IJDC | *Conference Paper*

Traveling on the same road while navigating different terrain: Institutional data services and repositories across the US

Mikala Narlock
Indiana University

Jake Carlson
University at Buffalo

Wind Cowles,
Princeton University

Leslie Delserone,
University of Nebraska, Lincoln

Joel Herndon,
Duke University

Wendy Kozlowski,
Cornell University

Jonathan Petters,
Virginia Tech

Abstract

This paper describes the Repository Readiness initiative and subsequent Summit for Academic Institutional Readiness in Data Sharing (STAIRS). These efforts examined the current state of institutional research data services and repositories at US academic institutions. Using federal memos and directives published in 2022 as a foundation, members of the Data Curation Network hosted a virtual learning series to identify areas of collaboration across institutions. The themes of this learning series led to STAIRS, which brought representatives from 32 institutions, to discuss the need for and potential benefits of deeper institution-wide engagement and increasing cross-institutional collaborations to support research data sharing efforts. Through discussions at the summit, three key themes emerged: difficulties in scaling services, need for shared resources and training materials, and importance of cross-institutional collaboration. The participant discussions, combined with data from pre- and post-summit surveys, suggest gaps in staffing, resources, and formal policies in data management in institutional settings. The authors conclude with recommendations for funding agencies to support institutional data services through incentives for collaboration, improved communication with program officers, and additional research into data sharing requirements across different institutional types.

Submitted 11 August 2024 ~ *Accepted* 20 February 2025

Correspondence should be addressed to Mikala Narlock, Email: mnarlock@iu.edu

This paper was presented at the International Digital Curation Conference IDCC25, 17-19 February 2025

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

http://dx.doi.org/10.2218/ijdc.v19i1.1060
DOI: 10.2218/ijdc.v19i1.1060

Introduction

As expectations and requirements around data management and sharing have grown over time, so too has the need for academic institutions to develop services and infrastructure to support researchers in addressing these expectations. Academic libraries in particular have recognized the need to provide support for researchers managing and sharing their data. Investments made by libraries have included developing research data management curricula and training programs, supporting consultation services for researchers needing to craft data management plans, and building institutionally based data repositories as a means to share and preserve data for their researchers.

Institutional Repositories (IRs) were introduced more than twenty years ago as essential infrastructure for research libraries (Lynch, 2003; Branin, 2005). IRs accept and preserve the research and teaching outputs from across their institutions, including research data. As data sharing requirements have proliferated, IRs have expanded their scope to become an important means for researchers who do not have access to discipline-specific repositories to meet the data sharing requirements of funding agencies and publishers. Several academic libraries have launched separate institutional repositories that focus solely on stewarding, sharing, and preserving research datasets. Examples of institutional data repositories (data IRs) include the Data Repository for the University of Minnesota and Deep Blue Data from the University of Michigan. These institutional repositories are meant to complement disciplinary repositories and generalist repositories. Each type of repository offers its own set of services, with different features, costs and benefits, but all strive to ensure data are findable, accessible, interoperable, and reusable (FAIR; Wilkinson et al., 2016). Recent research demonstrates that IRs and data IRs are hosting an increasing number of datasets and their rate of growth is increasing exponentially (Narlock et al., 2024b), suggesting that IRs and the services provided by libraries to support data management and sharing have become a critical component of the data sharing ecosystem.

Repositories strive to connect researchers to the global infrastructure of FAIR data through the adoption and use of tools like global persistent identifiers and shared practices like the application of common metadata practices. However, the design, development, and maintenance of data services and IRs are heavily influenced by their institution's culture, needs, and available resources. Although institutional data service and IR providers strive to base their work on accepted standards and best practices, local needs and culture often have led to uneven development of policies, standards, and practices across institutions. This has made it difficult for IRs to take advantage of economies of scale, to easily share new functionalities or practices, or to function as a community across many institutions which limits their overall effectiveness in supporting an open data ecosystem (Johnston et al., 2024).

A second challenge facing academic libraries is the increasing scope and scale of demand for research data services and support, including but not limited to IR services. In the United States, several recent federal directives have heightened demand for scholarly research outputs, including code and data, to be made publicly available. Data services are increasingly a common offering in academic libraries, but effectively staffing these services continues to be a challenge (Tenopir et al., 2019). Furthermore, librarians do not have the expertise or responsibility to address the full array of challenges presented by research data sharing, especially for data that is deemed sensitive or restricted, nor do they have the authority to certify on behalf of the institution that data have been shared successfully. Increased demand and expectations on institutions for managing and sharing data requires an institution-wide response. Therefore, there is a growing need for engagement and community building across institutional units that have responsibilities in supporting data management and sharing.

To address this growing need, the Data Curation Network (DCN) (Johnston et al., 2018) launched the Repository Readiness initiative, a research and advocacy agenda to advance the continued development of institutional data services and IRs as a component of a collaborative global research environment. We will describe the impetus for the Repository Readiness Project, the phases of our work therein, including the Summit for Academic Institutional Readiness in

Data Sharing (STAIRS), and lessons learned from the broader US data services landscape. While our efforts were initially scoped to IRs, we found cause to expand the scope to institutionally based data services as a whole. We will emphasize the importance of such data services and IRs in supporting researchers creating FAIR data outputs, including by providing the essential social and knowledge infrastructures to underpin technical solutions. We will articulate the importance of taking a more multi-institutional, rather than siloed, approach to developing data services and IRs in responding to evolving disciplinary and funder expectations. Finally, we will end with recommendations for future directions for institutionally based repositories and data services.

The Repository Readiness Initiative

During 2022, the White Office of Science and Technology Policy released a memo directing all federal funding agencies to begin requiring scholarly outputs be made publicly accessible (while recognizing legal and ethical considerations that may restrict data sharing) (Nelson, 2022). Additionally, the "Desirable Characteristics of Data Repositories for Federally Funded Research" (DC-DR) was released in 2022, which lays out a set of attributes to guide researchers and funding agencies in selecting appropriate repositories for data sharing and preservation (White House Office of Science and Technology Policy, 2022). While the DC-DR is a significant step forward in that it further articulates expectations for research data repositories and affiliated services, it lacks a clear roadmap for institutions, especially IRs, to translate the high-level guidance into practical implementation.

