

# A Study of Data-Focused Roles in Leading North American University Libraries

Ahmad R. Pratama

Data Literacies Lead, University Libraries, Stony Brook University, Stony Brook NY 11794

[ahmad.pratama@stonybrook.edu](mailto:ahmad.pratama@stonybrook.edu)

The published version of this manuscript in *The Journal of Academic Librarianship* is available at

<https://doi.org/10.1016/j.acalib.2025.103102>

## Abstract

In the past decade, academic libraries have evolved to support diverse data needs, creating a demand for specialized data-focused roles. North American academic libraries are widely regarded as a global benchmark for developing data services, owing to their strong research infrastructure and extensive administrative support. This study examines how libraries at leading members of the Association of American Universities (AAU) staff and structure these data-focused roles. Using a combined approach of directory reviews, site-specific searches, statistics, and text analysis, the study identifies 321 data-focused roles across 71 AAU libraries at the beginning of 2024–2025 academic year, averaging 2.30% of each library's total staff. Although generalist data librarian roles predominate, specialized positions, including non-librarian staff also play important parts in front-end data services. Results reveal no statistically significant differences between public and private AAU libraries in either the prevalence or percentage of data-focused roles, suggesting that administrative priorities may outweigh funding models in shaping data services capacity. Text analysis underscores the prominence of more general data terms within job titles, reflecting a broad focus on the entire data lifecycle in most institutions, while some more advanced and niche technical skills are supported in a few institutions. Furthermore, this study proposes a three-stage model of data services maturity: (1) basic Research Data Management (RDM) essentials, (2) general data support, and (3) specialized data expertise, with most AAU libraries occupying Stages 2 or 3, reflecting their readiness to address complex data needs. For libraries worldwide seeking to initiate or enhance data services, these results highlight strategic considerations such as flexible hiring practices, professional development, and dedicated web visibility. This study offers a replicable framework for other academic libraries worldwide aiming to progress from foundational RDM to advanced data specialization.

Keywords: academic libraries, data services, data librarianship, public vs. private, three-stage maturity model

## Introduction

We live in an age where data permeates every aspect of both our personal and professional lives. From social media interactions, e-commerce transactions, and electronic medical records, to data generated by sensors, wearable devices, and other types of internet of things (IoT), to sophisticated scientific research. Academic libraries, long seen as stewards of scholarly resources, are increasingly expected to support researchers across the entire data lifecycle, from planning and analysis to sharing and preservation (Nitecki & Davis, 2019). As the volume and complexity of available data continue to skyrocket, it has prompted universities to reassess their support systems. Academic libraries, in particular, are further expected to expand their traditional missions to provide comprehensive data services so they can better support researchers who generate, analyze, and preserve diverse datasets.

Yet, just a decade ago, the notion of “research data services” was relatively new among university libraries worldwide, and this was no exception in North America (Tenopir et al., 2014). However, some universities acting as early adopters were already beginning to pilot basic data management programs. This resulted in the shift from the “data librarian” term that was initially considered more as a buzzword to become a widely recognized job title in just a few years (Khan & Du, 2017). Since then, increasing demands for data expertise have driven the formal creation of data librarian and related specialist positions, particularly in North America. As one study found, of 104 data librarian job advertisements published online between 2013 and 2018, more than 95% were from American and Canadian universities (Eclevia et al., 2019).

Despite this growth, there is limited empirical evidence on how academic libraries staff and structure these data-focused roles. To make it even more complicated, many researchers themselves often have limited awareness of data services offered by their libraries. As one study in a single university setting shows, even as late as 2022, their researchers did not automatically associate data management plan (DMP) support with libraries (Sheffield & Burton, 2022). Furthermore, DMP requirement mandates from funding agencies like the National Science Foundation (NSF) or the National Institutes of Health (NIH) are often unfunded, leaving unclear responsibilities for their implementation and evaluation across the university (W. Bishop et al., 2021). Therefore, understanding how academic libraries staff, structure, and communicate their data services become essential to addressing these important gaps effectively.

Given this context, North American universities can serve as a valuable benchmark for academic libraries worldwide that aim to launch or strengthen data-related services. The Association of American Universities (AAU), established in 1900, brings together 71 high-profile research institutions in the United States and Canada (Association of American Universities, 2025). In North American context, AAU plays a pivotal role similar to the Russell Group in the United Kingdom (The Russell Group of Universities, 2025) or the Group of Eight in Australia (The Group of Eight, 2025). Analyzing the library structures within these prominent AAU members, therefore, can help us better understand how data support services are staffed and organized in their libraries. This, in turn, can help offer models for other institutions across the globe that wish to adopt similar practices.

Accordingly, this study addresses the following questions:

1. How prevalent are data-focused roles in these leading research university libraries?
2. What trends or patterns emerge from publicly available job titles and departmental structures?
3. How might these patterns inform the global academic library community's approach to implementing or expanding data services?

By answering these questions, this study aims to provide international readers with actionable insights into the evolving landscape of data librarianship, thereby fostering knowledge exchange and offering strategic guidance for academic libraries across the globe.

## **Literature Review**

Academic librarians have been experiencing a significant shift in their roles and responsibilities for some time now. While in the past their main roles and responsibilities were managing bibliographic resources, they are now expected to steward and disseminate research data and publication outputs. Yet, as of 2018, there was no established framework guiding librarians through this transition, leaving many of them, especially in smaller institutions, feeling overwhelmed and lacking confidence in supporting researchers with their needs for data services (Nitecki & Davis, 2019). This is where data-focused roles in libraries can fill the gaps.

For this study, data-focused roles in libraries refer to user-facing positions that support or provide services throughout the entire data lifecycle from planning and creation to storage, dissemination, and preservation. While many libraries employ individuals who work extensively with data in back-end capacities (e.g., metadata librarians), such roles are not considered data-focused roles in this study unless their job descriptions explicitly indicate engagement in front-end data services for library patrons. These data-focused roles are often held by individuals with librarian status (e.g., data librarians), but in some cases, they can also be designated as staff positions with more specialized job descriptions (e.g., data science specialists).

Recent literature highlights two critical factors influencing the success or failure of these emerging data-focused roles: (1) the professional capacities of library staff, and (2) the supportive leadership within library administration. On the one hand, librarians themselves must possess or develop the competencies to navigate and apply rapidly evolving data technologies. On the other hand, library leaders and administrators must provide the necessary infrastructural support, strategic direction, and advocacy to integrate data services seamlessly into broader library operations (Ashiq & Warraich, 2023).

Data librarianship is often portrayed as dynamic and interdisciplinary (Semeler & Pinto, 2020). This means the roles require professional development, certification, and training that bridge multiple domains (e.g., information science, computer programming, or data management). While some have argued that data librarians do not need to become expert programmers, statisticians, or database managers, they do benefit from familiarity with core concepts of computational logic, data science, and specialized research practices

(Semeler et al., 2019). The opportunities to collaborate with scientific data researchers, social scientists, and digital humanities scholars also highlight the breadth of potential engagement.

The scope of data services in academic libraries can vary significantly from one institution to another.

