Institutionally based research data services: Current developments and future direction

A report from the Summit for Academic Institutional Readiness in Data Sharing (STAIRS)

DATA CURATION NETWORK

DATA
CURATION
NETWORK

Institutionally based research data services: Current developments and future direction

A report from the Summit for Academic Institutional Readiness in Data Sharing (STAIRS)

Mikala Narlock, Jake Carlson, Wind Cowles, Leslie Delserone, Joel Herndon, Wendy Kozlowski, and Jonathan Petters

Acknowledgements

The authors and the STAIRS attendees are grateful for the financial and professional support of the following organizations and individuals:

- The National Institutes of Health, the National Library of Medicine, and MITRE, including Ishwar Chandramouliswaran, Andrew Hasley, Lisa Federer, David Tanenbaum, and Eldred Ribeiro.
- The University of Minnesota Liberal Arts Technologies and Innovation Services, for student assistance with data collection from ARL institutions.
- The University of Minnesota Libraries for their support of this event and project, including:
  - Logistics support from: Shaan Hamilton, Emma Molls, Khanh Tran, Katie Miller, Adelaide Paris, Dean Walczak, and Jody Gray.
  - Conversation Captains: Carol Kussmann, Kate Sheridan, Andy Herzog, Wanda Marsolek, Sarah Barsness, Kent Gerber, and Cody Hennesy.
- Dr. Kristi Holmes (Northwestern University), for her help crafting and refining the initial project proposal.
- Lisa Johnston (University of Wisconsin) and Cynthia Hudson Vitale (Johns Hopkins University), for their roles in the conceptualization of the project.
- The speakers at the 2023 Virtual Learning Series:
  - Session 1: Peter Arbuckle, Leighton Christiansen, and Ashley Farley
  - Session 2: Lisa Johnston, Wind Cowles, and Cynthia Hudson Vitale
  - Session 3: Heidi Imker, Susan Ivey, Moira Downey, and Tina Baich
  - Session 4: Amy Nurnberger, Sara Mannheimer, Kristi Holmes, and Wendy Kozlowski

The project team, as always, would also like to thank the members of the DCN, past, present, and future, for their constant support. Their willingness to share trials and tribulations freely inspired this event, as it is through the support of one another that we make our work better.

ISBN: 978-1-959870-10-4
E-ISBN: 978-1-959870-09-8
DOI: https://doi.org/10.24926/2024.70098

vii **Executive Summary**

1 **Introduction**
1 Background
4 Project description

6 **Preliminary assessment work**
6 Data sharing infrastructure growth in ARL libraries
8 Virtual learning series

12 **Summit for Academic Institutional Readinessin Data Sharing (STAIRS)**
13 Application process
15 Agenda development
16 Summit preparation
17 Summit structure

20 **Summit Conversations & Themes**
20 Broad topic conversations

26 **Overarching Summit Themes**
29 Post-summit
31 Recommendations
35 Conclusion

37 **References**

40 **Appendix A: STAIRS Application**

44 **Appendix B: STAIRS Application Rubric**

45 **Appendix C: STAIRS Pre-Summit Survey**

60 **Appendix D: Pre-Summit Response Visualizations**

63 **Appendix E: Full STAIRS Agenda**

67 **Appendix F: Conversation Captain Information**

69 **Appendix G: Template Institutional Action Plan**

70 **Appendix H: List of attending institutions**

71 **Appendix I: List of acronyms**

EXECUTIVE SUMMARY

The Summit for Academic Institutional Readiness in Data Sharing (STAIRS) was a multi-phased project that brought together a diverse group of representatives from academic institutions across the United States who support research data sharing efforts. Building off preliminary assessment work and a virtual learning series, this was a unique chance to discuss the opportunities and challenges in supporting researchers' data sharing needs within and across institutions. The centerpiece of STAIRS was an in-person gathering of more than 100 people from 32 different institutions who came together to discuss the current landscape of institutionally based data services, identify emerging best practices, tools and resources, and develop stronger connections and community.

The key themes that emerged from the summit provide a clear roadmap for continued investment in institutional support for research data sharing. The need for sustained resource investment in personnel, infrastructure, and shared services was a consistent refrain. Attendees emphasized the value of cross-institutional collaboration to maximize the impact of these resources, in particular through the sharing of templates, educational services, technologies, strategies, and best practices. Institutions also need to improve institutional data management and governance policies to increase transparency and compliance across their respective organizations.

Funding agencies and other research partners can play a pivotal role in supporting this work. In collaboration with academic institutions, funding agencies should consider further incentives to engage researchers in data sharing. Continued opportunities for community-driven conversations, the development of shared tools

and resources, and investments in relevant research were all highlighted as important next steps. By building on the momentum and relationships forged at STAIRS, the academic community can collectively strengthen the infrastructure necessary to meet researchers' evolving data sharing needs.

Based on our analysis of these themes, we have the following recommendations:

- Support additional opportunities for institutional data service providers and stakeholders to engage and learn from each other.
- Encourage institutional data service providers and stakeholders to build shared standards, norms, and structures across institutions.
- Strengthen communication and connections between program officers and other key stakeholders in funding agencies with institutionally based data service providers.
- Develop or maintain communal spaces for shared resources, templates, and case studies that are well-known and can be accessed, modified, and reused.
- Further define, articulate, and promote the role of institutionally based data services across disciplines and funding agencies.
- Invest in research relevant to understanding and responding to data sharing requirements at academic institutions.

Overall, STAIRS demonstrated the eagerness and dedication of academic institutions to rise to the challenge of ensuring open, equitable, and sustainable research data management and sharing. With strategic support and collaborative effort, academic institutions are well-positioned to continue to play a vital role in this domain. The STAIRS team looks forward to ongoing engagement and further productive discussions that build upon the insights gained at the summit.

INTRODUCTION

Background

Federal funding agencies and publishers, among others, increasingly expect that research data, code, and other scholarly outputs will be managed and shared in alignment with the FAIR Data Principles, which articulate that data should be findable, accessible, interoperable, and reusable (Wilkinson et al., 2016). The expectation that recipients of federal research awards will share their data took shape early in the 21st century with the 2003 announcement from the National Institutes of Health (NIH) that submissions for awards over $500,000 in direct costs must include a data sharing plan, detailing how the data would be made available to others outside of the project. The National Science Foundation (NSF) followed suit in 2011 with a stronger requirement that all funding submissions must include a two-page data management plan, which should include what data would be produced and how it would be shared. In 2023, the White House Office of Science and Technology Policy released the memo, "Increasing Access to the Results of Federally Funded Scientific Research," hereafter referred to as the *Holdren Memo*. This memo directed every federal agency with more than $100 million in annual research and development expenditures to develop a plan to support public access to the resulting products, including research data (Holdren, 2013).

In response, many US academic institutions recognized the need to support researchers in managing, preserving, and sharing their research data and related outputs by providing key infrastructure, including personnel for consultations, repositories for publishing

data, and redundant storage for preservation. As primary owners and stewards of the research data generated by their faculty, US academic institutions have the responsibility to meet their researchers' data sharing needs while securing the preservation of and access to these invaluable research materials.

Expectations surrounding making research data publicly available have continued to advance and become more stringent. In 2022, the National Science and Technology Council released a document titled *Desirable Characteristics of Data Repositories for Federally Funded Research* (DC-DR; National Science and Technology Council), which describes what services and functionalities a data repository should provide in order to be recommended to researchers as an appropriate means for sharing their data. A memorandum from the Office of Science, Technology, and Policy, *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research*, hereafter referred to as the *Nelson Memo*, was also released in 2022. The *Nelson Memo* went further than the *Holdren Memo* by requiring all federal agencies to come up with plans for sharing data immediately upon publication of the results, regardless of the agencies' amount of research expenditures (Nelson, 2022). And in 2023, the NIH extended their data sharing requirement to apply to most applications for extramural funding.

Academic institutions in the US are now tasked with considering how to extend their infrastructure to support a wide range of data sharing needs from different disciplines. Of critical importance in this effort to scale infrastructure will be connections to the global research ecosystem, manifested through alignment with community standards and connections to external systems that enable broader access.

Data repositories serve as an essential component of the emerging infrastructure that is needed for sharing, stewarding, and preserving research data at scale. However, the landscape of data repositories lacks a coherent organized structure, having developed and evolved in ad hoc and idiosyncratic ways. Some disciplines have well established, organized, and sustainable repositories specific to their scholarly domain, such as the Inter-university Consortium for Political and Social Research (ICPSR) for social, behavioral, and political science researchers. Some disciplinary repositories are maintained by grant funding or other small communities of practice, and are therefore more likely to be vulnerable (Kaiser, 2016; Strecker et al., 2023).

Other repositories are defined as generalist repositories, meaning they accept a wide variety of data formats and data types, covering a range of disciplines (e.g., Zenodo, Dryad, and figshare).

In the absence of widespread domain repositories, as well as uncertain long-term access to domain repositories (Strecker et al., 2023), many academic libraries stepped in to fill this gap by developing institutional repositories (IRs) to meet the research sharing needs of researchers affiliated with their host institution. For this report, IRs include those created for sharing traditional scholarly work as well as those purpose-built to support data sharing. These IRs serve as critical infrastructure in enabling academic libraries, and by extension their host institutions, to provide the support and services needed by researchers to meet the data management and sharing requirements of funding agencies, publishers, and others. Academic libraries have embraced the development and use of repositories as a component of the data services offered to researchers at their institutions (discussed more below under Preliminary Assessment Work).

Given increasing funder expectations of data sharing, as well as increasing demand for infrastructure and services to enable this, academic institutions of all sizes and specializations need to be prepared to meet researcher needs in managing, sharing, and preserving their research data and related outputs. However, institutionally based services and infrastructures have naturally focused on local needs and perspectives. The result has been uneven rates of growth and development across institutionally based data services and repositories, with minimal adoption of shared standards and few opportunities to scale services effectively or address common challenges as a community. To address this, the Data Curation Network (DCN) developed and hosted an exploratory program that culminated with an in-person summit in August 2024. As a network of academic institutions and non-profit data repositories that steward research data for future use, the DCN is shaping the future of data curation services while building community around data curation providers. The network approach, in which members benefit from the expertise of one another, meant the DCN was well positioned to facilitate conversations about research data sharing more broadly (Johnston et al., 2018). The Summit for Academic Institutional Readiness in Data Sharing (STAIRS) brought together representatives from US academic

institutions to discuss current strengths and opportunities, while articulating future research and collaboration possibilities. This program received generous support from the NIH Office of Data Science Strategy through the MITRE Corporation.

Project description

Although IRs have been built according to local needs and specifications, the DC-DR (National Science and Technology Council, 2022) and the *Nelson Memo* represent an opportunity to develop a common set of policies, standards, and practices to better connect data repositories. While both are important documents, neither describes a clear direction for repositories to implement the high-level guidance they provide. Bringing IRs into alignment with the DC-DR is a two-pronged sociotechnical challenge: repositories and other technical infrastructures need to be further developed and refined to effectively support researchers, and institutions need to review and revise their research services, policies, procedures, and staffing levels to ensure compliance with funding requirements.

Our multi-phased initiative sought to engage teams of collaborators to consider how their institutions could best address both aspects of this challenge. The STAIRS team launched this effort in Summer 2023 with preliminary assessment work. In particular, the team conducted a research project in June 2023 to better understand data sharing efforts in academic institutional repositories, and concurrently developed and launched a virtual learning series, running through October 2023, which expanded the focus from alignment with the DC-DR to data services more broadly (see Figure 1).

The culmination of this initiative was the STAIRS event, an in-person summit at which teams of colleagues from across academic institutions collaborated on three shared challenges:

- **OBJECTIVE 1:** Define the current landscape of institutionally based data services (including institutionally based repositories).
- **OBJECTIVE 2:** Identify current and emerging best practices, tools, and resources in data services.
- **OBJECTIVE 3:** Create a sense of connection and community for institutional data service providers, within an institution and across institutions.

Figure 1: Overview of the STAIRS project and different phases.

This report captures the details of the project, including the preliminary assessment work as well as the summit. Following a description of the broad themes and overarching takeaways from this multi-phased effort, we conclude with next steps and future directions for the academic data services community.

PRELIMINARY ASSESSMENT WORK

Data sharing infrastructure growth in ARL libraries

As a first step to documenting the landscape, our team initiated a research project to better understand the number of datasets being shared through institutional repositories. The analysis of this research is presented in the recently published paper, "Knowledge Infrastructures Are Growing Up: The Case for Institutional (Data) Repositories 10 Years After the *Holdren Memo* (Narlock et al., 2024).[1]

As described in this paper, in 2023 we examined the number of repositories and datasets hosted by libraries that are members of the Association of Research Libraries (ARL). In particular, we focused on academic members (i.e., excluded public or government libraries) and sought to answer the questions:

- Does the organization have an institutional repository?
- Does the organization have an institutional data repository (IDR) (i.e., a repository that has been purpose-built for collecting the research data outputs of the university)?
- If yes to either, how many datasets are shared via said repository?

We then combined this 2023 data with data from two previous studies (Hudson Vitale et al., 2017; Johnston & Coburn, 2020) showing the growth of data repositories and hosted datasets over time. The numbers overwhelmingly demonstrate that ARL members are building or adopting standalone data repositories, and that researchers are using institutional infrastructure to publish their research data and make it publicly accessible at increasing rates.

1 The data have been published and are accessible at: Narlock, Mikala R.; Priesman Marquez, Rachel; Herrmann, Heather; & Ibrahim, Maisarah. (2023). Data for "Knowledge Infrastructures Are Growing Up: The Case for Institutional (Data) Repositories 10 Years After the Holdren Memo." Retrieved from the Data Repository for the University of Minnesota (DRUM), https://doi.org/10.13020/w8nk-d131.

