




Data services at the academic library: a natural history of horses and unicorns

Jeffrey Oliver, University of Arizona Libraries, Tucson, AZ, USA, jcoliver@arizona.edu 

Fernando Rios, University of Arizona Libraries, Tucson, AZ, USA 

Kiriann Carini, Development Seed, Washington, DC, USA 

Chun Ly, Princeton Plasma Physics Laboratory, Plainsboro, NJ, USA 

Abstract

Objective: Increases in data-intensive research at colleges and universities is driving demand for data services provided by academic libraries. The current work investigates the distribution of library data services, how such services are offered, and the effect of resourcing on the amount of services offered by a library.

Methods: We used a web-based inventory of 25 academic libraries at U.S. Research 1 (R1) Carnegie institutions to assess the state of data services at university libraries. We categorized and quantified services, and tested for an effect of library resourcing on the size of library data service portfolios.

Results: Support for data management and geospatial services was relatively widespread, with increasing support in areas of data analyses and data visualization. There was significant variation among services in the modality in which they were offered (web, consult, instruction) and library resourcing had a significant effect on the number of data services a library offered.

Received: August 11, 2023 **Accepted:** February 23, 2024 **Published:** April 23, 2024

Keywords: research data, data management, data analysis, geospatial, library resourcing

Citation: Oliver, Jeffrey, Fernando Rios, Kiriann Carini, and Chun Ly. "Data services at the academic library: a natural history of horses and unicorns." *Journal of eScience Librarianship* 13 (2): e780. <https://doi.org/10.7191/jeslib.780>.

Data Availability: All data and analytical code are available at <https://github.com/UAL-RE/data-service-landscape-scan> and archived at <https://doi.org/10.25422/azu.data.22297177>. Appendices are available under the article [Supplementary Files](#).

The *Journal of eScience Librarianship* is a peer-reviewed open access journal. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited, and new creations are licensed under the identical terms.

See <https://creativecommons.org/licenses/by-sa/4.0>.

OPEN ACCESS

Abstract Continued

Conclusions: While a core subset of these data services are offered at most academic libraries, more specialized topics are restricted to well-resourced libraries. In light of the influence of resource scarcity on the number of services a library can offer, intra- and inter-campus partnerships will be critical to ensure campus support for data service needs.

Introduction

The growth of data-intensive scholarship across multiple research disciplines, in combination with increasing mandates from federal funders intended to foster open science and reproducible research, is placing increasing demands for research data services support at academic institutions. Academic libraries have long supported scholars' data needs, and those demands have resulted in quantitative and qualitative changes to the research data services portfolios at those libraries (Hooper 2023). The term "research data services" itself is still evolving, and here we use a broad definition that includes "traditional" data management support (Tenopir et al. 2012; Si et al. 2015) as well as support in other points of the research lifecycle, including data analysis and data visualization (Radecki and Springer 2020).

Several studies have reported on research data services offered through academic libraries, demonstrating how these services have evolved. A pair of studies provided a convenient snapshot of the growth in a subset of research data services (Tenopir et al. 2012; Tenopir et al. 2019). Although these studies used a narrow definition of research data services (focusing on research data management and repository services), they showed a growth in the percentage of large institutional libraries offering technical research data services from approximately 19% in 2011 to approximately 30% in 2016 (see Table 6B in Tenopir et al. (2012) and Table 6B in Tenopir et al. (2019)). With a similar focus on research data management services, Cox et al. (2019) surveyed academic libraries on the extent to which a variety of research data services were offered. Categorizing services as not offered, offered as a basic service, or offered as an extensive service, they found that most of the academic libraries surveyed provided data management services such as data management plan (DMP) support, data storage advice, and data publication advice at a level of basic service or better in 2018 (see Figure 4 in Cox et al. 2019). In contrast, most of the libraries surveyed in (Cox et al. 2019) provided no support in data analysis, data mining, or data visualization. These studies highlight an early emphasis by academic libraries in supporting storage, management, and archiving of data output from scholarly activities.

More recently the definition of research data services has expanded to include services supporting all parts of the research lifecycle. Radecki and Springer (2020) provided the broad definition of research data services: "any concrete, programmatic offering intended to support researchers ... in working with data." (p. 7). In the Radecki and Springer survey, the services were categorized by the type of service (e.g. consultation,

workshop) as well as the discipline of the service (e.g. digital humanities, geospatial, health sciences). Using web-based inventory, they found that academic libraries at R1 Carnegie institutions offered an average of 2.4 research data services. These results highlight how research data services offered at academic libraries encompass support in addition to data management services: 18 of 40 R1 libraries (45%) provided support for geospatial data services and 5 (12.5%) provided statistical support through either consultations or training events. This work further evidenced the growth in support for research data services in academic libraries: 80% of R1 libraries offered at least one type of research data service. This represents an increase in support for data services, compared with prior studies showing a minority of libraries with data services support (Tenopir et al. 2012; Tenopir et al. 2019; Cox et al. 2019), and may also be in part due to the broad definition used by Radecki and Springer (2020). For the remainder of the current work, we use the broad definition provided by Radecki and Springer in our investigation and discussion of research data services.