However, one model in the literature identifies three tiers: (1) education, which includes training, workshops, or research guides, (2) consultation, which is essentially individualized guidance on any data-related activities throughout the entire data lifecycle, and (3) infrastructure, the technical environment for data storage, curation, and dissemination (Ohaji et al., 2019). Other studies further indicate that while traditional library skills like cataloging and collection development may be less central for data librarians (Federer, 2018), soft skills like reference skills, relationship building, collaboration, listening, and facilitation, remain crucial (Henderson, 2020).

Reflecting the diverse nature of data librarian work, one study employing cluster analysis of survey data revealed two broad types of data librarians: (1) generalists, who offer a wide range of data services across multiple subjects, and (2) subject specialists, who focus on distinct disciplinary communities (Federer, 2018). In practice, many data librarians balance tasks such as locating secondary data, reviewing DMPs, conducting outreach, collaborating, and offering Research Data Management (RDM) training (B. W. Bishop et al., 2022). This mix of responsibilities can vary considerably, underscoring the adaptable nature of the role. As a result, libraries seeking to launch or enhance data services must anticipate both the technical and human elements necessary to make these initiatives sustainable and effective.

Additionally, there is ongoing debate among librarians and library administrators regarding whether a Master of Library and Information Science (MLIS) degree is necessary for data-focused roles in libraries. Some argue that MLIS programs do not always adequately prepare professionals for the specialized skill sets required in data services, as these programs may not cover data science, programming, or advanced analytics in depth (Thomas & Urban, 2018). A related study explores the potential benefits of eliminating the requirement of a Master of Library and Information Science (MLIS) for branch library management positions, as many of the skills necessary for library jobs (e.g., building community relationships and overseeing budgets) may derive from professional or academic experiences outside traditional MLIS programs (Huggins, 2022). This argument may also apply to data librarianship or any other data-focused roles. Waiving the MLIS requirement for these positions could allow libraries to tap into a more diverse candidate pool with direct expertise in data analysis, data science, or information technology.

Last, but not least, one study examining trends within AAU libraries from 2014 to 2019 reported that 54 out of 61 AAU libraries already offered dedicated data services with 39 of them employing dedicated data librarians by 2019 (Gowen & Meier, 2019). It is also important to note that many of them have RDM mentioned in their strategic plans. While it did not directly correlate with data services expansions, there was a noticeable trend toward integrating data into institutional repositories and developing dedicated data repositories. This is a strong indication that many AAU libraries have a strong institutional presence and support for data services for some time now.

## Methodology

### Data Collection

This study began by identifying North American research universities listed on the AAU website, from which each member university and its official website was identified (Association of American Universities, 2025). For each institution, the library's subdomain (e.g., library.university.edu) was visited, and staff directories were examined to identify the total number of library staff and the presence of any data-focused roles. Where possible, brief role descriptions or statements of responsibility were also collected. Many institutions provided online directories with sorting or search features based on department, subject expertise, or job titles. In such cases, relevant filters were applied to isolate positions containing relevant keywords such as "data", "digital", "geospatial", "GIS", "research", and "visualization". If directories did not offer such filters, the built-in browser search function was used to manually scan the directory for relevant keywords.

To ensure focus on front-end user services, roles primarily functioning as back-end support for library systems and resources were excluded. Specifically, any positions with "metadata" in their titles or any hybrid roles like "Digital Scholarship Librarian" or "Research Services Librarian" went through additional manual review by a single reviewer to determine whether they entailed significant user-facing data services or whether they were oriented toward cataloging and digital asset management. More specifically, they were included in the dataset only if their job descriptions explicitly mentioned substantial front-end data responsibilities; otherwise, they were excluded to maintain clarity and consistency in classification. The final dataset, therefore, emphasizes library roles that provide or support data services directly to patrons (e.g., students, faculty, staff, and external researchers).

Additionally, to capture any data services web pages or related positions that might not appear in staff directories, a Google site-specific search was conducted using the site: operator for each library subdomain (e.g., site:library.university.edu "data"). This approach helped locate the library data services landing page, departmental pages, staff profiles, and any announcements mentioning the term "data" while limiting results strictly to the library's official subdomain rather than the entire university's website. By confining the search scope, the study focused on library-based data support services, excluding broader data initiatives led by other campus units outside the library.

The entire dataset was collected in late August 2024, approximately aligning with the start of the Fall 2024 term in North America. This timing increased the likelihood of capturing recently updated staffing changes or new hires for the 2024–2025 academic year. In total, this process yielded a dataset of job titles, departmental affiliations, and, where possible, brief role descriptions or position statements. Additionally, the use of a single reviewer process does raise potential limitations of inter-rater reliability. To mitigate this, the manual review process was conducted in two rounds. The initial round identified data-focused roles and preliminary classifications. A second verification round revisited ambiguous cases, refining classifications to enhance consistency and reliability.

## Data Analysis

The collected dataset was analyzed using both quantitative and qualitative methods to extract meaningful insights. The analysis began with descriptive statistics to assess the prevalence of dedicated data roles, measuring (a) the total number of such positions per library and (b) their proportion relative to total library staff. This step also identified outliers and highlighted top-ranking institutions in both absolute numbers and percentage terms.

Next, inferential statistics were applied to compare public and private AAU libraries across various key metrics (i.e., total staff size, total data-focused roles, the proportion of data-focused roles among total staff size, total data librarians, and the proportion of data librarians among all data-focused roles). Appropriate hypothesis tests were selected based on the data's distribution and the type of variables involved.

To further clarify role definitions and institutional structures, text analysis was conducted. This included tokenizing job titles, filtering out stopwords, and performing n-gram extraction (unigrams, bigrams, and trigrams) to identify dominant phrases. Some keywords were used to identify specialized roles from more generalist data support, as well as to distinguish leadership roles within data services from non-leadership positions. In addition, the text analysis extended to libraries' web presence to determine whether the library hosted a dedicated data services page, and if so, the specific terminology used.

All analyses were conducted using Python scripts in a Google Colab environment. To promote transparency and facilitate future research, the code and associated data (after removing individual names and contact information) are openly available in the author's GitHub repository<sup>1</sup>. Researchers wishing to replicate or extend this work can access the source files, run the Python scripts, and adapt them for future longitudinal studies or cross-institutional comparisons.

## Results

### Prevalence of Data-Related Roles

By August 2024, a total of 321 data-focused roles were identified across 71 AAU libraries, with only three lacking a dedicated data-focused role. This is a testament to the near-universal adoption of data-focused roles within these institutions. Where such roles did exist, they constituted an average of 2.30% of total library staff (standard deviation = 1.41%; median = 2.08%). This figure indicates that while the proportion of data-focused roles is still relatively small, a majority of libraries now make explicit efforts to hire and maintain such expertise.

In terms of role classification, librarian roles dominated the dataset, accounting for 65.11% of all data-focused roles in these libraries. Some of these data librarians held faculty status (tenure-track or tenured), reflecting certain institutions' practice of treating librarians as faculty-equivalent. Non-librarian staff roles made up the

---

<sup>1</sup> <https://github.com/ahmadrpratama/data-roles-aau-libraries>

second largest category at 32.71%, often featuring more specialized or technical responsibilities (e.g., “Data Analyst”, “Data Specialist”, “GIS Specialist”, “Data Visualization Specialist”). Other positions, such as postdoctoral fellows, research assistants, or visiting scholars appeared in only four institutions, namely University of California, Davis (3 positions), University of Pennsylvania (2 positions), Carnegie Mellon University (1 position), and New York University (1 position), comprising 2.18% of total data-focused personnel overall. The results are summarized in Table 1.

