



Researcher Challenges and Experiences with Data Services

Chelsea McCracken
Ruby MacDougall

March 27, 2025



ITHAKA S+R

Ithaka S+R provides research and strategic guidance to help the academic and cultural communities serve the public good and navigate economic, demographic, and technological change.

Ithaka S+R is part of ITHAKA, a nonprofit with a mission to improve access to knowledge and education for people around the world. We believe education is key to the wellbeing of individuals and society, and we work to make it more effective and affordable.

Copyright 2025 ITHAKA. This work is licensed under a Creative Commons Attribution Non-Commercial 4.0 International License. To view a copy of the license, please see <https://creativecommons.org/licenses/by-nc/4.0/>.



Introduction

Research data services—support offerings which enable and improve data-intensive research—have garnered sustained attention from library research support service providers for nearly two decades.¹ Libraries have played a leading role in developing research data services, and on most university campuses they provide the largest and most diverse services.²

¹ Examples of relevant scholarship include: Elise Gowen, and John Meier, “Research Data Management Services and Strategic Planning in Libraries Today: A Longitudinal Study,” *Journal of Librarianship and Scholarly Communication* 8 (2020), <https://doi.org/10.7710/2162-3309.2336>; Julie Goldman, Jennifer Muilenburg, Andrea N. Schorr, Peace Ossom-Williamson, and C. Jeff Uribe-Lacy, “Trends in Research Data Management and Academic Health Sciences Libraries,” *Medical Reference Services Quarterly* 42, no. 3 (2023): 273-293, <https://www.tandfonline.com/doi/full/10.1080/02763869.2023.2218776>; Stephen Pinfield, Andrew M. Cox, and Jen Smith, “Research Data Management and Libraries: Relationships, Activities, Drivers and Influences,” *PLoS ONE* 9, no. 12 (2014), <https://doi.org/10.1371/journal.pone.0114734>; Bethany Latham, “Research Data Management: Defining Roles, Prioritizing Services, and Enumerating Challenges,” *The Journal of Academic Librarianship* 43 (2017): 263-265; Carol Tenopir, Dane Hughes, Suzie Allard, Mike Frame, Ben Birch, Lynn Baird, Robert 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) <http://dx.doi.org/10.7191/jeslib.2015.1085>.

² Jane Radecki and Rebecca Springer, “Research Data Services in US Higher Education,” *Ithaka S+R*, November 18, 2020, <https://sr.ithaka.org/publications/research-data-services-in-us-higher-education/>; Carol Tenopir, Ben Birch, and Suzie Allard, “Academic Libraries and Research Data Services,” *Association of College & Research Libraries*, June 2012; Nedelina Tchangalova et al., “Research Support Services in STEM Libraries: A Scoping Review,” *Issues in Science and Technology Librarianship*, no. 97 (May 7, 2021), <https://doi.org/10.29173/istl2574>; Sarah Barbrow, Denise Brush, Julie Goldman, “Research Data Management and Services: Resources for Novice Data Librarians,” *College & Research Libraries News* 78, no. 5 (2017); Elise Gowen and John J. Meier, “Research Data Management Services and Strategic Planning in Libraries Today: A Longitudinal Study,” *Journal of Librarianship and Scholarly Communication* 8, no. 1 (2020) <https://www.iastatedigitalpress.com/jlsc/article/id/12855/>; Andrew M. Cox, 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 (March 25, 2017): 2182-2200, <https://doi.org/10.1002/asi.23781>; Rebecca Bryant, “Cross-Campus Collaboration in Research Support: Insights from an RLP Leadership Roundtable,” *Hanging Together: the OCLC Research Blog*, August 13, 2024,

In addition, because of research data services' critical role in supporting research on college and university campuses across the country, they are also a central area of concern for stakeholders such as offices of research, campus IT, research computing, academic departments, and other units involved in the research enterprise.³ As foundations and federal governments in both the US and Canada have increasingly promoted an Open Science agenda—characterized by high expectations for data management and data publication—the need to develop an efficient infrastructure of research data services has become an even more urgent strategic priority.⁴ The introduction of generative AI also has the potential to transform the data services space, both in terms of

<https://hangingtogether.org/cross-campus-collaboration-in-research-support-insights-from-an-rlp-leadership-roundtable/>.

³ Alisa B. Rod, Biru Zhou, and Marc-Etienne Rousseau, "There's No 'I' in Research Data Management: Reshaping RDM Services Toward a Collaborative Multi-Stakeholder Model," *Journal of eScience Librarianship* 12, no. 1 (2023), <https://doi.org/10.7191/jeslib.624>; John Chodacki, Cynthia Hudson-Vitale, Natalie Meyers, Jennifer Muilenburg, Maria Praetzelis, Kacy Redd, Judy Ruttenberg, Katie Steen, Joel Cutcher-Gershenfeld, and Maria Gould, "Implementing Effective Data Practices: Stakeholder Recommendations for Collaborative Research Support," *Association of Research Libraries*, September 2020, <https://doi.org/10.29242/report.effectivedatapactices2020>; Sayeed Choudhury and Esmé Cowles, "Research Data Curation: A Framework for an Institution-Wide Services Approach," Educause Working Group Paper, May 2018; Rebecca Bryant, Annette Dortmund, and Brian Lavoie, "Social Interoperability in Research Support: Cross-Campus Partnerships and the University Research Enterprise," OCLC, October 11 2021, <https://www.oclc.org/research/publications/2020/oclcresearch-social-interoperability-research-support.html>; Shawna Taylor et al., "Public Access Data Management and Sharing Activities for Academic Administration and Researchers," *Association of Research Libraries*, November 22, 2022, <https://doi.org/10.29242/report.rads2022>; Jane Fry, James Doiron, Danny Létourneau, Laure Perrier, Carol Perry, Wendy Watkins, "Research Data Management Training Landscape in Canada: A White Paper," *The University of British Columbia*, 2017, <https://open.library.ubc.ca/soa/ciRcle/collections/ubccommunityandpartnerspublicati/52387/items/1.0372048>; Felicity Tayler and Maziar Jafary, "Shifting Horizons: A Literature Review of Research Data Management Train-the-Trainer Models for Library and Campus-Wide Research Support Staff in Canadian Institutions," *Evidence Based Library and Information Practice* 16, no. 1 (March 15, 2021): 78–90, <https://doi.org/10.18438/ebliip29814>.

⁴ John A. Borghi and Ana E. Van Gulick, "Promoting Open Science through Research Data Management," *arXiv preprint*, 2021, <https://arxiv.org/abs/2110.00888>.

increased data needs and increased automation of service provision.⁵

While there is general consensus that institutions should provide a coordinated research data services support infrastructure to their researchers, determining the most effective way to do this has proven more difficult, especially in light of the fast-paced technological changes that have precipitated new forms of research collaborations, methodologies, and discoveries. The development of data services has thus been largely *ad hoc*, lacking cohesive cross-campus collaborations or strategic frameworks. As a result, while many universities have made substantial investments in research data services and are likely to continue to make further investments, obstacles such as decentralization and inefficiency, insufficient staffing, lack of technical expertise, and ambiguity about the needs of researchers continue to limit the impact of these investments.⁶

While there is general consensus that institutions should provide a coordinated research data services support infrastructure to their researchers, determining the most effective way to do this has proven more difficult, especially in light of the fast-paced technological changes that have precipitated new forms of research collaborations, methodologies, and discoveries.

In order to design coordinated services to best meet the data service needs of researchers, it is necessary to understand researchers' existing data management practices and perspectives on data services. Studies of researcher perspectives indicate that researchers have not yet fully integrated best practices in data management into their workflows—following best practices often clashes with tight research timelines, especially when researchers lack adequate training. Furthermore,

⁵ Examples of AI applications for RDM include: OpenRefine, <https://openrefine.org/>; DataSynthesizer, <https://pypi.org/project/DataSynthesizer/>; and CEDAR.

⁶ Rebecca Bryant, Brian Lavoie, and Constance Malpas, "A Tour of the Research Data Management (RDM) Service Space: The Realities of Research Data Management, Part 1," *OCLC Research*, 2017, <https://doi.org/10.25333/C3PG8J>.

incentives for researchers to follow best practices are notoriously sparse, since researchers' data management practices are rarely assessed when they are evaluated for hiring, review, promotion, and tenure.⁷ Indeed, research data is rarely considered to be a primary research output and consequently traditional researcher assessment metrics may fail to recognize the time, labor, and expertise necessary to properly manage research data.⁸ Researchers who do spend time on data management are also likely to do so alone, without seeking out data services, which tend to be underutilized.⁹

In light of these persistent challenges, and in the interest of providing up-to-date data to inform university decision making, Ithaka S+R has collaborated for the past two years with 29 US and Canadian institutions to develop strategies for improving the coordination of research data support services offered across different campus offices. The project included the collection of institution-level data about the range and location of current data services across campus, design workshops geared toward creating new or strengthening existing data service infrastructure, and cohort-wide and individualized meetings to discuss topics and strategies for delivering effective research data services. In March 2024, Ithaka S+R published findings from a landscape survey of data services offered at colleges and universities in Canada and the US.¹⁰

⁷ Gaia Mosconi, Aparecido Fabiano Pinatti de Carvalho, Hussain Abid Syed, et al, "Fostering Research Data Management in Collaborative Research Contexts: Lessons Learnt from an 'Embedded' Evaluation of 'Data Story,'" *Computer Supported Cooperative Work* 32 (2023): 911–949 <https://doi.org/10.1007/s10606-023-09467-6>; Olivia Aguiar, "Rethinking Research Assessment for the Greater Good: Findings from the RPT Project," Scholarly Communications Lab, May 4, 2022, <https://www.scholcommlab.ca/2022/05/04/findings-from-the-rpt-project/>.

⁸ Iratxe Puebla and John Chodacki, "Make Data Count: Driving Metrics for the Meaningful Evaluation of Data," Zenodo, December 2, 2024. <https://zenodo.org/records/14261211>.

⁹ Jonathan Petters, Shawna Taylor, Alicia Hofelich Mohr, Jake Carlson, Lizhao Ge, Joel Herndon, Wendy Kozlowski, Jennifer Moore, and Cynthia Hudson Vitale, "Publicly Shared Data: A Gap Analysis of Researcher Actions and Institutional Support throughout the Data Life Cycle," *Association of Research Libraries*, March 2024, <https://doi.org/10.29242/report.radsgapanalysis2024>.

¹⁰ Ruby MacDougall and Dylan Ruediger, "The Research Data Services Landscape at US and Canadian Higher Education Institutions," *Ithaka S+R*, March 14, 2024, <https://doi.org/10.18665/sr.320420>; Ruby MacDougall, "Building Campus Strategies for Data Support Services Project Kicks Off," *Ithaka S+R*, February 2, 2023,

Now, this report shares additional findings from interviews with researchers exploring their practices and experiences using research data services, conducted by our university partners. The interviews provided a wealth of information about the data management needs and challenges of researchers as well as their current level of engagement with campus data services, guided by the following research questions:

- What needs and challenges do researchers encounter when creating or locating data for research? Managing research data? Analyzing or modeling data? Sharing data? Learning new data skills?
- What campus resources have researchers used to support their data services needs, and where are those resources housed?
- What was their experience using those resources? What could be improved about the process?

We are deeply grateful for the members of the cohort that made this report possible. See Appendix A for a full list of cohort participants.

<https://sr.ithaka.org/blog/building-campus-strategies-for-data-support-services-project-kicks-off/>.

Executive summary of findings

- The term “data services” has no fixed meaning, or has a contested meaning, among researchers and data service providers, and researchers are often unaware of data services as a form of assistance. Even when they use data services, researchers don't always know that they are doing so and are thus likely to underestimate their use of them.
- Researchers point to the need for considerable changes and support throughout the research enterprise to make widespread data publication and reuse possible.
 - Researchers perceive data publication as a significant new responsibility, and most are not well prepared for it, nor do they conceive of it as part of their scope of work.
 - Researchers also experience difficulties discovering and reusing published datasets.
- Researchers consider securing permissions to collect or reuse data to be a major administrative challenge and perceive the IRB/ethics office and legal services to be integral to this work.
- The cloud-based systems researchers use to share and transfer data between team members are often not sufficient for research collaborations that span different institutions and community partners.
- Researchers prefer data services that are 1) individualized, localized, or bespoke; 2) contain a strong instructional component; and/or 3) have memorable branding.