The DC-DR presents an opportunity to create an environment that better supports researchers in sharing and reusing data while enabling a connected repository landscape. Working through the DCN, an organization of academic and non-profit institutions who collaborate to make research data publicly accessible, the authors developed the multi-part Repository Readiness initiative. Although the impetus of this work was a desire to create shared, community-based approaches for the continued technical and organizational development of IRs, the project scope grew to include the essential social infrastructures and connections that enable data sharing services both within and across academic institutions.

The first phase of the project was to conduct a self-assessment among DCN members to gauge their IR's compliance to the DC-DR. Through this process, we recognized that our IRs are largely in broad alignment with the characteristics as stated (Data Curation Network, 2022; Reiff Conell & Wright, 2024). However, many of our members expressed hesitation at indicating whether they were in full compliance with the DC-DR as the characteristics often lacked sufficient depth or detail, or clear evaluative metrics, to make this determination. While the DC-DR are intentionally without such criteria in order to be broadly applicable, we quickly recognized that IRs would need to develop their own operational benchmarks to determine alignment and to be able to communicate the extent to which they are (or are not) able to demonstrate compliance in practice to our users. Our self assessment against the DC-DR was useful as a place to begin a discussion about the current state of institutionally based data services and IRs, but we recognized that DCN members only represented a portion of the larger landscape. If we really wanted to understand the current state of data services and IRs in order to identify where investments were really needed, then we would need to engage with more of our peers through larger venues.

In response, we developed a series of open virtual learning sessions to promote a shared understanding of current practices and identify shared issues in supporting data sharing through IRs. While our initial focus was on the DC-DR and IR alignment, the topics were intentionally designed to be broad and widely-applicable, even for institutions that do not host standalone data sharing infrastructure. Although many academic libraries are now making investments in providing data services, not everyone has access to the same level of resources or can support hosting data in an IR of their own. In order to understand the current landscape of institutionally based data services, we wanted to reach as many institutional data service providers as we could. To that end, we organized the sessions based on the themes that came out of our DC-DR discussion within the DCN. Each session consisted of a set of presentations delivered by experts

in their field and time for questions and discussion afterwards. The four sessions in the series were:

- Funding Agencies and the Desirable Characteristics
- Data Sharing Readiness in Academic Institutions
- Making the Case for Institutional [Data] Repository Services
- Developing and Maturing IR Technology Platforms to Support Data Sharing

These sessions were free and open to the public, recorded, and shared online after the event. This series was well attended by librarians, data stewards, repository managers, as well as federal funder representatives and researchers, with an average of 188 registrants per event.

While each session built on previous discussions, they were also designed to be standalone events. This structure was meant to encourage conversational threads throughout the series while recognizing that not everyone would be able to attend each session.

The series began by examining funding agencies' interpretations of the DC-DR: Representatives from the Department of Transportation, US Department of Agriculture, and The Bill & Melinda Gates Foundation shared their implementation strategies, compliance requirements, and how they are supporting, or anticipating supporting, researchers in meeting data sharing mandates. This was followed by an exploration of academic institutions' responses to increased data sharing expectations in session two. Presenters detailed infrastructure investments, policy development, and resource allocation strategies and spoke about challenges like staffing constraints, researcher engagement, and balancing institutional priorities with funder requirements. The focus of the third session was on IRs, with an emphasis on their unique role in the data sharing ecosystem. Speakers demonstrated how local repositories provide specialized support for institutional researchers, ensure compliance with local policies, and maintain long-term accessibility of research outputs, such as through integrations with existing research workflows and customization capabilities. The final session examined repository technologies and potential improvements. Topics included system interoperability, metadata standardization, storage solutions for large datasets, and user interface enhancements (Data Curation Network, 2023).

The virtual learning series helped to foster a better understanding of the broader US IR ecosystem, including the challenges and opportunities data stewards anticipate in the coming years, and enabled community connection. A key theme throughout the event series was the emphasis that a repository is more than a technical platform or even a combination of platforms. In order to be successful in meeting the needs of researchers, there need to be people, processes, and policies in place to support the technology and surrounding services. In other words, while the platform is important, the social and intellectual infrastructure is essential.

Summit for Academic Institutional Readiness in Data Sharing

The virtual learning series laid the foundation for the next phase of our project, in which we developed and hosted a summit to engage institutionally based data service providers. While the virtual learning series provided a solid foundational understanding of the current state of institutionally based data services, we were limited, due to the webinar format, in our ability to engage with service providers and to dig deeper into their approaches, strategies, and operations. We recognized that we would need to bring stakeholders from across institutions together to engage on the themes that we identified from the virtual learning series to get a representative understanding about the current state of the IR landscape in supporting research data. While there are similar efforts in the United States, such as HELIOS Open, we wanted to shift the narrative from high-level administrative perspectives to operational. In particular, we really wanted to explore where and how we could build up connections across institutional stakeholders and

develop a stronger sense of community among institutions. We also wanted to include institutional representatives outside of the libraries, including those in Offices of Research or IT units, to create or enhance intra-institutional relationships.

We crafted our approach for holding the Summit for Academic Institutional Readiness in Data Sharing (STAIRS) in late spring of 2024. Strong attendance in the virtual learning series led us to believe that demand for the summit would be high. We wanted STAIRS to be more representative of the full range of institutions offering data services, from those just launching data services to those with more mature support and everywhere in between. To accomplish this, we developed an application process that would enable us to gauge the current state of services, support, and resources of the institution as well as their development trajectory and anticipated investments. The application process consisted of a series of self-assessment questions for applicants to indicate their maturity level in providing support across the data lifecycle and two short answer questions: "How will attending this summit impact research data sharing at your institution?" and "How would your experiences and knowledge impact and bring value to other attendees?" Applicants were asked to respond to our questions from the perspective of their institution rather than as an individual or for their unit. Applicants were also asked to identify up to three people who would attend STAIRS and strongly encouraged to include personnel from the library, IT, Office of Research, and/or a Research Center or Institute in building their team. Applications were reviewed by no fewer than three team members against a publicly available rubric. We received 61 institutional applications and were able to accept 32 of them to attend STAIRS. Following here is a summary of the STAIRS event including the pre-event survey. Details about the event, discussion themes, and recommendations are covered in our final report which was published by the University of Minnesota (Narlock et al., 2024a).