As shown in Figure 2, every institution in our 2023 study had library-based infrastructure for data deposit, sharing, and stewardship (n=119). Of these 119 libraries, 54% (n=65) had both an IR and an IDR, while the remaining 46% (n=54) were leveraging their IR for data deposit, sharing, and stewardship.

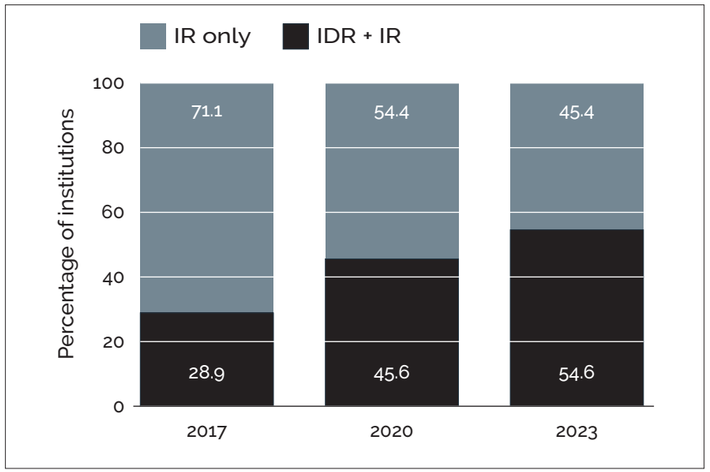

Figure 2: The percentage of ARL member institutions that have an institutional repository (IR) or both an institutional repository and a data-specific repository (IDR + IR in this image), reported as percentages.

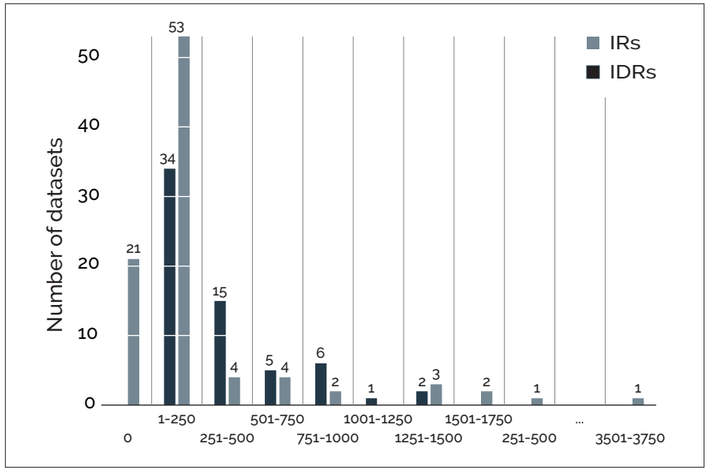

Figure 3: Number of datasets in institutional repositories (IRs) and data-specific repositories (IDRs in this image) in 2023, grouped by range.

While the full growth trends are detailed in the publication, it is notable that all ARL members are stewarding datasets, either in their IR or IDR. Most repositories host between 1–1,000 datasets, with a few outliers storing more. For institutions that have zero datasets in their IR, it is likely due to their data-specific repository hosting the datasets.

These data provide quantitative evidence for what many data librarians knew intuitively: that institutionally managed infrastructure provides critical support for sharing research data, and continues to do so at increasing rates. For this reason, it is essential that

researchers, funders, and other research partners take advantage of the data repository infrastructure provided by academic institutions.

Virtual learning series

Concurrently with this research, our team launched an initiative to better understand the current state of IRs in adopting the DC-DR and in their overall readiness to support data sharing. This initiative was composed of two separate, but interrelated efforts: an internal scan of DCN member alignment with the DC-DR, and an open virtual learning series.

Following the release of the 2022 *Nelson Memo*, the DCN conducted an internal assessment of member repositories against the DC-DR. We sought to better understand where we were in alignment and where there was room for improvement, in particular with an eye towards where we could collaborate to address gaps or amplify our impact via advocacy. At time of the assessment, the DCN had 17 members representing academic and non-profit data repositories. While DCN members are by no means representative of all data repositories, let alone those at academic institutions, the results provided a useful starting point for conversations. Key areas in which our members indicated strengths included characteristics related to digital object management and organizational infrastructure, with the most room for improvement in technical implementations.[2] This knowledge was useful when constructing the virtual learning series: we leaned into our areas of strength to share expertise, and recruited external experts in areas we needed to improve.

With the knowledge that other institutions were conducting similar internal assessments of their repositories, we developed a four-part virtual learning series on "Sustaining Open Research." We initially framed the virtual learning series with a specific focus on the DC-DR and why IRs should collaboratively develop a roadmap to demonstrate alignment with the characteristics; however, we soon realized that there was a need for engagement with the data service community on broader topics around infrastructure, services, and relationships. The series consisted of four 1.5 hour long sessions during the Summer/Fall of 2023 (Data Curation Network, 2023). Each session featured a series of expert speakers. The topics for 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

This series was well attended by librarians, data stewards, and repository managers, as well as by federal funder representatives and researchers, with an average of 188 registrants per event. All events were recorded, but the unstructured question-and-answer section at the end of every event was removed prior to making the recording publicly available. The slides and recordings from all events are available through the DCN collection in the University of Minnesota (UMN) Digital Conservancy (Data Curation Network, nd). All sessions were designed around the DC-DR to some degree to help familiarize participants with these characteristics and lay the foundation for future collaborative efforts. While the sessions built on previous discussions, each was designed to be a standalone event. This was to encourage conversational threads throughout the series while recognizing that not everyone would be able to attend all of them.

Below, we present key themes that emerged across all four sessions. While this is not an exhaustive list, it is indicative of the types of conversations the academic data service community is interested in pursuing further, and will likely be important for gathering future interest and information.

Data services and institutional repositories are more than technology

A key theme throughout the event series was the emphasis that a repository is more than a technical solution or even a combination of technologies. 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. Since initial adoption, and through migrations, some institutions have opted for purchased, or vended, IR solutions,

2 The full dataset is presented in Reiff Conell & Wright (2024).

while others have developed home-grown platforms, drawing from open source communities like DSpace or Samvera. However, regardless of the implementation of the IR infrastructure, personnel need to have the capacity and resources to effectively support researchers. As IRs look for areas to develop, it is of paramount importance that there are sufficient resources and staff to support the growing services, resolve any technical challenges, and adequately respond to changing researcher expectations.

Relatedly, many speakers and attendees emphasized the importance of adopting a platform that fits the resources and needs of the institution. While a home-grown repository may provide more flexibility and customization than a vended solution, the support cost may be too steep for the institution to bear, or perhaps the return on investment would be marginal and insufficient to justify any extraneous expenses. Additionally, there are opportunities for institutions to collaborate on both technical infrastructure (e.g., Samvera, DataVerse) and social infrastructure (e.g., DCN, Data Discovery Collaboration) to reduce the resource need for each institution.

What "Big Data" means for our local contexts

While the term "big data," especially in the era of machine learning and artificial intelligence (AI), is everywhere, including throughout the presentations, questions, and discussions in the learning series, each institution has a slightly different definition of what constitutes "big data." This includes quantifying the collective size of the dataset, the number of files in a submission, or some combination of the two. Many IRs leverage cloud-based storage solutions or tools to help with uploading and downloading big data (e.g., Globus). However questions remain as to how to best curate, provide access to, and preserve such datasets. While it may be difficult, if not impossible, to precisely and uniformly define the term "big data," it is important that the IR community continue to advocate for, create, and adopt tools and practices that will facilitate responsible big data sharing.

Data retention, review, and preservation

Related to discussions of preserving big data is the recognition that IRs will need to be proactive in preserving research data generally. Library-based IRs have been collecting scholarly output since the early 2000s. During that time, IRs have seen increasing deposits from different disciplines, in different formats, and with increasing sizes (see above). However, there are unique challenges to preserving research data. Examples include datasets made by, with, or in obsolete proprietary formats; data that were generated with versions of code or software that are no longer available; and the increasing size and complexity of datasets. Many attendees at the virtual learning series indicated an interest in preserving research data well — which would include retention, review, and even deaccessioning processes. A recent survey of DCN members revealed that many of our institutions want to be more active in reviewing content for deaccessioning, but are unclear about the best criteria for doing so (Luong et al., 2022). IR personnel would benefit from collaboratively developing review and deaccessioning guidelines.

Metadata

Library-based IRs have been built on a few common platforms and technologies designed for sharing and promoting research data. However, the implementation of these platforms and technologies, not to mention the policies, service, and support models that underlie IR operation, are hyper-customized to serve local needs. This siloed approach to developing and operating IRs means IRs sometimes cannot easily scale their work beyond the institution, hampering their ability to support a truly FAIR ecosystem. This is perhaps most evident in our approaches to and uses of metadata schema. While many of us are using a shared standard, such as Dublin Core or DataCite, each implementation comes with customizations. Attendees at our virtual sessions indicated an interest in learning more about how other institutions have mapped different metadata fields, and how our institutions may be able to collaboratively develop implementation guidelines or metadata aggregators to increase the discoverability of our materials.

SUMMIT FOR ACADEMIC INSTITUTIONAL READINESS IN DATA SHARING (STAIRS)

Drawing from the key themes of the virtual learning series as well as the internal scan of DCN repository alignment with the DC-DR, the project team developed plans for an in-person conference. A key strength of the virtual learning series was that attendees were from a variety of professional backgrounds. While librarians represented the majority of attendees, there were also funding agencies, research administrators, and others in the room that could not only hear the conversation and better understand the concerns of academic librarians in meeting the DC-DR, but also provide their expertise and experience for a more rounded conversation.

Our intent for STAIRS was to build on this virtual learning series discussion and leverage the multi-collaborator approach that was critical in its success. We sought to bring together data service providers, IR managers, data curation professionals, and others from across universities who support research data management and sharing. We designed the summit to build up our communities of practice for institutionally based research data services and repositories in academic libraries, identifying common areas of need and exploring ways to strengthen connections between institutions. We also intentionally sought to encourage a diversity of perspectives from organizations of different types and data service maturity.

We had three broad and interconnected goals for STAIRS. First, we sought to define the current landscape of institutionally based data services and repositories. Specifically, we wanted to get a better sense of what resources and services are being offered to address researcher needs in making their data publicly available. We also wanted a better understanding on how and to what extent different organizational units (Libraries, IT, Offices of Research, and Research Centers) were working together to provide support. Second, we wanted to identify current and emerging best practices, tools, and other resources in data services. We were particularly interested in understanding strategies for incorporating these resources into local services, as well as standardizing their use as a larger community of institutionally based service providers. Third, we wanted to strengthen a sense of connection and community for institutionally based data service providers, within and across institutions. How could we learn from each other, and where are there opportunities to work more closely together to address common challenges?

Application process

In order to ensure that there was an opportunity for a diversity of voices in the conversation, the STAIRS team decided to use an application process as opposed to an invitation-based meeting. As a part of this process, the team collaboratively drafted an application and a rubric for scoring to recruit a variety of institutions that are at different levels of data services, organizational capacity, and data sharing expertise. The full application was distributed through UMN's Qualtrics, and is available in Appendix A. Applicants were required to submit one application per institution — ensuring that selected institutions would likely have buy-in from other campus partners to effectively participate in the summit and effect local change afterwards — and to identify those institutional units sending representatives. After basic demographic information, applicants completed two key sections: two brief essays, and a series of self-evaluations against the Realities of Academic Data Sharing (RADS) Data Management and Sharing Activities (Kozlowski et al., 2023).[3]

3 Research team members of the RADS initiative, an NSF and IMLS funded project led by ARL in collaboration with the Data Curation Network, developed the RADS Data Management and Activities to identify expenses associated with a wide-range of data sharing activities across the research lifecycle. These activities were refined with feedback from funded researchers, administrators, and the research data librarian community; although originally developed to identify expenses, these comprehensive activities can also be used to identify institutional services.

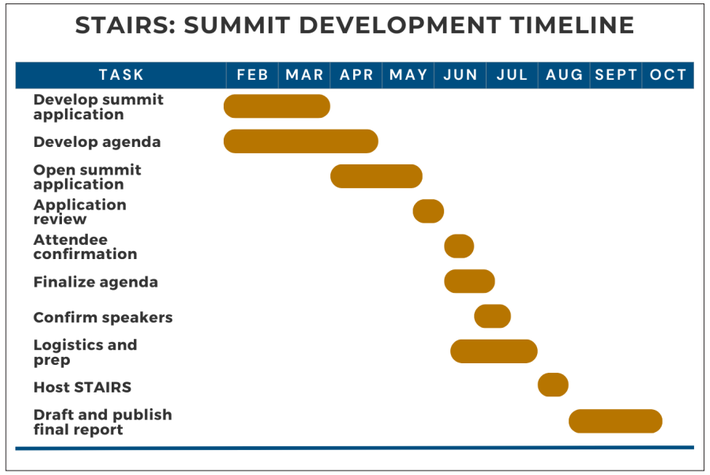

Figure 4: Timeline of the STAIRS application, review, and agenda creation processes.

For the first section,[4] applicants reflected on the impact they hoped attending would have and what expertise they would bring to the summit to benefit other institutions. Responses were evaluated against our publicly available rubric to ensure thoughtful and complete answers. For the self-evaluation portion, responses were given points to help disambiguate different institution's maturity levels. These were automatically totaled to divide applications into categories of institutional maturity. Our goal was to ensure as even a distribution of maturity levels as possible to allow attendees to learn from each other. No further review of applications was conducted to validate self-evaluation responses.