Methodology

Ithaka S+R trained cohort members to conduct interviews with researchers on their campus using a semi-structured interview guide that included questions on research focus, collaboration, data management, modeling, and sharing. The guide also included questions on how researchers currently use training and data support services, what additional services they need, and how they anticipate these needs will change in the future. Interview recordings were transcribed and de-identified in accordance with each institution's policies. Of the 29 participating cohort institutions, 27 submitted interview transcriptions, providing a total of 294 interview transcripts with metadata on each interviewee's disciplinary area and rank.¹¹ The median number of interviews submitted by each participating institution was 11. Ithaka S+R staff sorted interviewees' disciplinary areas and ranks into categories, making educated guesses in cases where metadata was incomplete. A stratified sample of 41 interviews was then selected for analysis.

Two Ithaka S+R analysts used a non-overlapping sample of five transcripts to develop qualitative codes in NVivo and check inter-analyst agreement. Qualitative thematic coding and analysis of the full sample was completed by a single analyst. We would like to thank the members of the institutional teams participating in our Data Services Assessment cohort project who collected the data for this report. We would also like to extend our gratitude to Lynda Kellam and Nina Exner for their insightful feedback on a draft of this report. Any errors or omissions remain our own.¹²

¹¹ See Appendix A for a list of institutions and cohort participants, Appendix B for demographics of the institutions; Appendix C for demographics of the interviewees; and Appendix D for the interview instrument.

¹² In addition to the typical limitations of qualitative research (e.g., findings that are directional rather than representative; interpretive bias), this study had the following limitations: 1) Missing context: We did not collect metadata on interviewers, and some metadata on interviewees was missing or incomplete. Automated transcriptions often contained transcription errors that could not be corrected. As a result of these limitations, it was sometimes not possible to determine exactly what data service participants were referring to in their interviews. 2) Definitional issues: As "data services" does not have a clear definition and neither interviewers, interviewees, nor the discussion guide clearly differentiated between "data services" and other types of

Researcher challenges

The researchers in this sample are engaged in cutting-edge, interdisciplinary work that advances the broader research enterprise. As their research often transcends disciplinary boundaries, the challenges they face—and the data services they need—likewise span multiple disciplines and are not confined to specific fields or even topics. For example, a European historian contracts with a drone technician to collect aerial photographs; a business researcher distributes electronic tags to hospital workers to track their movements; and an engineer uses data from a real estate listing company to determine the risk of flooding. The rich variety of research topics and methodologies across disciplines suggests that, where possible, data support services should be available to all researchers in an institution.

Different types of challenges tend to reinforce and overlap with one another. For example, a researcher might have difficulty accessing their own data (a logistical challenge) due to having to contract out data queries for security reasons (an administrative challenge) but might also struggle to write the code to do so (a technical challenge). Researchers at smaller schools were more likely to mention a lack of funding and other resources when they described research challenges. Integration of widely available data services would allow researchers to address interrelated and overlapping challenges simultaneously and holistically.

Researchers in this sample identified the following challenges.

- **Contested concept of “data services”:** Researchers were unfamiliar with the concept of data services and might not know where to turn for help when their established problem-solving techniques fail to resolve the issue.

support services provided by the institution, we have used the broadest possible definition of “data services” in this report. This may be a limitation in cases where a narrower definition is warranted. 3) Cross-tabulation: Throughout the report, we have noted where findings apply more to one subgroup of researchers than another. In general, however, the small and stratified sample prevented findings specific to individual Carnegie classifications, disciplines, and ranks.

- **Data custody:** Researchers struggle to store and access their own data while working on their projects, and to share that data with the members of their research team.
- **Data size:** Researchers experience difficulty when the size of data in their project exceeds the capacity of the available tools (e.g., storage, computation, etc.).
- **Technology and dataset access:** Researchers often struggle to discover and access pre-existing datasets, learn new technical skills, and access technology.
- **Managing research:** Although researchers acknowledge that administrative tasks such as project management and people management are their responsibility, many struggle to balance the administrative burden with their other duties.
- **Data publication:** Researchers express that data publication falls outside their expected job duties, and they struggle as a result.
- **Affective barriers:** Researchers may hesitate to ask for help when they need it due to the culture of academia.

Problems that data service providers commonly cite, such as “data curation” and “data management,” did not emerge as relevant categories of challenges for researchers, who may not conceptualize their experiences in this way. Research challenges that data service providers might include in these categories are instead discussed throughout the report.

Contested concept of “data services”

About three-quarters of researchers in our sample face at least one research challenge for which they do not see their institution as a potential solution. Instead, they turn to their scholarly communities—academic conferences, scholarly literature, colleagues, and online resources—for support and assistance when these challenges arise. While researchers may expect their institution to provide resources or infrastructure, they often fail to connect this provision with the idea of a “service” intended specifically to support them or to be responsive to their needs. Their mental model of what “data services” means is, essentially, undefined.

Around a quarter of researchers in the sample were unaware of the data services available at their institution—including researchers like a social

scientist working with census data who was unfamiliar with the very *idea* of data services. In fact, the interviews conducted for this project seemed to spark a paradigm shift for some researchers, surprising them with the realization that such support exists. One Alzheimer’s researcher remarked, “Let me just say that this interview is making me wonder whether I haven’t turned to [institution] enough for help, and whether I could turn to [institution] more for help.” Other researchers struggled to grasp the concept of data services as a category of support offered by multiple campus units—they might have been aware of one service provider (as, for example, one physical scientist was aware that IT provides services), but didn’t recognize that other services are available across the research life cycle or that these services can be connected under the umbrella of “data services.”

Because data services are not embedded into their typical workflow, researchers may follow a circuitous route to find the help they need. One biomedical engineer emailed a number of people before being directed to data service providers, but once they were able to connect, was “very impressed and actually surprised at the level of ignorance on my part, at how far things had moved forward” in terms of the services available.

The lack of familiarity with the concept of data services is understandable, as such support is not always available at every institution—one historian, for instance, previously worked at an institution with “zero” data support—and researchers may have become accustomed to a culture of self-reliance, so much so that they might not notice when help is available. As one education researcher noted: “There could be so much more out there that the campus does that I just am not aware of... Anything outside of your bubble you just ignore.” Just as studies on students’ use of student services show “gaps across availability, awareness, and utilization of supports,” and suggest that “explicit and intentional outreach” is “crucial to ensuring awareness,” researchers—once students themselves—are susceptible to the same awareness gaps.¹³

Although we know data service providers have made valiant efforts to communicate their services, the message that data service providers are available specifically to support data-intensive research (often at no cost to the researcher and often tailored to a specific project) has not yet

¹³ “2024 Listening to Learners,” Tyton Partners, <https://tytonpartners.com/listening-to-learners-2024/>.

effectively reached many researchers.¹⁴ One political scientist urged data service providers to use their existing advertising methods to explain the concept of data services and emphasize how it differs from researchers' expectations, especially "that you guys can help [with] basically anything and that [there's] no strings attached." As this researcher noted, alongside promoting the data services a school offers, it may be beneficial to raise awareness of the idea of data services itself as a type of support for scholars.

Meanwhile, researchers in the sample who were aware of data services were concerned that existing data services do not meet their needs. One major complaint about existing data services is that they are too slow. As one humanist noted, researchers usually need to find an answer to their problem quickly to keep project timelines on track and do not want to be delayed by the formal process of signing up and waiting for a consultation. Researchers have limited time and do not feel able to spend it on lengthy training processes, as this education researcher remarked: "If [workshops are] like three or four days of your time, [who] can do that[?] It's just so impossible which is disappointing."

Researchers were also concerned that existing data services do not offer the highly specialized expertise needed for their specific and often niche topics. At times, this concern stemmed from losing a trusted service provider due to staff turnover. As a health researcher described, "In my past job, we [had] IT personnel, somebody that actually takes care [of] updating libraries and versions... And they had experience to solve [problems] because they specialize in the [model] we use." But now, "[it] can be difficult to find the right person that has the expertise about the work."

If researchers lack a concept for data services, they may fall back on a general-purpose archetype of the service industry outside academia (e.g., food service, hospitality, trades) to contextualize their interactions with data service providers. Typically in the service industry, patrons pay more for speed and expertise, so researchers may be bringing these expectations to their interactions with data service providers. Educating

¹⁴ For examples of efforts data service providers have made see: Jessica Atkins, Kelsey Badger, Claire Jordan, Hannah G. Nelsen, Katerina Ozment, and Olivia Young, "Translating Liaison Librarians to the Scientific Community," *Journal of eScience Librarianship* 11, no. 1 (2022): e1229, <https://doi.org/10.7191/jeslib.2022.1229>.

researchers about the norms and behaviors of data service providers at their institution (just as faculty use a syllabus to educate their students about the expectations of their class) may help manage researchers' expectations and increase their willingness to engage services.¹⁵

At least a quarter of researchers in the sample explicitly recommended that institutional service providers change or increase the way they promote their services, often suggesting improvements to the services landscape or to marketing and communications. One genomics researcher recommended incorporating information about data services into orientation material for all new university members from undergraduates to graduates, PIs, and staff. In their view, “that's always the best time to capture folks, [when] they're walking [in] the door, they're figuring out all the systems from the start.” Providing this information early can be particularly helpful for researchers who—like one business researcher—are new to an institution, or who—like one ecologist—started during the pandemic and missed important trainings.

Many data service providers are already promoting their services in the ways that researchers recommend. For example, several researchers asked to be subscribed to “a mailing list or some way of being kept up to date on new information or new events,” as one health researcher put it. Several noted the need for a centralized system to connect to data services infrastructure, including a social scientist who requested a menu of services, a humanist who requested a directory of other researchers, and an engineer who requested a set of static institution-branded “explainer videos” combined with “someone who actually [picks up] the phone or the email when you're in trouble and [says], ‘How do you deal with X?’” These researchers' ideas for improvements match closely with the ideas data service providers already have (as demonstrated in Ithaka S+R-led workshops this spring), suggesting that providers and researchers are mostly in agreement on the types of marketing that appeal to researchers.¹⁶

¹⁵ Ashley Mowreader, “Teaching Tip: A More Strategic Syllabus Day,” *Inside Higher Ed*, June 26, 2024, <https://www.insidehighered.com/news/student-success/academic-life/2024/06/26/how-professors-can-make-first-day-course-engaging>.

¹⁶ Chelsea McCracken and Ruby MacDougall, “Draw New Directions for Research Data Services,” *Ithaka S+R*, May 14, 2024, <https://sr.ithaka.org/blog/drawing-new-directions-for-research-data-services/>.

Interestingly, these researchers' suggestions illuminate the disconnect between researchers' desire for services (which in many cases already exist) and their knowledge of existing services—researchers need to know that a service exists at the moment they need it, but they may not register the existence of a service before that moment.

Data custody

Almost three-quarters of the sample of researchers reported that they, or a member of their research team, have difficulties accessing data for their own projects. The main challenges in data custody involve problems transferring data to current and changing team members, inconsistent and insufficient data storage options, and, as one engineer noted, reconciling data security with data accessibility. These have significant effects on researchers' ability to work collaboratively, an issue with far-reaching impact as the vast majority of researchers in our sample work in collaboration with others.¹⁷ About half the sample collaborate with faculty at their own institution; about half collaborate with faculty at other institutions; about half work with community partners outside the academy; and about three-quarters work with student or post-doc researchers.

Data infrastructures have not fully adapted to the complex nature of contemporary research, which is often interdisciplinary, cross-institutional, and transnational, and involves collaboration with individuals with varying levels of access to data. For example, one geneticist complained that permitting access to IRB-secured data to off-campus collaborators is difficult and expressed frustration at the effort required to establish accounts on their system for team members working at other institutions. A nursing researcher related how they needed to go through the file system and remove a student's access to sensitive files when the student relocated from one country to another. If team members have access to large numbers of files, managing shifting permissions can grow cumbersome.