Table 1. Distribution of data-focused roles in AAU member libraries

No	University	Status	Total Staff	Data-focused Roles	Data-focused Librarian	Data-focused Staff	Other Data Roles
1	Harvard University	Private	346	18 (5.20%)	10 (55.56%)	8 (44.44%)	0 (0.00%)
2	University of Toronto	Public	516	16 (3.10%)	10 (62.50%)	6 (37.50%)	0 (0.00%)
3	New York University	Private	386	14 (3.63%)	8 (57.14%)	5 (35.71%)	1 (7.14%)
4	Princeton University	Private	370	11 (2.97%)	4 (36.36%)	7 (63.64%)	0 (0.00%)
5	University of Pennsylvania	Private	355	10 (2.82%)	4 (40.00%)	4 (40.00%)	2 (20.00%)
6	University of California, Davis	Public	156	9 (5.77%)	2 (22.22%)	4 (44.44%)	3 (33.33%)
7	Duke University	Private	224	9 (4.02%)	2 (22.22%)	7 (77.78%)	0 (0.00%)
8	University of Washington	Public	305	9 (2.95%)	9 (100.00%)	0 (0.00%)	0 (0.00%)
9	Stanford University	Private	371	9 (2.43%)	4 (44.44%)	5 (55.56%)	0 (0.00%)
10	John Hopkins University	Private	149	7 (4.70%)	2 (28.57%)	5 (71.43%)	0 (0.00%)
11	University of North Carolina at Chapel Hill	Public	226	7 (3.10%)	7 (100.00%)	0 (0.00%)	0 (0.00%)
12	Texas A&M University	Public	227	7 (3.08%)	3 (42.86%)	4 (57.14%)	0 (0.00%)
13	Cornell University	Private	300	7 (2.33%)	7 (100.00%)	0 (0.00%)	0 (0.00%)
14	University of Minnesota	Public	317	7 (2.21%)	4 (57.14%)	3 (42.86%)	0 (0.00%)
15	Yale University	Private	595	7 (1.18%)	5 (71.43%)	2 (28.57%)	0 (0.00%)
16	Carnegie Mellon University	Private	90	6 (6.67%)	3 (50.00%)	2 (33.33%)	1 (16.67%)
17	Purdue University	Public	138	6 (4.35%)	3 (50.00%)	3 (50.00%)	0 (0.00%)
18	University of Pittsburgh	Public	149	6 (4.03%)	4 (66.67%)	2 (33.33%)	0 (0.00%)
19	Washington University in St. Louis	Private	158	6 (3.80%)	2 (33.33%)	4 (66.67%)	0 (0.00%)
20	Arizona State University	Public	183	6 (3.28%)	3 (50.00%)	3 (50.00%)	0 (0.00%)
21	University of Colorado Boulder	Public	288	6 (2.08%)	6 (100.00%)	0 (0.00%)	0 (0.00%)
22	University of California, Los Angeles	Public	310	6 (1.94%)	3 (50.00%)	3 (50.00%)	0 (0.00%)
23	University of Michigan	Public	423	6 (1.42%)	5 (83.33%)	1 (16.67%)	0 (0.00%)
24	University of California, Santa Cruz	Public	107	5 (4.67%)	2 (40.00%)	3 (60.00%)	0 (0.00%)
25	University of Oregon	Public	134	5 (3.73%)	3 (60.00%)	2 (40.00%)	0 (0.00%)

26	University of Virginia	Public	220	5 (2.27%)	2 (40.00%)	3 (60.00%)	0 (0.00%)
27	University of California, San Diego	Public	231	5 (2.16%)	4 (80.00%)	1 (20.00%)	0 (0.00%)
28	University of Illinois Urbana-Champaign	Public	379	5 (1.32%)	5 (100.00%)	0 (0.00%)	0 (0.00%)
29	George Washington University	Private	107	4 (3.74%)	2 (50.00%)	2 (50.00%)	0 (0.00%)
30	Dartmouth College	Private	138	4 (2.90%)	3 (75.00%)	1 (25.00%)	0 (0.00%)
31	University of Miami	Private	145	4 (2.76%)	4 (100.00%)	0 (0.00%)	0 (0.00%)
32	Northwestern University	Private	175	4 (2.29%)	3 (75.00%)	1 (25.00%)	0 (0.00%)
33	Rutgers University	Public	177	4 (2.26%)	4 (100.00%)	0 (0.00%)	0 (0.00%)
34	University of Arizona	Public	185	4 (2.16%)	4 (100.00%)	0 (0.00%)	0 (0.00%)
35	University of Chicago	Private	204	4 (1.96%)	3 (75.00%)	1 (25.00%)	0 (0.00%)
36	Columbia University	Private	284	4 (1.41%)	3 (75.00%)	1 (25.00%)	0 (0.00%)
37	Penn State University	Public	564	4 (0.71%)	2 (50.00%)	2 (50.00%)	0 (0.00%)
38	Tufts University	Private	58	3 (5.17%)	3 (100.00%)	0 (0.00%)	0 (0.00%)
39	University of California, Riverside	Public	79	3 (3.80%)	2 (66.67%)	1 (33.33%)	0 (0.00%)
40	University at Buffalo	Public	132	3 (2.27%)	3 (100.00%)	0 (0.00%)	0 (0.00%)
41	University of California, Irvine	Public	137	3 (2.19%)	3 (100.00%)	0 (0.00%)	0 (0.00%)
42	Vanderbilt University	Private	152	3 (1.97%)	3 (100.00%)	0 (0.00%)	0 (0.00%)
43	Massachusetts Institute of Technology	Private	155	3 (1.94%)	3 (100.00%)	0 (0.00%)	0 (0.00%)
44	Michigan State University	Public	202	3 (1.49%)	3 (100.00%)	0 (0.00%)	0 (0.00%)
45	University of Southern California	Private	202	3 (1.49%)	3 (100.00%)	0 (0.00%)	0 (0.00%)
46	Emory University	Private	220	3 (1.36%)	3 (100.00%)	0 (0.00%)	0 (0.00%)
47	University of Utah	Public	270	3 (1.11%)	2 (66.67%)	1 (33.33%)	0 (0.00%)
48	University of Wisconsin-Madison	Public	283	3 (1.06%)	3 (100.00%)	0 (0.00%)	0 (0.00%)
49	Brandeis University	Private	56	2 (3.57%)	2 (100.00%)	0 (0.00%)	0 (0.00%)
50	McGill University	Public	98	2 (2.04%)	1 (50.00%)	1 (50.00%)	0 (0.00%)
51	Rice University	Private	100	2 (2.00%)	2 (100.00%)	0 (0.00%)	0 (0.00%)
52	University of Rochester	Private	102	2 (1.96%)	2 (100.00%)	0 (0.00%)	0 (0.00%)
53	Indiana University	Public	113	2 (1.77%)	2 (100.00%)	0 (0.00%)	0 (0.00%)
54	University of California, Berkeley	Public	120	2 (1.67%)	2 (100.00%)	0 (0.00%)	0 (0.00%)
55	Brown University	Private	133	2 (1.50%)	2 (100.00%)	0 (0.00%)	0 (0.00%)
56	University of Kansas	Public	135	2 (1.48%)	1 (50.00%)	1 (50.00%)	0 (0.00%)
57	University of Iowa	Public	187	2 (1.07%)	2 (100.00%)	0 (0.00%)	0 (0.00%)
58	University of Texas at Austin	Public	207	2 (0.97%)	0 (0.00%)	2 (100.00%)	0 (0.00%)
59	Ohio State University	Public	211	2 (0.95%)	2 (100.00%)	0 (0.00%)	0 (0.00%)