Pre-summit Survey

After confirming attendance from the 32 institutions admitted into STAIRS, we developed a pre-summit survey to gather additional information about the state of data services and support at each institution. We used the information collected from the survey to structure the agenda, inform the development of the presentations, and determine the discussions of the summit. We asked that each of the institutions accepted into the STAIRS program submit a single response to the survey rather than responding as individual attendees. As with the application process, our intention was that structuring responses in this way would lead to initial conversations across units before they arrived at STAIRS. All of the 32 accepted institutions completed the survey for a response rate of 100%.

We adopted the list of "Public Access Data Management and Sharing (DMS) Activities (v3)" developed by the Realities of Academic Data Sharing (RADS) project as a means to structure the survey. The purpose of the Public Access DMS activities is to "describe what researchers and administrators at institutions that support research data sharing may need to consider when managing and sharing data to comply with funder or other data sharing and public access policies", which nicely dovetailed with our goals for this survey (Kozlowski et al., 2023). The activities are loosely organized chronologically across a generic data lifecycle and categorized into five phases:

- Planning, Design, and Start Up of Projects
- Data Collection, Storage, and Management
- Making Data Broadly Available
- Data Retention, including Preservation, Archive, and Long-Term Access
- Project Closeout and Compliance

Each of the activities in this list was rewritten as a possible service that could be offered by an institution. Survey respondents were asked if their institution was currently offering each of the

possible services, if they were planning on offering the service in the next year or two, or if they were not planning on offering the service. If the respondent selected that they were offering the service, or planned to, they were taken to an additional screen and asked how developed the service was currently (fully developed, planned or in progress, or not yet in progress). Finally, for the services that they were currently offering or planned to offer, respondents were asked who in the institution was involved or would be involved: Libraries, Office of Research, IT or Academic Research Computing, Campus Research Centre, or someone else. Respondents could select more than one option from this list to indicate that multiple types of people were involved in offering the service.

The results of the survey were used to structure the content and activities at the STAIRS summit. Most of the 32 institutions which attended STAIRS were currently offering, or planned to offer, some kind of services or support at every stage of the data lifecycle. In aggregate, services in the "Planning, Design, and Start Up of Projects" phase were the most fully developed, with room for further investment in budgeting and policy compliance planning. The "Making Data Broadly Available" portion of the lifecycle was also an area where institutions were investing time and resources, though more institutions were in the progress of developing or planning their services than in the "Planning..." phase. Fewer institutions offered or planned to offer support in the "Data Collection, Storage and Management" stage in comparison, though there were particular services that were receiving higher levels of attention, such as support for managing in-progress datasets and recommending data analysis tools. Respondents also indicated a notable investment in developing institutionally-based training and educational programs. Services in the latter two phases "Data Retention" and "Project Closeout and Compliance" were more likely to be in the partially implemented or planning phases than services in the other phases and fewer institutions said that they were doing or considering services there, suggesting future opportunities for development.

The results of the survey also provided insights into which institutional stakeholders are offering or planning to offer support across the data lifecycle. As shown in the charts below:

- Librarians are particularly engaged in the "Planning..." phase, though IT is the primary agency for consultations on data storage and security issues, and personnel from the Office of Research take the lead in consultations on budget and compliance issues. (Figure 1)

- Support is more balanced across the four stakeholder types in the "Collections..." phase. (Figure 2)

- Librarians are more heavily engaged in services in the "Making Data Broadly Available" phase on average, with the exception of compliance checks, which falls to the Office of Research when available. (Figure 3)

- Libraries and IT are the primary agencies involved in providing services in the "Data Retention and Preservation" stage, with the exception of "Security", which is offered through IT or the Office of Research. (Figure 4)

- Finally, the Office of Research is the primary agency involved in providing support at the "Project Closeout" phase, with support from Libraries in providing training. (Figure 5)

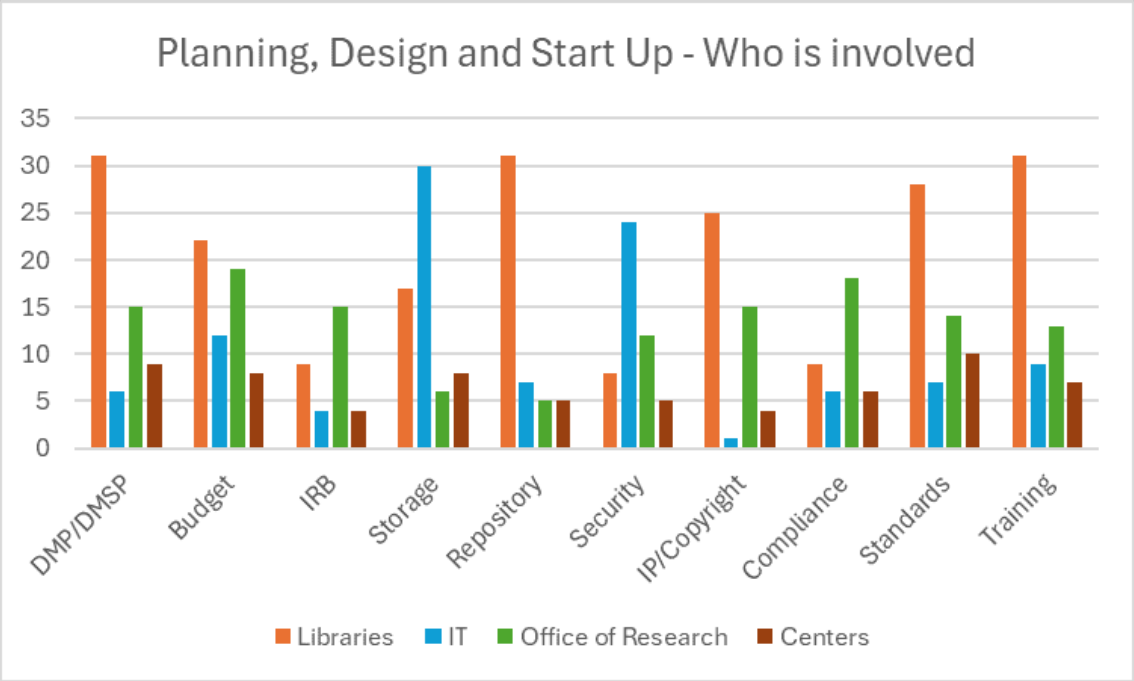

Figure 1. STAIRS Pre-summit survey responses to which units are involved in different activities of the "Planning, Design, and Start Up" phase.