¹⁷ Only one researcher in the sample does not mention any collaborators.

Several researchers reported that file sharing systems at different institutions are incompatible, leading many to use Google Drive despite concerns about security and some university policies prohibiting its use. One anthropologist found it “tricky” to coordinate between members of their research group whose institutions use Dropbox, Box, and Google Drive: “Google Drive, we use by default, often because it is the most accessible, but it's actually the worst for data management and data delivery.” Many are finding creative workarounds, or as one engineer put it, “macgyvering their own solutions” for the problems created by file sharing system incompatibility. These include editing the registry of their computer to enable access to Google Drive, as one social scientist notes, and physically transporting data to their collaborators: “[It's] a mess [and] it's a problem,” one engineer said, acknowledging that meeting government representatives in a parking lot to exchange hard drives appeared “kind of shifty,” but was nonetheless the most suitable solution they could identify. As Cory Doctorow argues, interoperability standards and transferability between platforms would prevent many of researchers’ data custody concerns.¹⁸

Limited access to secure data creates extra steps for research teams, especially when teams access data remotely, ultimately slowing down the analysis. A business researcher related how a government agency required them to ensure data security by outsourcing data management to a third-party private company, which would only allow them to access the data through a virtual “clean” remote access environment. “We cannot take the data outside there, we can only [what they call ‘egress’] reports on the data,” meaning the researcher’s team cannot transfer raw data out of the secure environment, only aggregated data. The researcher confessed that this security process prevents them from directly interacting with the data or observing their post-doc’s interaction with it: “We agree on a report design and my post-doctoral fellow will process and [egress] the [report] for me to see.” While some institutions build infrastructure to support secure data, the data provider—in this case, a government agency—has the prerogative to make their own policies with regard to access. If the data supplier’s policies restrict institutional access, there is little that data service providers can do to help.

¹⁸ Michael Nolan, “Cory Doctorow: Interoperability Can Save the Open Web,” *IEEE Spectrum*, September 5, 2023, <https://spectrum.ieee.org/doctorow-interoperability>.

Version control is another major data custody challenge for researchers. For those working in teams, it is difficult to monitor shared data used by multiple team members—especially when some are at different institutions. Researchers struggle to ensure that everyone has continued access to the same working file “without you having to send around a new Excel file every time” as an ecologist described, and to keep track of what has been changed with each version. For individual researchers, version control can become an issue when institutional cloud storage systems fail to sync local versions of a file: “One day [OneDrive] stopped saving [and some of my files] just disappeared. A colleague down the hall lost four months of work, and [my institution] was unable to retrieve it... Total nightmare. [So] I'm not willing to trust my research data [to] desktop support.”

Researchers reported a variety of other concerns with data storage systems. It is often difficult to keep track of data migration from one platform to another—for example, a public policy researcher “quite frankly [lost] track” of their institutional data backups after a transition to another platform. Monitoring when data will expire is also a challenge: a researcher at a medical center where data is deleted at the five-year mark said that they are trying to “put some processes in place with research computing [to] send automated email alerts to users” to let them know their data is expiring. There are cost barriers to securely storing data; one researcher who needed to purchase a HIPAA-compliant storage environment believes “there's not really a very cost-effective way to [do] data management with HIPAA data.” And changes in data storage policies over time create ongoing challenges. As one humanist noted, “I'm not happy that we are a Microsoft-only institution; every time there's some new restriction in place, it seems to affect some part of my research.”

“The most pressing need we have is some sort of simple user-friendly, cloud-based, secure data storage and sharing platform.”

While cloud-based storage has clear benefits for collaboration, the lack of international and cross-institutional standards and policies for interoperability, security, and accessibility make it difficult for researchers to fully reap these benefits. One psychologist described going “back and forth with the IRB [over] the years” to ask for “a good secure cloud-based data storage and sharing platform.” The psychologist submitted a letter to

support the IRB's advocacy with administrators and eventually was able to secure access to REDCap for one offsite student. This psychologist succinctly stated what many researchers across disciplines express: "the most pressing need we have is some sort of simple user-friendly, cloud-based, secure data storage and sharing platform." Of course, such a solution would also have to be acceptable to data providers, such as government agencies.

Data size

About half of researchers in the sample reported challenges related to the size of data in their project.¹⁹ Researchers whose data consists of images, video, simulations, or models tend to experience this challenge more frequently—meaning that, though data size problems are most common for STEM researchers, researchers in the arts, humanities, and social sciences struggle with the size of their data as well. Handling oversized data requires distinct data management practices which researchers may or may not be equipped for. Researchers in this sample reported challenges at every step in the research process—including funding, data storage, security, sharing, cleaning, analysis, archiving, etc.—as they struggle to adapt to the new procedures that are necessary to handle big data. As the examples in this section demonstrate, researchers with data size challenges tend to feel alone due to the institutional expectation that they will learn how to handle big data without sufficient institutional support.

Several researchers in the sample reported solving data size challenges without institutional support, including by storing data off-campus or purchasing their own clusters with grant funding. One geoscientist described a departmental culture of "each PI is supposed to figure it out on their own" by designing and managing their own system for handling big data, but this means researchers who are less equipped to do so are at a disadvantage. "The downsides are I'm not an expert in maintaining and storing large data sets," the geoscientist elaborated. "[I] really do

¹⁹ Ithaka S+R's project on big data provided a definition of big data and investigated big data research practices and support needs. See See Dylan Ruediger et al., "Big Data Infrastructure at the Crossroads: Support Needs and Challenges for Universities," Ithaka S+R, December 1, 2021, <https://doi.org/10.18665/sr.316121>.

worry about, you know, is the disk going to fail on this thing[?] Those are the things that keep me up at night.”

Researchers who have sought institutional assistance with oversized data are not always successful. A computational linguist sought help to use a cluster, but was only given a “document that [walks] you through how to set up the basic working environment on a cluster and then you need to take care of your own things,” with the implication that this was the only level of help that was available. A cell biologist reported that their allocated cluster storage is limited to two terabytes, which is not enough to manage their working data—they accumulate terabytes of data within weeks. Their colleagues regularly request special allocation to store a larger amount of working data just for a short period of time since it is “cost-prohibitive” to purchase more space on the cluster.

Researchers also grapple with moral quandaries when handling big data. Reluctant to use private company storage options like AWS for projects funded by taxpayer money, but without many alternative options, they may find themselves relying on options that don’t align with their personal values. Indeed, there are few, if any non-profit options, that can sustainably meet researchers’ needs at scale. “I would like to be a thoughtful steward of public money,” one biostatistician explained. “We’re funded by [federal agency], and the thought of funneling taxpayer dollars to AWS doesn’t feel good to me.” In this researcher’s “data analytic utopia” dreams, a nonprofit “that isn’t pegged to major tech companies” would offer scalable, reasonably priced computing and use any income it generated for long-term public benefit: “I want to have a system that works, that enables me to do my research, and then never have to think about it again.”

Data size can be a problem for researchers even when their data is not “big.” Several researchers in the sample noted challenges with “medium” data—data that is too big for traditional research procedures, but nonetheless inappropriate for “big data” research procedures. One researcher who works on English literature related prior failed attempts to approach a national research computing provider for help with data that was too big to work with on a personal computer: “And they just kind of laugh [at] this little piddly tiny bit of data, it’s [too] big for my laptop, too small for them.” Researchers like this who deal in “medium” data “just seem to fall between the cracks.” Campus data service providers should

be aware of this resource gap and the need for assistance in solving problems with data of all sizes.

Technology and dataset access

Researchers frequently need specific technology to find, store, collect, access, analyze, or share their data. About three-quarters of the sample use a highly specialized instrument or software to generate or analyze data, and about three-quarters use pre-existing datasets with unique structures and access procedures. Indeed, nearly all researchers in the sample currently use either specialized instruments, software, or pre-existing datasets. However, a majority of these (about three quarters) experience challenges accessing technology or datasets, typically because they lack some technical skill or technology resource.

About half of researchers in the sample have experienced challenges related to discovering, accessing, and processing pre-existing datasets. Dataset discovery is a major challenge that researchers are not trained for, and it can be “a bit of a slog” to review the catalog of available datasets in a repository, as one social scientist noted. Discovery is difficult because, as one biologist described, the platforms where the datasets are indexed require “expert knowledge to navigate.” One civil engineer worried that they may not have been able to discover all the available datasets on their research topic and wishes they had talked to a librarian to make sure they had “got it all.”

Several researchers mentioned the need for a directory of which researchers on campus have access to which datasets, which would make discovery easier and also, as one business researcher noted, allow researchers to maximize the use of resources. One oceanographer urged librarians to focus their collections on datasets that can be used by multiple researchers on campus, since having each researcher download it on their own is “a waste of resources.” One economist had to discontinue a “strand of research” using a dataset at a prior institution because the dataset was not available at their new institution. Researchers expect their libraries to assist them with dataset discovery and coordinate access to pre-existing datasets across the institution and likely have little awareness of the limitations data providers place on usage and access.

Once researchers have discovered a dataset, they often experience challenges accessing it. Dataset access issues are most common with health data, which needs to be stored in a protected environment (as a biologist noted) and often must be heavily processed before they can be used. One clinical researcher remarked that their “biggest need” is software that will process health data into a usable format. When an institution does not have an existing secure environment to store health data, this can create barriers to dataset access. One nursing researcher pointed to examples of two peer institutions that ensure “seamless access” to secured health datasets, wishing their own institution would offer such a “clear and efficient” process. Researchers in this sample demonstrated a clear need for an investment of collection resources in the discovery and access of pre-existing datasets.

Researchers also described challenges learning to use specialized software and equipment. It can be difficult for researchers to keep up with the rapid pace of technological change—one humanist complained that “We still don't have good [professional development] training embedded in universities”—and many find themselves needing to use new technology that was not covered by their training. An engineer mentioned needing help to onboard and use their high-performance computing (HPC) cluster; a policy researcher needed help learning analysis software; an architect needed help identifying the right equipment to generate data; and a psychologist needed help writing or translating code.

Even when researchers know how to use technology, they may have difficulty accessing it. An English researcher found it “frustrating” that software they need to process data is “blocked” by their institution, so they have used workarounds to install it. The scope of technology challenges suggests a need for an inter-institutional network for exchanging expertise on specialized technology and troubleshooting technology access, so that not every institution needs to maintain specialists on every technical subject.

Managing research

More than half the researchers in our sample reported that the administrative burden of running a large project is a major challenge, including training and managing junior collaborators; managing budgets;

and managing contracts and regulatory and reporting requirements. Researchers who are tenure-track are most likely to report these challenges, though tenured professors also report a high volume of administrative challenges.²⁰ Researchers recognize managerial duties as part of their job but often express concern about whether these duties make the best use of their time. “It’s hard to be a researcher, and it’s not getting easier. [A]dministrative requirements [take] up bandwidth [which] prevents you from just spending time, thinking... I’m not saying that [these tasks are] not worth doing. But [it’s] another burden on the researchers.” This asthma researcher believed “the administrative burden... will continue to rise” and called for institutions to invest in faculty productivity by “unloading them of these sorts of administrative tasks.”

Producing a research output begins by identifying a potential source of data and securing permissions to collect or use it. Then researchers collect or access the data, prepare and clean it, and analyze it. Managing this “data pipeline,” as described by one engineer, is perhaps the biggest administrative challenge for researchers, who find it “overwhelming,” as another engineer put it, especially considering the increasing amount of data flowing along the data pipeline. “I feel like a lot of the time my head is just exploding,” the second engineer said. A biologist agreed, noting that “student projects are getting increasingly data dense” with more need for management and administrative support.