60	University of Florida	Public	248	2 (0.81%)	2 (100.00%)	0 (0.00%)	0 (0.00%)
61	California Institute of Technology	Private	42	1 (2.38%)	0 (0.00%)	1 (100.00%)	0 (0.00%)
62	Stony Brook University	Public	53	1 (1.89%)	1 (100.00%)	0 (0.00%)	0 (0.00%)
63	Georgia Institute of Technology	Public	89	1 (1.12%)	0 (0.00%)	1 (100.00%)	0 (0.00%)
64	Case Western Reserve University	Private	117	1 (0.85%)	0 (0.00%)	1 (100.00%)	0 (0.00%)
65	University of South Florida	Public	117	1 (0.85%)	0 (0.00%)	1 (100.00%)	0 (0.00%)
66	University of California, Santa Barbara	Public	122	1 (0.82%)	1 (100.00%)	0 (0.00%)	0 (0.00%)
67	University of Missouri	Public	175	1 (0.57%)	1 (100.00%)	0 (0.00%)	0 (0.00%)
68	University of Maryland	Public	206	1 (0.49%)	1 (100.00%)	0 (0.00%)	0 (0.00%)
69	Boston University	Private	91	0 (0.00%)	-	-	-
70	Tulane University	Private	94	0 (0.00%)	-	-	-
71	University of Notre Dame	Private	132	0 (0.00%)	-	-	-
<b>Total</b>			-	<b>14,470</b>	<b>321 (2.22%)</b>	<b>209 (65.11%)</b>	<b>105 (32.71%)</b>
						<b>7 (2.18%)</b>	

When focusing on raw counts, Harvard University reported the highest number of data-focused roles, with 18 roles out of 346 total library positions (5.20%). The University of Toronto followed with 16 such roles among 516 positions (3.10%), and New York University had 14 out of 386 positions (3.63%). However, based on the percentage of data-focused roles relative to overall library positions, Harvard placed third behind Carnegie Mellon University that led with 6.67% (6 out of 90), followed by University of California, Davis at 5.77% (9 out of 156). These patterns illustrate that even smaller libraries (in terms of total library positions) can demonstrate a high relative focus on data services.

### Public vs. Private Comparisons

To test whether funding models influence staffing for data-focused roles, a comparison between public and private AAU libraries was made across key indicators, and the result is presented in Table 2. Using Mann-Whitney U tests due to non-normal distribution of the data, no statistically significant differences emerged, whether in total library staffing ( $p = .496$ ), number of data-focused positions ( $p = .493$ ), proportion of data-focused roles ( $p = .181$ ), number of data librarians ( $p = .840$ ), or the percentage of data librarians among all data-focused roles ( $p = .294$ ). On average, public AAU libraries reported 206.12 total staff ( $SD = 110.17$ ) vs. 200.63 ( $SD = 129.23$ ) in private ones. While public AAU libraries maintained a slightly lower mean percentage of data-focused positions (2.13% vs. 2.55%), the gap was neither substantial nor statistically significant. Overall, institutional priorities and leadership appear more decisive than funding models in shaping libraries' commitment to data services.

Table 2. Differences in data-focused staffing between public and private AAU libraries

Metric	Mean (SD)	Median	p-value
Total library staff			.327
• All	203.80 (117.74)	175.00	
• Public	210.48 (112.47)	186.00	
• Private	195.19 (125.56)	152.00	
Total data-focused roles			.632
• All	4.52 (3.54)	4.00	
• Public	4.20 (2.95)	3.50	
• Private	4.94 (4.19)	4.00	
Percentage of data-focused roles			.169
• All	2.31% (1.40%)	2.08	
• Public	2.12% (1.24%)	1.99	
• Private	2.55% (1.57%)	2.33	
Total data librarian roles			.776
• All	2.94 (2.19)	3.00	
• Public	2.92 (2.18)	2.50	
• Private	2.97 (2.24)	3.00	
Percentage of data librarian roles among all data-focused roles			.497
• All	70.25% (32.01%)	75.00%	
• Public	72.20% (31.81%)	81.67%	
• Private	67.47% (32.67%)	75.00%	

## Text Analysis of Job Titles

Subsequent text analysis of the collected job titles via n-grams underscored the prominence of phrases containing the word “data”. The unigram in Figure 1 reveals that 73.52% of the job titles contained the word “data”, and only 42.68% contained the word “librarian”. The next three common words in the job titles are “research” at 27.10%, “services” at 19.94%, and “specialist” at 16.82%, reflecting the overarching emphasis on research data support. Meanwhile, terms like “digital”, “scholarship”, and “GIS” appeared less commonly but signaled historically rooted (e.g., “Digital Scholarship Librarian”) or more niche roles (e.g., “GIS Specialist”). Notably, the occurrence of “director” in the top 10, appearing in 7.48% of job titles indicates that some data-focused roles are positioned within library leadership.

As unigrams alone may overlook context, bigram (Figure 2) and trigram (Figure 3) analyses underscored “research data”, “data services”, and “data librarian” as the dominant pairs while “data services librarian”, “research data management”, and “research data services” as the most common triplets. Collectively, these findings demonstrated a robust focus on the data lifecycle. Leadership titles, such as “director research data” or “head research data” surfaced at lower frequencies (at 1.56% each) yet illustrated that some libraries house dedicated data departments or divisions with formal oversight structures.

