

Strengthening Research Infrastructure: Data Curation + Institutional Repositories

By Jen Darragh, Shawna Taylor, and Mikala Narlock

In 2024, advocacy emerged as a focal point of external DCN activities, with members drawing on years of collaboration and expertise to emphasize the essential role of institutional repositories and professional data curation expertise. From surveying researchers and repository managers to publishing impactful articles and white papers, DCN members have provided compelling evidence underscoring the importance of local data stewardship and the value of contributions of data curators and curation in data sharing.

One key focus has been an examination of the value curation brings to data sharing and institutional goals. Surveys conducted across six academic institutions found that 97% of researchers reported that data curation enhances data sharing, while 90% indicated increased confidence in data dissemination following the curation process ([Maroslek et al., 2023](#)). Similarly, a related survey of repository managers and curators found that 82% strongly agreed that data curation adds value to the data sharing process ([Johnston et al., 2024](#)). While curators have long recognized the value of their work, these surveys provide concrete evidence to support those beliefs and highlight the measurable impact of curation services on the research lifecycle.

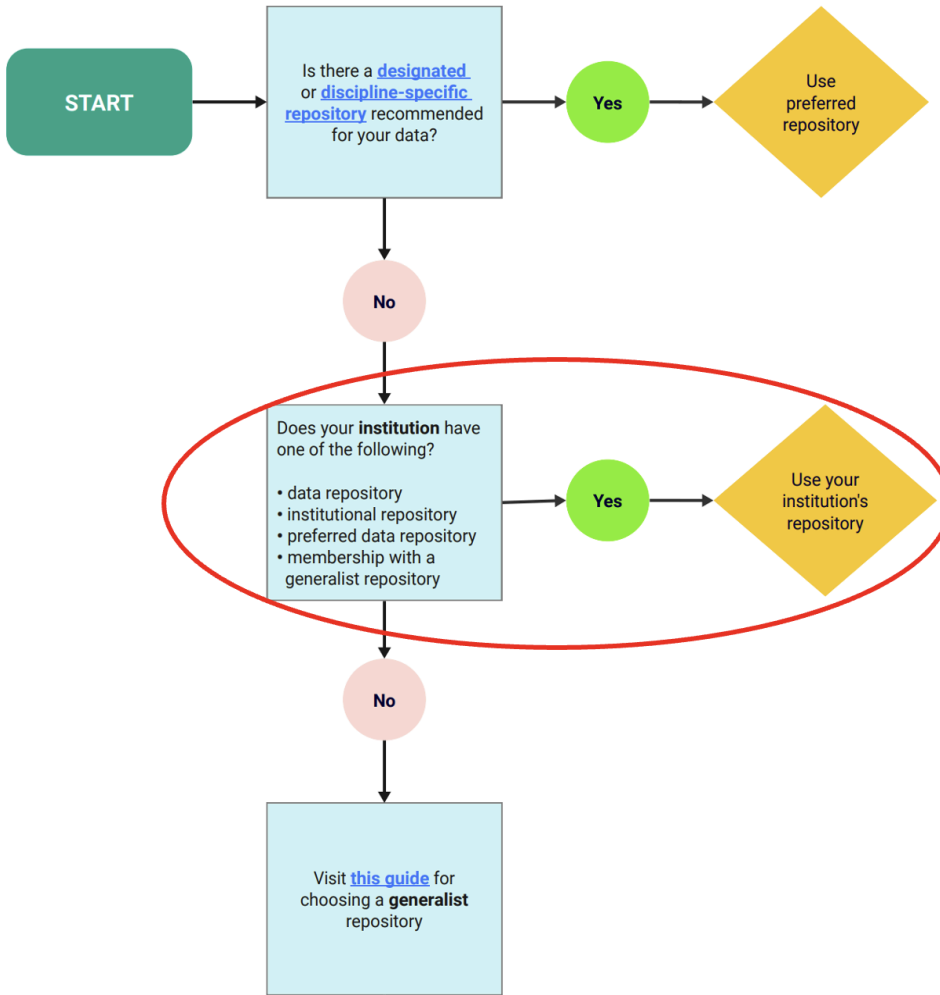
Another important effort comes from the ARL-based [Realities of Academic Data Sharing initiative](#), currently funded by the Institute of Museum and Library Services (IMLS), with all participating institutions as members of the DCN. Notably, the first of its kind expense report ([Hofelich Mohr, et al., 2024](#)), provides detailed insights into the costs researchers and institutions incur to meet data sharing requirements. This work also showed that sharing in institutional infrastructure *may* be correlated with lower direct costs for researchers. The gap analysis report ([Petters et al., 2024](#)) identifies opportunities to better align institutional services with researcher needs. Together, these reports provide a clearer understanding of the resources and support needed to sustain effective data sharing practices. Paired with evidence that curation enhances researcher confidence, they underscore the importance of strengthening institutional services to foster successful data sharing outcomes.