In addition to the amount and identity of research data services offered at academic libraries, it is useful also to understand the modalities in which services are offered. Prior work indicates the type(s) of research data services may influence the modality through which the service is offered. For example, Tenopir et al. (2019) found 40% of the libraries at doctoral-granting institutions offered data management consultation while 54% offered data management information through web resources. In the sample of 40 R1 university libraries, Radecki and Springer (2020) found geospatial and statistical data services were provided more frequently as consultations (geospatial: 45%, statistical: 17.5%) than as instructional training events (geospatial: 15%, statistical: 2.5%). These comparisons suggest differences and warrant formal evaluation of potentially statistically significant differences in modalities in research data service offerings. Different modalities require different levels of investment and vary in scalability, and identifying how different data services are supported may provide guidance for determining future resource allocation.

The suite of skills necessary to provide research data services often requires additional training or experience on the part of library personnel (Burton et al. 2018; Tenopir et al. 2019; Cox et al. 2019). This demand for “data savvy” skills implies a relationship between library resourcing and the amount of research data support a library can provide. Wages, one measurement of library resourcing, have been shown to affect the amount of support libraries can provide in other service areas. Comparing special district libraries to municipal libraries, Schattman and Liu (2022) found special district libraries had higher wages as well as more library services than did municipal libraries. In a qualitative assessment, Hamad et al. (2022) found financial restrictions reduced the number of “smart services” (user-oriented services based on individual-level data from mobile devices) offered at academic libraries. In regards to research data services, Radecki and Springer (2020) explicitly identify staffing as a potential limiting factor in determining the amount of research data service support that an academic library can provide.

Building on the work of Radecki and Springer (2020) and others, this work provides a more granular view of research data services support at a sample of academic libraries, with investigations on how support is

provided and how resources affect the amount of support a library provides. The potential link between library resources and data services has received little attention, and warrants investigation given the significant role personnel resources appear to play in provisioning data services, as alluded to in prior work (Radecki and Springer 2020). This information is intended to serve decision makers at academic libraries when developing, expanding, or assessing research data services. Using a web-based inventory, we categorize and quantify research data services at U.S. academic libraries and test for potential differences in how support is provided (i.e. the modality) among different types of services. Based on publicly available data on libraries' total wage expenditures, we also test the relationship between library resourcing and the quantity of research data services. We close with recommendations for supporting research data services in the absence of increased library resourcing.

Methods

Data Collection

This landscape scan was designed to determine the prevalence of a variety of data services at a sample of R1 universities in the U.S. To define this sample of universities, we used the 25 institutions described as peers to our own institution, the University of Arizona Libraries (UAL), as defined by the Arizona Board of Regents (Appendix 1). We included all 25 institutions in our survey, including the University of Arizona. We used a systematic web survey of the institutions' library web sites to assess which services were supported directly by the library. Determination of whether a service was present or absent had three phases: first, two library personnel not involved in providing data services surveyed each of the institution's library website for each service (see below) using a Google Form to record results (Appendix 1). This initial survey took place during June and July of 2020. Following this initial scan, the authors first checked each URL that was collected in the first phase indicating a service was present to ensure the service was described at the URL; this was required due to the specialized nature of some services not being correctly classified in the initial scan. Each URL indicating the presence of a service was independently checked by two of the authors. In some cases, this resulted in some services being re-classified as absent. This second phase took place August through December of 2020. Finally, for each service still listed as being absent from an institution, the authors searched the institution library's website to confirm that the service was not offered. In some cases, this resulted in some services being re-classified as present. This final check took place January through February of 2021.

A pre-selected set of services (Table 1) was generated from a prior, informal landscape scan of services offered at R1 university libraries. For a subset of the services, we also recorded the mode(s) in which the service was offered (web resource, instruction, or consult). Only those services that were (1) offered by library personnel (as opposed to guest instructors or outside consultants), and (2) in the case of instructional sessions or events, offered during or after January 2019 were considered present at an institution's library (see Appendix 2 for detailed criteria). Because the focus of this work is on academic libraries developing,

expanding, or assessing research data services, the set of services in Table 1 is more granular than the broad categories used by Radecki and Springer (2020).

Table 1: Data services included in this landscape scan; asterisks (*) indicate those services for which delivery mode (web resource, instruction, or consultation) was also assessed.