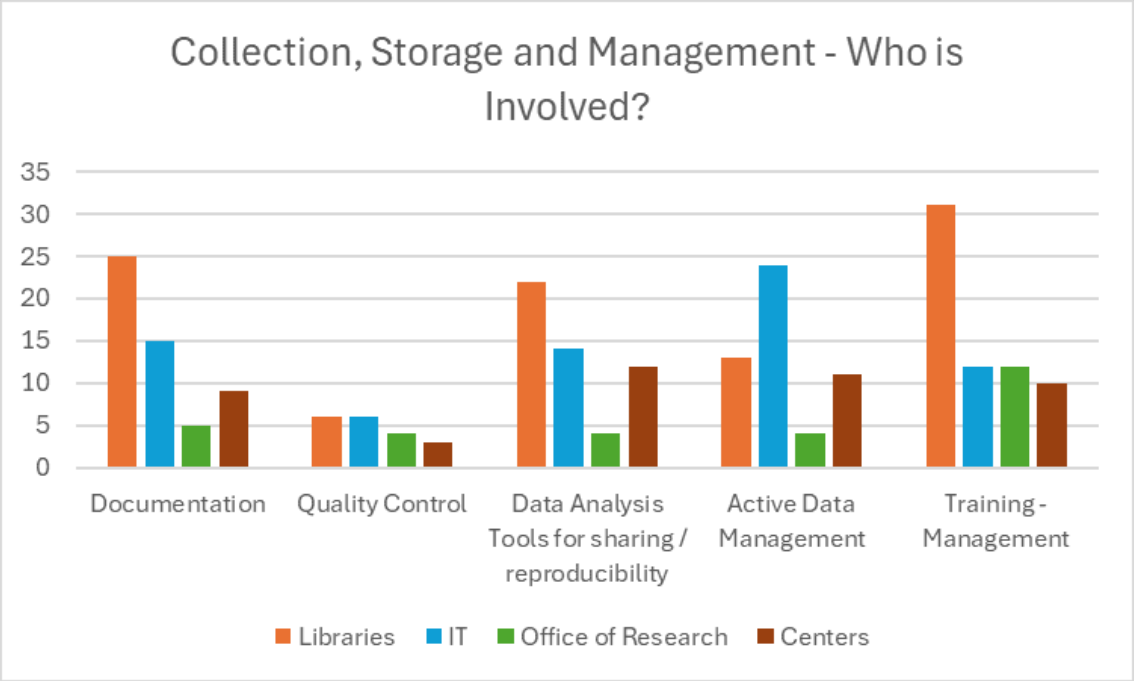

Figure 2. STAIRS Pre-summit survey responses to which units are involved in different activities of the "Collection, Storage, and Management" phase.

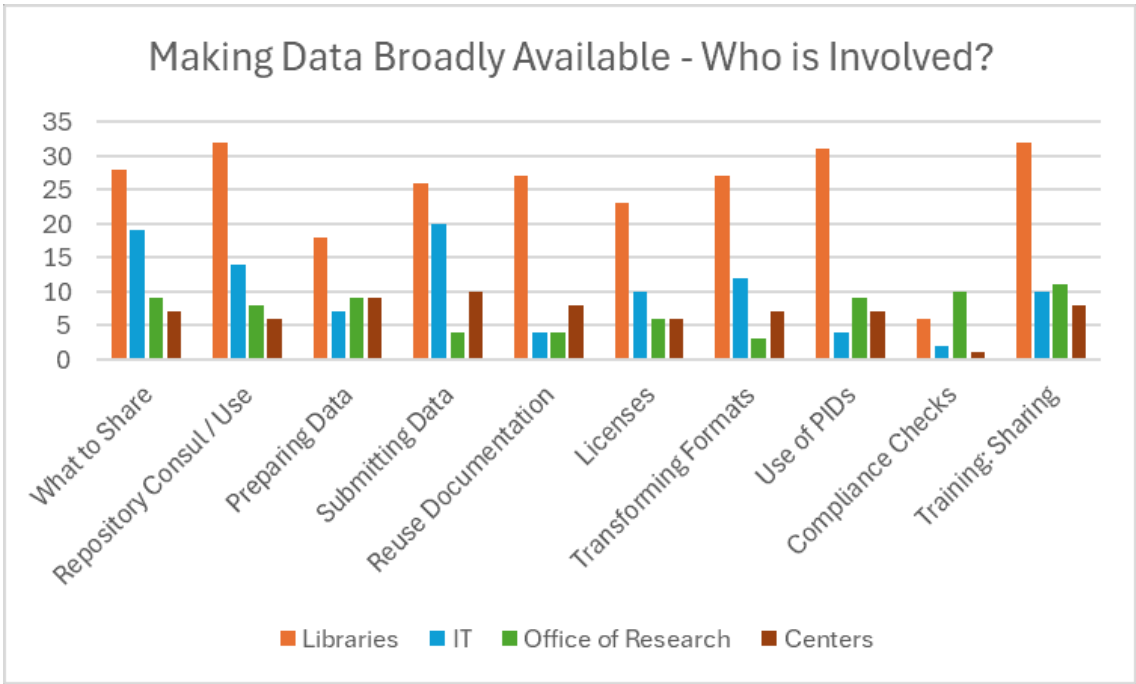

Figure 3. STAIRS Pre-summit survey responses to which units are involved in different activities of the "Making Data Broadly Available" phase.

