities, such as history and languages. Current areas of expansion include creating and piloting lessons in new areas such as the digital humanities and intermediate level workshops that build upon the existing workshops, as well as developing training for lesson development.

THE NEED FOR DATA PEOPLE

These detailed histories of grassroots action and growing networks of research support lead to a larger conversation on the need for investing in the people, or “data people,” with both the technical and interpersonal skill sets required to do this work. At broader institutional levels, there has been a continued lack of centralized, or at the very least coordinated and research-focused, data support. At large institu-
People Who Need Data People

tions such as UW–Madison, services are often siloed and decentralized, creating an environment that is more frequently reactive rather than proactive to emerging researcher needs. This breeds opportunities for duplication of efforts and costs across individual units and increases the obstacles to investing in shared infrastructure such as hardware, storage, institutionally licensed tools, and dedicated staff support roles. While the campus has been investing in data science support and computing infrastructure in very recent years, there is still a reluctance to coordinate or sustain training, consultative, and curatorial or preservation efforts. Focus remains on the concrete infrastructures rather than the human infrastructures and services needed to help researchers make the best use of campus investments and resources for their research. At UW–Madison, there has been established investment in technical infrastructures, and often, staff capacity has had minimal or no investment from campus leadership. However, investment in technology infrastructures, while critical to supporting the needs of researchers, forgets that human infrastructures are also critical, not only to stewarding and caring for the technology but also for enabling the use of, understanding of, and critical engagement with those technologies.

In decentralized environments, researchers often struggle to locate support services or know where to start. There are also large learning gaps in using new technologies or reworking workflows that are not supported by campus. Researchers often want IT staff or research support staff to simply tell them what to do, so that they can get back to doing their most important job: research. Having to search for assistance, make infrastructure and workflow decisions, and account for set-up time builds frustrating delays into their research (Cook et al., 2021). Researchers would benefit from a data person who can help them navigate the service environment and make informed decisions on their pipelines. In another vein, many researchers also rely on student labor to help them administer infrastructure, acquire or analyze data, or leverage new tools. As student labor turns over, this causes roadblocks in their work as they try to address orphaned code and data as well as fill in stewardship roles for their computing infrastructure (Bloom et al., 2018; Cook et al., 2021). Students are learning these technical skills as part of their research and may have little support from their peer researchers. Both researchers and students would benefit from having people who can assist them in learning the needed skills to power their research and would make turnover and reliance on students for computational skills less of a roadblock. Researchers are attempting to fill gaps in their needs where and how they can, and they would welcome assistance from campus resources. However, that requires having dedicated personnel—data people with interpersonal skill sets—charged with teaching computational skills aimed at closing these gaps, consulting with researchers to identify specific needs, and finding solutions.

Interpersonal Skills and Librarians in Data Support

In the above environment, with an absence of broader campus goals and a broader leadership vacuum around research data support, librarians have taken up the role of attempting to address the big picture of research data support needs on campus. The authors think librarians are well-poised to be one of the partners to fill such a gap. The library literature provides examples of the need for data people with interpersonal skill sets, which Federer called personal attributes, and includes teamwork, communication, and creativity (2018, pp. 297–298). In a survey of 82 data specialists who work in libraries, Federer found that personal attributes was the highest-rated category of traits that respondents considered important to succeed in data support. This was chosen out of nine categories, the others being: two categories of data management skills, technology skills, evaluation and assessment skills, teaching skills, library skills, networking and outreach skills, and education. Each category contained individual items that
People Who Need Data People

the respondents rated. Out of the 47 individual items, the following five were ranked the highest: “Developing relationships with researchers, faculty, etc.,” “Oral communication and presentation skills,” “Teamwork and interpersonal skills,” “Written communication skills,” and “One-on-one consultation or instruction” (pp. 297–298). These findings align with the interpersonal skill sets that, the authors propose, need increased attention when building the staffing infrastructures necessary to connect researchers to emerging technologies.

Looking at 104 advertised job descriptions for library professionals in the area of research data support, Eclevia et al. found that interpersonal and communication appeared the most frequently as a required skill in 68 job descriptions, as opposed to software and tools, which appeared as a required skill in twenty descriptions (2019, p. 283). The most listed personal traits were innovative, creative, service-oriented, and collaborative (p. 280). Although knowledge of research data management and technical skills also appeared heavily in the job descriptions, interpersonal skills were given considerable weight. When examining the duties listed in the job descriptions, they found those that were related to people, such as collaborating with researchers, providing consultations, and the development of data services, had the highest number of occurrences, 443, as opposed to 251 occurrences of duties that were related to tools such as collections, technologies, and repositories (p. 286). The job descriptions showed a strong desire for interpersonal skill sets in hiring for research data support in libraries.