Managing the various aspects of the data pipeline is time consuming, and researchers cited time constraints as one of their biggest challenges. A health scientist who operates an animal lab noted that time is “the biggest hurdle for investigators,” who have to balance data gathering with teaching and the administrative tasks that make it possible to gather data, such as ordering materials, calibrating instruments, and keeping the lab clean. Researchers who work with pre-existing datasets also experience

²⁰ Existing literature on the topic suggests that most academics receive little preparation for taking on leadership roles and that administrative tasks negatively impact job satisfaction in academia. For more see: Tracy L. Morris and Joseph S. Laipple, “How Prepared Are Academic Administrators? Leadership and Job Satisfaction within US Research Universities,” *Journal of Higher Education Policy and Management*, 37, no. 2 (2015) 241-251, <http://dx.doi.org/10.1080/1360080X.2015.1019125>; Joan V. Gallos and Lee G. Bolman, *Reframing Academic Leadership*, 2nd ed., (San Francisco: John Wiley & Sons, Jossey-Bass, 2021).

time constraints, like a research staff member who waited eight months to gain access to data from a state agency. Even once raw data is obtained through data collection or a pre-existing dataset, data cleaning and preparation is “time-intensive,” according to a psychologist. Learning new data skills also takes time.

Some researchers enlist students for help, directing research while students do most of the “on-ground” work, as an arts researcher noted. A biostatistician who relies on a former student to manage their data storage platform recognized that this is “a totally unsatisfying, unscalable solution.” Other researchers who rely on volunteer labor from colleagues have found creative but imperfect ways to complete their work. For example, because their funding agency will not fund data analysis, a historian relied on their co-PI on soft money to “do an awful lot of analysis of the data on a voluntary basis which I’m not very comfortable with.” Researchers who have students or colleagues fulfill data pipeline tasks also struggle with training those collaborators in “data analysis techniques,” according to one biologist. In some sense researchers are caught between taking time to do a task themselves and taking time to manage someone else to do the task, and neither option is a good solution. Demands on researchers’ time can only be expected to increase.

Several researchers stated that adding a data manager to their team is their preferred solution to addressing time constraints caused by managing the data pipeline. A health researcher explained that having the institution supply a dedicated data manager would increase efficiency so that other team members could focus on their own jobs and projects wouldn’t fall behind. An ecologist argued that the institution should provide them with a data manager so that they don’t have to try to support constantly increasing data needs out of their finite grant. Meanwhile, a health economist didn’t expect the institution to solve this problem but planned to write three or four “dedicated data people” into their next grant to ensure they would have enough support required to deal with increased data needs. As John Wilbanks pointed out in a September 2024 panel, the “hard questions” about data management are “who pays, who decides, and who does the work?”²¹ Researchers in well-funded fields may have

²¹ “Access to Science and Scholarship Workshop: Research Data Access, Curation, and Storage Panel,” hosted at the Association for the Advancement of Science headquarters in Washington, DC, on September 20, 2024. The workshop was conceived and

the opportunity to hire a dedicated data manager, but this option is not always available in other fields, leaving researchers more reliant on their institution for help.

However, challenges persisted even among the handful of researchers who did have access to data managers. For example, data managers at a school research center declined to assist one policy researcher with de-duplicating merged datasets because they “decided that that is in fact [an] excessive amount of work.” If data managers themselves are understaffed, they may not be able to assist researchers in a timely manner, as in the case of one nursing researcher: “I know one person who's been dealing with trying to work with [institutional service core] staff to hone in on the variables of interest [for] a year” because “there’s so few” staff and they have to work closely with the researcher to determine which data fields contain the information they need.

Document preparation and relationship management—the “permitting process” part of the “data pipeline” that must take place before data is collected or used, as one biologist noted—is another major administrative challenge. Negotiating with partners (e.g., agencies, companies, organizations, etc.) for permission to collect data or access existing datasets requires—in addition to contract specialists—people skills and time to establish trust. One geneticist working with NIH data explained that it took “three years [to] get the lawyers to even agree to [the] data use agreement.” Gaining trust and permission to collect or use data—a process a corrections researcher called “challenging and laborious and slow”—often requires familiarity with field-specific language, knowledge of community-specific norms, and understanding and respect of cultural practices. These tend to come with experience so, as one business researcher noted, they must be handled by the researcher rather than by students. Nor can researchers guarantee that their relationship-building efforts will be successful. For example, a health researcher worked with an Indigenous community for over a year to gain permission to administer a survey, but in the end not everyone in the community agreed to participate.

sponsored by the MIT Press and supported by funding from the National Science Foundation. Video available at: <https://www.youtube.com/watch?v=COZdCqN63Wg>.

While researchers may consume data services, more than a third in this sample also provide data services, both formally and informally. These researchers often consider themselves to be the go-to people that colleagues ask for help, as one information scientist noted. The group includes researchers developing new methodologies (like an engineer at “the forefront of analysis”) and scholars who teach research methods, where they cover some of the same content offered by institutional data service providers, such as coding languages. One researcher in the sample is the director of a research center, and a few others are or have been journal editors. One of these journal editors, a statistical methodologist, provides other researchers with data-related methodological advice. Institutional data service providers should consider how researchers who are also data service providers fit into the data services landscape, and whether there are any particular data services that should be tailored to them (e.g., resources on teaching research methods, editing a journal, etc.).

Data publication

About two-thirds of researchers in the sample reported challenges with data publication. Most researchers in this sample support the philosophy of open research, yet the skills needed for effective data publication—organizing data so that others can navigate it, selecting which data to share, and preparing it for deposit in a repository; ensuring discoverability with metadata; maximizing data longevity and accessibility; fulfilling data sharing requests—are skills associated more with academic librarianship than academic research. Many researchers thus interpret increased requirements for data sharing in federally funded research as a new job responsibility for which they are not trained.²² They see data publication as valuable, but don’t see it as part of their job, and are reluctant to add it to their list of responsibilities: “we’re asking the researchers to spend a lot

²² “Memorandum for the Heads of Executive Departments and Agencies,” Office of Science and Technology Policy, August 25, 2022, <https://web.archive.org/web/20250102041624/https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>; Jocelyn Kaiser and Jeffrey Brainard, “Ready, Set, Share!” *Science*, January 25, 2023, <https://www.science.org/content/article/ready-set-share-researchers-brace-new-data-sharing-rules>.

more time, so that other people can do research,” one health researcher noted. “It’s a good goal, but it also is going to make people miserable enough that they’re going to leave science.”

While researchers use a variety of affective language throughout the interviews, they tend to use more shame- and fear-based language to describe their challenges around data publication than for other types of challenges. Examples of language related to **shame** include: “It’s **embarrassing** to say, but I don’t have a data plan;” “the records management librarian would be **horrified**.” Examples of language related to **fear** include: “I’m getting a little bit **scared** [about data preservation];” “data transfer is the thing that [I] would be **concerned** [about with data sharing];” “[data backup is] a **worry**.” Researchers’ shame- and fear-based language is consistent with the various data publication challenges they describe, many of which were previously identified in Ithaka S+R’s report on big data infrastructure.²³ These will be discussed in the rest of this section and include lack of preparation, lack of prioritization, and lack of clarity on data sharing requirements.

Researchers in this sample described a pervasive lack of preparation or training for data publication at both a theoretical and practical level. At the theoretical level, they were often unable to imagine that other researchers might find their data useful other than “for the sole purpose of transparency” (as one engineer noted) and didn’t understand what data should be shared. Researchers do not believe that federal data sharing requirements are clear in this regard, pointing out that the guidelines lack clarity on exactly “what equals data” (according to one health researcher). Furthermore, at a practical level, researchers are trained to streamline the analysis process, but making that process transparent requires an entirely different approach: “I’ve never written code that [would] be consumed by others[, which] really requires some adjustment... It’s a different type of communication,” one physics researcher explained. Certain data formats may be necessary to produce the desired findings but are unacceptable for archival purposes; one biologist noted that “things get serious” when they need to convert data from a proprietary format to an ASCII file for archiving.

²³ Dylan Ruediger et al., “Big Data Infrastructure at the Crossroads,” *Ithaka S+R*, December 1, 2021, <https://doi.org/10.18665/sr.316121>.

Researchers' lack of preparation for data publication often becomes evident suddenly at the end of a project when it is too late to solve problems. For instance, a qualitative researcher did not realize they could archive their interview transcripts when they designed the consent form at the beginning of the project, meaning they would have to "go back and ask my 100-some [interviewees] to sign a new consent form, which I know [is] not gonna happen." Researchers may also find out about data publication resources too late for them to be useful, as one health researcher noted: "I tried to use [the DMPTool but] to be honest, we were getting down to the wire when I found out about it, so I didn't have the time to thoughtfully use it." Given that researchers are expected to perform expert-level data publication that they feel unprepared for, the shame they express makes sense. Solving this problem will require not just a comprehensive investment in training and preparation throughout the entire educational researcher pipeline, but also that researchers make changes in their methodologies that prioritize data publication over other factors like speed and effectiveness.

Researchers believe that those with the power to require increased data publication (e.g., funders, publishers, institutions, etc.) do not acknowledge the trade-off involved in prioritizing data publication above other considerations. A few researchers pointed out that they do not have the time available to prioritize data publication. For example, a researcher working with Indigenous health data designed an application form to handle data sharing requests: "We've [had] easy data requests, [where] we've been able to simply use our existing resources from our grant to process them. But difficult data requests [we're] not able to comply with at this moment" because "we actually don't have the person power to provide whatever random data somebody might want." An environmental scientist noted that complying with data sharing requests could "be a full-time job," and that "[data publication] is a new burden on scientists in this generation that didn't used to exist. [With the] freedom to explore all these data comes the big responsibility to be able to document what you're doing." This researcher described getting researchers to think in terms of data publication as "more of [a] psychological than a technological problem in some ways."

"[With the] freedom to explore all these data comes the big responsibility to be able to document what you're doing."

Reiterating the idea that data publication should not be their responsibility, several researchers stated that data publication should be handled by the institution. "The grants usually are given to [the institution], not to the individual PI. So officially [the institution] should be helping store that data. Not the investigators," one biologist opined. A few researchers also suggested that funders should support this work "as part of [a] budget line item" (according to a biologist), especially to ensure project "longevity" in repositories when universities are unable to "pick up the slack" (as a humanist noted). One engineer recommended that the solution to the problem of responsibility for data publication is to provide professional rewards to researchers who prioritize it: "Even though people say we need to do [data publication], you're not being rewarded for it right now... If you publish your data and you get a DOI and there are citations, [then] it should be included in your h-index. [But] very few people do that." A geoscientist thought the solution might be making data sharing a prerequisite for publication: "We wouldn't be able to publish [data] so promptly [without] the pressure" to publish.

Another major locus of data publication challenges centers on a communication breakdown between researchers and funders with respect to recent changes in federal data sharing requirements. While federal agencies' data sharing policies allow for exceptions "to protect trade secrets, confidential commercial information, personally identifiable information, and other information which is protected under law or policy," messaging around the exceptions (which are mentioned in a footnote of the memo) has not reached researchers, who express fear about perceived pressure to share sensitive data.²⁴

Researchers in this sample were supportive of the philosophy of open access. However, as Stephen Pinfield observed in a 2024 study of open access critiques, "There are good reasons for not sharing knowledge: personal privacy or commercial confidentiality are often quoted examples. Exploitative appropriation and misuse should be added to the list of

²⁴ "Memorandum for the Heads of Executive Departments and Agencies," Office of Science and Technology Policy, August 25, 2022, <https://web.archive.org/web/20250102041624/https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>.

reasons for not sharing.”²⁵ Researchers in this sample echoed this, reporting a variety of legal and ethical concerns about sharing their data without realizing that, in many cases, they are exempt from data sharing requirements. Most researchers’ concerns related to the need to protect research participants and community research partners from the potential negative consequences of their data becoming public.

Specific concerns about protecting their research participants inform many researchers’ perspectives. For a qualitative methodologist whose recorded videos are governed by HIPAA, patient health data should not be shared: “I don't think there's any pathway for [health data] becoming commonly shared data. [Nor] do I think there should be.” Respecting the security and sensitivity of data from vulnerable groups, such as the records of children who passed away in Indigenous boarding schools or survey data from incarcerated people, is an important priority and serves as a “counter-consideration when it comes to the spirit of open science” for one psychologist.

