Evolving research data management services in academic libraries

Sadoqat Raimjonova
Scientific supervisor: Rashid Turgunbayev
Kokand State University

Abstract: The transformation of academic libraries from repositories of published knowledge to active participants in the research data lifecycle represents one of the most significant shifts in the profession since the adoption of online public access catalogs. Research data management has emerged as a critical service area, driven by funder mandates, publisher requirements, and the growing recognition that data are primary research outputs deserving of curation, preservation, and reuse. This article traces the evolution of research data management services in academic libraries, examining how librarians have moved from peripheral advisory roles to embedded partnerships with researchers. It analyzes the competencies required for effective data services, including metadata, preservation, ethics, and domain-specific literacy, while acknowledging the persistent challenges of staffing, infrastructure, and researcher engagement. The discussion further explores the tensions between generalist liaison models and specialized data curation expertise, the role of library publishing platforms in data dissemination, and the emerging frontier of data citation and scholarly credit. Ultimately, the article argues that research data management is not a temporary add-on to existing library work but a fundamental reconceptualization of what academic librarianship means in an open science ecosystem.

Keywords: research data, academic libraries, digital literacy, digital navigators, data management, digital divide

Introduction

For much of library history, the research output that librarians concerned themselves with took a familiar set of forms: the monograph, the journal article, the conference proceeding, the dissertation. These were finished, polished, static objects that entered the library collection after a long chain of peer review, editing, and commercial publishing. The library's role, while vital, was largely passive. It selected, acquired, organized, preserved, and provided access. The messy, iterative, generative work of inquiry that produced those final artifacts remained invisible, housed in researchers' filing cabinets, laboratory notebooks, or, increasingly, on hard drives and cloud storage platforms. That invisibility has ended. Over the past two decades, a quiet revolution has unfolded in how research is conducted, shared, and evaluated. Funder mandates from organizations such as the National Science Foundation, the National Institutes of Health, and Research Councils UK have required that data underlying published findings be preserved and made accessible. Journals have begun requiring data availability statements as a condition of publication. And researchers themselves, particularly in the sciences and social sciences, have come to recognize that their raw and processed data are valuable scholarly assets in their own right, capable of generating new questions, enabling replication studies, and accelerating discovery across disciplinary boundaries. Academic libraries, responding to these pressures and opportunities, have found themselves in an unfamiliar position. They are no longer merely curators of the published record. They have become active partners in the research process itself, helping researchers plan for data management, advising on file formats and metadata standards, providing storage and backup infrastructure, and ensuring that valuable data are not lost to decay, obsolescence,

or simple neglect. This article examines how research data management services have evolved in academic libraries, where they stand today, and the unresolved questions that will shape their future.

From Ad Hoc Advising to Structured Services

The earliest library involvement in research data management was reactive and informal. A graduate student would appear at the reference desk with a corrupted spreadsheet and ask for help. A principal investigator would panic upon realizing that a funder's data management plan was due in forty-eight hours. A department chair would wonder aloud where to deposit fifteen years of longitudinal survey data. Librarians with technical aptitude or personal curiosity would step in, offering whatever assistance they could. These individual efforts, while valuable to the researchers involved, were unsustainable and invisible to library leadership. They did not constitute a service so much as a series of small mercies. The transition to structured services began in earnest around 2010, when several research-intensive universities established dedicated data services positions and units. These early adopters quickly learned that research data management could not be grafted onto existing reference or instruction roles without significant additional training and workload redistribution. A librarian who spent fifteen minutes helping an undergraduate find peer-reviewed sources was not equipped to spend three hours advising a genomics lab on metadata schemas for sequencing data. Structured services required dedicated personnel, often called data librarians or research data management specialists, whose primary responsibility was understanding the data lifecycle and building relationships with research groups. These specialists developed a core set of service offerings. Consultations on data management plans became a standard entry point, with librarians learning to translate funder requirements into practical guidance on file naming conventions, version control, documentation, and backup strategies. Instruction sessions on data management best practices were integrated into graduate student orientations and doctoral seminars. And curation services emerged for data that had no obvious disciplinary repository, with libraries offering trusted digital repositories that promised long-term preservation, persistent identifiers, and controlled access where necessary.

The Competency Challenge

No discussion of evolving research data management services would be complete without confronting the uncomfortable question of librarian competencies. Most academic librarians entered the profession through humanities or social science backgrounds, with master's degrees that emphasized descriptive cataloging, reference interviewing, collection development, and library administration. These are honorable and useful skills, but they do not automatically translate to the world of data. Understanding the difference between a relational database and a flat file, knowing why one chooses XML over CSV for certain kinds of data, being able to advise on de-identification of human subjects data, and recognizing the signs of impending file format obsolescence are not skills acquired in typical library science curricula. The profession has responded with continuing education, graduate certificates, and even specialized master's programs in data curation. Organizations such as the Digital Curation Centre and the Research Data Alliance have produced extensive training materials and communities of practice. Yet a persistent gap remains between the demand for data services and the supply of librarians qualified to provide them. Some libraries have addressed this gap by hiring from outside traditional library channels, bringing on board individuals with degrees in information science, computer science, or domain sciences who also possess the service orientation that libraries value. Others have upskilled existing staff through intensive mentorship and released time for training. A third model involves distributed expertise, where a small number of highly trained data specialists serve as internal consultants to subject librarians, who in turn maintain primary

relationships with academic departments. Each model has strengths and weaknesses, and the optimal approach likely depends on institutional size, research intensity, and existing organizational culture.