Commentaries in the library literature also stress the importance of interpersonal skills in research data support. Henderson (2020) outlined five important areas that use interpersonal skills—including reference, relationships, collaboration, listening, and facilitation—and made the case that they are crucial to support researchers in identifying exact needs, forming communities of support, and guiding them to best tools and practices. Rinehart examined the strong emotional connections that researchers can have with their data due to the pressures of acquiring funding and meeting scholarly expectations (2015). These emotions can create barriers for researchers when faced with unfamiliar data management and sharing requirements or learning new tools for computational analysis. Rinehart shared how active listening was instrumental in guiding a principal investigator (PI) through the process of writing a data management plan. The PI was initially uncommunicative but later admitted to struggling, finding the process to be new and unfamiliar. By using interpersonal skills in communication and listening, Rinehart was able to identify the data needs and navigate the PI to an institutional repository (pp. 438–439). Without these interpersonal skills and the data people to apply them, the technical infrastructures become less accessible and underutilized; their potentials remain unrealized.

This case study is framed within a larger commentary on building a network of people, and particularly, research support roles focused on interpersonal and translational skills. These roles, and the interpersonal skill sets that tend to define them, have been historically undervalued within research, yet they are critical for helping researchers navigate ever-increasingly complex and fast-paced research environments, for advocating for and developing new services based on known needs, and for providing leadership in emerging areas. To articulate more fully the roles that research support staff members fill, the authors advocate for broadening the notion of infrastructure within the research space. While hardware products have traditionally defined research infrastructure, communities and research support staff are infrastructure, too. A community of research and data support staff, including librarians, can provide researchers a first point of contact before they pursue the use of a specific data collection, analysis, or storage tool.

In building these communities, the authors encourage campus-wide data science units to collaborate with libraries. In addition to valuable collaborative efforts, such partnerships can also provide campus units with access to extensive domain knowledge for understanding the departmental research needs
and to campus networks within departments. Academic research librarians are a valuable asset to any research support team, both as collaborative partners and as directly hired personnel for a project or data science unit.

From the perspective of the author who works as a facilitator for the Data Science Hub, working with individuals from the library has emphasized how much external campus units can learn from libraries in supporting research. Researchers often misunderstand the role of data librarians and think that it is limited to the management of book and journal collections. This misperception then makes it difficult for individuals to recognize possible collaborations in the areas of data management, privacy and security, publication, and sharing and archiving, which have long been important aspects of librarianship. Non-library units often seek to hire team members who have experience doing research, software engineering, or other technical activities but frequently have little to no training or experience in research support itself and the interpersonal skills needed to do the work. Directly hiring librarians with functional expertise as a part of non-library units can fill in this gap in experience and equip the team to better support researchers. If researchers value diversity of thought as a way to improve research outputs, support needs to be provided for people with diverse learning styles, backgrounds, and identities to access data science tools and skills. This requires creating collaborative teams with a diversity of interpersonal and technical skill sets.

At many institutions, librarians are among the first to provide support for data sharing requirements. As such, they have experience filling gaps in campus information needs with minimal resources. On top of this, to be able to provide data management and sharing support, they also have to have a well-rounded view of the research and data lifecycle. They also have established and deep ties to departments, are well aware of researchers’ needs in situ and at a practical level that campus and IT administrators may not be, and often have broad cross-campus functional and cross-disciplinary networks from which to draw support and collaborations. These strengths make libraries well-prepared both to understand researcher needs and to call upon their wide networks to fill gaps. As libraries already exist as interdisciplinary and centralized services at many campuses, they are also well-suited to helping alleviate siloed support areas. The authors recommend that library and campus leadership look to their data and subject librarians for their expertise and ideas on building collaborations for data support, and then prioritize and advocate from their recommendations. It is also recommended that data and subject librarians get involved with campus interest groups, communities of practice, committees, and other forums at which one can add their voice in order to help build awareness of data support needs and to start building a coalition of collaborators that bring unique lenses to the work.