Applications opened on April 10, 2024, and closed on May 17, 2024. The application was distributed to several research data specific and adjacent listservs, including:

- The Research Data Access and Preservation Association (RDAP) Listserv
- The Digital Library Federation Forum Listserv
- The Digital Curation Google Group
- The DataCure Slack Channel
- The US Repository Network
- American Indian Library Association
- The US Agricultural Information Network Listserv
- ARL Day in Review

We also promoted it through a pre-recorded project briefing through the Coalition for Networked Information (CNI; Lynch et al., 2024; Carlson & Narlock, 2024)

We received 62 applications from a wide range of organizations. The application list was then divided in half, with three team members reviewing the first 31 applications, and four team members reviewing the other 31.[5]

After every reviewer completed their scores, the results were averaged, and the top ten institutions of each category were identified. To make room for as many institutions as possible, we asked institutions that were closely affiliated to attend as one unit (e.g., Hofstra University and the Zucker School of Medicine at Hofstra/Northwell and Northwell Health). During the official registration process, we also asked institutions to specify how many representatives would attend, to maximize the number of institutions who could participate. The full list of institutions that attended STAIRS can be found in Appendix H.

The invitees represented different institutional types (public universities, private universities, medical schools, and specialized centers) with a wide range of R&D expenditures in 2023. While these factors were not used to invite institutions, they were useful metrics to confirm representation from a variety of research organizations.

Ultimately, we accepted 104 individuals (including speakers and conversation captains; see below) from 32 institutions to attend STAIRS.

Agenda development

The team intentionally identified topics from the virtual learning series that would be of relevance to individuals across an academic institution, regardless of the institution's size, focus, repository platform (if any), or data service maturity. We drew from feedback on the virtual sessions, information from successful institutional

4 Full rubric available in Appendix B.

5 UMN, the fiscal home of the Data Curation Network and site location of the summit, applied and was accepted. This did not require any travel funding, and therefore did not prohibit any other institutions from attending.

applications, and our own experiences and gaps to create a lengthy set of topics, which were eventually grouped into four broad topic categories:[6]

- Training, Consulting, and Education
- Technologies, Metadata, and Repository Platforms
- Building Community Internally within the Institution
- Building Community Externally across Institutions

This was done not in an attempt to solve all four broad topics in two days, but instead to provide a sampling of current challenges and opportunities to identify areas of collaboration. There were also many additional topics we were unable to cover during the summit, some of which emerged organically during the conversations, others which will need to be addressed in subsequent endeavors (discussed below under Next Steps and Future Directions).

Summit preparation

To help the project design team better understand the needs and perspectives of attending institutions, as well as to prepare attendees for the summit, the team created and distributed a pre-summit survey through UMN Qualtrics.[7] Again drawing from the RADS Data Management and Sharing Activities (Kozlowski et al., 2023), respondents were asked to indicate which services their institution offers, or would like to in the near future, how developed those services are, and which unit(s) on campus offer support for researchers in completing said activity. The survey was only sent to the primary institutional contact, or the individual who applied and confirmed their attendance. Individuals were encouraged to collaborate with their colleagues to fill out the survey, and were sent a separate PDF of the survey to help prepare responses prior to entering the Qualtrics form. Institutional contacts could contact the STAIRS team to receive a PDF copy of their responses. Several contacts noted that the process was helpful for them in better understanding their institutional landscape, and therefore wanted a copy of their answers to help with future planning and cross-campus collaborations.

6 The full agenda is available in Appendix E.

7 The full pre-Summit survey is available in Appendix C.

We received responses from all of the 32 institutions who attended STAIRS, for a 100% response rate. Select aggregate results are shared via visualizations in Appendix D.

Summit structure

To provide structure for attendees, we developed a framework for the summit that was applied to each broad topic. In other words, over the course of two days, attendees followed this agenda four times:

1. Current state of the topic, based on responses to the pre-summit survey.[8]
2. Conversation starters: Similar to lightning talks, presenters were asked to provoke conversation by presenting on recent successes, challenges or gaps, or even future directions.[9]
3. Two rounds of table discussions:
    a. Round 1: Attendees were invited to reflect on the current state as presented as well as what topics, successes, or challenges they would like to bring back to their institution.
    a. Round 2: Attendees were to consider how we could collaborate to advance capabilities in that particular broad topic. At the end of each broad topic, there was time for table report-outs and general reflections.

Attendees from the same institution were encouraged to sit at different tables, but this was not enforced. At the end of the first day, there was time for the attendees to provide feedback to the organizers, which resulted in some minor changes to the structure of the second day. During the final day, STAIRS attendees could nominate a topic for discussion, which would be the focus of the final two discussion sections; five tables ended up topic-based (deaccessioning, de-identification, research data governance, outreach, and data with special considerations) while the rest remained focused on the broad topics. The summit concluded with an hour-long opportunity

8 The "Current State" presentations can be accessed in the complete STAIRS Slide Deck (Narlock et al., 2024).

9 To facilitate knowledge sharing with institutions and individuals that could not attend the Summit, all slides have been archived through the DCN collection in the UMN Digital Conservancy: https://hdl.handle.net/11299/264208

for institutional teams to begin drafting an action plan together (see Appendix G).

To ensure that conversations were captured and recorded in a shared notes document for team members to use after the summit, each table had a designated Conversation Captain (see Appendix F). This individual took notes during the conversations, but would also do more advanced facilitation (e.g., engaging individuals who might not have contributed to the conversation) as needed. We also asked the captains to not hold too tightly to set conversation topics; since research data topics are inextricably linked, it was important to allow attendees to discuss what was coming to mind for them at that moment.

To support note taking and post-summit reflections without capturing specific names or institutions, we asked each attendee to select a lanyard color that most closely corresponded to their current unit. For this, we used the categories of campus units that we defined in the application process (Librarian, IT specialist, Office of Research personnel, or Research Center staff). Attendees were allowed to select more than one, and it was left to their discretion. While we did not ask for this information in the registration form, based on publicly available information, we approximate that in attendance were:

- 64 representatives from Libraries[10]
- 11 representatives from Research Offices
- 13 representatives from IT & Computing
- 6 representatives from campus Institutes & Centers

Attendance leaned heavily towards librarians and archivists — perhaps unsurprisingly, given our primary distribution channels. While we had hoped for a greater diversity in terms of unit representation, we were satisfied that approximately one-third of attendees were not from the library. Our team felt it was important to achieve this diversity for several reasons. First, to ensure cohesion within an institution: if a librarian is suggesting one mechanism, the Office of Research another, and IT a third, not only is it confusing to researchers, but it also duplicates efforts and wastes invaluable resources. Second, we attempted to gather as many perspectives as possible to make our own work stronger by pooling our attendees' respective expertise. And last, we wanted to be sure that any outcomes of the summit could be useful and implementable. Many of us had previously attended conferences or workshops where only one campus unit was represented, which did not support the implementation of cross-campus changes following the event. Our hope for the STAIRS attendees is that their conversations and discoveries during the workshop might result in a deeper connection among individuals from the same institution and greater collaborative action upon returning to their institutions.

10 Categories created by the RADS team. See Hofelich Mohr et al., 2024.

SUMMIT CONVERSATIONS & THEMES

Following the summit, recurrent themes were identified from shared notes documents captured by the Conversation Captains and other note-takers in the room. Identifying information was removed and the notes were manually reviewed by Narlock for themes. The STAIRS organizing team then reviewed the analysis for clarity, organization, and accuracy. Thematic analysis of the workshop notes was assisted by Claude 3.5 Sonnet,[11] an AI language model developed by Anthropic, which helped identify and categorize recurring themes.

Broad topic conversations

In the following section, we briefly summarize the themes that emerged during table discussion sessions, referencing specific conversation starters where appropriate. While these are presented in discrete sections, these conversations often overlapped, with themes repeating throughout the two days. Below, the bolded sentences and phrases are the key themes as identified by the STAIRS organizing team.

Training, consulting, and education

Two of the brief presentations, "Opportunities for education at the institution level" (Wham, 2024) and "How curation can function to educate students and researchers" (Woodbrook, 2024), inspired discussions on ***collaboration and sharing***, both within and across institutions. In particular, the possibility of sharing curricula, templates, best practices, expertise, and even services across institutions resonated with many attendees. For example, the README template developed at Cornell University (Kozlowski, 2024) has been adopted by many institutions offering research data services and can be adapted to suit institutional needs. A similar idea of sharing resources across institutions was to develop a matrix of institutional data services maturity options and levels, which may help to guide early-stage institutions as they build and expand their portfolios to match institutional demand. Along the same lines, many attendees expressed an interest in providing or expanding training and education for both staff and researchers in data management. This is another area where institutions can collaborate to upskill and share resources, as data management training easily transcends institutional boundaries. While mechanisms to collaborate across institutions exist already (e.g., the Carpentries, DCN, the Data Management Clearinghouse) they require an ***investment of time and resources*** to sustain the work.

Nearly every attendee emphasized the need for sustained and increased funding, particularly to support personnel. With growing demands of funding agencies and publishers to enable FAIR data, scaling services up will require a renewed financial commitment to these services from the institution. This tied in with another key topic from this section: how to effectively demonstrate the value of data services to researchers and administration. This was discussed alongside the often invisible work that goes into robust data management and sharing and the need to incentivize data sharing in ways that motivate investment in data sharing beyond compliance.

Technologies, metadata, and repository platforms

Unsurprisingly, a key discussion point for this topic was ***metadata standardization and interoperability***. After a presentation on the challenges of enhancing metadata at a global level (Johnston, 2024; see also Johnston et al., 2024), many attendees spent time discussing the challenges for improving metadata across repositories, and expressed a desire for practical strategies that could be implemented locally. In particular, when considering ***persistent identifiers*** (PIDs) ***beyond Digital Object Identifiers*** (DOIs), attendees emphasized that this would require not only time to develop an institutional strategy, but also both technical personnel to implement and an investment

11 Team members used the public version of Claude 3.5 using anonymized and summarized notes from the event.

from researchers and research administrators to adopt and follow best practices in their application and use. While some institutions are beginning to implement multiple types of persistent identifiers (Cowles, 2024), it will require continued investment at the national and international level to be successful and sustainable.

This led into the next key challenge and opportunity discussed by summit attendees: that "technical challenges" are often more than just technical in nature. For example, even after repositories are developed, researchers may not understand how to use them, or even see value in depositing their data into a repository. Overall, ***researchers need more incentives to share their data***, specifically with an eye towards incentivizing both the development and use of ***local infrastructure and services*** to maximize the impact of organizational investment, ensure institutional compliance with funder mandates, and better handle the nuances of a particular dataset. The push for these incentives has to come externally from funding agencies, as well as internally from promotion and tenure committees.

In this discussion, shared resources again emerged as a key need for attending institutions —including a ***toolkit for communicating the value*** of data management and ***strategies for matching researchers with appropriate repositories*** based on an understanding of their needs. There was also ***cautious excitement about AI*** during this session, including the potential for AI to help in the collection and augmentation of metadata (Lafia, 2024). While it seems clear from the experience of attendees that researchers appreciate the human-in-the-loop during research data management (see also Marsolek et al., 2023), creating AI tools could help facilitate the curation process, reduce the amount of labor required, and augment data interoperability and sharing across institutions.

Building community internally within the institution

During the second day, our conversations revolved around building communities of practice. In terms of what this would look like within a single institution, most attendees wanted more attention on strengthening ***institutional data policies***. In particular, following a presentation about an effort to develop and implement an institution-wide data policy (Herndon, 2024), attendees expressed interest in learning more about the nuts and bolts of this work to adapt for their own institutional contexts, including which stakeholders to include and when. For institutions that have fewer resources, the cost of policy implementation was another concern. Key elements of developing an institutional data policy, which participants expressed both excitement for and apprehension about, were the questions of ownership of the different phases of the research lifecycle and which organizational units in the institution have decision-making authority.

Relatedly, participants wanted additional guidance on ***cross-departmental collaboration***. Building relationships among libraries, IT, research offices, campus centers, and other institutional units is essential for stewarding research data effectively while reducing costs. While several challenges were identified, one in particular was how often collaboration relies on the strength of individual relationships, and therefore turnover can be a significant setback. There was interest in understanding how to build robust collaborations that do not rely on individuals, but are more ingrained in organizational structures. The presentation from North Carolina State University on their research facilitation services, a joint endeavor from their Libraries, Office of Information Technology, and Office of Research and Innovation, provided a great example of these kinds of collaborative organizational structures (Downey & Ivey, 2024).

This further ties into ***leadership and organizational changes***, which many institutions in attendance experienced in recent years. In particular, the ability to navigate transitions and build resilient structures that can allow for continued collaboration is essential. These structures require not only time, but also robust leadership, communication, and marketing to demonstrate the value and impact of data services. Assessment of this work to demonstrate the benefit to the institution is crucial, and will require a combination of quantitative and qualitative approaches. Toolkits for supporting institutions in developing these kinds of collaborative structures would be invaluable.

Lastly, and inextricably linked to the above points, was the discussion around ***cost and resource allocation*** at an institution. This remains a tricky topic that varies at every institution, especially when considering grant funding that includes direct and indirect costs. While research continues to uncover the true costs of making data publicly available (Hofelich Mohr, 2024; Hofelich Mohr et al., 2024), it is absolutely clear that sharing and preserving data is an expensive

endeavor. To be effective at the institutional levels and to ensure their long-term sustainability, the units offering these services need to be appropriately funded.

Building community externally across institutions

In the final phase of the summit, the discussion shifted from an internal focus within our individual institutions to an external one, where we discussed how we might work together across institutional boundaries. Perhaps unsurprisingly, several of the topics that were previously discussed were brought up once again, only from a different perspective. Attendees reflected broadly on ***research data governance***. Many institutions represented at STAIRS expressed an interest in developing institution-wide data policies, and implementing corresponding governance in support of these policies. Inspired by the previous topics, as well as presentations on leveraging new tools and resources such as the DC-DR (Petters, 2024) and the National Institute of Standards and Technology's Research Data Framework (NIST RDaF; Stollar Peters, 2024), there was a clear desire for expert guidance with a practical perspective on structuring institutional bodies for effective data governance. In addition, attendees reiterated the need for outreach and communication efforts targeted towards engaging researchers and promoting institutional data services, especially in light of current or future data governance policies and infrastructure. Given that outreach is a common challenge for all institutions, there was interest in exploring opportunities for collaboration on broad-based communication initiatives to raise the profile of data management and sharing. In other words, there was a desire to collectively amplify the work of data support offered at institutions. Of note, there were several requests for outreach resources that already exist, highlighting gaps in awareness, communication, and maintenance of services that may require additional investigation.