### Top 10 Unigrams

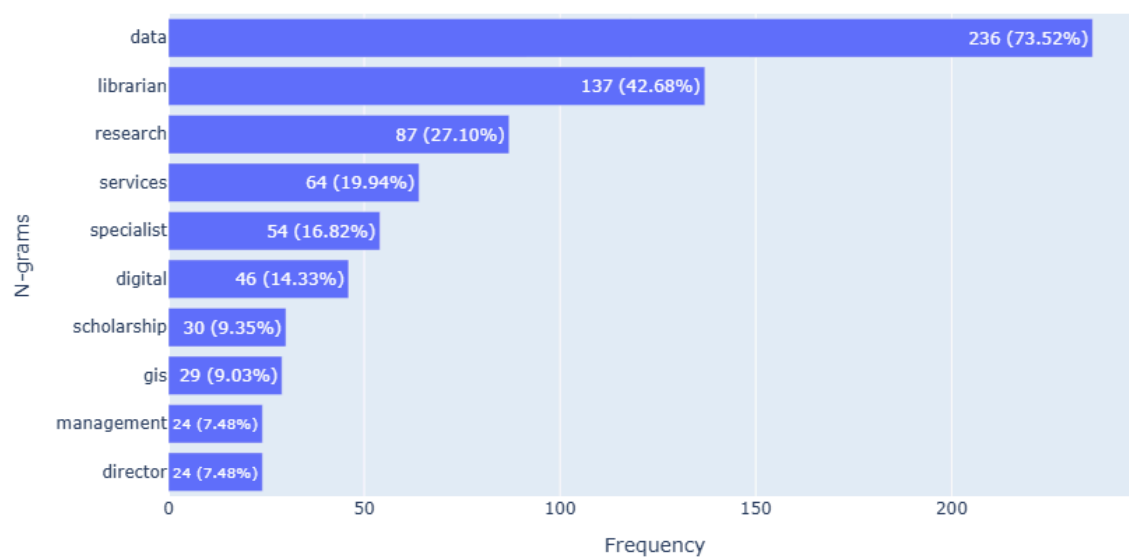


Figure 1. Top 10 unigrams of data-focused job titles in AAU libraries

### Top 10 Bigrams

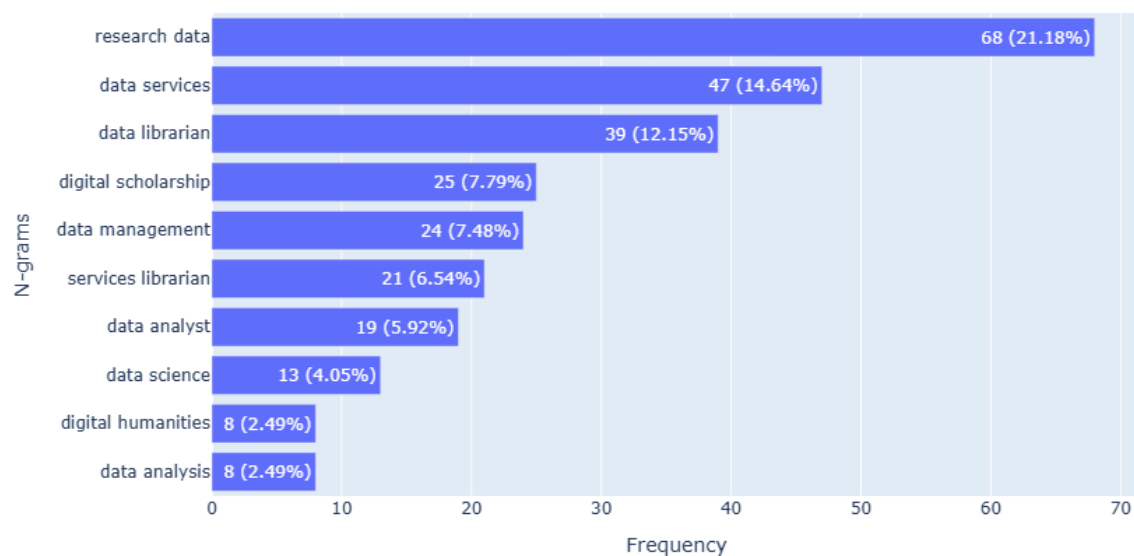


Figure 2. Top 10 bigrams of data-focused job titles in AAU libraries

### Top 10 Trigrams

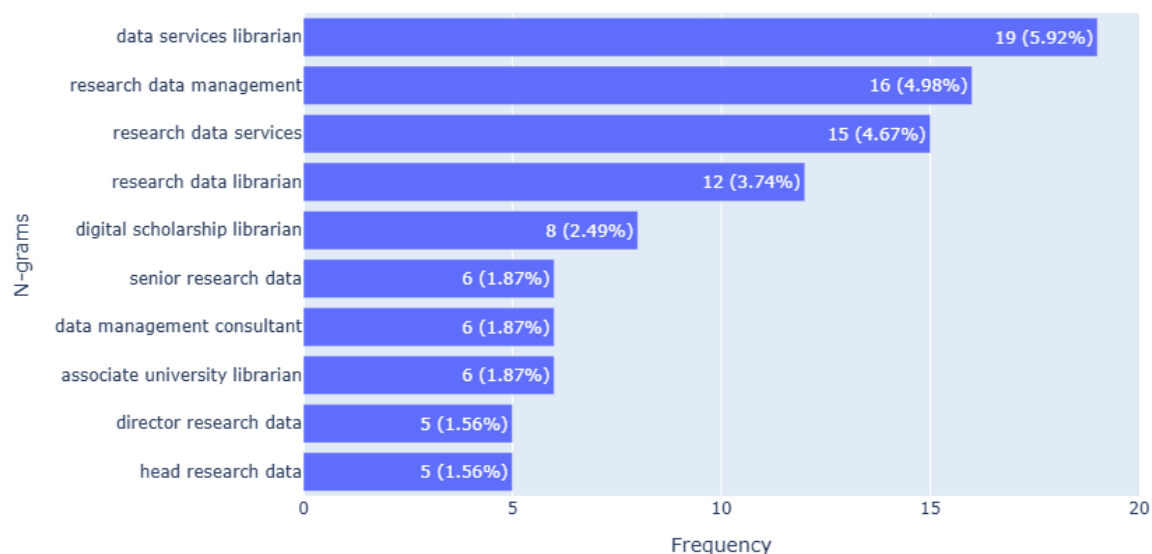


Figure 3. Top 10 trigrams of data-focused job titles in AAU libraries

### Generalists vs. Specialists

By further examining keywords tied to specialization (e.g., “GIS”, “geospatial”, “visualization”, “qualitative”, “quantitative”), a categorization emerged distinguishing generalist from specialist positions (Figure 4). Over half of the roles surveyed leaned toward generalist descriptions, while roughly one-third indicated specialized expertise. Leadership roles, identified by “director”, “head”, or “chief”, accounted for about 15% of data-focused positions, divided between generalists and specialists at a 1:5 ratio. Fellows and similar non-traditional titles composed a small minority (i.e., “other” category).