**RDS and the Data Science Hub Working Together as Data People**

The Data Science Hub and RDS have leveraged their relationship to provide a safe and creative space for the units to collaborate and experiment with projects. The Carpentries provided a starting point that allowed the units to grow a strong partnership centered around the shared goals of improving campus data literacy, teaching reproducible research skill sets, and supporting existing gap areas in research data support. Both units see a natural alliance between each other’s skill sets, interests, and values, and strive to be data people. A crux of the success of working together is that both units share a number of the same values. The Hub follows the Nine Core Values of the Carpentries, and library staff in RDS follow the UW–Madison Libraries Values statement as well as professional library values. There is much overlap in Carpentries and Libraries values of how the units work together with emphasis on collaboration, inclusion,
People Who Need Data People

and empowerment (Carpentries, n.d.b; UW–Madison Libraries, n.d.); as well as library professionals’ values of access, diversity, and lifelong learning, as stated in the American Library Association’s Core Values of Librarianship (ALA, 2006). Pedagogical approaches have also aligned in striving to create welcoming and learner-centered environments that include hands-on learning techniques, such as code-along sessions in the Carpentries workshops and guided database searching during library instruction sessions. The instructor development meetings have had guided discussions on other shared interests, such as stereotype, threat active learning, and feminist pedagogy.

In addition to the Carpentries community, the units have taken opportunities to partner in low-resource, extensible ways. One example has been the Hub’s Coding Meetup, a twice-weekly opportunity for drop-in consultations hosted by RDS members and other Hub stakeholders. Coding Meetup regularly brings together individuals with questions about how to apply new technical skills to their work and those with expertise to answer questions. One researcher who attended a Carpentries workshop reported finding the follow-up help in Coding Meetup instrumental to applying skills learned in the workshop (Data Science Hub, 2021). The availability of data people in Coding Meetup provides a uniquely supportive environment for applying new technical skills compared to attending a workshop alone, attending online courses, or using self-guided learning.

RDS has regularly sat on the planning committee and participated as presenters at the Hub’s Data Science Research Bazaar, a yearly campus conference to build a community for “researchers, data scientists, entrepreneurs and community members, including students,” and is modeled after an international event “promoting the digital literacy emerging at the center of modern research” (Data Science Hub, n.d., paras. 1-2). Again, data people are using interpersonal skills to bring people together for information sharing and knowledge creation to support computational research on campus and in the community.

The Data Science Hub also contributes to the data science support efforts of RDS consultants and library staff. Members of the UW–Madison Carpentries Community volunteer as instructors and helpers for a Python and R programming workshop series organized by the Libraries once each semester. Without this pool of volunteers, the Libraries would not have the staff to hold these workshops. These workshops use a mix of original and Carpentry-based lessons broken up into two-hour sessions instead of the full-day Carpentry workshop modules. The Data Science Hub also promotes this series in its newsletter. During the COVID-19 pandemic, the Hub provided helpful guidance in moving the series to an online format. Learners in the series have been from all campus departments, and workshops always hit registration limits, showing the great need on campus for this training.

The units have also collaborated, along with a partner from the university’s interdisciplinary research institute and a partner from the Research Cyberinfrastructure Initiative, to revamp a campus resource guide called the Researcher Toolkit (https://researchertoolkit.wisc.edu). This resource guide was aimed at bringing all campus research-related resources and services together to make them easier to find and navigate. The updated guide not only included more items, but also provided context of how to incorporate the resources and services into the research process by mapping them to the phases of research data lifecycle. Pathways for feedback and a plan for long-term maintenance were also added.

Knowing that there is ample opportunity between RDS and the Hub, they have made attempts to formalize their connection to enable more concerted efforts and time for special projects. In an initial attempt to establish more formal connections, the RDS chair and lead data science facilitator drafted a joint position that would have split time between the Hub and RDS. The position was intended not only to support the growing needs of each unit but also to put effort into shared initiatives. Fortunately, the position was untenable at the time of the proposal, but the authors encourage other institutions to
explore such a position. It would have allowed for more concerted efforts and time dedicated to sustaining shared work as well as expanding into new exploratory projects and services for both units. Such a position would also formalize the Libraries and the Hub as partners in providing research data support to campus and would promote more stable collaboration and centralization of research support. Since this proposal, the units have been able to maintain semi-formal relationships through (a) the chair of RDS sitting on the steering committee for the Data Science Hub and (b) one of the Data Science facilitators serving as a consultant for RDS.