Other key themes that emerged in this session were ***data deaccessioning*** and the ***handling of sensitive data***. With regards to the former, there is a need to consider how long research data will be kept, whether it is public or private, who will pay for long-term storage of the data, and when data will be deaccessioned based on which evaluative criteria — and which units shoulder which responsibilities in this endeavor. There are many variables in this topic as well as current unknown factors; all will need to be watched and addressed in the coming years. Guidance from funding agencies on long-term retention expectations and criteria would be welcomed.

Data with special considerations was already a key topic — in particular, ***managing sensitive data***, such as Personal Health Information (PHI), responding to deidentification requests, and navigating data use agreements. This topic was queued by the work currently happening at Arizona State University with regards to implementing the CARE Principles for Indigenous Data Governance (Carroll et al., 2020) in their research data services (Fernandez, 2024). Attendees expressed many needs in this area that are worth further, and focused, investigation to better understand how funding agencies may support institutions, and by extension researchers, in managing sensitive data.

Lastly, this final session provided space to reflect on the discussions at STAIRS as a whole, to look for connections between topics and themes to consider next steps for this work. For example, we discussed the adoption of PIDs as an area of collaboration: with increased PIDs to augment the findability, accessibility, interoperability and reuse, our repositories can further enable links to other research organizations (RORs), to other materials by the same authors (ORCIDs), to other research projects. This in turn maximizes the reusability of data which can help reduce duplicative efforts and inspire future research projects. Such potential can only be actualized through cross-institutional collaboration that improves the use of PIDs in repositories and metadata; other collaborations could have similarly impactful effects if supported.

OVERARCHING SUMMIT THEMES

As a reflective exercise, and to better understand the next steps for this effort and similar ones, the STAIRS project team synthesized the above themes into broad categories. Our intent is to capture the high-level takeaways from the summit and provide a roadmap for future efforts, which will be discussed in detail in the Next Steps and Future Directions section.

1. **RESOURCE INVESTMENT:** There is a need for investment in personnel, infrastructure, and shared resources across institutions. While such investment primarily needs to happen on a local level, institutions should continue community-building efforts to understand where investments can be made across institutions to maximize impact.
2. **COLLABORATION AND SHARING:** While institutionally based services are naturally shaped by and developed for the local needs of affiliated researchers, we would all benefit from sharing templates, educational services, technologies, strategies and best practices. Perhaps more importantly, these must be shared shared openly and candidly — especially the failures — so we can learn from one another. Collaborations in this space would require coordination and advocacy and would likely need additional funding to be sustainable. Some key challenges that may benefit from cross-institutional collaboration include storage costs, data deaccessioning, and sensitive data considerations.
3. **INSTITUTIONAL RESEARCH DATA MANAGEMENT AND GOVERNANCE:** With the increasing recognition that data are institutional assets, and that researchers can directly benefit from sharing their data, there remains a need to move researchers and administrators past the mandate/compliance mentality to an understanding of the positive results that can come with open scholarship. Alongside this challenge is the opportunity for academic institutions to build out better data stewardship, governance, and de-identification practices.
4. **INCENTIVES:** The importance of incentivizing data sharing and creating value for researchers to use local services emerged, with a recognition that this would need to come from both internal (promotion and tenure committees) and external (funding agencies, publishers) sources to be effective in enticing researchers to make substantial investments in data sharing efforts. This is balanced by the need for compliance efforts from institutions and funding agencies.
5. **TRAINING AND EDUCATION:** There is a national need for training curators, research support staff, researchers, and graduate students on research data management best practices. There is also a need to expand training programs at the institutional level that account for local infrastructures, policies, and services. While many institutions are already taking this on individually, there are the opportunities to maximize impact by developing shared curricula and training materials that can be applied across institutions and adapted for local usage.
6. **TECHNICAL INFRASTRUCTURE:** Investment is needed in modular components of repositories, aggregators, and PIDs to promote connectivity between repositories. Developing workable modular components could also make it easier, more rewarding, and more impactful for institutions to adopt repository technologies, especially those at earlier stages of data service maturity.
7. **AI AND AUTOMATION:** There is cautious excitement about AI's potential in automating data management and curation processes, while recognizing the continued need for healthy skepticism about AI's capabilities and for human intervention in the curation process.

8. **FUNDING AND SUPPORT:** Funder support is critically needed, not just financially but in setting standards, defining procedures for compliance, and providing feedback to researchers. Closer collaboration between program officers and research data managers at institutions could have a larger impact and a greater return on investment of time and funding.
9. **INSTITUTIONAL ADVOCACY:** Nearly every attendee affirmed the importance of communicating the value of data services to campus leadership and demonstrating transparency across institutions.

Across themes and topics, STAIRS attendees emphasized the importance of cross-institutional collaboration as well as a desire to identify clear and well defined starting points. There was an eagerness to get started, a frustration at feeling behind the curve, and a hope of learning from the successes and failures of other institutions. Continued opportunities to openly discuss the different components of research data management, services, administration, and governance would be welcomed by this community.

In addition, STAIRS attendees expressed grand concerns that transcend individual topics or themes. While the latter were useful for grounding conversations, they swirled around larger concerns of data stewardship that will need to be addressed. Making research data publicly accessible is a relatively new concept for many researchers and most administrators, so there are still many questions about how to respond effectively at the scale of the institution. These concerns were particularly acute when considering data governance, and who at institutions has the responsibility and authority to make and implement decisions around data management, retention, etc. It is imperative that institutions identify the people currently involved in the research data lifecycle, who else needs to be involved at each stage, and what investments need to be made in order to be successful. A few attendees expressed concern that, in the pre-summit survey, libraries were represented at every stage of the research process when perhaps they should not be, and when they may even be asked to do more. As was captured in the collective notes document:

> [We are b]eing asked to do more in this space. What is the role for librarians? Teaching python or R? Maybe not. Also need to ask what we don't do and don't offer. Some things are really labor intensive and we are reaching a limit to what librarians can contribute to this space.

Understanding where different units can most effectively contribute would not only allow for better resource allocation overall, and therefore a better return on investment, but also clearer boundaries and more effective working relationships. Interestingly, throughout the summit but particularly when discussing this topic, attendees used the term "existential crisis" — to nods and applause across the room. It is clear that this is a source of uncertainty and frustration to many in the library community, and one without a ready solution.

A final reflection on the summit relates to openly sharing our successes, opportunities, challenges, and setbacks. One attendee noted that they felt constantly behind the curve, lagging behind peers, and thinking that they could be doing more. They shared that, through this summit, they were able to more objectively see where they stood in comparison with their peers: where they were succeeding and where they could improve. Continued national opportunities to share and discuss current states would be appreciated. A few attendees noted that some institutions conduct an annual facilities management survey, and that may translate well to an institutional review of data management that could be shared across institutions to help one another learn and grow. As was captured in the shared notes document:

> The nature of growth is iterative. We always feel like we are behind. But also, we are always somewhere in the cycle of identifying and filling gaps, growing and reflecting. Wherever you are in that cycle is a good place to be.

Post-summit

The team will continue to analyze the themes and topics presented at the summit and how best to support the efforts of our colleagues. Approximately three months from the event, we will be reaching out

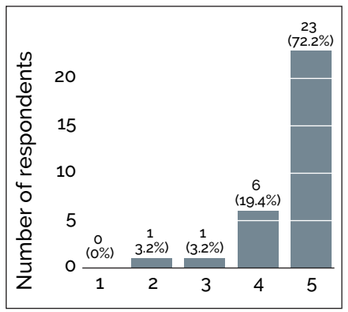

Figure 5: Responses to the question, "Overall, how satisfied were you with the event?" A value of 1 represented not at all satisfied, and 5 represented entirely satisfied.

to STAIRS participants to better understand how the action plans they developed at the summit were useful: what was accomplished, what wasn't, and reasons for both. We hope to better understand how we could support future similar efforts in creating a useful and engaging environment.

We distributed a survey immediately after the summit to better understand how attendees perceived the event. We received largely positive feedback, with suggestions for improving the accessibility of future events. Thirty-one individuals responded to the anonymous form.

- "I had a fantastic time at the STAIRS conference. I'll fill out the feedback form you sent, but wanted to let you know personally I thought it was very well run — great balance of presentations, conversations, and breaks. I liked that you kept the conversation topics consistent so we all knew what to expect. Also, all of the people were fun to talk to and so engaged! I learned a lot and we have a lot of ideas going forward. Thank you for all the work you put into making this happen."
- "Thank you so much for organizing the STAIRS Summit. I appreciate your incredible efforts that went into planning the event and in every way was a success from my vantage point (it was certainly the most productive conference I have attended, and my Office of Research colleague and director, who also attended, have already saved dates for us to continue working on our action plan!)"
- "I found this conference more useful than anything I've been to in a long time. It was great to talk with so many people in similar situations and share our thoughts and ideas on how to improve our institutional readiness for data sharing."
- "The event was fantastic. Even though I am in the research office, I found the information to be quite useful. Look forward to seeing if this summit becomes annual!!"
- "I would like to see more engagement with technology solutions and research ideas if there are future events."

Based on feedback from attendees, we will consider restructuring the format of the event: the Conversation Starters were too short for those newer to the content, and the report outs tended to be somewhat repetitive between tables. Additionally, while many appreciated the opportunity to work with their colleagues on an institutional action plan, there was a desire to work on a community-driven resource, toolkit, or other output after the event.

Overall, it is clear that events like this one fill a critical gap. While meeting in person can be prohibitive in terms of cost and accessibility, attendees valued the chance to gather, discuss, and share freely. We are grateful for the support of NIH and MITRE in making this a reality, and look forward to continued collaboration.

Recommendations

Based on the themes that arose from the STAIRS initiative, we propose the following recommendations to inform future initiatives and investments in advancing institutionally based data services.

Support additional opportunities for institutional data service providers and stakeholders to engage and learn from each other.

Offering a full range of research data services at scale will require the development of new partnerships across the institution and collaborations using new and innovative approaches. Fostering these relationships will require dedicated time and resources to be successful.

STAIRS provided an opportunity for different representatives to come together to learn more about our roles, practices, and perspectives. It was deliberately structured to let participants engage in broad, high-level discussions. The success of the STAIRS initiative as a means of connecting institutional collaborators and building community should be analyzed and replicated to foster deeper connections. Additional opportunities like STAIRS can help build understanding and normalize these important connections.

A weakness of the STAIRS initiative was that it was largely driven by librarians and informed by the library perspective. The summit did not include broad representation from some key perspectives, such as other federal funders or researchers themselves. Future events that focus on community development in this space could benefit from incorporating a wider variety of perspectives.

The STAIRS event was also designed to capture the current state of institutionally based data services at a high-level. There are opportunities to conduct further explorations of specific aspects of data management and sharing, such as training or technical infrastructure development, which could serve as the focus of subsequent initiatives or programs for this community.

Encourage and support additional opportunities for institutional data service providers and collaborators to build shared standards, norms, and structures across institutions.

By their very nature, institutional data services are hyper-focused on local needs. Although they are informed by best practices, their inward focus has led to isolation and duplication of efforts. Initiatives and organizations like the Generalist Repository Ecosystem Initiative (GREI)[12] and the DCN are working to create communities where collaborators can leverage each other's knowledge, expertise, and resources to foster and sustain a sense of community. The institutional data services community would benefit from a similar effort focused on institution-spanning initiatives and partnerships.

Strengthen communication and connections between program officers and other key partners in funding agencies with institutionally based data service providers.

Some STAIRS attendees expressed a lack of confidence that our institutions have a firm understanding of what funding agencies are asking them to do to demonstrate researcher compliance with data sharing mandates. We also heard from NIH representatives that they appreciated the opportunity to connect with the individuals who support data sharing as an operational service. As expectations, administration, and enforcement of the data sharing requirements are still taking shape, there is an opportunity for these groups to connect to share information and learn from each other, which in turn can normalize data sharing practices and expectations, easing the burden on researchers.

Develop or maintain communal spaces for shared resources, templates, and case studies that are well-known and can be accessed, modified, and reused.

One common refrain from the table discussions was the need for a publicly accessible set of resources and tools that would provide a common foundation for supporting research data service providers. Suggestions included tools that could be updated and edited for local nuances, such as education materials or a vocabulary for data management across the research lifecycle. Similarly, attendees expressed interest in tools to self-evaluate institutional data repositories based on the DC-DR as a means to demonstrate institutional infrastructures as viable options for depositing data.

Many of the desired resources, however, have already been developed in some form. The Data Management Clearinghouse[13] provides a catalog of freely available educational materials, and the CODATA RDM Terminology Working Group stewards a glossary of terms relevant to data management work (CODATA, 2024). Further research is needed to better understand the disconnect between these resources and institutional application. While it appeared that people were simply not aware of these tools, particularly people just entering this community, it is also possible that the resources have not been maintained due to funding constraints or personnel turnover. It is also possible that community members know of these resources, but are uncertain about their relevance to institutional perspectives and needs. There is an opportunity to collaborate with resource providers to offer workshops and incentivize adaptations of the tools.

Incentivizing the reuse of existing resources, and developing a shared set of resources that are easy to find and well-maintained by dedicated staff with the resources necessary to ensure that information is regularly reviewed and up to date, would be a boon to this community.

12 For more information about the Generalist Repository Ecosystem Initiative: https://data-science.nih.gov/data-ecosystem/generalist-repository-ecosystem-initiative

13 https://dmtclearinghouse.esipfed.org/home

Further define, articulate, and promote the role of institutionally based data services across disciplines and funding agencies.