And speaking of showing the importance of local infrastructure, DCN members produced three works that show the value of institutional repositories for stewarding and preserving research data. As funding agencies and publishers increasingly expect data and code to be shared to support scientific claims, the effort data curators put into ensuring scholarly outputs are of a high quality cannot be discounted. In September, a letter to *Science* titled "[Institutional Data Repositories Are Vital](#)," authored by 14 DCN members, outlines why these repositories are essential for data sharing and preservation and called for their broader recognition. Lead author Jen Darragh expanded on these themes in her Duke Libraries blog post, "[Local Data Repository](#)

[Infrastructure Matters](#)," emphasizing how institutional repositories integrate expert curation and local infrastructure to help researchers meet funder and publisher requirements while supporting sustainable and reproducible research.

Additionally, inspired by the need to demonstrate the value of institutional data repositories to senior administrators, beyond compliance, Scout Calvert guided the foundation for an article by DCN authors published in the *Data Science Journal* titled "[Knowledge Infrastructures Are Growing Up: The Case for Institutional \(Data\) Repositories 10 Years After the Holdren Memo](#)" (Narlock et al., 2024). This work built on earlier DCN research from 2017 and 2020, examining the growth of infrastructure and datasets shared through institutional repositories in 2023. Authors visualized the increasing adoption of local infrastructure and the expanding number of stewarded datasets, illustrating how institutional repositories provide significant benefits for both researchers and the institutions that support them.

On another note, the [Generalist Repository Selection Flowchart](#), developed under the the National Institutes of Health (NIH) based [Generalist Repository Ecosystem Initiative](#) (GREI), provides researchers with guidance on selecting the most appropriate repository for their data. Notably, the flowchart emphasizes institutional repositories as the preferred option when no discipline-specific repository is available. By highlighting the role of institutional repositories, the tool reinforces their importance as trusted, local infrastructure for data sharing, preservation, and compliance with funder mandates. This resource supports institutional efforts to showcase the value of their repositories in the broader research data ecosystem.



Cropped image (and added red oval) from: Barbosa, S., Carson, M., Curtain, L., Diggs, S., Gonzales, S., Herzog, J., Olson, E., Snowden, T., Sweet, S., Van Gulick, A., & Wood, J. (2024). Generalist Repository Selection Flowchart. Zenodo. <https://doi.org/10.5281/zenodo.11105430>.

Finally, tied to the broader goal of advancing institutional readiness, the NIH-funded [Summit for Academic Institutional Readiness in Data Sharing](#) (STAIRS) brought together representatives from 32 U.S. academic institutions to address research data sharing needs. Through assessments, a virtual learning series, and an in-person summit, the initiative highlighted the need for sustained investment in personnel, infrastructure, and shared services. Participants emphasized the importance of cross-institutional collaboration and resource-sharing to advance data-sharing practices. The outcomes of STAIRS underscore the role of institutional readiness and collective action in supporting open science and meeting funder and publisher expectations.

As we look to the future of research data management, the evidence is clear: institutional repositories and professional data curation services are not just valuable—they are essential components of the research ecosystem. Through our collaborative efforts in the DCN, we have demonstrated both quantitatively and qualitatively that curated data sharing enhances research quality, accessibility, and impact. The growing adoption of institutional repositories, combined

with the expertise of dedicated curatorial staff, provides a sustainable foundation for meeting evolving funder mandates and advancing open science principles. As research practices continue to evolve, investing in and strengthening institutional data infrastructure will be crucial for ensuring that today's research data remains accessible, usable, and valuable for generations to come.