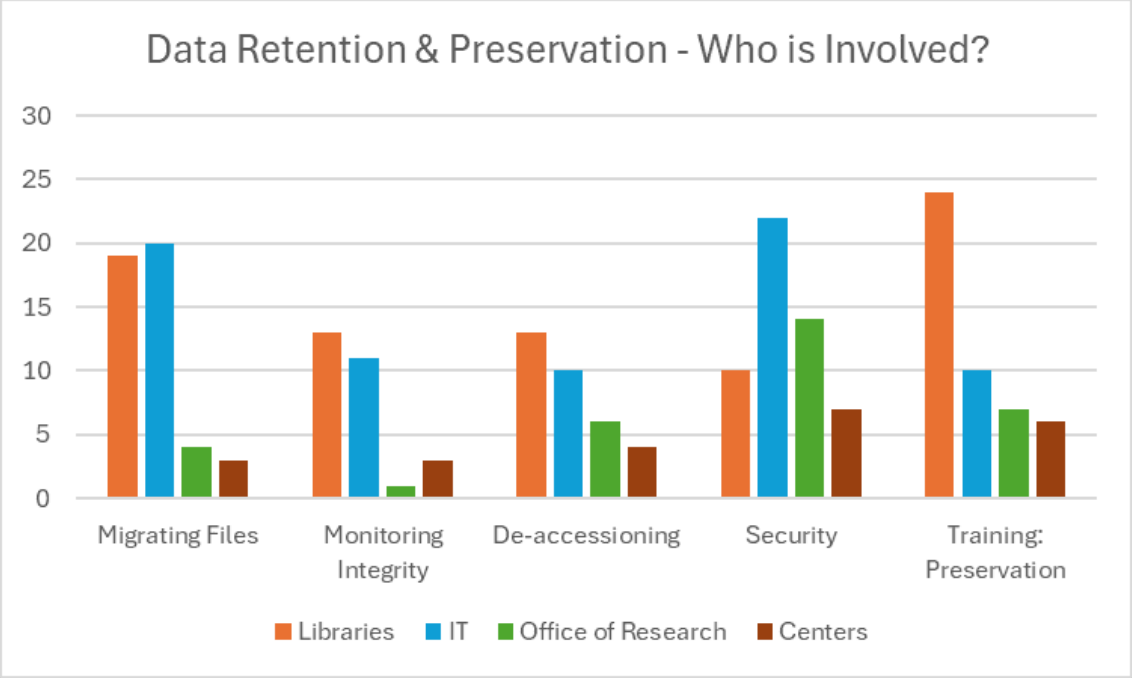

Figure 4. STAIRS Pre-summit survey responses to which units are involved in different activities of the "Data Retention & Preservation" phase.

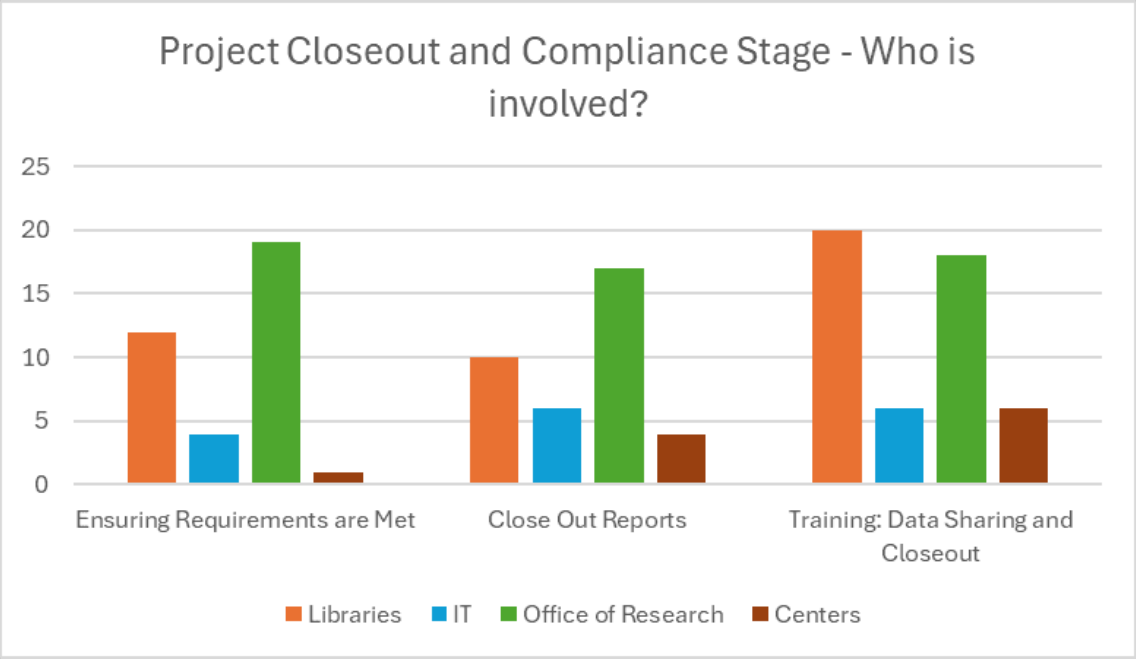

Figure 5. STAIRS Pre-summit survey responses to which units are involved in different activities of the "Project Closeout and Compliance" phase.

Finally, institutions were prompted at various portions of the pre-summit survey to provide open-ended feedback. Through this, we gained invaluable insight about institutional service provision across STAIRS attendees' institutions. First, not all services provided were necessarily available to everyone at their institution: Some services were offered by offices that are only dedicated or provisioned to serve a part of the institution, such as the medical school. Second, not all of the support provided at an institution is offered formally as a service but may be given on an ad hoc basis or due to a special relationship that exists between particular groups. Third, institutional governance, policies, and workflows for managing, sharing, and preserving data are generally underdeveloped, if they exist at all, which leaves a lot of researchers and service providers left to best determine how to proceed without official guidance. These factors and others make it difficult for researchers to know what services and support are actually provided across the institution and by whom, presenting a significant barrier to gaining access to services and resources even when they are offered.