While research is driven by disciplinary knowledge, practices, and cultures, the on-the-ground support needed to conduct research is provided at the institutional level. Researchers depend upon their institution to provide space, equipment, and resources to enable their research. This tension between disciplinary practices and institutional support presents a challenge to researchers who are trying to manage, share, and preserve their data. Although some disciplines have made significant investments in defining standards or building repositories, many others are still developing these practices. Institutional data services provide critical support to researchers who lack clear disciplinary guidance or a disciplinary repository in which to submit their data. Even researchers affiliated with robust disciplinary data sharing guidelines may depend on institutional data services for assistance in creating, analyzing, or sharing their data.

Despite the importance of institutionally based data services to the larger ecosystem of data sharing, opportunities to connect with disciplinary communities are sporadic at best. This results in limited awareness of the institutional data services available to researchers. Collaborating with funding agencies and disciplinary societies to establish a presence for institutional data services would increase awareness and improve services. This would have additional benefits, such as supporting disciplines in establishing their data-specific guidelines and reducing duplication of efforts between disciplinary communities and institutional data services.

Invest in research relevant to understanding and responding to data sharing requirements at academic institutions.

The *Nelson Memo* and other data sharing requirements have elevated institutional awareness of data sharing needs and sparked a recognition that additional investments are required to fully support researchers in aligning with funder expectations. Despite the investments that have already been made, we still lack a clear understanding of the impact that data sharing mandates have had or are likely to have in the future. This makes it difficult for institutions to know what is needed, what resources to allocate, and where to invest.

Initiatives like the RADS project are beginning to explore aspects of research data sharing requirements such as the costs that researchers and institutions incur in meeting them. However, there are additional aspects of data sharing requirements that are also important to investigate, such as:

- How and to what extent are graduate students learning how to develop and manage their data? What approaches in teaching graduate students are most effective?
- Under what circumstances and to what extent are datasets being recognized as research outputs worthy of recognition in promotion and tenure cases? What models or metrics could be used to evaluate data sets as scholarly outputs?

Conclusion

STAIRS was a multi-part research and advocacy project that culminated in a gathering of representatives from academic institutions across the US, aimed at addressing the sociotechnical challenges facing institutionally based research data services and repositories as they work to align with emerging standards and guidelines, such as the DC-DR, and to meet the needs of the broader research ecosystem.

The themes and takeaways from the STAIRS event underscore the importance of institutionally based services for research data management in enabling researchers and their institutions to demonstrate compliance with the data sharing mandates of funding agencies. Unlike many external service providers, institutional service providers are co-located with researchers and have the ability to engage and collaborate with researchers where they work. Thus, they can see first-hand how researchers are handling their data. This enables institutional service providers and administrators to develop a deep understanding of the practices, cultures, and needs of their institution's researchers, which can be essential for developing specific services and supporting local needs. Institutional support for data sharing is still evolving, with significant variation in the types and levels offered by different institutions. Creating stronger working relationships across institutional boundaries may help create a more

unified approach to research data sharing that would support greater connectivity and standardization.

There is an urgent need for continued investment, collaboration, and knowledge-sharing within the academic data services community. STAIRS participants emphasized the importance of resource allocation for personnel, infrastructure, and shared tools/services; the value of openly sharing successes, failures, and best practices across institutions; the criticality of developing robust institutional data governance and management frameworks; and the necessity of incentivizing data sharing among researchers.

Moving forward, funding agencies and other key stakeholders must continue to foster spaces for these crucial cross-institutional conversations and provide support for the development of shared resources, training programs, and research relevant to data sharing challenges faced by academic institutions. Equally important is the need to directly engage with and enable institutions that have fewer resources, supporting more equitable access to the knowledge and capabilities required to build effective research data management, and sharing infrastructure and services.

By heeding the lessons and recommendations from STAIRS, the research data community can work collectively to strengthen the national infrastructure for responsible data stewardship, ultimately accelerating scientific discovery and innovation. The momentum and goodwill generated through this event must be leveraged and built upon to drive transformative change in how academic institutions support researchers in making their data FAIR.

REFERENCES

Carlson, J. & Narlock, M. R. (2024). Championing institutional data sharing efforts. CNI Spring 2024 Recorded Project Briefing. https://hdl.handle.net/11299/262382.

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for Indigenous data governance. *Data Science Journal*, 19(43), 1-12. https://doi.org/10.5334/dsj-2020-043

CODATA RDM Terminology Working Group. (2024). CODATA RDM Terminology (2023 version): overview (Version 2023). Zenodo. https://doi.org/10.5281/zenodo.10626170

Cowles, W. (2024). Technologies, metadata, and repository platforms: Persistent identifiers beyond articles. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264886.

Data Curation Network. (2023). Sustaining open research: Virtual learning series. Data Curation Network website. https://datacurationnetwork.org/sustaining-open-research-virtual-learning-series/

Data Curation Network. (nd). *Data Curation Network Outputs*. [Repository Collection]. Retrieved from the University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/202807.

Downey, M., & Ivey, S. (2024). Building community internally within the institution: North Carolina State University's Research Facilitation Service. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264884.

Fernandez, R. (2024). Building community externally: Operationalizing CARE at ASU. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264880.

Herndon, J. (2024). Building community internally within the institution: Developing an institution-wide data policy. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264887.

Hofelich Mohr, A. (2024). Building community internally within the institution: Costs of data management and sharing. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264888.

Hofelich Mohr, A., Carlson, J., Ge, L., Herndon, J., Kozlowski, W., Moore, J., Petters, J., Taylor, S., & Hudson Vitale, C. Making research

data publicly accessible: Estimates of institutional and researcher expenses. Association of Research Libraries, February 2024. https://doi.org/10.29242/report.radsexpense2024.

Holdren, J. P. (2013). Memorandum for the heads of executive departments and agencies: Increasing access to the results of federally funded scientific research. The White House Office of Science and Technology Policy. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Hudson Vitale, C., Imker, H., Johnston, L. R., Carlson, J., Kozlowski, W., Olendorf, R., & Stewart, C. (2017). *SPEC Kit 354: Data curation*. Association of Research Libraries. https://doi.org/10.29242/spec.354

Johnston, L. (2024). Technologies, metadata, and repository platforms: Metadata interoperability. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264889.

Johnston, L. R., & Coburn, L. (2020). *Data sharing readiness in academic institutions*. Data Curation Network. http://hdl.handle.net/11299/211358

Johnston, L., 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., & Coburn, E. (2018). Data Curation Network: A cross-institutional staffing model for curating research data. *International Journal of Digital Curatio*n, 13(1), pp.125-140. https://doi.org/10.2218/ijdc.v13i1.616.

Kaiser, J. (2016). Funding for key data resources in jeopardy. *Science*, 351,14. DOI:10.1126/science.351.6268.14

Kozlowski, W. (2024). Training, consulting and curation services: README call to action. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264890.

Kozlowski, W., Herndon, J., Taylor, S., Hudson Vitale, C., Carlson, J., Hofelich Mohr, A., Moore, J., Narlock, M., & Petters, J. (2023). Realities of Academic Data Sharing (RADS) initiative public-access Data Management and Sharing (DMS) activities, v3. Association of Research Libraries. https://doi.org/10.29242/radsdmsactivities2023.

Lafia, S. (2024). Technologies, metadata, and repository platforms: AI and data curation. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264891.

Luong, H., Narlock, M., & Petters, J. (2022). "We're all doing the best we can with what we've got": Preservation practices of Data Curation Network members. Proceedings of the 18th International Conference on Digital Preservation iPRES 2022. Glasgow, Scotland. https://doi.org/10.17605/OSF.IO/HMZER.

Lynch, C., Goldenberg-Hart, D., & Pope, P. (2024). Pre-Recorded Project Briefing Series, May 2024 Edition Guide. Coalition for Networked Information. https://www.cni.org/news/ppbs_editionguide_may24

Marsolek, W., Wright, S. J., Luong, H., Braxton, S. M., Carlson, J., & Lafferty-Hess, S. (2023). Understanding the value of curation: A survey of researcher perspectives of data curation services from six US institutions. *PLoS ONE*, 18(11): e0293534. https://doi.org/10.1371/journal.pone.0293534.

Narlock, M., Carlson, J.; Cowles, W., Herndon, J., Petters, J., Kozlowski, W., Delserone, L., Chandramouliswaran, I., Federer, L., Wham, B., Woodbrook, R., Johnston, L., Lafia, S., Ivey, S., Downey, M., Hofelich Mohr, Al., Stollar Peters, C., & Fernandez, R., (2024). Summit for Academic Institutional Readiness in Data Sharing (STAIRS) Complete Slide Deck. Retrieved from the University Digital Conservancy, https://hdl.handle.net/11299/264875.

Narlock, M., Taylor, S., Calvert, S., Priesman Marquez, R., & Parkman, A. (2024). Knowledge infrastructures are growing up: The case for institutional (data) repositories 10 Years After the Holdren Memo. *Data Science Journal* 23(1), p. 46. Available at: https://doi.org/10.5334/dsj-2024-046.

The National Science and Technology Council. (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

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

Petters, J. (2024). Building community externally: Using the Desirable Characteristics to evaluate our repository. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264892.

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

Stollar Peters, C. (2024). Building community externally: Applying the NIST RDaF to SUNY. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264893.

Strecker, D., Pampel, H., Schabinger, R., & Weisweiler, N. L. (2023) Disappearing repositories: Taking an infrastructure perspective on the long-term availability of research data. *Quantitative Science Studies*, 4(4): 839–856. https://doi.org/10.1162/qss_a_00277

Wham, B. (2024). Training, consulting and curation services: Penn State case study. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264894.

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

Woodbrook, R. (2024). Training, consulting and curation services: Curation as education. Proceedings of the Summit for Academic Institutional Readiness in Data Sharing (STAIRS). https://hdl.handle.net/11299/264895.

APPENDIX A: STAIRS APPLICATION

The following is the institutional application to attend STAIRS. Applications opened on April 10, 2024, and closed on May 17, 2024. The application was made available through Qualtrics.

Q1 Thank you for your interest in the **Summit for Academic Institutional Readiness in Data Sharing (STAIRS) to be held August 5-6 at the University of Minnesota-Twin Cities.**

Detailed project information is available on the Data Curation Network's website.

The STAIRS will be capped at 30 institutions. Each institution will be invited to bring up to three representatives to the workshop. These representatives may include individuals who support data management and sharing from across your institution. Attendees will receive a travel stipend from the DCN thanks to funding from the NIH Office of Data Science Strategy's program, Data Management Center of Excellence (DMCOE). The DMCOE is a project of HHS's Health FFRDC, which is operated by MITRE.

We strongly encourage applicants from a range of institutions that vary in size, research activity, and level of development of services and infrastructure for research data management and sharing, including Historically Black Colleges and Universities, Hispanic-serving institutions, or other Minority-serving institutions. Even if your institution has just begun planning for research data management and sharing, we invite you to apply.

Your institution does NOT need to have an institutional repository or data repository to apply to this workshop. We are not assessing or evaluating your institution's research data services, or particular repository solutions. The questions asked on this application are to ensure that a diverse group of institutions are in attendance at STAIRS.

Each institution should submit **one** application. The application is for the institution, not the individual.

All identifying information will remain within the project team. Data will be anonymized and shared in aggregate in publications and presentations.

Q2 Name of Institution

Q3 Applying Institution's Contact Name

Q4 Applying Institution's Contact Email

Q5 Each institution can send up to three representatives to the summit. We encourage institutions to send representatives from different units on campus, such as the Office of Research, Library, Information Technologies, etc. If accepted, which units do you anticipate will be represented? (Note: This response is not used to evaluate your application and is not binding.)

Q6 How will attending this workshop impact research data sharing at your institution? Please note: responses will be evaluated using our publicly available rubric. (Maximum: 250 words)

Q7 Consider the expertise your institution brings to the summit. How would your experiences and knowledge impact and bring value to other attendees? Please note: Responses will be evaluated using our publicly available rubric. (Maximum: 250 words)

Q8 In this section, we'd like to get a better sense of the different research data support offered by your institution's library and broader campus. The following research phases are adapted from the Realities of Academic Data Sharing Data Management and Sharing activities.

Responses will be evaluated using our publicly available rubric. We will also use the responses to ensure we have a wide-range of expertise in attendance.

Q9 For the following research data management phases, indicate the level of support offered by your institution:

	Yes, fully developed support (1)	Yes, in development (2)	Want to offer (3)	No, and n o plans to offer (4)	Unclear (5)
Planning, design, and start up of projects (1)	○	○	○	○	○
Data collection, storage, and management (2)	○	○	○	○	○
Making data broadly available, including selecting data for sharing and assigning persistent identifiers (3)	○	○	○	○	○
Data retention, including preservation, archive, and long-term access (4)	○	○	○	○	○
Project closeout and compliance (5)	○	○	○	○	○

Q10 Do you currently operate or provide access to a repository where your researchers can make their research data or code publicly available (e.g., Dataverse, Dryad, TIND / Invenio)? Note: Having a repository is not a prerequisite for attending this workshop.

- ○ Yes (1)
- ○ No, but we have plans to develop, acquire, or license one (2)
- ○ No (3)

Q11 Which repository platform(s) do you leverage, or are considering leveraging? (If none, please enter n/a)

Q12 This concludes the application. The application closes May 17, 2024, at 11:59pm Eastern Daylight Time. The STAIRS project team will send an update on the applications, including acceptances, in early June. If you have any questions, contact the STAIRS project lead, Mikala Narlock, at mnarlock@umn.edu.

Press the continue arrow in the bottom right of the screen to submit your application. Thank you for your interest!