Some researchers noted a conflict between data sharing requirements and respect for community partners, where access to information is predicated on trustworthiness to receive it. One researcher mentioned the importance of following the CARE principles, acknowledging Indigenous data sovereignty and the history of extractive research practices in Indigenous communities: “[We] have established committees with [Indigenous community partners] to evaluate data sharing requests [and] we're also requesting that people make some contribution to the populations’ well-being in return for data.”²⁶ Other researchers were concerned with protecting resources from misuse since sharing geotagged data “can put people at risk, it can put resources at risk, you can basically be showing looters where a big archaeological site is” (as an anthropologist noted) or “where cougars are, what location at what time” (according to an environmental scientist). They went on to explain that researchers spend “considerable time [in] trust building” with community

²⁵ Stephen Pinfield, “Epistemic Openness and Constructionism,” in *Achieving Global Open Access: The Need for Scientific, Epistemic and Participatory Openness* (1st ed.), (London: Routledge, July 2024), 47, <https://doi.org/10.4324/9781032679259>. The full text of this chapter is available at:

<https://www.taylorfrancis.com/reader/download/5aafb4a-cb8b-4d69-9bdd-af0be23c57c1/chapter/pdf?context=ubx>.

²⁶ The CARE Principles for Indigenous Data Governance emphasize “Collective Benefit, Authority to Control, Responsibility, Ethics.”

partners to gain access to sensitive data, and they fear that data sharing requirements will put that trust at risk.

Some researchers were concerned about sharing data with strategic military or economic value. A mechanical engineer working on classified military data noted “competing interests” in terms of data sharing: “[Government agencies] want to make this data and research results accessible to the [people] who pay for it with their tax dollars. But at the same time [military intelligence] is going to be looking at [it] and saying, no, but we don't want this to get outside [the country].” Similarly, a geneticist with an active patent application for “tech transfer” has had to keep the data “blocked off” from public view. Finally, in some cases, disciplinary factors discourage data sharing: “I think journalists have moved in [the] direction [of data publication but by] the nature of human subjects it just feels icky to disclose all of that,” stated a journalism researcher.

Researchers’ stated concerns around data sharing are evidence of the need for more explicit and targeted communication from funders and open science advocates about what types of data are excluded from data sharing requirements. Discipline-specific norms and policies around sharing certain types of data are also needed. One biostatistician opined that it is easy for researchers to claim data cannot be shared due to privacy concerns, but it’s actually “quite straightforward” to sanitize genetic data; “there’s standard pipelines to do this.” The researcher went on to recommend that “journals begin to say ‘data on request from the authors’ is an unacceptable data sharing statement.” A political science researcher’s request that IRBs provide guidance on ethical sharing of data also suggests the need for legal services and ethics boards to serve as a robust part of the data services ecosystem.

Affective barriers

Researchers may hesitate to ask for help when they need it. About a third of the sample voiced some reluctance to use data services that appears to be affective in nature, linked to an academic culture of “heroic stamina” as described by Elaine Beretz. As students hesitate to ask for help, so too do researchers; many of the reasons identified in the literature as preventing students from asking for help (“desire for autonomy and self-

reliance,” “fear help provider will [lack] ability to understand the situation,” “overconfidence,” “preservation of self-image and self-esteem”) are also evident in researchers’ statements about their unwillingness to use data services, as demonstrated in the rest of this section.

When asked about their experience using data services, one psychologist’s confession demonstrates their desire for autonomy: “[I] tend to be a little bit of [a] one-man band. I have my research team, but I tend to be the head of it, and I tend to do my thing. [So] I’m not someone who’s really constantly reaching out to the university to help me with that and to manage that and to give me staff support for stuff like that.” A political science graduate student described their fear that data service providers will not understand them if they are from a different discipline: “[I] perhaps think twice [about using data services because] it’ll probably take a little more explaining. [But] I think that’s, you know, a prejudice on my end. That’s laziness of not wanting to go through that.”

A few other researchers who see themselves in the position of helping others were uncomfortable with the idea of accepting help themselves. When asked where they go for professional development, a researcher building a protein database stressed that their team can meet their own professional development needs, then pivoted to highlight their role in serving others: “the goal is for our research to serve the scientific community.” One English researcher told the interviewer they have a lot to learn from librarians but that the reason for this is “so that we can pass that on to our students, integrate that into our classrooms” rather than to learn themselves. In the past few decades, institutional infrastructure has been built around student services, to provide help to students automatically without requiring them to seek it out. Researchers’ challenges speak to the need for a similar approach to support those who may be reluctant to seek help.

Researcher experiences with data services

In some sense it is natural for researchers not to spend time thinking about how their needs are met so long as they are met, but it is useful for administrators to remember that researchers are likely to underreport their use of data services.

As discussed above, many researchers do not know what “data services” are, can’t conceptualize the idea of data services, and often do not understand when they are using a service. As a result, they may unknowingly use data services. This makes understanding the scope of researcher engagement tricky because researchers report not using services when in fact they have. For example, when asked, “Have you or people from your research team used campus resources or services to help support your research data needs?” one researcher replied, “No, not yet at [institution].” Yet elsewhere in the interview, the same researcher described storing data from a health data platform they have access to through the institution on an institutional HPC cluster; having “good relations with the IT people” assigned to their academic unit, where they have access to statistical software, Tableau, and ILOG “through an academic or educational license”; and praised the “resource allocation” in their academic unit. Researchers’ lack of awareness of the data services that they use makes it difficult to gain a clear picture of their full engagement with institutional data services. In some sense it is natural for researchers not to spend time thinking about how their needs are met so long as they are met, but it is useful for administrators to remember that researchers are likely to underreport their use of data services.

The campus data service provision units that emerge from this study match fairly closely with how providers are represented on institutional websites. As in the data services inventory, researchers in Canadian institutions report more robust data services than researchers in US institutions.²⁷

²⁷ Ruby MacDougall and Dylan Ruediger, “The Research Data Services Landscape at US and Canadian Higher Education Institutions,” *Ithaka S+R*, March 14, 2024, <https://doi.org/10.18665/sr.320420>.

Researchers in this sample identified the following sources of research assistance:

- **Peer support:** While peer support is not a data service, most researchers rely on peers for help and consider this their first line and most natural source of support. Data service providers should support peer networks and provide a ready alternative when they fail.
- **IT support:** Researchers are most likely to recognize IT as a service provider and prefer IT services that are personalized or localized, services that have a strong instructional component, and services with clear branding like HPCs.
- **Research office:** Researchers think highly of the trainings they receive from research units.
- **Academic departments:** The training provided by academic departments is not sufficient to support all researchers. While departments succeed at supporting their own faculty, researchers outside a department cannot depend on it to provide bespoke data services.
- **Libraries:** Researchers highly value the bespoke assistance of individual librarians but lack a full understanding of the scope of library services.
- **GIS and Statistics:** Researchers typically rely on peer support for statistics and data service providers—often librarians—for GIS assistance.
- **Other campus data services:** Researchers take a more expansive view of data services than providers might and consider the ethics office and legal services to be essential data services.
- **Extramural alternatives to campus data services:** Researchers rely on services outside the campus to meet their data needs, especially in terms of data analysis and branded repositories.

Peer support

Researchers in this sample were most likely to report seeking assistance with research challenges from colleagues. “I guess the very first thing we usually do as humans [is we] run to the nearest friend you have at hand [that] may have some knowledge on” a topic, noted a public health researcher. This practice is well-documented in the literature. Ming Ju’s

survey of the literature on collegiality notes that the nature of academic knowledge is to build upon other knowledge, including that of one's colleagues, and that research productivity is linked to collegial "exchange of information and collaboration between high commitment and high-level colleagues."²⁸ One computing researcher described the processes and systems their lab has set up to facilitate an ideal environment of collegiality, where faculty and students meet officially twice a week. One of the meetings is devoted to specific projects, while the other is focused on discussing "recent readings or recent [code] libraries or anything to keep up with the field," and someone might lead a tutorial. It's "self-motivated learning[; this] is how we learn."

While the laboratory creates a certain degree of built-in structure for peer support, researchers in non-laboratory disciplines also create structures for collegiality. One English researcher described a project team composed of four people in "a flat structure" where everyone is equal. The team "did a skills inventory exercise when we first started working together and we figured out what everybody's good at, what everybody likes to do," and now assigns project tasks based on that inventory. "So [we] split up the work according to what works."

When a problem is too complicated for a researcher's immediate colleagues, they may reach out to another expert in their discipline for help. A neurosurgeon described how they usually "just exchange information among ourselves," but were interested in learning more about how to analyze EEG recordings. So because they are "very connected within the worldwide epilepsy research network," they sought help from a particular expert who introduced them to a new software for the purpose. "If you are knowledgeable in the field, you always know someone somewhere [who] can help," the neurosurgeon explained. This researcher's success depended upon their dense network of collaborators and personal connections.

Data service providers should acknowledge the primacy of peer support and should not attempt to replace it, but rather to facilitate, nurture, and augment it, and to provide a ready alternative when it fails. As one social

²⁸ Ming Ju, "The Impact of Institutional and Peer Support on Faculty Research Productivity: A Comparative Analysis of Research Vs. Non-Research Institutions," Seton Hall University Dissertations and Theses (ETDs), 2010, <https://scholarship.shu.edu/cgi/viewcontent.cgi?article=2611&context=dissertations>.

scientist pointed out, the entire system of collegiality is built on “generosity” which can be “abus[ed].” They related how, before consulting with their colleague who is their “go-to person” for questions on coding and methods, they try to learn what they can from YouTube and the internet. “I wanna come to them with some knowledge [so that I’m] not abusing [their] generosity.” On the flip side, peers can fail to provide that generosity—such as researchers who don’t publish their data online, “so you need to send an email to them and ask for the data. But I tried before. Sometimes they didn’t reply to my email. So this is a little bit not so convenient,” explained a physical scientist. Some researchers lack access to collegial relationships and networks that can provide help; data service providers are often well-connected and may be able to use their connections to help researchers find community. Furthermore, the accessibility of data services as a *service*—that is, as an alternative to a favor from a colleague or a paid consultant—gives it a unique value proposition for researchers in situations where neither of those options will work. Data services providers could lean into this distinction in their marketing materials.

IT support

Information technology support is the most recognized data service in this sample; more than half of researchers described interactions with technology staff on their campus, including at IT, Research Computing, and HPC clusters. Overall, it seems that researchers are turning to tech support services to provide them with professional development by teaching them the technical skills they need to learn as part of their job. In general, researchers are most pleased with services that are personalized or localized, services that have a strong instructional component, and services with clear branding.

When describing the services offered by IT, researchers mentioned help with software and hardware, storage and security, training workshops, and individualized problem solving. A nursing researcher used REDCap with the assistance of a lab manager at a core facility on campus who manages that software: “They’ve been very helpful. [They] have a really nice website. You know, they explain [how] the process works.” A different nursing researcher relied on health sciences IT staff to manage security procedures for an international team member. A bioengineering

researcher appreciated that IT “gives us [the] rolling training on Python and Linux, and [the campus cluster], and all those basic training opportunities.”

Some researchers distinguished between services offered by “central” IT and services offered more locally by staff who work in their academic unit. They tended to positively evaluate these “distributed” staff. For example, for a researcher at a research institute, internal IT capacity is an “added benefit” because if they need bespoke tech services, it can happen more quickly than with central IT. This researcher observed a “close working relationship” between the “IT-heavy positions” at the institute and central IT, who work together “on a daily or a weekly basis.” The researcher described a long process to evolve this working relationship and develop the skills of the internal IT staff members “so that we don't have to rely on [central IT] as heavily,” a situation that benefits both parties in the relationship. On the other hand, researchers did not spend a lot of time describing services offered by central IT. Local or distributed IT staff, who have the capacity to understand researchers’ needs on an individual level, appear to be a more appealing model for researchers.