Both RDS and the Data Science Hub benefit from having each other on campus for support and feedback. Both units emphasize relationship building at the core of their work, seeking to cultivate networks of resources and colleagues for referrals. Both units also attempt to serve as centralized, free resources for all campus researchers regardless of their campus unit. The two units also share a common approach of using a combination of interpersonal and technical skills to help support researchers, and members of each unit see themselves as data people. These shared goals and values allow the units to lean on each other and also to dream and innovate together—whether discussing new pedagogical approaches, pivoting in-person events to online, or strategizing for expanding workshop offerings. The authors hope that this example of connecting with another campus unit through a program or service and expanding on shared goals and values will be useful to librarians looking to establish cross-campus partnerships.

AREAS OF OPPORTUNITY

Internally, libraries often face continued challenges around data support. Locally, the authors have found shifting traditional perceptions on the scope of research services, empowering librarians in learning data science skills, and the implementation of library-wide data support services challenging. As research becomes more interdisciplinary and more disciplines take advantage of big data and computational methods, the scope of librarians’ work is also widening. The authors want to caution libraries against perpetuating the phenomena of creating and hiring for “unicorn” functional or specialist roles, which are often created in reaction to this widening scope. Often, data librarians have the whole scope of RDS shoehorned into their responsibilities even though these librarians are often at full capacity running a small segment of their organization’s data services. Sustainability is only possible with the addition of more staff or accepting negative growth, which is undesirable and unrealistic.

Adding staff requires investment in either hiring new roles or re-skilling and reshaping the understanding of librarians’ responsibilities in the research services space. This can be challenging, as some librarians display hesitancy in engaging with data skills across the spectrum. The authors think there are many opportunities for libraries to help facilitate engagement by fostering deeper relationships and opportunities for collaborations that draw on the unique expertise of both functional specialists and subject specialists. Subject specialists are critical to the success of data services in libraries, given their deep understanding of their disciplines’ standards and workflows. Libraries could also empower librarians to address data science needs by creating opportunities for librarian-led communities of practice and supporting librarians interested in leading technical workshops. Locally there has been success with librarian-led groups communally learning Python and high turnout to libraries-hosted Python and R workshops. An avenue to reshaping services can be to identify ways to shift traditional research support, such as incorporating data sharing compliance into grant support or branching into visualization of publication metrics as part
of impact services. Pushing the needle and taking small steps with traditional services can help reshape the view that librarians and researchers have of library services.

The authors have also found that locally, librarians who do engage with data skills may feel isolated from other technologists on campus and not feel confident in those spaces. In Carpentries instructor training, one author has observed that oftentimes librarians come in with more experience and training with teaching as well as strong perspectives on the needs of researchers compared to their research computing support colleagues but lack confidence in their technical skills. Providing small groups or technical skill-up workshops specifically for the instructional community is one possible way to address this observed issue.

There is room for improvement in mass communication methods, especially coordinating need-to-know information and services across critical units. It is unfortunately still the case that researchers on the campus struggle to locate services and units among the swarm of information and resources available to them on a decentralized campus. People do not always find RDS or the Hub at their point of need, leading to confusion, frustration, and potentially not locating the right assistance. An opportunity the authors see in this area is building better relationships with units that are core touch-points for researchers. Stronger relationships with such units would ensure data sharing and data science services are highlighted on the research office website or that research office staff are aware of data service providers.

In this section, the authors have discussed some of the local challenges and opportunities faced but want to acknowledge that there are greater existing challenges for all libraries that can make it difficult to meet basic operational needs, let alone grow nascent service areas. The authors want to take this opportunity to recognize that libraries find themselves in a moment of continuing transition from traditional services to sustained efforts in computational and data-intensive research support. On top of this transition, many libraries also find themselves in a moment of uncertainty within the profession as we continue to see libraries grappling with growing costs and shrinking budgets, staff capacity, and burnout, as well as changing services, priorities, and spaces. These challenges are only amplified at state institutions with funding dependent on the political climate of the state legislature during each biennial budget. Supporting and participating as a campus partner in data science activities can be challenging even to the best well-resourced institutions, but it can be even more challenging for those who face budgetary challenges, staffing challenges, and changing environments. The following section expands upon observed challenges, but it is important to recognize that these are often highly complicated situations. The authors hope these lessons provide a learning opportunity for other institutions.