We also saw several concerns expressed in the survey results around how services and support are (or are not) staffed. Some comments indicated that even identifying who in the institution *should* be responsible for providing support can be difficult to determine. Furthermore, separating out researchers' responsibilities versus those of other professionals is still under discussion at some institutions. As demonstrating compliance with funding agency requirements becomes increasingly important for institutions, determining responsibility is crucial. Finally, several respondents expressed serious concerns about offering their services at scale given their current resources. Staffing is a significant impediment that caps the extent to which services can be provided, or limits expanding the service even in response to increasing demand. These survey results strongly suggest additional investments in resources and infrastructure will be needed by institutions to address anticipated demand for services.

The Summit

STAIRS was held in August 2024 at the University of Minnesota-Twin Cities campus with generous support from the National Institutes of Health's Office of Data Science Strategy. Attendees represented a wide variety of institutions, with a range of data service maturity, institutional types, and sizes. Additionally, the 102 attendees spanned research support units and

included libraries (64), information technology departments (11), offices of research (11), and centers and institutes (6). Although attendance was skewed toward library and archival professionals, we were pleased to welcome a wider variety of perspectives and expertise as it led to a richer set of discussions and a stronger sense of connectedness.

Based on the themes of the virtual learning series and the pre-summit survey results, we developed STAIRS around four broad topic areas:

1. Consulting, Training, and Education
2. Technologies, Metadata, and Repository Platforms
3. Building Community Internally Across the Institution, and
4. Building Community Externally.

For each topic area, we began with a presentation to provide some initial context to lay the foundation for discussions. The content of the introductory presentation was based on the results of the pre-summit survey as well as other recent research relevant to the area. Then, three speakers presented a conversation starter—a 5-7 minute presentation designed to raise questions of interest and provoke discussion. Attendees were encouraged to react to what they had heard in the presentations and engage in discussion at their tables. Each table had a "captain" who lightly moderated the discussion and reported out a summary of it to the rest of the room after each session. At the end of the event, we dedicated time for attendees to begin developing an institutional action plan, which included identifying their institutional goals and priorities, what resources they would need to be successful, as well as any potential barriers they might need to consider.

Summit themes

While all the institutions have local challenges, nuances, and strengths, there was a remarkable consistency across discussions by STAIRS attendees that underscore a broader picture of institutionally based research data services in the US. From reviewing each table's discussion notes we were able to identify several themes that were prevalent among attendees.

Scaling up services across the institution

Institutionally based data services are challenged by both internal and external factors. The rapid changes in government policies, publisher requirements, and other pressures to share the data generated from funded research by external agencies has prompted many institutions to figure out how to respond by considering what kind of services and support may be needed for their researchers to be successful. However, datasets come from a variety of disciplines, formats, sizes, etc. and have an array of functions depending upon the field of study, expectations, and the cultures of practice developed by these fields. As all different kinds of research are conducted at an institution, institutionally based data services may encounter a myriad of data—some of which may be unfamiliar to the service providers. Being able to work with and curate any kind of data presents an enormous challenge to data service providers and to institutional repositories.

At individual institutions, data services depend upon the support of the unit that hosts them for resources, infrastructure, and advocacy. There was widespread concern that the services provided are not appropriately provisioned to support increasing demand for assistance, particularly as the expectations of the Nelson Memo (2022) begin to take effect. Overall, STAIRS attendees were hesitant about their ability to scale their services, particularly given that data sharing is a relatively new expectation for many researchers.

This is further complicated by the idiosyncrasies of collaborations. Although nearly every institutional delegation at STAIRS included someone from outside of the library, most of these relationships were still nascent. There was also considerable concern from many that their institution did not have adequate data governance or policies on research data management, sharing, or retention in place. Without mutual understanding between institutional units, it is not

clear who has the authority or responsibility to make decisions at their institution. The need for institutions to develop their social infrastructure to support research data was seen as critically important, perhaps even more so than building technical infrastructure.

Shared resources

There was a strong push for the creation of shared resources to support data service professionals in handling the diversity of data generated at their institutions. Attendees were particularly eager for training materials, templates, and curricula that could be used as a foundation for teaching researchers about data management and sharing concepts and practices, or for skilling up data service professionals. Summit attendees also expressed interest in learning more about the strategies, approaches, and experiences of institutional data service providers as a means of informing their own work. Although these materials could not address the particular needs of the local institution, there was an eagerness on the part of STAIRS attendees to have a common foundation to work from and resources that could serve as a starting point and be modified based on localized needs and perspectives.

Although the appeal of developing and offering these shared resources is readily apparent, the practical aspects of implementation pose challenges of their own. Developing a community-based set of templates, guides, and resources would require coordination and a commitment to review, ingest, update, and maintain materials from a trusted organization. It would also require regular promotion to ensure that the influx of data service professionals are aware of what resources and tools are available to them. Some of the resources requested by STAIRS attendees already exist but were not widely known. For example, the Data Management Training Clearinghouse (DMTC) run by the Earth Science Information Partners (ESIP) and others contains more than 500 resources on teaching research data management and data stewardship ("Learn More About the DMTC", nd). However, even when a community resource like the DMTC exists, it can be difficult to sustain and update it.

Cross-institution collaboration

The Summit provided a key opportunity for attendees to look across institutions. During conversations, it became clear that our organizations encounter many of the same challenges in responding to researchers' needs for managing and sharing their data; however, institutional data services have understandably responded based on local cultures, capacities, and organizational structures. This has led to two interrelated challenges. First, although we know that supporting researchers adequately will require a wide variety of knowledge and expertise from across the university, we lack the strong relationships needed to offer and run services that span organizational units. Moreover, many institutions are in the early stages of articulating where responsibilities in data management, sharing, and retention lie. Second, although we understand the advantages and capabilities that are enabled by building infrastructures and services based on standards and shared practices, service and IR development largely has been conducted within institutional silos, limiting our ability to interoperate effectively and support each other across institutions. Repeated opportunities to connect across both institutions and roles (e.g., IT, Libraries, Office of Research, Campus Centre and Institutes) will be essential for identifying and building strong working relationships to support data management and sharing.