APPENDIX B: STAIRS APPLICATION RUBRIC[14]

This rubric is for evaluating applications to the Summit for Academic Institutional Readiness in Data Sharing (STAIRS) to be held August 5-6 at the University of Minnesota-Twin Cities. Additional project information is available on the Data Curation Network's website.

The rubric is broken into two sections: free text responses and Data Management and Sharing Activities responses. We will use the information provided and the rubric to identify institutions and ensure we have a diverse mix of institutional preparedness in attendance.

Free Text Responses

	1 point	3 points	5 points
Guidelines	Responses in this category are incomplete or a non-answer	Responses in this category are complete but could potentially be more specific	Responses in this category are complete and specific
Example	"Our institution should attend."	"Attending this summit will improve our institutional data services and provide additional support to researchers."	"Attending this summit will expand our ability to provide a repository service to data creators. Our institution is interested in expanding our support of data management and sharing plans."

Data Management and Sharing Activities Responses

Responses in the matrix will be quantified to roughly sort applicants into different categories of data services.

Level of Support	Points
Unclear	0
No, and no plans to offer	0
Want to offer	1
Yes, in development	3
Yes, fully developed support	5

The responses will then be totaled and roughly divided into:

- 1–9 points: Interested in developing data services
- 10–17 points: Developing data services
- 18–25 points: Refining data services

This categorization will help us ensure that attendees represent different categories. Our goal will be to have approximately 10 institutions from each category in attendance, but that will depend on the applicants we receive.

14 Published April 9, 2024.

APPENDIX C: STAIRS PRE-SUMMIT SURVEY

The following survey was distributed to all attending institutions to better understand the current services and needs of STAIRS attendees. The survey was distributed via Qualtrics.

Q1 We are excited to see everyone in Minneapolis in just a few short weeks!

In preparation for the Summit for Academic Institutional Readiness in Data Sharing (STAIRS), we ask that each institutional representative respond to the questions in this survey. Responses should reflect the current situation and perspectives of the institution as a whole to the extent possible. You may want to coordinate with institutional colleagues in responding. Each institution should only submit one response.

Your responses to this survey will be used to inform presentations to introduce and frame discussions during the Summit. Information collected from this survey will only be presented in aggregate: no individual institutions will be named. We may also include information from this survey in the final report from STAIRS in aggregate. We will not associate data from this survey to individuals or institutions without their permission.

We use the Realities of Academic Data Sharing (RADS) Initiative Public Access Data Management and Sharing (DMS) Activities, v3 as the basis for the data lifecycle questions with permission.

Please complete the survey no later than by the end of day Wednesday, July 24. We expect this will take a minimum of 15 minutes to complete. Please answer to the best of your ability, but do not overextend yourself in responding.

Questions may be directed to Mikala Narlock (mnarlock@umn.edu) or Jake Carlson (jakecarl@buffalo.edu). Thank you for taking the time to complete this survey.

Q2 Please provide your institution name for tracking purposes.

Q3 For this set of questions, we would like you to consider a set of possible data services that could be offered in the first stage of the data lifecycle: "Planning, Design and Start Up of Projects."

Please indicate if you offer the service at your institution or would like to, if you do not offer the service, or if you are unsure if you offer the service or not.

Q4 Planning, Design, and Start Up of Projects	**Yes, or would like to offer in the next year or two (1)**	**No (2)**	**Unsure or unclear (3)**
Consulting on or preparing DMPs / DMSPs including considerations for ethical and privacy-related practices (1)	○	○	○
Consulting in data management and sharing costs and expenses to be included in grant budgets (2)	○	○	○
Reviewing of institutional review board (IRB) protocols and informed consent language (3)	○	○	○
Developing, building, providing or recommending storage solutions for active research data (4)	○	○	○
Recommending an appropriate repository(-ies) for making research data broadly available (5)	○	○	○
Assessing data security needs and recommending solutions (6)	○	○	○
Supporting intellectual property and copyright considerations (7)	○	○	○
Checking for compliance with existing policies and/or federal requirements (for example, HIPPA, FERPA, Data Use Agreements, material transfer agreements, etc.) (8)	○	○	○
Referring to disciplinary, funder, and institutional standards, and/or good practices for handling, collecting, documenting, and sharing data (9)	○	○	○
Developing or creating training or educational materials in support of data sharing (10)	○	○	○

Display This Question:

If Planning, Design and Start Up of Projects = Yes, or would like to offer in the next year or two

Carry Forward Selected Choices from "Planning, Design and Start Up of Projects"

Q5 These are the services that you offer or would like to offer as indicated from your responses to the previous question. How well developed are each of these services?	**Fully developed support (1)**	**Partially developed support (planned or in progress) (2)**	**Planned support (not yet in progress) (3)**
Consulting on or preparing DMPs / DMSPs including considerations for ethical and privacy-related practices (x1)	○	○	○
Consulting in data management and sharing costs and expenses to be included in grant budgets (x2)	○	○	○
Reviewing of institutional review board (IRB) protocols and informed consent language (x3)	○	○	○
Developing, building, providing or recommending storage solutions for active research data (x4)	○	○	○
Recommending an appropriate repository(-ies) for making research data broadly available (x5)	○	○	○
Assessing data security needs and recommending solutions (x6)	○	○	○
Supporting intellectual property and copyright considerations (x7)	○	○	○
Checking for compliance with existing policies and/or federal requirements (for example, HIPPA, FERPA, Data Use Agreements, material transfer agreements, etc.) (x8)	○	○	○
Referring to disciplinary, funder, and institutional standards, and/or good practices for handling, collecting, documenting, and sharing data (x9)	○	○	○
Developing or creating training or educational materials in support of data sharing (x10)	○	○	○

Display This Question:

If Planning, Design and Start Up of Projects = Yes, or would like to offer in the next year or two

Carry Forward Selected Choices from "Planning, Design and Start Up of Projects"

Q6 These are the services that you offer or would like to offer as indicated from your responses to the previous question. Please indicate who is or will be involved in providing the support.	**Libraries (1)**	**Office of Research Adminis-tration (2)**	**Information Technology / Advanced Research Computing (3)**	**Insti-tutes and Research Centers (4)**	**Other (please specify in comments) (5)**
Consulting on or preparing DMPs / DMSPs including considerations for ethical and privacy-related practices (x1)	○	○	○	○	○
Consulting in data management and sharing costs and expenses to be included in grant budgets (x2)	○	○	○	○	○
Reviewing of institutional review board (IRB) protocols and informed consent language (x3)	○	○	○	○	○
Developing, building, providing or recommending storage solutions for active research data (x4)	○	○	○	○	○

Q6 These are the services that you offer or would like to offer as indicated from your responses to the previous question. Please indicate who is or will be involved in providing the support.	**Libraries (1)**	**Office of Research Adminis-tration (2)**	**Information Technology / Advanced Research Computing (3)**	**Insti-tutes and Research Centers (4)**	**Other (please specify in comments) (5)**
Recommending an appropriate repository(-ies) for making research data broadly available (x5)	○	○	○	○	○
Assessing data security needs and recommending solutions (x6)	○	○	○	○	○
Supporting intellectual property and copyright considerations (x7)	○	○	○	○	○
Checking for compliance with existing policies and/or federal requirements (for example, HIPPA, FERPA, Data Use Agreements, material transfer agreements, etc.) (x8)	○	○	○	○	○
Referring to disciplinary, funder, and institutional standards, and/or good practices for handling, collecting, documenting, and sharing data (x9)	○	○	○	○	○
Developing or creating training or educational materials in support of data sharing (x10)	○	○	○	○	○

Display This Question:

If Planning, Design and Start Up of Projects = Yes, or would like to offer in the next year or two

Q7 Comments

Q8 For this set of questions, we would like you to consider a set of possible data services that could be offered in the second stage of the data lifecycle: "Data Collection, Storage, and Management."

Please indicate if you offer the service at your institution or would like to, if you do not offer the service, or if you are unsure if you offer the service or not.

Q9 Data Collection, Storage, and Management	**Yes, or would like to offer in the next year or two (1)**	**No (2)**	**Unsure or Unclear (3)**
Developing or advising on data documentation, including version control documentation (1)	○	○	○
Creating or reviewing established quality control mechanisms or procedures (2)	○	○	○
Evaluating or recommending data-analysis tools and processes to support sharing and reproducibility (3)	○	○	○
Supporting or providing for the management of active data throughout project (for example, storage, security, backup, lab notebooks, including considerations for managing and storing large or specialized datasets). (4)	○	○	○
Developing or creating training or educational materials for data sharing; implementing data sharing training (5)	○	○	○

Display This Question:

If Data Collection, Storage, and Management = Yes, or would like to offer in the next year or two

Carry Forward Selected Choices from "Data Collection, Storage, and Management"

Q10 These are the services that you offer or would like to offer as indicated from your responses to the previous question. How well developed are each of these services?	**Fully developed support (1)**	**Partially developed support (planned or in progress) (2)**	**Planned support (not yet in progress) (3)**
Developing or advising on data documentation, including version control documentation (x1)	○	○	○
Creating or reviewing established quality control mechanisms or procedures (x2)	○	○	○
Evaluating or recommending data-analysis tools and processes to support sharing and reproducibility (x3)	○	○	○
Supporting or providing for the management of active data throughout project (for example, storage, security, backup, lab notebooks, including considerations for managing and storing large or specialized datasets). (x4)	○	○	○
Developing or creating training or educational materials for data sharing; implementing data sharing training (x5)	○	○	○

Display This Question:

If Data Collection, Storage, and Management = Yes, or would like to offer in the next year or two

Carry Forward Selected Choices from "Data Collection, Storage, and Management"

Q11 These are the services that you offer or would like to offer as indicated from your responses to the previous question. Please indicate who is or will be involved in providing the support.e support.	**Libraries (1)**	**Office of Research Adminis-tration (2)**	**Information Technology / Advanced Research Computing (3)**	**Insti-tutes and Research Centers (4)**	**Other (please specify in comments) (5)**
Developing or advising on data documentation, including version control documentation (x1)	○	○	○	○	○
Creating or reviewing established quality control mechanisms or procedures (x2)	○	○	○	○	○
Evaluating or recommending data-analysis tools and processes to support sharing and reproducibility (x3)	○	○	○	○	○
Supporting or providing for the management of active data throughout project (for example, storage, security, backup, lab notebooks, including considerations for managing and storing large or specialized datasets). (x4)	○	○	○	○	○
Developing or creating training or educational materials for data sharing; implementing data sharing training (x5)	○	○	○	○	○

Display This Question:

If Data Collection, Storage, and Management = Yes, or would like to offer in the next year or two

Q12 Comments

Q13 For this set of questions, we would like you to consider a set of possible data services that could be offered in the third stage of the data lifecycle: "Making Data Broadly Available."

Please indicate if you offer the service at your institution or would like to, if you do not offer the service, or if you are unsure if you offer the service or not.

Q14 Making Data Broadly Available	**Yes, or would like to offer in the next year or two (1)**	**No (2)**	**Unsure or Unclear (3)**
Consulting on what data to share or host, including addressing proper levels of access and security (1)	○	○	○
Consulting on, providing and/or hosting repositories for making data available (2)	○	○	○
Preparing or consulting on preparing data for sharing (for example, de-identification, checking for privacy/personally identifiable information [PII]/ protected health information [PHI], considering ethical impacts, data selection, data curation*, data cleaning, validation, reproducibility checking, and quality control) (3)	○	○	○
Submitting or supporting submission of data into a data sharing platform/ repository, including considerations for sharing and moving large or specialized datasets, both on local and high performance computing (HPC) resources (4)	○	○	○
Creating or reviewing documentation for research data (for example, supporting development of structured metadata and README files) (5)	○	○	○
Consulting, selecting, or applying licenses to data and software/code (6)	○	○	○
Recommending or transforming data file formats to be open or more accessible (7)	○	○	○
Creating, recommending and/or ensuring use of persistent identifiers (PIDs) (for example, digital object identifiers [DOIs], ORCIDs, RORs, etc.) (8)	○	○	○
Developing or checking for compliance with DUAs or material transfer agreements (9)	○	○	○
Developing or creating training or educational materials for data sharing; implementing data sharing training (10)	○	○	○

Display This Question:

If Making Data Broadly Available = Yes, or would like to offer in the next year or two

Carry Forward Selected Choices from "Making Data Broadly Available"

Q15 These are the services that you offer or would like to offer as indicated from your responses to the previous question. How well developed are each of these services?	**Fully developed support (1)**	**Partially developed support (planned or in progress) (2)**	**Planned support (not yet in progress) (3)**
Consulting on what data to share or host, including addressing proper levels of access and security (x1)	○	○	○
Consulting on, providing and/or hosting repositories for making data available (x2)	○	○	○
Preparing or consulting on preparing data for sharing (for example, de-identification, checking for privacy/personally identifiable information [PII]/protected health information [PHI], considering ethical impacts, data selection, data curation*, data cleaning, validation, reproducibility checking, and quality control) (x3)	○	○	○

Q15 These are the services that you offer or would like to offer as indicated from your responses to the previous question. How well developed are each of these services?	Fully developed support (1)	Partially developed support (planned or in progress) (2)	Planned support (not yet in progress) (3)
Submitting or supporting submission of data into a data sharing platform/repository, including considerations for sharing and moving large or specialized datasets, both on local and high performance computing (HPC) resources (x4)	○	○	○
Creating or reviewing documentation for research data (for example, supporting development of structured metadata and README files) (x5)	○	○	○
Consulting, selecting, or applying licenses to data and software/code (x6)	○	○	○
Recommending or transforming data file formats to be open or more accessible (x7)	○	○	○
Creating, recommending and/or ensuring use of persistent identifiers (PIDs) (for example, digital object identifiers [DOIs], ORCIDs, RORs, etc.) (x8)	○	○	○
Developing or checking for compliance with DUAs or material transfer agreements (x9)	○	○	○
Developing or creating training or educational materials for data sharing; implementing data sharing training (x10)	○	○	○