RECOMMENDATIONS

Building upon the known challenges shared above, the authors also want to share forward-looking recommendations for those libraries looking to start developing staff into data people and building collaborations to support data science. As a core recommendation, libraries should develop a concrete vision for their participation in data science efforts at their institution that they are able to convey both internally to their staff and externally to campus. Vision is critical to establishing norms and expectations and helping staff find purpose in their roles. Does the library see itself as a partner in data science research? As critical to helping students learn computational skills? As a proponent for equity in data science methods and tools? Library leadership clearly communicating a vision can help units align their work in the short term and help ensure the development of proactive services and engaged librarians in the
long term. Concrete vision and goals create alignment by helping set motivation and direction for data librarians while also creating agency within their roles, helping them scope work and know where to put their efforts and creativity. With external campus partners, a communicable vision enables leadership to signal interest and willingness to invest in the short-term and enable opportunities for partnership and a deeply embedded library role in data science in the long term. Staking a claim in data science support on campus will help partners understand how they can collaborate with the library and make it easier for them to see the value of what the library can bring to the work. With a driving vision established, the authors encourage the development of concrete goals for both research services and staff professional development, so that the vision is measurable and achievable and moves both the human and technical infrastructures forward together.

The authors also encourage libraries to cultivate partnerships with other units such as IT and research units to develop community and to build this into their vision. Libraries should not feel pressure to solve campus data science needs but instead see themselves as one of the existing services upon which a campus should draw to create robust data science support. Leveraging existing services and using collaboration to improve or complement them is one of the easiest and most effective ways to bolster support. In fact, building coalitions for data science support among campus units is one of the best ways to fill infrastructure gaps. At large institutions, it can be common for campus units to compete in services or accidentally create redundancies as they build out services. Taking time to build invested coalitions can help ensure that units are leveraging each other’s strengths, centralizing efforts, and avoiding some of these challenges. In line with this, the authors also encourage libraries to invest time and effort into advocating for the librarian’s role in data science support with campus partners. While libraries spin their wheels trying to figure out how to support data science efforts at their institutions, other campus units are filling the gaps that libraries could have been well-suited to fill. Leadership should look to their data librarians, who likely already have campus networks on which to rely and ideas for ways to fill gaps in the research support landscape.

Another overarching recommendation that sits at the heart of this case study is that investment in supporting data science, whether as a library or as a campus, requires an equal commitment to investing thoughtfully in the human infrastructure to accompany the technical infrastructures. Creating holistic research infrastructures involves investing in data people. Such infrastructures leverage the technical roles of back-end staff, system administrators, and developers and place the same weight of importance on staff who are facilitators of learning, using, and ethically engaging with new research methods and technologies. While strategic hiring of such data people is a major element of this recommendation, another often overlooked component is the professional development available to data librarians and other data-intensive research support specialist roles once hired.

Specifically, data librarianship does not fit within traditional library roles or other traditional research paths at an academic institution. These roles are often called out in library strategic plans to showcase how the libraries are fulfilling goals for supporting “emerging” or “innovative” research. However, these roles and their associated small units often struggle to receive the same recognition for their work or to acquire the same funding and administrative support as other technical or long-established units within the library. The support and funding given to traditional technical roles both reveals and perpetuates the devaluation of interpersonal skills and the patron-facing or instructional work that data support roles often include. This encourages data librarians to develop and wield disparate skills, being both competent practitioners of computational skills on top of the ability to interface and translate across a wide constituency of patrons and stakeholders. Devaluation of interpersonal skills, in turn, contributes to long-term
burnout as data librarians often spend great amounts of energy advocating for both their emerging roles and the rapidly changing needs of campus researchers. However, their initiatives frequently remain slow to gain traction and growth.

The authors encourage library leadership and unit heads to question what progression within each of these roles would look like. How are librarians enabled to grow their skills and find opportunities for data projects or new work? How do units operationalize services in order to allow for professional growth and sustainability? One way to enable this within roles is to build in professional development through the ability to participate in deep-dive or collaborative projects that can help practice and build computational skills. Often librarians build these skills in order to teach workshops or assist with consultations, but they are not able to apply them to the benefit of research projects or to showcase work within libraries. Without these opportunities being built into roles by administration and supervisors, data librarians may not be able to find the extra time in their often-overloaded responsibilities. Another opportunity in this area would be to formalize relationships or dedicated efforts with data science or computational support units into library staff position descriptions. Ensuring there is significant dedicated time in positions for this work also means that librarians are able to prioritize it, establish yearly goals towards it, and count it towards their review processes. Finally, the authors also encourage further discussion within professional organizations and conferences. Discussion of data librarianship and its work with data literacy and computational skills is often sequestered to data librarian-specific conferences, rarely making appearances at the larger professional or disciplinary conferences. This would help not only to cross-pollinate ideas between units but also to embed data librarianship more firmly within established disciplinary librarian roles.