Post-summit Survey

Three months after STAIRS, we distributed a post-summit survey to better understand the impacts of the event. We asked each attendee to consider the most important next steps in developing or refining their institutional data support, what they have been able to advance since STAIRS, and how STAIRS was useful in their efforts. We also asked attendees to indicate what future initiatives they would be most interested in participating in (see Appendix A for survey instrument). In total, we received 40 unique submissions for a response rate of approximately 40%, with approximately 80% of respondents representing Libraries, followed by IT (10%), Office of Research (5%) and Campus Institutes and Centers (5%). We wanted to allow STAIRS

attendees to be in control of the sharing of their institution's plans and thus will not publicly share the individual results of this survey. However, we have provided aggregated findings below.

In response to institutional priorities following STAIRS, many respondents emphasized that additional resources, primarily in terms of personnel, will be essential for their efforts. Additionally, 16 respondents indicated that their next steps were to finalize, continue, or even begin cross-unit conversations and efforts across campus. For the respondents outside of Libraries, there was a particular interest (n=6) in data governance and related policies, demonstrating an awareness from campus partners that there needs to be better clarity between units offering data services. Lastly, 7 respondents flagged the importance of needs assessments and environmental scans to better understand their current offerings and identify gaps for growth.

In response to what actions the attendees have taken since the event, 4 reported that no progress has been made. Of the remaining respondents, 17 indicated that they have been able to continue conversations with STAIRS attendees as well as other campus partners, including the launch of three task forces. In response to whether STAIRS helped attendees prepare for conversations at their institution, 92.5% indicated that the interactions at STAIRS were useful.

Lastly, in looking to the future, respondents were invited to select what future engagement and collaboration opportunities they would like to see. Respondents were allowed to select multiple options. The two most requested future engagement opportunities were annual in-person events and the creation of shared templates (n=31). The least requested were mechanisms for engaging stakeholders outside the institution (n=22) and periodic virtual events (n=24). While there are no immediate plans for future collaborations or funding opportunities, the results of the survey overwhelmingly suggest that there is an appetite for events like STAIRS. One attendee remarked:

> 'STAIRS was an intense and meaningful experience, but it was the start of the conversation and the beginning of work toward institutional readiness. I hope there is indeed more to come.'

Recommendations

Following STAIRS, our team reviewed the extensive collection of shared notes as well as the pre-summit results. Using this data, the STAIRS team drafted a series of recommendations primarily designed to strengthen connections within and across institutions as well as between institutions and funding agencies. Below, we offer a few of the recommendations from our project for funding organizations.

1. Offer incentives and opportunities for institutional data service providers and partners to engage and learn from each other. This includes defining shared standards, creating templates and resources for reuse, and structures for collaboration across institutions.

2. Foster connections between program officers and institution-based data service providers to ensure researchers receive robust support and consistent guidance throughout the research lifecycle.

3. Fund additional research into data sharing requirements and services at academic institutions to better understand challenges and opportunities across different institutional types.

While there were several limitations to STAIRS, detailed in full in our final report (Narlock et al., 2024a), these recommendations will likely resonate with individuals across the US and beyond. The STAIRS event provided the first opportunity for many individuals at the operational level to come together, across units and institutions, to discuss challenges and potentials. There are further means through which this approach could be refined, including focusing on different institutional types, repository infrastructure, and data services offered. We are hopeful these conversations will continue, informally and formally, to strengthen the landscape of US academic institutional data service providers.

Conclusion

What began as a discussion on applying the DC-DR to IRs and data IRs among members of the DCN turned into a much larger exploration of the state of institutionally based support for managing and sharing data. The STAIRS event was an opportunity to ascertain the current state and needs of institutionally based data services and institutional repositories as they support data management and sharing, and to capture a snapshot in time. Although libraries have often taken the lead in considering researchers' needs and developing services to address them, the expected increase in demand and the need for more comprehensive support requires a broader investment by institutions and greater involvement from other units. The Summit demonstrated a real need to consider how the policies developed by funding agencies are impacting institutions and the importance of not just local technical infrastructure, but the social infrastructure that will be needed to ensure that data are managed and shared at scale. The event led to three main recommendations for funding agencies, as described directly above. Feedback from attendees both during and after STAIRS indicates an eagerness for stakeholders to work together as a community in developing common understandings, shared resources, and a sense of connectedness in doing this work. Going forward we seek to leverage what we have learned from the Repository Readiness virtual learning series and the STAIRS summit to continue to foster community development across a range of institutional service providers, as well as motivate intra-institutional collaborations.

Acknowledgements

This work was supported in part by an award from the National Institute of Health's Office of Data Science Strategy in collaboration with MITRE.

References

Branin, J. J. (2005). Institutional repositories. In M. A. Drake (Ed.), *Encyclopedia of library and information science* (pp. 237–248). Taylor & Francis Group.

Data Curation Network. (2022) "DCN is Ready to Support Policies Resulting from OSTP Public Access Memo." DCN Blog. https://datacurationnetwork.org/2022/09/06/dcn-is-ready-to-support-policies-resulting-from-ostp-public-access-memo/

Data Curation Network. (2023) "Recap: Sustaining Open Research: Virtual Learning Series." DCN Blog. https://datacurationnetwork.org/2023/10/23/recap-sustaining-open-research-virtual-learning-series/

Johnston, L.R., Hofelich Mohr, A., Herndon, J., Taylor, S., Carlson, J., Ge, L., Moore, J., Petters, J., Kozlowski, W., & Hudson Vitale, C. (2024) Seek and you may (not) find: A multi-institutional analysis of where research data are shared. *PLOS ONE 19*(6): e0302426. https://doi.org/10.1371/journal.pone.0302426

Johnston, L.R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., Stewart, C., Blake, M., Herndon, J., McGeary, T.M., Hull, E., and Coburn, E. (2018). Data curation network: A cross-institutional staffing model for curating research data. *International Journal of Digital Curation, 13*(1), pp.125-140. https://doi.org/10.2218/ijdc.v13i1.616.