In our sample, research computing was not mentioned frequently as a data service provider—possibly because fewer researchers are aware of the name of this unit. When research computing was mentioned, researchers were about equally likely to report that it solved their data size issue as they were to say that it did not solve their problem. Those researchers whose problems were solved by research computing tended to use more *effusive praise* than for general IT services, possibly because the problems research computing solves are somewhat larger scale than those solved by general IT. For example, research computing set up an internal lab share and a Globus data storage system with multiple clusters for a geography researcher that “just *changed our lives*. [Like,] literally, I know that sounds *dramatic*. But it really has.” A cancer researcher who receives help from the health unit’s research computing office found “being able to authenticate [with the institution] domain [to be] *hugely helpful*,” and “being able to give instrument MAC addresses and serial numbers [so] our machines can talk to the [institution] network [has] been *awesome*. So yeah, that's an easy to overlook thing, but that's been *really, really huge*.”

Several researchers mentioned using the instructional materials hosted by research computing, especially when those materials are accessible

online. One biologist received the research computing newsletter to keep up to date on instructional offerings, “which always look fabulous” if difficult to fit into their schedule. Another biologist mentioned the convenience of research computing’s online workshop calendar and the fact that workshops are “mainly through Zoom, so it makes it easy to attend.” They also find research computing’s instructional offerings to be useful for students, although not specialized enough for their needs. On the flip side, a social scientist thought research computing’s instruction is too advanced for them: “I did a session..., I think, several years ago, I have to admit, it wasn't helpful. I left there more confused than I did going in. [Those] tutorials were good for people to understand the system” but assumed a lot of background knowledge and were not practical. It appears that there is a need for better targeted and scaffolded instruction from research computing.

HPC clusters as a data service are highly salient to researchers. Almost as many researchers described using HPC clusters or supercomputers as mentioned general IT services. Some HPC clusters are campus-wide, while others are administered by particular academic units or labs. Researchers often know the brand name of the cluster they use and refer to it by name (e.g., Palmetto, Polaris, RedHawk, Rivanna, PACE, Amarel, WAVE, Caviness, Darwin, HiPerGator). One biologist stated that it’s a dealbreaker for them if an HPC cluster is not available: “People get recruited because of this resource. I've been on search committees and they didn't have it. They wouldn't come, right?... I think this is, I'll just say, as big as having high performance computing at [institution], like having [that] capacity.” Data service providers may want to consult with their campus marketing department on naming and promoting their HPC clusters for maximum visibility.

This high degree of brand awareness aligns with the tight packaging of HPC clusters as a service—researchers described robust infrastructure around onboarding to the cluster, including websites and trainings. One oceanographer praised their HPC’s weekly live Q&A drop-in session for troubleshooting and “different categories [of trainings] from the very entry level to [the] very high level” which are available on video if researchers are unable to attend.

Researchers whose HPC onboarding was less robust asked for increased support. One civil engineering postdoc started their journey with HPC staff telling them to “Go look at the documentation,” which they did not feel

was an adequate level of support. The postdoc then consulted a grad student friend who gave them some code, which they took back to HPC staff. With code to work with, the postdoc had a more “helpful experience” but eventually reached a point where the postdoc did not understand: “I would have never been able to figure that out on my own.” This same researcher described the emotionally difficult aspects of navigating HPC onboarding without enough infrastructure: “The HPC thing was a challenge for me. Honestly, I felt happier once I figured that out then when I finished my PhD itself. [It] would have been helpful to have more examples or templates, or something, or just like babysitting me through each step.” While the researcher recognized the value in learning the system and that HPC staff are busy, they still thought that “there's some middle ground where there could have been less pain for me.”

In some instances, researchers access data service technology—primarily HPCs—through their academic units. This technology is often shared or borrowed across academic departments. An economist bemoaned how they use a remote server in the business school to handle big data, but it’s “shared for research and teaching purposes, and there's other faculty using it. [That's] something that I hope we would be able [to] get better access to in the future—if there's funding, obviously.” A mechanical engineer will “just kinda go and knock on [my colleagues’ door] and say, ‘Hey, are you using your cluster for this weekend?’” The engineer couldn’t justify purchasing their own cluster access “because I don't need that service all the time. But when I do need it, I really do need it.” Several researchers mentioned the distributed or local tech support staff who manage this technology in their academic units, and one researcher specifically asked for local IT support staff to enable easier use of the computer science department’s HPC: “When [IT is] centralized and there are no dedicated people working with a specific department or program or research lab, [it's] really hard to help because they don't know the needs of each individual unit or department.”

Research office

About half the researchers in the sample have received data services from a named research unit on campus. These units are highly varied, as in Ithaka S+R’s data services inventory where this is a “composite category that includes the interdisciplinary institutes, research cores, research

facilities, data science institutes, as well as the actual research office.”²⁹ (Due to transcript limitations, it was not always possible to determine the boundaries between research computing and research cores.) Research units were most commonly described as providing instructional services (trainings and newsletters) and being hired to conduct some portion of data management and analysis.

Researchers had positive feedback for the trainings they have attended from research units. Unsurprisingly, the type of training offered depends on the type of research unit. A health sciences researcher participated in trainings at the campus research center on “how to learn entrepreneurship for your research[,] how to improve communication for your research[, and] initial statistic training.” A nursing researcher felt “very supported” by trainings at the VPR’s office and did a “SciENCv workshop” at the office of proposal development. A qualitative methodologist encouraged students to attend trainings at the campus research support center on “basic uses of qualitative data analysis software packages, how to conduct a qualitative interview, [and] how to do coding and thematizing.” Data service providers at campus research units might consider coordinating a joint survey of campus training needs and divvying up the subject matter accordingly. Also, research units that are not already doing so should consider marketing their data services; one health researcher only found out about a research institute’s newsletter because their family member works there.

Researchers who engage research units to conduct data management and analysis for their projects had more mixed experiences. Some reported frequent and successful collaboration, such as a policy researcher who engaged three separate campus research units to help them with surveys and data mergers. However, several researchers reported that research units were unable to provide them with the data services they needed. For example, a geographer offered to pay the regional research center on campus for their data, but they had no regular process to provide the data, thus delaying the project by several months. The geographer considers this to be a “real red flag” about data sharing

²⁹ Ruby MacDougall and Dylan Ruediger, “The Research Data Services Landscape at US and Canadian Higher Education Institutions,” *Ithaka S+R*, March 14, 2024, <https://doi.org/10.18665/sr.320420>; Yuzhou Bai and Roger Schonfeld, “What Is a Research Core? A Primer on a Critical Component of the Research Enterprise,” *Ithaka S+R*, December 16, 2021, <https://doi.org/10.18665/sr.316205>.

on campus, where “we've got this great resource there. But getting access to that data in a way that's usable, and, you know, friendly to access, I'd say is definitely a problem.” A health researcher reported that, though a campus research unit was willing to provide a service, it did not have the resources to do so: “They have a small group of people who you can hire to do [qualitative analysis. But] everybody's backed up specifically [for that service]” due to high demand. If scarcity of resources is impacting research units' ability to offer data services, VPR offices and other administrators with budgets could consider using their own resources to assist research units in collaborating with other researchers on campus, perhaps by paying for additional staff hours to respond to data sharing requests.

Academic departments

Academic departments and colleges are another common provider of data services; about half the researchers in the sample described soliciting or receiving data services from an academic unit. The most common data services academic units provided were courses and other forms of instruction; tech support such as storage, software, and equipment; and personalized assistance with data management or analysis.

As units specifically designed to provide instruction, academic departments often see themselves as uniquely equipped to teach technical skills and data management and analysis methodologies. One researcher explained that they teach best practices for data management in their courses: “[At our engineering school], **we teach people how to program...** When I teach a class, [I] require and enforce and grade for documented code and licensing and, you know, copyright statements,” in spite of skepticism from other faculty who think it is unnecessary. While defining all university coursework as research data services would essentially make the category meaningless, it's important to note that the boundaries between pedagogy and research data support services are fuzzy in the context of graduate education. One biostatistician explained how the instructional environment of academic departments can serve as “a back door to formal trainings” in the skills graduate students need—essentially an alternative to research data services. They enrich their graduate students' education by bringing them to sit in on classes outside their discipline: “And in the process of being in a room with 50 microbial

ecologists and listening to intro lectures, they'll get maybe a better picture of what these typical analyses are and how one might carry them out.”

As with other forms of peer support, whether instructors in an academic unit actually provide the data management training that researchers need depends on many factors, including whether a researcher's skill level is aligned to the course's design. One nursing researcher “signed up for a class [on data visualization] that was [like] \$60[;] I forget which department sent it out.” The researcher hoped the course would help them visualize their data, but didn't realize they were required to know a coding language to take the course. While academic departments should design their curriculum so that their students gain the needed data management skills in their field, and should seek co-curricular credit for these courses in other departments where the students will be well-prepared to succeed in their courses, the courses offered by academic departments will not be a one-size-fits-all solution for every researcher on campus.

Only a few researchers mentioned recruiting an outside department to provide more intensive research data assistance, and most of these reported negative experiences with this type of data service. In one case the assistance took too long because the assisting department did not have enough resources: a clinical researcher who enlisted researchers in another department to extract data waited more than a year “because they were so short on resources, and communication was poor, and they didn't have priority on our project. [We] had to pay about \$100 per hour for data extraction [and] there is not any accountability for that [- it's just] not a good process.” In another case, a research unit “started engaging with [the statistics department] on campus, and then the person who we were working with [left] the university;” after this staff turnover, the research unit was unable to re-establish the relationship. These experiences emphasize the importance of dedicated data service professionals on campus who can prioritize meeting researchers' needs for individualized consultations, since relying on other departments to do so often has negative outcomes.

Ithaka S+R's data service inventory found that the most common data service offered by academic departments was statistical consulting.³⁰ In this study, it is interesting that researchers did not reference departmental statistical consulting in proportion to how frequently it is available. It is possible that researchers are using statistical services but are not associating them with the statistics department, or there may be another explanation that is beyond the scope of this study.

Libraries

Although academic libraries have shifted some of their focus from collections to services, faculty continue to perceive collections as the main function of libraries.³¹ Researchers in our sample seldom described libraries as data service providers. Several researchers demonstrated a lack of understanding of the full scope of libraries' capabilities. "I will say I feel like I [don't] tap into the library resources enough, cause I don't know what's available," an education researcher explained. Though this researcher stated that they will use library services in their data management plan, they "honestly have no idea" what it means to state this, or what supports the library provides. A biologist stated that they have never used the library and "don't know what they offer," while an arts researcher described taking a circuitous path to using library data services: "My first thought wasn't [to] reach out to the libraries;" instead they spoke to a series of people who eventually directed them there. When researchers did describe library data services, they sometimes mentioned workshops and occasionally mentioned other services but were most likely to cite the name of a particular librarian who helped them.

The individualized assistance provided by *librarians* was the most salient aspect of library data services for researchers in this study. A few researchers had heard of library workshops, but they rarely had anything clear to say about them. (This is in contrast to, for example, the trainings

³⁰ Ruby MacDougall and Dylan Ruediger, "The Research Data Services Landscape at US and Canadian Higher Education Institutions," *Ithaka S+R*, March 14 2024, <https://doi.org/10.18665/sr.320420>.

³¹ Melissa Blankstein, "Ithaka S+R US Faculty Survey 2021," *Ithaka S+R*, July 14 2022, <https://doi.org/10.18665/sr.316896>; Ioana G. Hulbert, "US Library Survey 2022: Navigating the New Normal," *Ithaka S+R*, March 30 2023, <https://doi.org/10.18665/sr.318642>.

offered by research units. Perhaps researchers are more easily able to identify with specialized trainings from a research unit than with general workshops offered by libraries). A graduate student in political science was unable to attend a library workshop on NVivo, so they contacted the **data librarian** who had run the workshop. She sent them slides and suggested a meeting to answer the student's questions, which the student found helpful. The student still received services from the library, but they essentially used a consultation rather than a workshop.