Embeddedness Versus the Service Desk Model

A critical evolution in research data management services has been the movement from a centralized, desk-based model to an embedded, partnership-based model. In the traditional reference model, patrons come to a physical or virtual service point with a question, and a librarian provides an answer. This transaction works reasonably well for factual queries or navigation of known resources. It works poorly for research data management, where the questions are complex, the stakes are high, and the relevant context is dispersed across years of a research project. A researcher who needs help with a data management plan does not want a thirty-minute conversation with a stranger at a desk. They want a trusted advisor who understands their disciplinary practices, their laboratory's workflows, and the specific requirements of their current grant application. Achieving this level of partnership requires librarians to leave the library. Data librarians increasingly attend laboratory meetings, sit on research ethics boards, collaborate on grant proposals as co-investigators, and embed themselves in the daily rhythms of academic departments. This embeddedness builds the trust and mutual understanding necessary for effective data stewardship. A researcher who knows that the data librarian has attended three of their lab's weekly meetings is far more likely to ask for help before a crisis, rather than after. Yet embeddedness carries costs. It is time-intensive, scaling poorly to large institutions with hundreds of research groups. It requires librarians to accept a degree of unpredictability and boundarylessness that can lead to burnout. And it raises questions about impartiality: if a librarian is embedded in a particular research group, can that librarian also provide neutral advice about data sharing or repository selection that might conflict with the group's preferences? The most successful programs have balanced embedded relationship-building with a clear articulation of the librarian's professional obligations to the institution and the broader scholarly community.

Infrastructure and the Politics of Storage

Underlying all research data management services is a deceptively simple question: where will the data live? This question quickly becomes complex, touching on institutional strategy, information technology politics, and significant financial investments. Academic libraries have historically controlled their own infrastructure for traditional collections. The integrated library system, the discovery layer, the institutional repository for publications, and the digital preservation system all fall within the library's administrative domain. Research data management disrupts this tidy arrangement. Data storage requirements for a single research project can dwarf the entire library's existing digital collections. Active data, which researchers need to access quickly for ongoing analysis, demands different infrastructure from archival data, which can be stored on slower, cheaper media. And many campuses already have centralized research computing and data centers that hold budgetary and political power far exceeding that of the library. Libraries entering the data management space have had to navigate these politics carefully. Some have ceded infrastructure entirely, focusing their efforts on policy, training, and curation of selected high-value datasets while relying on campus information technology partners for bulk storage. Others have built their own library-managed data repositories, arguing that long-term preservation requires the kind of dedicated attention to metadata, fixity checking, and format migration that general-purpose research computing groups rarely prioritize. A middle path has emerged around the concept of the trusted digital repository. Libraries do not need to own the physical storage hardware to provide repository services. They can contract with consortial partners, commercial cloud providers, or even other campus units, as long as they maintain control over curation decisions and certification standards. The key is for

libraries to articulate what only they bring to the data infrastructure conversation: a preservation mandate that transcends individual grant cycles, expertise in metadata and discovery, and a commitment to open access and equitable reuse.

Data Curation as Distinct From Data Archiving

A persistent confusion in discussions of research data management is the conflation of archiving with curation. Archiving, in the traditional library sense, implies ingestion, storage, and occasional retrieval. Curation implies active intervention throughout the data lifecycle. The distinction matters profoundly for library services. A library that merely offers to accept researchers' data at the end of a project, store it on a server, and mint a persistent identifier has performed an archival function. This is useful but limited. It does nothing to ensure that the data are interpretable by anyone other than the original researchers. It does nothing to improve the data's structure, documentation, or metadata. It does nothing to link the data to related publications, code, or materials. True curation involves work that begins long before deposit. A data curator advises researchers on file formats and folder structures during the active research phase, making later deposit smoother. They review documentation and metadata for completeness and clarity, requesting additional information from researchers when necessary. They transform data from proprietary to open formats where permissible. They create data dictionaries and readme files. They assess data for disclosure risk and advise on appropriate access controls. And they establish preservation plans that account for likely technological changes over a ten, twenty, or fifty-year horizon. This curatorial work is labor-intensive and requires specialized judgment. It cannot be automated or outsourced to graduate students. Libraries that have built successful data curation services have been honest about these costs, securing ongoing operational funding rather than relying on soft money or volunteer effort. They have also been selective, curating deeply the datasets with highest long-term value rather than attempting to accept every file that a researcher offers.