Kozlowski, W., Herndon, J., Taylor, S., Hudson Vitale, C., Carlson, J., Hofelich Mohr, A., Moore, J., Narlock, M., and Petters, J. (2023) "Realities of Academic Data Sharing (RADS) Initiative Public-Access Data Management and Sharing (DMS) Activities," v3. Washington, DC: Association of Research Libraries, November 6, 2023. https://doi.org/10.29242/radsdmsactivities2023.

"Learn More About the DMTC." (nd) Data Management and Training Clearing House. https://dmtclearinghouse.esipfed.org/about

Lynch, C. A. (2003). Institutional repositories: essential infrastructure for scholarship in the digital age. *portal: Libraries and the Academy, 3*(2), 327-336. https://doi.org/10.1353/pla.2003.0039

Narlock, M., Carlson, J. Cowles, W., Delserone, L., Herndon, J., Kozlowski, W., and Petters, J. (2024a). Institutionally based research data services: Current developments and future direction: A report from the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). University of Minnesota Libraries Publishing Services. https://doi.org/10.24926/2024.70098

Narlock, M., Taylor, S., Calvert, S., Priesman Marquez, R., and Parkman, A. (2024b) Knowledge Infrastructures Are Growing Up: The Case for Institutional (Data) Repositories 10 Years After the Holdren Memo. *Data Science Journal 23*(1), 46. https://doi.org/10.5334/dsj-2024-046

Nelson, A. (2022, August 25). Memorandum for the heads of executive departments and agencies: Ensuring free, immediate, and equitable access to federally funded research. http://bidenwhitehouse.archives.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

Reiff Conell, S., & Wright, S. (2024). Desirable Characteristics and Trust in Repositories: A Cross-institutional Comparison. 18th International Digital Curation Conference (IDCC24), Edinburgh. Zenodo. https://doi.org/10.5281/zenodo.10728405

Tenopir, C., Kaufman, J., Sandusky, R.J. & Pollock, D. (2019). The Time Has Come... To Talk About Why Research Data Management Isn't Easy. In *Proceedings of the Charleston Library Conference.* http://dx.doi.org/10.5703/1288284317185.

White House Office of Science and Technology Policy (OSTP). (2022). Desirable characteristics of data repositories for federally funded research. Executive Office of the President of the United States. https://doi.org/10.5479/10088/113528

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J. Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data, 3*(1), 1-9. https://doi.org/10.1038/sdata.2016.18

Appendix A: Post-Summit Survey

STAIRS 2024 Post Summit Survey distributed through Qualtrics, managed by the University of Minnesota.

Start of Block: Default Question Block

Q1 Thank you for attending STAIRS in August 2024. We are grateful for your participation and candor, which made the summit a success. We want to better understand what effect this event had on your institutional practice, policies, or development direction. Individual answers will be kept within the project team, but anonymized quotes may be shared. Aggregated answers may also be shared via reports, presentations, and grant proposals.

We are asking **every individual** to respond to this survey. We expect the survey will take you approximately 15 minutes to complete.

If you have any questions, please let us know by contacting Jake (jakecarl@buffalo.edu) or Mikala (mnarlock@umn.edu).

Page Break

Q2 **Section 1: Demographic Information** We are collecting this information to compare with pre-summit responses. We will aggregate responses here, and throughout, with other similar organizations.

Q3 Individual name

_____

Q4 Institution name

_____

Q5 Select the unit types that best describes your current position

- ☐ Information Technology (IT) and Computing (1)
- ☐ Libraries (2)
- ☐ Office of Research (3)
- ☐ Campus Institutes and Centres (4)

Page Break

Q6 Section 2: Impact of STAIRS Locally
In this set of questions, we ask you to consider the impact of STAIRS on your institution since the event. If your institution created an action plan at the summit, it may be useful to revisit the plan, which may be available in the STAIRS Attendee folder.

Q7 Having attended STAIRS, what do you see as the most important next steps in developing your institution's data services?

_____
_____
_____
_____
_____

Q8 Regarding creating or sustaining RDM services, what have you been able to move forward on since STAIRS (including thinking / planning)? To what extent are different units involved in these efforts and conversations?

_____
_____
_____
_____
_____

Q9 Do you feel like the interactions at STAIRS helped prepare you for similar interactions and conversations at your institution?

- ○ Definitely not (17)
- ○ Probably not (18)
- ○ Might or might not (19)
- ○ Probably yes (20)
- ○ Definitely yes (21)

Q10 Following the STAIRS event have you continued conversations with representatives from other institutions?

- ○ Yes (5)
- ○ Not yet but plan to (4)
- ○ No (6)

Q11 What types of conversations have you had?

_____
_____
_____
_____
_____

Page Break

Q12 Section 3: Future Efforts (STAIRS or Otherwise) As we conclude this effort, we want to consider future engagement opportunities that may be valuable to the community.

Q13 Of the following, what do you think are most needed to support institutionally based research data services? Please provide additional information.

- ☐ Annual in-person events like STAIRS (1)

_____

- ☐ Periodic virtual events like STAIRS (5)

_____

- ☐ Shared standards and norms (2)

_____

☐ Shared tools and resources (6)

_____

☐ Templates (e.g., README template) (3)

_____

☐ Mechanisms for fostering engagement with other stakeholders within institution (e.g., libraries, IT, office of research) (7)

_____

☐ Mechanisms for fostering engagement with other stakeholders outside of the institution (e.g., funders, publishers, etc.) (8) _____

☐ Other (4) _____

Q14 What communication or marketing channels would be most effective for future efforts?

☐ LinkedIn (1)

☐ CARCC (2)

☐ Educause (3)

☐ Other (4) _____

Q15 If we were to offer a similar event in the future, would you be willing to contribute?

o Yes (1)

o Maybe (2)

o No (3)

Q16 Is there anything else you'd like to share with the organizers?

_____

_____

_____

_____

_____

Q17 This concludes the survey. Thank you for your time. Clicking "Next" in the bottom right will submit your responses.

End of Block: Default Question Block