Rather than describe standardized services at the library, researchers who mentioned the library tended to name specific librarians who have helped them with bespoke issues, e.g., “We're using scanners with [**technology librarian's**] team, we use their scanners.” While this may be because most researchers were being interviewed by librarians—so the researchers could be sure the interviewers knew who they were talking about—it also evokes a collegial framing, where librarians are positioned as team members providing peer support. In their descriptions, researchers also speak collegially of the librarians who are helping them. For example, one **health science librarian** helped a researcher construct a database on tissue engineering by “[sorting] through this enormous database [of sources] to make sure we have the most relevant information,” while a **data management librarian** participated in a “mini data working group” to construct a database with an arts researcher. It seems that many researchers are finding help from librarians in much the same way that they would find help from peers—on the occasion of a pressing need—rather than seeking out the library proactively as a service provider. In advertising library data services, providers should consider what features distinguish it from peer support and highlight those.

Few researchers seemed to have an understanding of the larger role of libraries in data management. Those who do tend to become evangelists, like this health science researcher who first found out about the library's research data management team over a year ago: “I tell all the students I teach now about [the library's] team and about how important it is that they connect with [them. I] think it's an important part of being a responsible researcher to have those connections.” This researcher's comments speak to the importance of institutionalizing general data management services as a habitual part of project workflow.

GIS and statistics

GIS and statistics are common analytical approaches for which researchers need support. About one fifth of researchers in this sample described a need for support to use each of these approaches. Yet they seek support from different providers: researchers typically rely on peers for statistical support, while they often turn to data service providers for GIS support.

Most researchers who need statistical support seek that support from a colleague rather than from an existing service. Often the statistical consultant is hired and paid (as in the case of a researcher collecting clinical data), although in some cases the statistician is brought on as a team member (as in the case of a different health researcher). When the topic occasionally comes up, researchers voiced their support for statistical consulting as a data service. As one physical scientist noted, “I know that some universities have help with statistical analysis. It's a group, it's an office. [I] don't think we have a resource like that. So that's probably useful for a lot of projects on campus.” If researchers are used to considering statistical consulting as a paid service, institutions wishing to establish statistical data services will need to carefully consider their funding model.

In contrast, most researchers who need GIS support interact with data service providers, who are often GIS librarians. One social scientist related that they typically rely on online training videos from Esri for GIS support but will attend virtual trainings provided by their GIS librarian if they get an email about it. They also suggested that workshops on new GIS tools “could be helpful for a lot of researchers.” In addition to GIS librarians, researchers seek GIS help from peers, IT services, and academic departments. For example, an architecture researcher noted that their academic unit offers courses on GIS and BIM. Our inventory found that, while most R1 institutions have a GIS librarian, other institutions are less likely to have them.³² Institutions without a GIS librarian should consider adding one to their roster, as this method of delivering data services

³² Ruby MacDougall and Dylan Ruediger, “The Research Data Services Landscape at US and Canadian Higher Education Institutions,” *Ithaka S+R*, March 14, 2024, <https://doi.org/10.18665/sr.320420>.

seems fairly successful. Institutions may also want to consider whether a similar model for “statistics librarian” is plausible.

Other campus data services

In this sample, researchers took a more expansive view of data services than providers may anticipate. Many researchers mentioned the institution’s provision of data storage and software without naming the provider; researchers may perceive these benefits as accruing at an institutional level. For example, a physics researcher was grateful for the “campus-wide license” for Matlab, which saves them a lot of money, in contrast with Mathematica, for which the researcher has to buy their own license. In the same vein, a few researchers mentioned institutional memberships in consortia that allow them to access data and data services. For example, a health science researcher recognized the benefit of access to a database of health data that is shared across institutions. In each of these cases, it is not clear exactly which data service providers negotiated the benefit on behalf of the institution. This trend speaks to the need for data services to have a single, well-publicized point of contact for researchers so that requests such as these can be routed to the appropriate provider.

Researchers also mentioned various *unexpected campus offices* as important providers of data services. One arts researcher credited the *undergraduate research office* with supplying research labor and the *business incubator office* with consulting on how to make research sustainable. But two campus units come up frequently enough that it is clear that researchers consider them to be part of data services: the *IRB or ethics office*, and *legal services*.

Ethics office and legal services

Two-thirds of researchers in this sample collect human or animal data, meaning that applying for and receiving ethics approval is a required step in their data pipeline. Of those researchers, about half described the IRB or ethics office as an integral part of data services. In a few cases, this is portrayed as a negative, as researchers complained that a mismanaged IRB can slow down or prevent data collection. Most commonly, however, researchers mentioned the IRB’s role in governing how their data is secured and whether it is shared, both vital services in the data pipeline.

Researchers gave several examples of **data services they desire or expect** from their IRBs. One researcher seemed to expect **concrete guidance on data sharing** from IRBs, such as “knowing how to prepare data for [data sharing]. What ethical, you know, IRB considerations to take into account?” A few researchers described relying on the IRB for data services including **ethics training** through “Ask IRB” sessions and online resources and **procedural review**: “[When] I was in my PhD program, some faculty would specifically request [the IRB] come” and “check all your procedures and processes.” One researcher who works in the criminal justice system doesn’t ask for help to **coordinate multiple IRBs** simultaneously on the same project (though this seems like a needed service), but they do ask for training on data sharing because it is unclear which IRB has jurisdiction: “If I’m collecting data at a county jail, who’s the IRB there? Is it county level? Is it some sort of state level review?” Overall the picture that emerges is that researchers consider the IRB to be part of data services and expect it to provide services to support their research.

Several researchers similarly referenced legal services as a data service provider that primarily assists with contracts with outside organizations. To obtain pre-existing data from an outside company, one engineer “had to reach out to an attorney at [institution] to get the final paperwork.” An arts researcher went on a “wild goose chase” to locate a draft “data use agreement framework” for their flagship public university (though no researchers explicitly mention oversight units for campus data use agreements). A social scientist described getting in trouble in the past with the contracts office for not having had the proper paperwork for a collaboration with a community partner. They have since learned to distinguish between different types of contracts and explain to the interviewer how service agreements and research agreements differ. The increasing frequency of use of pre-existing datasets and community-engaged research suggests the need for vastly expanded support for researchers contracting with outside organizations.

Legal assistance may also be necessary to negotiate increasing cross-institutional and international collaborations. One humanities researcher described their extensive use of legal services for ensuring that the data from their many outside collaborators will not be lost: “The contracts that are grinding their way through [institution] legal right now specify that if [collaborators] drop the project, if they disappear, if they die, we will have the right to use [what] they’ve done already as a starting point for

somebody else to come on board.” This researcher engaged a number of legal services over the lifecycle of the project, including a personal lawyer, the contract office, the vice provost for research’s partnerships office, and campus legal services.

Several Canadian institutions and at least one US institution in our cohort already incorporate IRB and legal services into their data service ecosystems. Even in these institutions, researchers noted the need for closer engagement. A health science researcher in Canada believed that the ethics office “could be steering students and researchers to [the library’s RDM services. And] the same with [legal services. I previously] thought ethics and legal were the two groups to check in on and it was a great surprise to find out about [the library’s RDM team].” This researcher’s point of view suggests that requiring researchers to consult with the library on data management could be added as part of an IRB application checklist. Consideration of eventual data publication during ethics review could eliminate the major challenge of researchers realizing too late that they need to share their data.

Extramural alternatives to campus data services

Rather than turn to campus data services, some researchers turn to outside resources first for assistance. Other researchers turn to outside resources only when their institution lacks the appropriate resource. External resources primarily include professional development and training in new data skills, contracts with an outside party to conduct data management and analysis, repositories, and external datasets and storage.

Most researchers in the sample have received training in new data skills from an outside resource. YouTube and “my best friend, Dr. Google” are mentioned as common sources of information along with ChatGPT, GitHub, Reddit; software trainings like Esri, Covidence, Mathworks; and online schools like Khan Academy, Udemy, W3schools, and Coursera. Discipline-specific books, journals, forums, listservs, conferences, and professional organizations are also important resources. A few researchers asked for more integration between these outside resources and their institution, like a health communication researcher who wanted

data service providers to make NIH trainings available to campus researchers “[so] we don't have to reinvent wheels.” This researcher suggested data service providers could work to index the external training resources available to facilitate access. In order to do this, subject librarians could, for example, regularly survey researchers at their institution about the external training resources they use and update library guides with that information.

Data analysis

Around a fifth of researchers in this sample contract with an outside party to conduct data management or analysis. In a few cases this is a requirement for working with an outside dataset. For the most part, however, researchers turn to these outside contractors because they see them as the best or only person for the job. As one psychologist noted, they can get most of their research needs met at their institution, but “finding contracted consultants” is often necessary for “things that are super specialized.” Researchers often view this “esoteric” assistance (as a health researcher put it) as something that can best be accomplished by “just one or two or three experts on this globe” (as another health researcher noted) or by someone who specializes in a niche type of analysis. For example, a biostatistician subcontracted with a collaborator in bioinformatics to process their data into a particular format: “And he's got a streamlined process through which he's able to do that.” For the most part researchers are satisfied with the services provided by these outside analysts. Vice provosts for research may want to consider providing a standing fund for such “out-of-network” data services which cannot be provided on campus.

Repositories

Researchers in this sample were about four times as likely to use an external repository for their data as a campus repository. Researchers often identified their repository by brand name, regardless of whether it is a campus or external repository; in addition to HPC clusters, repositories are another area where data services have strong branding that is reflected by researchers’ narratives. The most common repository mentioned was GitHub (used by these researchers for both datasets and code); other brands mentioned by name included Dryad (one interviewer tells a health researcher “you could use Dryad for free”), Zenodo, the Syracuse Qualitative Data Repository, NIH UniProt, NCBI GEO (Gene

Expression Omnibus), DesignSafe Data Depot Repository, IEEE DataPort, and tDAR (the Digital Archaeological Record).

Conclusion

In order to design research data services to provide maximal support to researchers, it is necessary to understand researchers' support needs. This study has found evidence for several major challenges that researchers experience.

The first, a meta-challenge, is that researchers lack awareness of the idea of data services. This is different from just not knowing what services are available on their campus. If they don't have a mental concept that something called "data services" could exist, they will not even think to investigate what data services are available on their campus. Researchers who do know that data services exist are often concerned that these services are too slow or do not provide the specialized expertise necessary to help them with esoteric problems. What this means for the library, then, is that while librarians may be embracing the service model, they are not yet doing so in a way that is clearly legible from the researchers' perspective.

Researchers also experience challenges accessing their own working data and sharing it with members of their project teams. Almost all researchers in this sample work on collaborative teams using cloud-based file-sharing systems, yet issues with security, version control, interoperability, and reliability of these systems threaten data integrity and force researchers to find time-consuming workarounds.

Data size can be an issue for researchers in all disciplines since handling both "big" and "medium" data requires distinct data management practices researchers may not be equipped for. Researchers described feeling alone when institutions expect them to figure out these practices without support. They also noted the tension inherent in being dependent on the for-profit technology industry for solutions.

Researchers experience challenges related to accessing the technology they need to interact with their data. They reported struggling most with

discovering and reusing pre-existing datasets and request more assistance with this process from their libraries. Learning to use specialized software and equipment is also a challenge.

Administrative tasks related to managing the data pipeline are time-consuming for researchers who have many other responsibilities, and they often lack training in skills like project management or people management. Several asked for the assistance of a data manager, but this is unlikely to be a workable solution for most researchers. Negotiating with partners for permission to collect or reuse data is one especially challenging step in the data pipeline.

Researchers expressed shame and fear around what they perceive to be the new job responsibility of data publication. They are not prepared or trained to incorporate data publication into their workflows and as a result may scramble to fulfill data publication requirements at the end of a project. They also do not prioritize data publication, opining that it comes at the expense of other priorities and that it should be someone else's responsibility. They seem unaware of exceptions to data sharing requirements and are concerned about the ethics and legality of publishing sensitive data.

Finally, researchers experience affective barriers to seeking help, such as a desire for autonomy or fear that service providers will not understand them. Researchers may experience dissonance with their sense of self when they seek or accept help due to academia's culture of finding solutions to problems on your own.