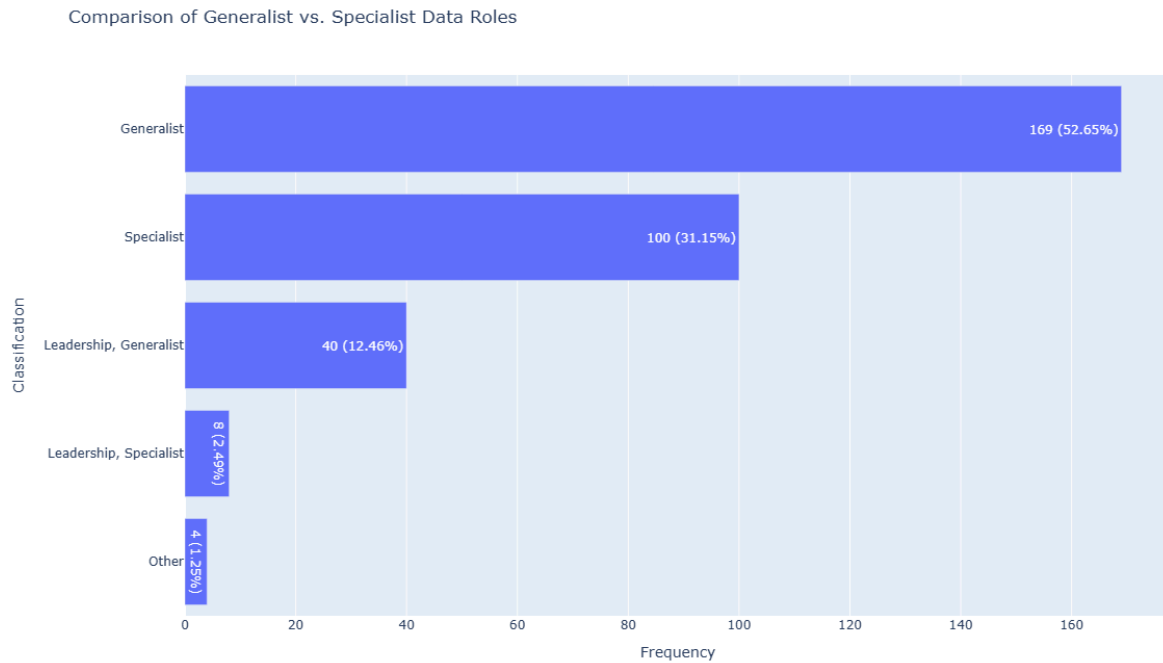


Figure 4. Comparison of generalist vs. specialist data-focused roles in AAU libraries

### Visibility of Data Services Web Pages on Library Web Sites

Finally, as summarized in Table 3, a notable subset of 30 institutions (42.25%) featured a dedicated “Data Services” web page on their university library’s website, although this term itself was not always used verbatim. Half used the exact phrase “Data Services” while the other half listed variations, including but not limited to “Research Data Services”, “Data Support Services”, and “Data Science Services”. Another 16 libraries maintained a “Data” webpage (e.g., “Data Lab” and “Data & Statistics Research Guides”) without explicitly naming it as “services”. Meanwhile, three university libraries offered some data-related support under broader (e.g., “Digital Scholarship Services”) or alternative labels (e.g., “Visualization Support”). The remaining 22 libraries had no dedicated data-focused landing page on their website, suggesting that their data services could be folded into general research or technology units, or possibly managed by external campus entities outside the library. A series of chi-square tests of independence found no significant association between the terms used and public/private status, indicating that naming and web visibility varied independently of institutional funding models. The varying levels of web visibility underscore potential institutional differences in how data services are communicated to patrons. Libraries with explicit “Data Services” or similar pages might more actively promote these offerings, while others could rely on more dispersed or ad hoc references elsewhere on their websites.

Table 3. Institutional use of data services terminology on library websites

Classification	Frequency (%)	p-value
Exact phrase “Data Services” <ul style="list-style-type: none"> <li>• All</li> <li>• Public</li> <li>• Private</li> </ul>	15 (21.13%) 7 (9.86%) 8 (11.27%)	.577
Includes both “Data” and “Services” but not as an exact phrase <ul style="list-style-type: none"> <li>• All</li> <li>• Public</li> <li>• Private</li> </ul>	15 (21.13%) 7 (9.86%) 8 (11.27%)	.577
Includes “Data” without “Services” <ul style="list-style-type: none"> <li>• All</li> <li>• Public</li> <li>• Private</li> </ul>	16 (22.54%) 9 (12.68%) 7 (9.86%)	1.00
Alternative phrasing used <ul style="list-style-type: none"> <li>• All</li> <li>• Public</li> <li>• Private</li> </ul>	3 (4.23%) 1 (1.41%) 2 (2.82%)	N/A*
No dedicated page found <ul style="list-style-type: none"> <li>• All</li> <li>• Public</li> <li>• Private</li> </ul>	22 (30.99%) 16 (22.54%) 6 (8.45%)	.108

\* p-value was not calculated due to expected frequency < 5

## Discussion

The findings of this study suggest that data services have become a firmly established facet of North American academic libraries, echoing earlier observations in the literature that libraries are expanding beyond traditional print-focused roles to meet emerging research needs. In line with the introduction, which noted the explosion of data and the evolution of “data librarian” from buzzword to recognized profession, the near-universality of data-focused roles at these major research institutions signals a maturation of data librarianship. On average, AAU libraries now devote around 2.30% of their staffing to data-focused roles, and many of them maintain dedicated “Data Services” pages, paralleling the three-tier model of education, consultation, and infrastructure discussed in the literature review.

Interestingly, public versus private status did not yield significant differences in the prevalence or percentage of data roles. This might partly reflect the high baseline of research investment and infrastructure common to all AAU members, minimizing disparities in funding or staffing. Nonetheless, this finding implies that institutional priorities and administrative vision may carry greater weight than funding models alone in shaping data capacities, reinforcing prior work that success depends heavily on top-level support (Ashiq & Warraich, 2023). The text analysis underlines the centrality of “research data” in these positions, suggesting that many

positions seek to support the entire data lifecycle, from planning and analysis to preservation and sharing, rather than honing in on specific, niche technical skills (e.g., GIS or visualization). For instance, job titles such as “Research Data Librarian” or “Data Services Specialist” typically encompass consultation on data management plans, data wrangling, and collaboration with diverse campus units.

### **Redefining “Data Librarian”**

Another theme resonating with the literature is the shifting definition of what it means to be a data librarian. As prior studies suggest, data librarians are increasingly defined by their ability to provide data services rather than traditional library skills, with an emphasis on understanding computational logic, databases, and information retrieval rather than becoming programmers or statisticians (Federer, 2018; Semeler et al., 2019). Some existing librarians already have the skills, background, or aptitude needed to transition into data librarian roles, particularly those with experience in data-intensive research or a strong proficiency with digital tools. This can be a practical approach for expanding data services, especially for basic data support that does not require hiring specialists with deeper technical expertise. However, not all data librarians follow this path, as some enter the field from outside traditional librarianship.

Anecdotally, based on the author’s background and professional network, several individuals holding data librarian or similar job titles do not have traditional MLIS credentials or any prior library experience. Instead, these data librarians come from either academic or professional backgrounds, often holding advanced degrees such as a PhD or master’s in data-related fields or having established careers in data-intensive industries. While many transitioned from academia, having worked as professors, researchers, or postdoctoral fellows in data-intensive fields, others moved from industry roles in information technology, data science, or related areas into data-focused roles within academic libraries. These pathways bolster the idea of interdisciplinary professional development: newcomers bring valuable domain expertise or advanced technical know-how, while on-the-job training helps them develop the reference, outreach, and collaboration skills essential in librarianship. From an administrative standpoint, recruiting such professionals may expedite the growth of data services, albeit requiring more robust integration into library culture and user-focused principles.

To build on this idea, two main strategies have emerged for libraries seeking to fill the gap in data services: (1) train existing librarians, or (2) recruit non-library data professionals. Each has its pros and cons. On the one hand, leveraging existing librarians who already understand library systems, user needs, and campus culture can potentially foster deeper loyalty and continuity. However, these existing library professionals may have limited incentives to invest time and effort into mastering complex data skills unless they foresee career advancement or institutional support (e.g., dedicated professional development funds). On the other hand, hiring data professionals can bring immediate technical competence (e.g., advanced programming, domain expertise) and potentially broaden the library’s service profile, but they often require onboarding into library culture (e.g., reference, user privacy norms, open access principles) and might need additional training for effective instruction or liaison work.