Display This Question:

If Making Data Broadly Available = Yes, or would like to offer in the next year or two

Carry Forward Selected Choices from "Making Data Broadly Available"

Q16 These are the services that you offer or would like to offer as indicated from your responses to the previous question. Please indicate who is or will be involved in providing the support.	Libraries (1)	Office of Research Adminis-tration (2)	Information Technology / Advanced Research Computing (3)	Insti-tutes and Research Centers (4)	Other (please specify in comments) (5)
Consulting on what data to share or host, including addressing proper levels of access and security (x1)	○	○	○	○	○
Consulting on, providing and/or hosting repositories for making data available (x2)	○	○	○	○	○
Preparing or consulting on preparing data for sharing (for example, de-identification, checking for privacy/personally identifiable information [PII]/protected health information [PHI], considering ethical impacts, data selection, data curation*, data cleaning, validation, reproducibility checking, and quality control) (x3)	○	○	○	○	○

Q16 These are the services that you offer or would like to offer as indicated from your responses to the previous question. Please indicate who is or will be involved in providing the support.	Libraries (1)	Office of Research Adminis-tration (2)	Information Technology / Advanced Research Computing (3)	Insti-tutes and Research Centers (4)	Other (please specify in comments) (5)
Submitting or supporting submission of data into a data sharing platform/repository, including considerations for sharing and moving large or specialized datasets, both on local and high performance computing (HPC) resources (x4)	○	○	○	○	○
Creating or reviewing documentation for research data (for example, supporting development of structured metadata and README files) (x5)	○	○	○	○	○
Consulting, selecting, or applying licenses to data and software/code (x6)	○	○	○	○	○
Recommending or transforming data file formats to be open or more accessible (x7)	○	○	○	○	○
Creating, recommending and/or ensuring use of persistent identifiers (PIDs) (for example, digital object identifiers [DOIs], ORCIDs, RORs, etc.) (x8)	○	○	○	○	○
Developing or checking for compliance with DUAs or material transfer agreements (x9)	○	○	○	○	○
Developing or creating training or educational materials for data sharing; implementing data sharing training (x10)	○	○	○	○	○

Display This Question:

If Making Data Broadly Available = Yes, or would like to offer in the next year or two

Q17 Comments

Q18 For this set of questions, we would like you to consider a set of possible data services that could be offered in the fourth stage of the data lifecycle: "Data Retention, Including Preservation, Archive and Long-Term Access."

Please indicate if you offer the service at your institution or would like to, if you do not offer the service, or if you are unsure if you offer the service or not.

Q19 Data Retention, Including Preservation, Archive and Long-Term Access	**Yes, or would like to offer in the next year or two (1)**	**No (2)**	**Unsure or Unclear (3)**
Consulting on or migrating files to new formats or across systems as needed (1)	○	○	○
Monitoring integrity of preserved data (2)	○	○	○
Making decisions about de-accessioning and removal of research data (3)	○	○	○
Ensuring data security when appropriate (for example, PHI/ HIPAA, export controls, Federal Information Security Management Act [FISMA], student data, and intellectual property) (4)	○	○	○
Developing or creating training or educational materials for implementing preservation, archiving, and Long-Term Access of data (5)	○	○	○

Display This Question:
If Data Retention, Including Preservation, Archive and Long-Term Access = Yes, or would like to offer in the next year or two
Carry Forward Selected Choices from "Data Retention, Including Preservation, Archive and Long-Term Access"

Q20 These are the services that you offer or would like to offer as indicated from your responses to the previous question. How well developed are each of these services?	**Fully developed support (1)**	**Partially developed support (planned or in progress) (2)**	**Planned support (not yet in progress) (3)**
Consulting on or migrating files to new formats or across systems as needed (x1)	○	○	○
Monitoring integrity of preserved data (x2)	○	○	○
Making decisions about de-accessioning and removal of research data (x3)	○	○	○
Ensuring data security when appropriate (for example, PHI/ HIPAA, export controls, Federal Information Security Management Act [FISMA], student data, and intellectual property) (x4)	○	○	○
Developing or creating training or educational materials for implementing preservation, archiving, and Long-Term Access of data (x5)	○	○	○

Display This Question:
If Data Retention, Including Preservation, Archive and Long-Term Access = Yes, or would like to offer in the next year or two

Carry Forward Selected Choices from "Data Retention, Including Preservation, Archive and Long-Term Access"

Q21 These are the services that you offer or would like to offer as indicated from your responses to the previous question. Please indicate who is or will be involved in providing the support.	**Libraries (1)**	**Office of Research Adminis-tration (2)**	**Information Technology / Advanced Research Computing (3)**	**Insti-tutes and Research Centers (4)**	**Other (please specify in comments) (5)**
Consulting on or migrating files to new formats or across systems as needed (x1)	○	○	○	○	○
Monitoring integrity of preserved data (x2)	○	○	○	○	○
Making decisions about de-accessioning and removal of research data (x3)	○	○	○	○	○
Ensuring data security when appropriate (for example, PHI/HIPAA, export controls, Federal Information Security Management Act [FISMA], student data, and intellectual property) (x4)	○	○	○	○	○
Developing or creating training or educational materials for implementing preservation, archiving, and Long-Term Access of data (x5)	○	○	○	○	○

Display This Question:
If Data Retention, Including Preservation, Archive and Long-Term Access = Yes, or would like to offer in the next year or two

Q22 Comments

Q23 For this set of questions, we would like you to consider a set of possible data services that could be offered in the fifth and final stage of the data lifecycle: "Project Closeout and Compliance."

Please indicate if you offer the service at your institution or would like to, if you do not offer the service, or if you are unsure if you offer the service or not.

Q24 Project Closeout and Compliance	**Yes, or would like to offer in the next year or two (1)**	**No (2)**	**Unsure or Unclear (3)**
Ensuring funding agency, institutional, and/ or publisher requirements for data sharing and retention have been met (1)	○	○	○
Providing compliance support around research project reports and project closeout (2)	○	○	○
Developing or creating training or educational materials for data sharing in closing out a project or demonstrating compliance to the requirements of a funding agency (3)	○	○	○

Display This Question:

If Project Closeout and Compliance = Yes, or would like to offer in the next year or two

Carry Forward Selected Choices from "Project Closeout and Compliance"

Q25 These are the services that you offer or would like to offer as indicated from your responses to the previous question. How well developed are each of these services?	**Fully developed support (1)**	**Partially developed support (planned or in progress) (2)**	**Planned support (not yet in progress) (3)**
Ensuring funding agency, institutional, and/ or publisher requirements for data sharing and retention have been met (x1)	○	○	○
Providing compliance support around research project reports and project closeout (x2)	○	○	○
Developing or creating training or educational materials for data sharing in closing out a project or demonstrating compliance to the requirements of a funding agency (x3)	○	○	○

Display This Question:

If Project Closeout and Compliance = Yes, or would like to offer in the next year or two

Carry Forward Selected Choices from "Project Closeout and Compliance"

Q26 These are the services that you offer or would like to offer as indicated from your responses to the previous question. Please indicate who is or will be involved in providing the support.	**Libraries (1)**	**Office of Research Administration (2)**	**Information Technology / Advanced Research Computing (3)**	**Institutes and Research Centers (4)**	**Other (please specify in comments) (5)**
Ensuring funding agency, institutional, and/ or publisher requirements for data sharing and retention have been met (x1)	○	○	○	○	○
Providing compliance support around research project reports and project closeout (x2)	○	○	○	○	○
Developing or creating training or educational materials for data sharing in closing out a project or demonstrating compliance to the requirements of a funding agency (x3)	○	○	○	○	○

Display This Question:

If Project Closeout and Compliance = Yes, or would like to offer in the next year or two

Q27 Comments

Q28 Please answer the next set of questions based on the repository platform that you would consider to be your primary institutional or generalist repository.

This could be a repository that is maintained and operated "in-house" by the library or other unit at your institution. This could also be a "generalist repository" such as Dryad, figshare or Vivli, that is maintained by a 3rd party but with whom you have a membership or other arangement to deposit data from your institution.

If you do not have a primary repository, please select N/A.

Q29 Which of these repository platforms would you consider to be your primary institutional / generalist repository for data? (select one)

- ○ Atmire (1)
- ○ bepress (2)
- ○ Dataverse (3)
- ○ Digital Commons (4)
- ○ Dryad (5)
- ○ DSpace (6)
- ○ Esploro (7)
- ○ figshare (8)
- ○ Invenio / TIND (9)
- ○ Open Science Framework (OSF) (10)
- ○ Samvera / Hyrax (11)
- ○ Vivli (12)
- ○ A "home grown" platform (13)
- ○ N/A (My institution does not have an institutional or generalist repository) (14)
- ○ Other (please describe below) (15)

Q30 If you selected "other," please provide additional information about your primary repository.

Q31 Does your primary institutional / generalist repository support the use of the following persistent identifiers?

- ○ Digital Object Identifier (DOI) (1)
- ○ Global Research Identifier Database (GRID) (2)
- ○ Open Researcher and Contributor ID (ORCID) (3)
- ○ Research Organization Registry (ROR) (4)
- ○ Other (please provide details in the comment box below) (5)

Q32 If you selected "Other," please provide details about the persistent identifer(s) supported by your primary institutional / generalist repository

Q33 What would consider to be the strengths of your primary institutional / generalist repository?

Q34 What would you consider to be the weaknesses of your primary institutional / generalist repository?

Q35 This concludes the survey. Clicking "Next page" submits your responses. Thank you for your time!

APPENDIX D: PRE-SUMMIT RESPONSE VISUALIZATIONS

Below are the aggregated responses to the pre-summit survey that was completed by all attending institutions.

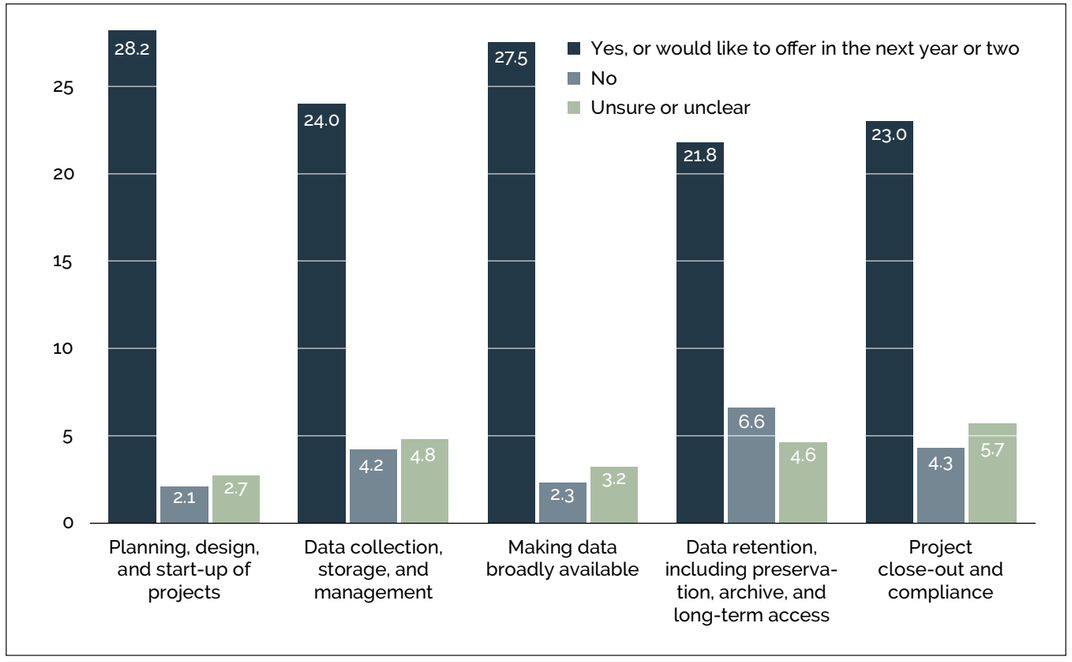

Figure 6: Number of institutions with support in each data management and sharing phase, normalized by the total number of responsess

Institutions were asked to indicate which of the RADS Data Management and Sharing Activities they offer (or would like to soon), do not offer, or are unsure. Our team aggregated the responses, averaged across phases, and found:

- Support tends to be focused in the "Planning, Design, and Start up of Projects," and the "Making Data Broadly Available" phases.
- Support tends to be needed in the "Data Retention, including Preservation, archiving, and Long-term Access" phase.

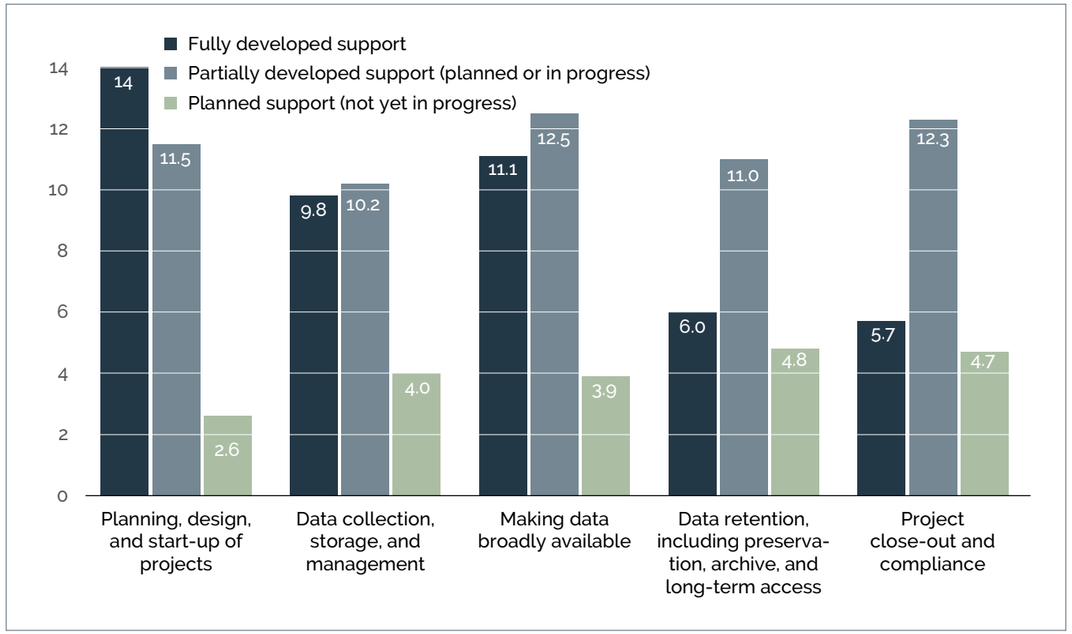

Figure 7: Number of institutions and level of support in each data management and sharing phase, normalized by the total number of responses

Institutions were asked to indicate how developed the support for each of the RADS Data Management and Sharing Activities they offer. Our team aggregated the responses, averaged across phases, and found that:

- Support is fully or partially developed in the "Planning, Design, and Start up of Projects," and the "Making Data Broadly Available" phases.
- Support is developing, or planned, for the "Data Retention, including Preservation, archiving, and Long-term Access" and "Project Closeout and Compliance" phases.