In the next section, we present our recommendations for data service providers to address these challenges. We acknowledge that some of these recommendations may not be actionable for lesser-resourced institutions but represent aspirational goals.

Recommendations

Funders

1. Continue investing in shared infrastructure to facilitate cross-institutional data management and preservation tools and platforms.
2. Convene disciplinary communities to build consensus about expectations for data management and sharing, as well as the value and lifespan of research data to address researchers' concerns that the labor involved in data management is disproportionate to its value.
3. Consider developing funding mechanisms to better support long-term cloud storage costs.

Data Providers

1. Consider updating policies around data security and use to better reflect researchers' collaboration practices.

Universities

1. Create cross-unit governance structures and protocols to support periodic, systematic assessment of the university's research data support infrastructure to identify gaps, redundancies, and promote cross-unit coordination and collaboration.
2. Develop, maintain, and promote a central directory of research data services to improve the visibility of research data services offered across the university, and explore opportunities to leverage AI to improve the interactivity of the directory.
3. Consider staffing a concierge service to help researchers navigate research data service options.
4. Leverage existing peer-to-peer faculty support networks to create communities of practice.

5. To improve researchers' awareness of existing research data service options as well as data services as a category, socialize existing offerings at high leverage events and milestones such as new faculty or grad student orientation or annual reviews.
6. Seek extra-curricular and curricular opportunities to expose graduate students, who in many cases are heavily involved in research data management for faculty projects, to research data service offerings and consider tailoring workshops and programming to this constituency.
7. Seek opportunities to invest in shared infrastructure to facilitate cross-institutional data management and preservation.
8. At the department level, explore ways to include data as a research output in retention, promotion, and tenure.

Research Offices

1. Create and maintain structured opportunities for researchers to improve their understanding of and skills with para-research skills such as budgeting, supervising personnel, and reporting, compliance, and security policies that faculty frequently report struggling with.
2. Explore opportunities to leverage the sponsored projects/research development office to connect research data service providers and researchers as part of the pre- or post-award process. In addition to connecting researchers with services, this practice could also surface opportunities to write providers into grant budgets.

University Libraries

1. Increase collections of research datasets and develop programming or resources to assist researchers with discovery of and access to datasets held by the library and available elsewhere.
2. When feasible, build capacity for individualized consultation and on-demand assistance with research data management.
3. Conduct outreach to help researchers better understand the specific requirements and goals of funders' data sharing requirements, about which there are considerable misconceptions.

4. Develop ties with the sponsored projects office to market relevant services to researchers and, when appropriate, find opportunities to include allocations to the library in grant proposals.
5. For libraries that have not already done so, coordinate with the research office to prioritize services aligned with the institution's strategic plan and existing research strengths.
6. In addition to the guidelines around data privacy and security that IRBs already enforce, consider building in a data publication consultation as part of an IRB application.

IT and Research Computing

1. Coordinate with libraries and other service providers to streamline research data service offerings.
2. Seek opportunities to invest in or host shared and community and community infrastructure to facilitate cross-institutional data management and preservation.
3. Consider developing an active outreach plan for establishing relationships with researchers and better recognition of unit capabilities for providing research data services.

Appendix A

These 29 institutions participated in the cohort project. Twenty-seven of them submitted interviews for this report.

Institution	Cohort Members
Brandeis University	Margarita Corral, Ford Fishman, Laura Hibbler, Jennifer Perloff
Carnegie Mellon University	Lencia Beltran, Emily Bongiovanni, Alfredo Gonzales, Brian Matthew, Emma Slayton
Chapman University	Anna Alber, Doug Dechow, Andrew Greenman, Jana Remy
Clemson University	Becky Ligon, Nalinee Patin, Stacie Powell, Megan Sheffield, Elias Tzoc
Dartmouth College	John Bell, Lora Leligdon, Lilly Linden
Florida State University	Neelam Bharti, Renaine Julian, April Lovett, Mila Turner, Nick Ruhs
Georgia Institute of Technology	Karen Glover, Cynthia Kutka Fred Rascoe, Matt Sanders
Harvard University	Emre Keskin, Ardys Kozbial, Yuan Li, Scott Yockel
Indiana University	Katie Chapman, Ethan Fridmanski, Ryan Hedrick, Emily Meanwell, Theresa Quill, Esen Tuna
Montclair State University	Stefanie Brachfeld, Klavdiya Hammond, Siobhan McCarthy, Danianne Mizzy, Danielle Richardson
Northwestern University	Tobin Magle, Kelsey James Rydland, Pamela Shaw, Sarah Thorngate
Ohio State University	Kelsey Badger, Anna Biszaha, Tanya Berger-Wolf, Sandy Shew, Alexander Davis
Queen's University	Alex Cooper, Elise Degen, Meghan Goodchild, Rebecca Pero, Nevil Joseph Silverius
Rutgers University	Diane Ambrose, Joseph Deodato, Mei Ling Lo, Victoria Wagner, Ryan Womack
San Diego State University	Michael Farley, Margaret Henderson, Mark Reed, Scott Walter

Santa Clara University	Nicole Branch, Mary-Ellen Fortini, Benjamin Hall, Carol Jordan
SUNY Albany	Spencer A Bruce, Kathleen Flynn, Angela Hackstadt, Emily Kilcer, Sandra McGinnis, Terrell D. Rabb
SUNY Binghamton	Amy Gay, Mike Jacobson, David Schuster, Nick Walling, Alexander Carter
SUNY Stony Brook	Susan Gasparo, Jessica Koos, Mona Ramonetti,
Towson University	Songyao Chen, Samuel Collins, Joyce Garczynksi, Carrie Price, Patricia Westerman
University of Chicago	Greg Fleming, Jen Green, Jenny Hart, Adrian Ho, Anna Jackson, Barbara Kern, Cecilia Smith
University of Delaware	Sarah Katz, Michael Kyle, Daniel Peart, Michael Stewart, Alison Wessel
University of Florida	Erik Deumens, Cassandra Farley, Natya Hans, Kevin Hanson, Emily McElroy, Carol McMahon, Carl Moritz, Hannah F. Norton, Trey Shelton, Laura Spears
University of Manitoba	Jordan Bass, Jackie Cooney-Birch, Janet Rothney, Dawn Sutherland, Huy Tran, Wei Xuan
University of Pittsburgh	Renea Elaine Barger, Dominic Bordelon, Aaron Brenner, Michael Colaresi, Christopher Lemery, Melissa Anne Ratajeski
University of Victoria	Lisa Goddard, Monique Grenier, Sarah Huber Shahira Khair
University of Virginia	Lucy Carr Jones, Andrea Denton, Jacalyn Huband, Jennifer Huck, Ricky Patterson
University of Washington	Xiaosong Li, Jacob A Morris, Jenny Muilenburg, Sarah A. Stone
Yale University	Lisa D'Angelo, Barbara Etsy

Appendix B

The majority of participating institutions are located in the eastern United States (Table 1), and most have a very high level of research activity (Table 2).

Table 1

Census Region	Number of Participating Institutions
Northeast	9
South	7
Midwest	4
West	4
Canada	3

Table 2

Carnegie Classification	Number of Participating Institutions
Doctoral Universities Very High Research Activity	18
Doctoral Universities High Research Activity	4
Doctoral Universities Doctoral/Professional Universities	1
Master's Colleges & Universities Larger Programs	1
Doctoral Universities Canada	3

Appendix C

Tables 3 and 4 show the demographics of all interviewees and of the sample.

Table 3

Disciplinary Area	Total #	Total %	Sample #	Sample %
Arts & Humanities	21	7%	4	10%
Biological Sciences, Agriculture, & Natural Resources	33	11%	4	10%
Business	10	3%	1	2%
Communications, Media, & Public Relations	3	1%	0	0%
Education	8	3%	1	2%
Engineering	37	13%	5	12%
Health Professions	70	24%	10	24%
Physical Sciences, Mathematics, & Computer Science	41	14%	6	15%
Social Sciences	56	19%	8	20%
Social Service Professions	15	5%	2	5%
Total	294	100%	41	100%

Table 4

Rank	Total #	Total %	Sample #	Sample %
Administrator	9	3%	1	2%
Full Professor	102	35%	15	37%
Tenured	67	23%	9	22%
Tenure-track	62	21%	9	22%
Research Staff	16	5%	2	5%
Postdoc	9	3%	1	2%
Instructor	2	1%	0	0%
Grad student	8	3%	1	2%
Rank Unavailable	19	6%	3	7%
Total	294	100%	41	100%

Appendix D

Below is the interview guide that cohort participants used to interview researchers on their campus.

Semi-Structured Interview Guide

Introduction

As data-intensive research becomes the norm in an increasing number of disciplines, universities are investing in providing support services to help researchers manage data across the research lifecycle. [Insert name of institution] is conducting a study to understand how well existing support services align with the needs of researchers and to coordinate future offerings across campus. I'd like to ask you questions today about your experiences engaging with research data support services on campus.

Before we begin, I'd like to briefly define what research data support services mean in the context of this study. Research data support services are programmatic offerings such as trainings, workshops, and consulting that support data-intensive research. These services are typically offered by the university library, an academic department, an institute or research center, an IT department, research computing units or cores, but may also be offered by other entities on campus. Research data support services can range from instruction in specific software such as GIS or Python, data management, or analysis.

Our focus today is on services that support your research: services that support your teaching are generally out of scope. Otherwise, I encourage you to think broadly about the services you are aware of on campus and especially those which you have used to support your research. We're also interested in data support needs that the university does not yet fulfill. Do you have any questions about the study and/or your participation before we get started?

Do I have your consent to begin the interview? And do I have your consent to record our interview?

Introduction

1. Briefly describe your research focus.
 - a. » Can you share an example of a specific research project you've conducted or participated in?
2. Do you conduct your research alone or as part of a lab or research team?
 - a. If they conduct research alone, skip ahead to the next section.
 - b. If they work as part of a team, how are research responsibilities divided up among members of your team?

Data Practices

1. Do you generate most of your own research data or work mostly with secondary data?
 - a. » What challenges do you face in creating or locating data for research?
2. Walk me through your typical process for managing research data.
 - a. » How confident are you with your data management process?
 - b. » What are the biggest challenges you face in managing research data?
3. How do you analyze or model data in the course of your research?
 - a. » What software or computing infrastructure do you use?
 - b. » How do you keep up with new tools and methods for analyzing or modeling data?
 - c. » What challenges do you face in analyzing or modeling data?
4. Do you make your data available to other researchers after a project is completed?
 - a. » What factors influenced your decision to make/not to make your data available?
 - b. If yes, where do you deposit or publish your data? What steps, if any, do you take to prepare data for sharing with others?

Training and Support Needs

1. When you need help learning a new data skill, where do you turn? Examples: informal help from colleagues, formal training opportunities like classes or workshops, online tutorials or videos.
 - a. If they work with a team/lab, To the best of your knowledge, where do members of your team go when they need help learning a new data skill?

2. Have you used campus resources to help support your research data needs?
 - a. Feel free to use examples of specific services available at your institution to the interviewee if you desire.
 - b. If no, why not?
 - c. If yes, when you've used resources on campus to help you with your research data needs, what has that experience been like?
 - i. » What types of services have you used?
 - ii. » What campus unit offered those services?
Examples: the library, an HPC, IT department, a research core.
 - iii. » How easy was it to find the resource you were looking for?
 - d. Did you learn what you hoped to learn or solve the problem you were hoping to solve? If they work with a team/lab
 - i. » Have you ever suggested that a team member make use of a research data support service offered on campus?
 - ii. » Are you aware of members of your team seeking out these resources on their own?

3. Tell me about a time when you looked for a data-related resource and couldn't find what you needed.
If they say that has never happened before, go on to the next question.

Looking Ahead

1. Looking to the future, what types of training or support will be required for researchers in your field five to ten years from now?
2. Do you anticipate that funder and/or publisher requirements for

data management or data sharing will affect your future support needs, and if so, how?

3. Is there anything else from your experience or perceptions as a researcher that I should know about your data support needs or your experiences using campus services?