## Library “Data Services”

To ensure the effectiveness of their data services offerings, academic libraries must proactively promote these services and clearly communicate their value to the entire campus community. For example, linking RDM to tangible academic incentives such as career advancement opportunities, increased citations, and scientific progress can help raise awareness and engagement among researchers within the university (Joo & Schmidt, 2021). It is equally important to clearly outline the specific types of data services currently available at the library, whether they include consultation and advisory services (Joo & Schmidt, 2021; Yu, 2017), DMP and metadata assistance (Nitecki & Davis, 2019; Ohaji et al., 2019; Yu, 2017), help locating and acquiring datasets (Ohaji et al., 2019; Sheffield & Burton, 2022), or guidance on data analysis and visualization tools and techniques (Ashiq & Warraich, 2023; Eclevia et al., 2019; Wang et al., 2025). Showcasing these elements on a single, well-maintained “Data Services” landing page not only clarifies current capabilities but also signals how the library intends to expand them. This staged messaging prepares stakeholders for a roadmap of growth, outlined in a three-stage maturity model as follows:

1. **Stage 1 – RDM Essentials:** Libraries provide basic data repository services, minimal consultation, and primarily a focus on data curation and storage.
2. **Stage 2 – General Data Support:** Libraries expand into data consultation, data literacy workshops, and campus outreach, often through newly hired or upskilled generalist data-focused roles.
3. **Stage 3 – Specialized Data Expertise:** Libraries introduce advanced, niche and more specialized data-focused roles (e.g., data visualization specialists, geospatial librarians, or qualitative data experts) and heavily integrate data services into the institution’s research workflows.

For example, three AAU libraries in this dataset (i.e., Boston University, Tulane University, and the University of Notre Dame) are considered Stage 1. These institutions have no dedicated data-focused roles, and their data-related web pages primarily address RDM or similar research data services. Other AAU libraries with a small number of generalist data roles (e.g., California Institute of Technology with a “Research Data Specialist” and Stony Brook University with a “Data Literacies Lead”) are categorized as Stage 2. In contrast, a few other AAU libraries (e.g., Duke University and the University of Washington) exemplify Stage 3. They have multiple specialized data roles (e.g., “Data Visualization Librarian” and “Map and Geospatial Data Specialist”) as well as established data-focused units or departments (e.g., the “Center for Data and Visualization Sciences” or “Clinical Research and Data Services”), reflecting a high level of maturity in their data services offerings.

It is important to note that the actual classification of each institution may vary, given potential discrepancies between what is publicly available on library web pages and the actual conditions in the field. As such, this study refrains from assigning definitive classifications to each AAU member library, particularly in cases where their placement is not clear-cut. Nonetheless, the three-stage maturity model presented here serves as a useful framework for AAU libraries to reflect on their current state, compare themselves with peer institutions, and make informed decisions about staffing and structuring their data services moving forward.



## Global Perspectives

Although this study focuses on North American libraries, data librarianship is a rising trend in other world regions as well. For instance, European institutions, fueled by Open Research Data mandates, have begun introducing data librarian roles to facilitate research data curation, training, and institutional policy compliance (Watek, 2019). More recently, an environmental scan shows that libraries in the South and Southeast Asia region often prioritize RDM services (e.g., maintaining data repositories or providing data management training) while giving less attention to data mining, visualization, or reproducibility. Most librarians in these regions emphasize metadata standards, planning skills, and basic training resources (Sinha et al., 2023).

A newly published study in 2025 highlights a growing emphasis on data visualization services among some leading Chinese academic libraries, demonstrating an expanded role for librarians as stewards, trainers, collaborators, and advocates for data-driven innovation (Wang et al., 2025). While many AAU institutions in North America have long operated at Stage 3 with advanced data services, China's rapid investment in research data infrastructure has led to significant progress in just a few years. However, it is important to note that these developments in China reflect an emergent trend based on select recent case studies rather than a generalizable shift across the entire country's academic library landscape.

These examples illustrate that data librarianship is evolving worldwide, though models and job titles vary based on regional mandates, funding structures, and institutional priorities. Many institutions in South and Southeast Asia remain at Stage 1, while many European and North American academic libraries are already at Stage 2, with some, particularly among AAU members, having reached Stage 3 maturity. That said, AAU institutions, while influential, are not necessarily representative of the broader North American library landscape, especially among smaller or less research-intensive universities where funding and staffing for data services remain limited. Some leading Chinese academic libraries, on the other hand, are beginning to demonstrate Stage 3 capabilities, aligning with developments seen in long-established AAU institutions, although further studies are needed to understand the broader, country-level landscape.

## Job Security, Compensation, and Faculty Status

Another significant consideration is job security and compensation, particularly in attracting high-caliber data professionals. Some North American universities treat librarians as faculty, complete with tenure lines or continuous appointment pathways (Hoggan, 2003; Weng & Murray, 2020), which can offer long-term stability and potential for professional growth. Others hire data librarians as staff, either on contracts lacking tenure or with potential for permanent employment. Librarians with faculty status typically juggle teaching, publishing, service, and committee obligations (Hoggan, 2003), whereas staff roles might concentrate on hands-on data support. All these differences can affect workload, responsibilities, and salary scales.

Additionally, some university libraries occasionally hire experienced candidates at higher ranks (e.g., associate librarian) with an expedited tenure process. Such models can incentivize talented data professionals to choose academic libraries over other academics (i.e., professors or postdoctoral positions) or even private sector

roles, but they also introduce variability in hiring practices, as not all libraries offer faculty status or tenure options. This variability can heavily influence libraries' ability to recruit and retain individuals with specialized data expertise.

## Limitations and Future Work

Despite the broad scope of the dataset, this study has several limitations. First, the snapshot approach captures roles listed at a single point in time (i.e., end of August 2024). Future updates to library staffing could quickly shift the numbers. Second, self-reported staff directories may not always be current or fully transparent about job scope. Some institutions bury data support in broader divisions or departments, making it possible that this study either under- or over-counted relevant roles. While some of them, including "Digital Scholarship" and "Research Service" were already scrutinized, some roles may still be overlooked if they are buried in less transparent departmental structures. Third, "metadata" positions were largely excluded unless their job descriptions specifically indicated front-facing data services, reducing potential overcounting but leaving ambiguity for hybrid roles. This means there is potential inflation or undercounting of roles due to dual appointments or naming conventions, particularly if roles related to data services do not explicitly include the term "data" in job titles but still involve substantial data responsibilities. Lastly, organizational nuances (e.g., shared roles or joint-appointment with other campus departments) may have eluded detection if the library website lacked thorough staff listings.

Future research might deepen this analysis by incorporating qualitative interviews with data librarians, focusing on skill gaps, job satisfaction, and service impact. Another avenue involves longitudinal studies of how data roles evolve over time or cross-comparison with smaller, non-AAU institutions in North America or other institutions in other regions to see if these patterns hold in less resource-intensive environments.