The Open Science Turn

Research data management is not an end in itself. It serves a larger purpose: enabling open science. The open science movement seeks to make not only publications but also data, code, methods, and peer review openly available by default. Data management is the scaffolding that makes open science possible. Without well-managed, well-documented, accessible data, the principle of openness collapses into mere availability of uninterpretable files. Academic libraries, with their long tradition of open access advocacy, have naturally positioned themselves as champions of open science. Yet tensions arise when open science principles encounter researcher realities. A geneticist studying a rare disease may be willing to share summary statistics but not individual-level genomic data due to privacy concerns and informed consent restrictions. A social scientist studying political violence in an authoritarian context may face genuine risks to research participants if raw data are released. An industry-funded engineer may be contractually prohibited from sharing certain data at all. Responsible data management services must navigate these tensions, helping researchers make their work as open as possible and as closed as necessary. This requires nuanced understanding of data sensitivity, anonymization techniques, controlled access mechanisms, and the legal frameworks governing data sharing. The library's role is not to enforce a maximalist open data policy but to provide the expertise and infrastructure that allow researchers to make informed, ethical, and compliant decisions about their data.

Measuring Success and Demonstrating Value

As research data management services mature, academic libraries face increasing pressure to demonstrate their value to institutional stakeholders. Traditional library metrics are ill-suited to this task. Counting the number of data management plan consultations provided or the terabytes stored

tells administrators little about whether those services actually improved research outcomes or reduced risk. More meaningful indicators include the success rate of grant applications that received library support, measured against a baseline of unsupported applications. They include longitudinal tracking of datasets to see whether they are downloaded, cited, or reused by other researchers. They include surveys of researcher satisfaction and self-reported reduction in data-related anxiety or loss incidents. They include compliance audits showing that library-curated data meet funder and journal requirements for preservation and accessibility. Some libraries have gone further, conducting qualitative interviews with researchers who have experienced data emergencies, such as a corrupted hard drive or a lost laptop, to understand whether prior library training or infrastructure mitigated the damage. The most sophisticated assessment frameworks align library metrics with institutional strategic goals. If a university prioritizes interdisciplinary research, the library can demonstrate how shared data repositories lower barriers between departments. If a university prioritizes undergraduate research training, the library can show how data management instruction for early-career researchers builds institutional capacity. The key is moving from activity counts to outcome narratives, from what the library did to what changed because of what the library did.

Conclusion

The evolution of research data management services in academic libraries is incomplete, and its final destination remains uncertain. Yet one conclusion is already clear: the old metaphor of the librarian as steward of a finished collection no longer suffices. Stewardship implies care for something complete and static, something that exists independently of the steward's intervention. Data are not like that. Data are living, growing, changing. They acquire meaning through documentation, structure through description, and value through connection to other research outputs. The librarian who works with data cannot stand at a distance, preserving what already exists. That librarian must step into the research process itself, becoming a participant, a collaborator, and sometimes a critical friend. This is a demanding identity shift, requiring technical skills that many current librarians do not possess and comfort with ambiguity that many never sought. It is also an exhilarating one. To help a doctoral student design a data management plan that will safeguard years of dissertation research, to assist a research team in selecting a repository that will make their findings discoverable for decades, to recover a dataset that a principal investigator thought was lost forever, to receive an email from a researcher on another continent who has built upon data that your library curated - these are not transactions. They are contributions to the scholarly enterprise itself. Academic libraries that embrace research data management as a core service will not merely survive the open science transition. They will help define it, shaping the norms, infrastructure, and practices that determine how knowledge is created, shared, and preserved for generations to come. That is a future worth building.

References

1. Andrikopoulou, A., Rowley, J., & Walton, G. (2022). Research data management (RDM) and the evolving identity of academic libraries and librarians: A literature review. New Review of Academic Librarianship, 28(4), 349-365.

2. Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. Library & information science research, 36(2), 84-90.

3. Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. Journal of librarianship and information science, 46(4), 299-316.

4. Sheikh, A., Malik, A., & Adnan, R. (2025). Evolution of research data management in academic libraries: A review of the literature. Information Development, 41(2), 305-319.

5. Xu, Z. (2022). Research data management practice in academic libraries. Journal of Librarianship and Scholarly Communication, 10(1).

6. Flores, J. R., Brodeur, J. J., Daniels, M. G., Nicholls, N., & Turnator, E. (2015). Libraries and the research data management landscape. The process of discovery: The CLIR postdoctoral fellowship program and the future of the academy, 2010, 82-102.

7. Pinfield, S., Cox, A. M., & Smith, J. (2014). Research data management and libraries: Relationships, activities, drivers and influences. PLoS one, 9(12), e114734.

8. Corrall, S., Kennan, M. A., & Afzal, W. (2013). Bibliometrics and research data management services: Emerging trends in library support for research. Library trends, 61(3), 636-674.

9. Subaveerapandiyan, A. (2023). Research data management practices and challenges in academic libraries: A comprehensive review. Library Philosophy and Practice, 1-106.