For institutions that may not have dedicated data librarians and are interested in reallocating staff efforts to library-led RDS, empowering staff requires re-skilling librarians and reshaping services. It is critical that library leadership value the time and investment to transition librarian skill sets to support data science research and then commit to the proactive change management process of doing so. To ensure the success of the transition, leadership needs to dedicate both the time and the resources for librarians to learn these skills. The pace of research at many institutions, along with tenure requirements, often leaves librarians little time to dedicate to growing their skill sets. Accountability should be built in through dedicating a realistic percentage change in a librarian’s position description for re-skilling alongside shared goals and regular check-ins. Critically, there should also be emotional support available from supervisors and administration as re-skilling may feel overwhelming or frustrating at times. Forming communities of practice is another great way to build emotional support and collaborative learning into re-skilling. Library leadership should foster these communities and encourage librarians to prioritize attending them. The authors also want to caution that re-skilling staff requires the aforementioned vision and goals to be clearly disseminated and regularly related to daily work. This is important for helping librarians see their efforts as contributing towards significant progress for their organization and sustaining their motivation.

Finally, while many institutions have invested heavily in data science research institutes and technical infrastructure in the past few years, there continues to be a disconnect between the availability of computational skill training, both at the broad institutional level and within the classroom, and the larger professional expectation that researchers and students have these technical skills. With the rise of interest in big data, machine learning, visualization, artificial intelligence, and other data science methods and tools, it is critical that foundational computational skills and foundational ethical concepts are introduced to students early on in their academic careers. Libraries have the opportunity to assist with the advocacy
CONCLUSION

Librarians are well-positioned to be partners in data science support and have a wealth of skills, both interpersonal and functional, to offer both researchers and campus partners that make them indispensable contributors to a thriving research data support landscape. This case study is a potential learning opportunity for how to develop cross-campus partnerships in data science support. Every institution will have its own challenges and areas for growth. Having been in this space for several years, the authors have described ongoing efforts at UW–Madison, which will likely be applicable for many R1 institutions. Many of the efforts described may also be useful strategies for the smaller libraries where perhaps only one librarian or several librarians with partial percentages in data support can work collaboratively to implement similar services. Being able to collaborate with non-library groups on campus with similar goals and values will help further this effort, spread the workload, and provide different perspectives to support the same mission.

The collaborative efforts between RDS and the Data Science Hub at UW–Madison have enriched the data and data science support for many campus researchers struggling to manage large amounts of data. Human infrastructure and interpersonal skills needed in communication, training, and consultation are integral to meeting researcher needs and require additional investment to meet increasing demand. The authors hope that the described shared experiences can assist others in promoting library skills at their institutions and in building robust collaborations to support the people who need data people.

REFERENCES


KEY TERMS AND DEFINITIONS

**Carpentries**: A global, inclusive community of learners and instructors to cultivate data and computing skills in research.

**Computational Research**: The use and development of software tools for data analysis.

**Data People**: Research staff with technical and interpersonal skills who assist researchers in developing computational research skills, advocate for researcher needs across campus silos, and build broad networks of support throughout the research lifecycle.

**Data Science Hub**: A team of facilitators with a mission of community engagement and learning opportunities around data science and computing for UW–Madison campus researchers through a variety of services.

**Data Repository**: A third-party archive that stewards deposited data, making it accessible and findable in the long term. There are multiple forms of data repositories, including disciplinary repositories,
People Who Need Data People

in institutional repositories, and generalist repositories. This chapter mentions both an institutional repository, MINDS@UW, and a generalist repository, Dryad.

**Grassroots:** An effort built from the ground up, often by people who are invested in the purpose of the effort. Instead of being driven by those with significant power or formal leadership positions in an institution, these efforts are initiated and driven by individuals working in a particular field.

**Holistic Research Infrastructure:** A research infrastructure that leverages both technical resources to enable the day-to-day functions of the research process and human resources to enable the human understanding and use of those technical resources.

**Interpersonal Skills:** A suite of person-facing and empathetic skills, such as communication, instruction, emotional intelligence, leadership, teamwork, and collaboration.

**Research Data Management:** The organization, description, and stewardship of research data throughout the research lifecycle.

**Research Data Services (RDS):** A term typically used to denote a suite of research services focused on research data management and sharing typically provided by an academic library. In this chapter, RDS is both the name of the UW–Madison service and used as a descriptor for similar services broadly.

**Technical Skills:** A suite of technology-facing skills, such as software engineering, data management, system administration, and database administration.