## Conclusion

This study underscores three key findings that are particularly relevant to an international readership:

1. **High Prevalence:** Data-focused roles are nearly universal among leading research university libraries in North America, with about 2.30% of library staff dedicated to data services.
2. **Emerging Patterns and Trends:** Public vs. private status does not significantly affect the number or percentage of these roles, suggesting that institutional priorities matter more than funding models alone. Moreover, common job titles like "Research Data Librarian" or "Data Services Specialist" emphasize a broad, research-centered approach, rather than niche technical expertise.
3. **Global Implications:** These North American examples reflect a broader global trajectory of library data services maturity, where libraries progress from basic RDM to specialized data expertise. Regions such as South and Southeast Asia may still be building foundational RDM services, while others, like China, have already begun demonstrating more advanced capabilities, including support for data visualization. For university libraries worldwide aiming to implement or expand data services, these findings highlight the importance of administrative buy-in, flexible hiring practices (e.g., MLIS vs.

non-MLIS backgrounds), and clear online visibility (e.g., “Data Services” web pages). Whether through training existing librarians or recruiting data professionals, institutions benefit from clarifying their data support strategies, aligning them with user needs, and offering robust career pathways.

Overall, this study highlights the growing momentum of data librarianship, which is no longer a fringe experiment but a growing specialization, as reflected in staffing proportions, departmental structures, and web presence. Further research (e.g., in-depth case studies, longitudinal surveys, long-term staffing trends, exploring salary and job satisfaction, or comparing outcomes of different hiring models) is required, not only to assess service effectiveness and evolving role definitions, but also to further illuminate how libraries can effectively develop, sustain, and evolve their data services in the years to come.

## Acknowledgment

The author would like to thank Chris Kretz, Janet Clarke, and Karim Boughida for their support and guidance throughout this project.

## References

- Ashiq, M., & Warraich, N. F. (2023). A systematized review on data librarianship literature: Current services, challenges, skills, and motivational factors. *Journal of Librarianship and Information Science*, 55(2), 414–433. <https://doi.org/10.1177/09610006221083675>
- Association of American Universities. (2025, February 14). *Association of American Universities (AAU)*. <https://www.aau.edu/who-we-are-americas-leading-research-universities>
- Bishop, B. W., Orehek, A. M., Eaker, C., & Smith, P. L. (2022). Data Services Librarians’ Responsibilities and Perspectives on Research Data Management. *Journal of eScience Librarianship*, 11(1), Article 1. <https://doi.org/10.7191/jeslib.2022.1226>
- Bishop, W., Collier, H., Orehek, A. M., & Ihli, M. (2021). Potential Roles for Science Librarians in Research Data Management: A Gap Analysis. *Issues in Science and Technology Librarianship*, 98, Article 98. <https://doi.org/10.29173/istl2602>
- Eclevia, M. R., Fredeluces, J. C. L. T., Eclevia, C. J. L., & Maestro, R. S. (2019). What Makes a Data Librarian? An Analysis of Job Descriptions and Specifications for Data Librarian. *Qualitative and Quantitative Methods in Libraries*, 8(3), 273–290.
- Federer, L. (2018). Defining data librarianship: A survey of competencies, skills, and training. *Journal of the Medical Library Association*, 106(3), Article 3. <https://doi.org/10.5195/jmla.2018.306>
- Gowen, E., & Meier, J. (2019). *AAU University Libraries and Data Services in 2019* [XLSX]. <https://doi.org/10.26207/r3xm-kh34>
- Henderson, M., (2020) “Why You Need Soft and Non-Technical Skills for Successful Data Librarianship”, *Journal of eScience Librarianship* 9(1): 2. doi: <https://doi.org/10.7191/jeslib.2020.1183>
- Hoggan, D. B. (2003). Faculty Status for Librarians in Higher Education. *Portal: Libraries and the Academy*, 3(3), 431–445.

- Huggins, M. (2022). MLIS Required? Rethinking the Skills and Knowledge Necessary for Managing in a Public Library. *Journal of Library Administration*, 62(6), 840–846.  
<https://doi.org/10.1080/01930826.2022.2102384>
- Joo, S., & Schmidt, G. M. (2021). Research data services from the perspective of academic librarians. *Digital Library Perspectives*, 37(3), 242–256. <https://doi.org/10.1108/DLP-10-2020-0106>
- Khan, H. R., & Du, Y. (2017). *What is a Data Librarian?: A Content Analysis of Job Advertisements for Data Librarians in the United States Academic Libraries*. <https://library.ifla.org/id/eprint/2255/>
- Nitecki, D. A., & Davis, M. E. K. (2019). Expanding Academic Librarians' Roles in the Research Life Cycle. *Libri*, 69(2), 117–125. <https://doi.org/10.1515/libri-2018-0066>
- Ohaji, I. K., Chawner, B., & Yoong, P. (2019). The role of a data librarian in academic and research libraries. *Information Research*, 24(4). <https://informationr.net/ir/24-4/paper844.html>
- Semeler, A. R., & Pinto, A. L. (2020). Data librarianship as a field study. *Transinformação*, 32, e200034. <https://doi.org/10.1590/2318-0889202032e200034>
- Semeler, A. R., Pinto, A. L., & Rozados, H. B. F. (2019). Data science in data librarianship: Core competencies of a data librarian. *Journal of Librarianship and Information Science*, 51(3), 771–780. <https://doi.org/10.1177/0961000617742465>
- Sheffield, M., & Burton, K. B. (2022). Research Data Management needs assessment of Clemson University. *Journal of Librarianship and Scholarly Communication*, 10(1), Article 1. <https://doi.org/10.31274/jlsc.13970>
- Sinha, P., A., S., & Sinha, M. K. (2023). Research data management services in academic libraries: A comparative study of South Asia and Southeast Asia | Emerald Insight. *Global Knowledge, Memory and Communication*, ahead-of-print(ahead-of-print). <https://doi.org/10.1108/GKMC-01-2023-0033>
- 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. <https://doi.org/10.1016/j.lisr.2013.11.003>
- The Group of Eight. (2025). *Group of Eight*. The Group of Eight: Australia's Leading Universities, Leading Excellence, Leading Debate. <https://go8.edu.au/>
- The Russell Group of Universities. (2025, February 21). *Russell Group | Our universities*. The Russell Group. <https://russellgroup.ac.uk/about/our-universities/>
- Thomas, C. V. L., & Urban, R. J. (2018). What Do Data Librarians Think of the MLIS? Professionals' Perceptions of Knowledge Transfer, Trends, and Challenges. *College & Research Libraries*, 79(3), Article 3. <https://doi.org/10.5860/crl.79.3.401>
- Wałek, A. (2019). Data librarian and data steward – new tasks and responsibilities of academic libraries in the context of Open Research Data implementation in Poland. *Przegląd Biblioteczny*, 497–512. <https://doi.org/10.36702/pb.634>
- Wang, G., Noorhidawati, A., Idaya, A. M. K. Y., & Yuan, Y. (2025). Navigating data visualization: Service landscapes and librarian roles in academic libraries. *The Electronic Library*, ahead-of-print(ahead-of-print). <https://doi.org/10.1108/EL-09-2024-0272>
- Weng, C., & Murray, D. C. (2020). Faculty perceptions of librarians and library services: Exploring the impact

of librarian faculty status and beyond. *The Journal of Academic Librarianship*, 46(5), 102200.  
<https://doi.org/10.1016/j.acalib.2020.102200>

Yu, H. H. (2017). The role of academic libraries in research data service (RDS) provision: Opportunities and challenges. *The Electronic Library*, 35(4), 783–797. <https://doi.org/10.1108/EL-10-2016-0233>