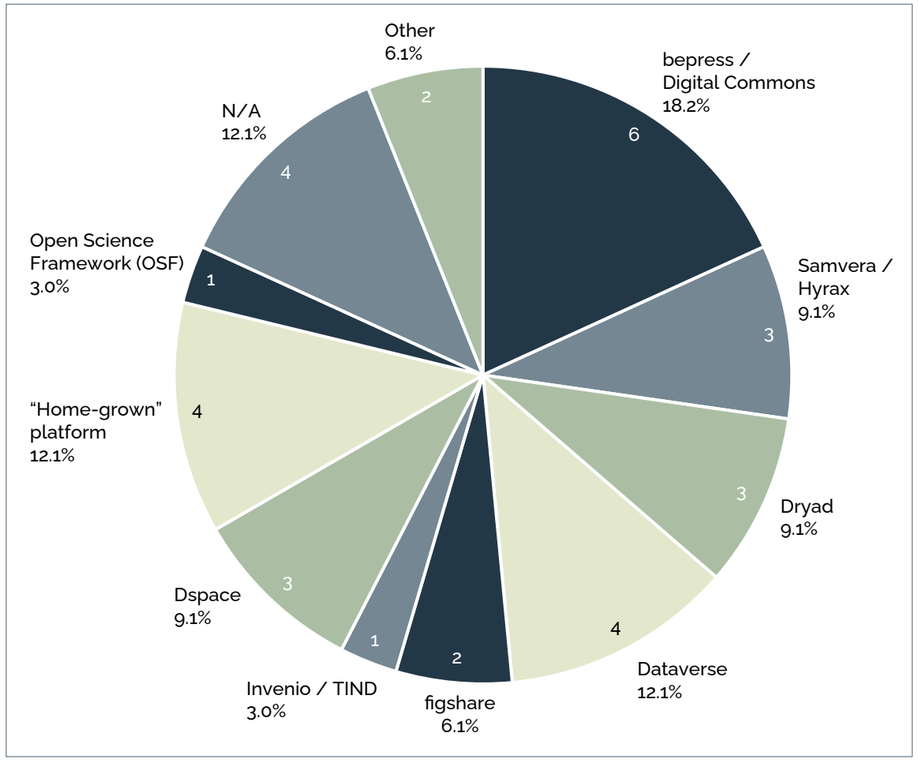

Figure 8: Repository solutions used by institutions

Institutions were asked to indicate the technology underpinning their primary institutional repository. As is demonstrated in the figure, there are a wide variety of vended and locally-managed solutions each institution is using.

APPENDIX E: FULL STAIRS AGENDA

Data Curation Network, August 5–6, 2024

Monday August 5

8:00-8:30 am Breakfast
8:15 - 8:30 am Introduction to STAIRS and "Housekeeping"
8:30 - 8:45 am Introductory remarks by **Lisa Federer**
8:45 - 9:00 am Introductory remarks by **Ishwar Chandramouliswaran**

Broad Topic #1: "Training, consulting and curation services"

9:00 - 9:10 am Current State of Training, Consulting and Curation Services
Speaker: Jake Carlson, University at Buffalo

9:10 - 9:40 am Conversation Starters. Potential topics include:

- Shared / Common curricula for Data Training / Education Programs
  Speaker: Briana Wham, Penn State
- README file Templates for specific data formats / types
  Speaker: Wendy Kozlowski, Cornell University
- Curation as education
  Speaker: Rachel Woodbrook, University of Michigan

9:40 - 9:45 am Set up next exercise
9:45 - 10:15 am First round discussion

- What is your take on the current state of things? What are your institutional strengths with these topics?
- What stuck out for you from the presentations? Are there things you would like to bring back to your institution?

10:15 - 10:45 am Break / Set up next rotation
10:45 - 11:15 am Second round discussions

- Starting with a recap of the first discussion...
- Where might we work more closely together as a community to advance capabilities with Training, Consultation, or Curation Services?

11:15 - 11:45 am Reporting out / Sharing ideas
11:45 am - 1:00 pm Lunch

Broad Topic #2: "Technologies, Metadata, and Repository Platforms"

1:00 - 1:10 pm	Current State of Technologies, Metadata, and Repository Platforms **Speaker: Mikala Narlock, Data Curation Network**
1:10 - 1:40 pm	Conversation Starters. Potential topics include: • How can we improve metadata interoperability between repositories? • Speaker: Lisa Johnston, University of Wisconsin-Madison • The potential for automating data curation work using AI • Speaker: Sara Lafia, NORC • Persistent Identifiers beyond Articles: The wonderful world of PIDs • Speaker: Wind Cowles, Princeton University
1:40 - 1:45 pm	Set up next exercise
1:45 - 2:15 pm	First round discussion • What is your take on the current state of things? What are your institutional strengths with these topics? • What stuck out for you from the presentations? Are there things you would like to bring back to your institution?
2:15 - 2:45 pm	Break / Set up next rotation
2:45 - 3:15 pm	Second round discussions • Where might we work more closely together as a community to advance capabilities with Technologies, metadata, and repository platforms?
3:15 - 3:45 pm	Reporting out / Sharing ideas
3:45 - 4:00 pm	Final Thoughts
4:00 - 4:15 pm	Quick write! Minute papers. • Assignment: Write one take away from today and one thing that is still unclear. Is there anything that could be addressed or improved upon for tomorrow?
4:15 - 4:45pm	Socialize, explore campus!
4:15 - 4:45 pm	Facilitators remain to review quick write papers
5:00 - 7:00 pm	Dinner

Tuesday, August 6

8:00-8:30	Breakfast
8:30-9:00	Recap of day one / summary of the minute papers

Broad Topic #3: "Building community Internally within the Institution"

9:00 - 9:10 am	Current State of Building community Internally within the Institution **Speaker: Jake Carlson, University at Buffalo**
9:10 - 9:40 am	Conversation Starters. Potential topics include: • How will institutions need to change to support data management, sharing and preservation at scale? **Speaker: Susan Ivey and Moira Downey, North Carolina State University** • Developing an Institution Wide Data Policy **Speaker: Joel Herndon, Duke University** • The costs of managing and sharing data **Speaker: Alicia Hofelich Mohr, University of Minnesota**
9:40 - 9:45 am	Set up next exercise
9:45 - 10:15 am	First round discussion • What is your take on the current state of things? What are your institutional strengths with these topics? • What stuck out for you from the presentations? Are there things you would like to bring back to your institution?
10:15 - 10:45 am	Break / set up next round
10:45 - 11:15 am	Second round discussions • Starting with a recap of the first discussion... • Where might we work more closely together as a community to advance capabilities / supporting one another in building community at our institutions?
11:15 - 11:45 am	Reporting out / Sharing ideas
11:45 am - 1:00 pm	Lunch

Broad Topic #4: "Building community externally across institutions"

1:00 - 1:10 pm	Current State of Building community Externally across Institutions **Speaker: Mikala Narlock, Data Curation Network**
1:10 - 1:40 pm	Conversation Starters • Applying the DC-DR to our services and repositories. Where are we doing well and where do we need to invest more? **Speaker: Jon Petters, Virginia Tech**

	• Applying the NIST RDaF to our work **Speaker: Catherine Stollar Peters, SUNY Office of Library and Information Services** • Operationalizing the CARE Principles in data sharing **Speaker: Rachel Fernandez, Arizona State University**
1:40 - 1:45 pm	Set up exercise
1:45 - 2:15 pm	First round discussion • What is your take on the current state of things? What are your institutional strengths with these topics? • What stuck out for you from the presentations? Are there things you would like to bring back to your institution?
2:15 - 2:45 pm	Break
2:45 - 3:15 pm	Second round discussions • Where might we work more closely together as a community to advance capabilities by building community with one another?
3:15 - 3:45 pm	Reporting out / Sharing ideas
3:45 - 4:00 pm	Overview of final activity
4:00 - 4:45 pm	Institutions start their action plans
4:45 - 5:00 pm	Final thoughts and wrap up
5:00 - 7:00 pm	Dinner

APPENDIX F: CONVERSATION CAPTAIN INFORMATION

This information sheet was distributed to STAIRS facilitators prior to the event. It has been shared here to both clarify the role of the "Conversation Captain" and to allow for reuse.

STAIRS Conversation Captain Information Sheet

Or, information that will be helpful during our event

Thank you for agreeing to be a Conversation Captain for STAIRS! We are excited that you will join us, and are incredibly grateful for your participation. Below is information that may be helpful during the event.

Please note that all attendees are expected to follow the DCN's Code of Conduct. Report any violations to Mikala Narlock or Jake Carlson.

What is a Conversation Captain?

You have two roles as a Conversation Captain. First, is to start the conversation. You will likely need to start the conversation with each new group. They may not know each other and may be timid about speaking out. We have added prompting questions to our shared notes document that you can use.

You will take notes in a shared notes folder that corresponds to your table number. While attendees will rotate throughout the event, you'll be stuck at one table per day.

Your second role is to take as many meaningful notes as you can!

1. Do not try to capture everything. We aren't looking for a transcript! Try to capture themes, big ideas, and +1s.
2. All attendees will be wearing colored lanyards to indicate the unit their position is in:
    a. Blue = Libraries
    b. Red = Information Technology
    c. Green = Office of Research
    d. Yellow = Centers (e.g., Center for Digital Scholarship)
    e. Black = Other campus unit
    f. Attendees may select more than one.
    g. Try to capture the perspective in the notes (e.g., Attendee from IT thinks IRs are the best. Office of Research attendee agrees).
3. If possible, avoid names of individual participants. Institution names are ok, but not necessary.
4. You are welcome to use acronyms — just make sure to define those early in the notes!

The conversation will meander — do not feel like you need to force participants into certain discussions.

Do ask questions about anything you don't understand — don't assume others at the table know. If you're confused, someone else probably is, too.

Pro-tip: If you're nervous about asking for clarification, you can always say "Quick question, to make sure I capture this in the notes, what does X mean/stand for/etc?"

If you're comfortable, you can do more facilitation, such as calling on individuals to participate. This is totally not necessary!

Lastly, we will tell participants that they, too, can take notes. We hope this will help!

Thank you again– we are incredibly grateful for your time and energy.

APPENDIX G: TEMPLATE INSTITUTIONAL ACTION PLAN

This was provided to all participating institutions on the final day of the event to help encourage the adoption of tools, resources, or services at their institution.

Goals:

Define 1-3 near term (>12 month) goals for your institution.

Resources:

What resources do you need to be successful? Think about personnel, tools, services, etc.

Collaborators:

Who do you need and WANT to work with to make this a reality? Add more rows if needed, and clarify which goal corresponds to which collaborators.

Internal (to the institution) Collaborators	External Collaborators

Your local context

What is going to help you locally? What do you need to work around?

Drivers (helps to achieve your goal)	Barriers (keeps you from achieving goal)

APPENDIX H: LIST OF ATTENDING INSTITUTIONS

Arizona State University
Boston University
Brown University
Bucknell University
Children's Hospital of Philadelphia. Department of Bioinformatics and Health Informatics. Arcus Program- Library Science Team
Dartmouth College
Duke University
Harvard University
Hofstra University
Kansas State University
Louisiana State University
North Carolina State University
Oregon Health & Science University
Penn State University
Pittsburgh Supercomputing Center
Carnegie Mellon University
Rosalind Franklin University of Medicine and Science
San Jose State University
The George Washington University
The University of Tennessee Health Science Center
University at Buffalo
University of Arizona
University of Cincinnati
University of Georgia
University of Illinois Chicago
University of Michigan
University of Minnesota
University of New Hampshire
University of New Mexico
University of North Dakota
Virginia Commonwealth University
Wayne State University
West Virginia University
Zucker School of Medicine at Hofstra/Northwell and Northwell Health

APPENDIX I: LIST OF ACRONYMS

AI	Artificial intelligence
ARL	Association of Research Libraries
CNI	Coalition for Networked Information
DC-DR	Desirable Characteristics of Data Repositories for Federally Funded Research
DCN	Data Curation Network
DOI	Digital Object Identifier
GRID	Global Research Identifier Database
ICPSR	Consortium for Political and Social Research
IDR	Institutional data repository
IR	Institutional repository
NIH	National Institutes of Health
NIST RDaF	National Institute of Standards and Technology's Research Data Framewor
NSF	National Science Foundation
PHI	Personal Health Information
PID	persistent identifier
RADS	Realities of Academic Data Sharing
RDAP	Research Data Access and Preservation Association
ROR	research organization
STAIRS	Summit for Academic Institutional Readiness in Data Sharing
UMN	University of Minnesota