Service
Aerial imagery
Carpentries
Data analysis *
Data management services (general) *
Data management plans *
Data repository
Data visualization software *
DMPTool
Electronic lab notebooks
Geospatial services (general) *
Geospatial software *
Open Science Framework (OSF)
Reproducible research
Statistical consulting
Text data mining
Version control
Web scraping

We were also interested in testing how library resourcing might affect the number of services offered by an academic library. Given the assertion that libraries require qualified staff in order to provide data services (Burton et al. 2018; Cox et al. 2019; Tenopir et al. 2019), we chose to investigate how labor expenditures affect the data services of a library. As a measure of library labor resourcing, we used library combined salary and wage expenditures for 2019, defined as the “total salaries and wages expenditures for all full-time and part-time staff, student assistants, and Work-Study students, if paid from the library budget.” These data were retrieved from the Integrated Postsecondary Education Data System (IPEDS; U.S. Department of Education 2021). While this measures total salary and wage expenditures for a library, it is the best publicly available proxy for measuring resourcing to research-oriented services at academic libraries.

Analysis

Following the quality control checks described above, we measured the frequency of services offered by institutions' libraries. We counted the total number of services offered by each library, regardless of which modalities through which they were offered. We also counted the number of libraries offering each service, to identify which services are common among academic libraries and which services are rare or unavailable.

For a subset of the services, we analyzed if the general area of service (data management, geospatial, or data science) influenced the mode (web resource, instruction, or consult) by which the service was offered. We analyzed each of the three areas separately with logistic regression to determine if there were significant differences in the likelihood of a service being offered in each modality. We used Tukey *post-hoc* pairwise comparisons to determine if modes were offered at significantly different rates.

We used the total count of services offered at each library to test for an influence of library resources. Briefly, we used generalized linear regression to test whether the total combined library salary and wage expenditures predicted the total number of services offered, using a Poisson model for the count of total services. We tested to see if data from any individual institution had disproportionate influence on regression results, but none were identified as having undue influence. We used Cook's distance (Cook 1977) and DFBETA (Belsley et al. 1980) for outlier detection.

All analyses were performed using the R software package (R Core Team. 2022) and dplyr (Wickham et al. 2022a), ggplot2 (Wickham 2016), multcomp (Hothorn et al. 2008), readr (Wickham et al. 2022b), stringr (Wickham 2019), and tidyr (Wickham and Girlich 2022) packages. All data and analytical code are available at <https://github.com/UAL-RE/data-service-landscape-scan> and archived at <https://doi.org/10.25422/azu.data.22297177>.

Limitations

The systematic web survey method has been used in prior work (Radecki and Springer 2020), but it is important to note some caveats with the approach. First, surveying web sites relies on efficient navigational design—some services may have been categorized as absent by our scan if library web sites had poor search optimization or non-intuitive menus and navigation. In some cases it was difficult to determine whether or not a service listed on a library's web sites was still being offered. This approach also restricts detection of those services that an institution chooses to advertise—it is possible that a library provides a service but does not include such information on the library's web sites. Finally, as this scan took place during the COVID-19 pandemic, library service offerings were undoubtedly evolving rapidly. Because of this, we acknowledge that our results represent a snapshot in time that may have changed. For example, changes in modalities that arose out of pandemic lockdowns may have persisted as universities returned to more normal operations. Despite these limitations, we believe our results still capture general underlying trends because the primary influencing factors such as increasingly stringent mandates from federal funders for open and reusable

science (e.g., the Office of Science and Technology Policy “Nelson Memo” (Nelson 2022), the 2023 NIH Data Management and Sharing Policy (Collins 2020)) and wage limitations remain largely unchanged.

Results

Data services offered at peer academic libraries broke into two distinct groups: those services offered at a clear majority of the libraries surveyed and those “rarities” that were only offered at a minority of libraries (Figure 1). Support for data management in general as well as data management plans was ubiquitous, found at all surveyed libraries. Geospatial-oriented services, including geospatial software and aerial imagery services were also found at a majority (72% and 92%, respectively) of the libraries investigated here. In contrast, support for electronic laboratory notebooks and web scraping were found in fewer than 25% of libraries we surveyed. Libraries provided support for an average of 10.6 (± 0.5 SE) of the 17 services in our survey (Figure 2).

The mode in which a service was offered differed among the different areas of data services (Figure 3). Data management services (data management support in general and support for data management plans) were more likely to be offered through consultations and web resources than through formal instructional events ($p = 0.001$ and $p = 0.009$, respectively). In contrast, data science services (data analysis and data visualization software support) were more likely to be offered through instructional opportunities than web resources ($p = 0.003$). Geospatial services (general geospatial data support and geospatial software support) were equally likely to be offered via any of the three modes (web, consult, or instruction).

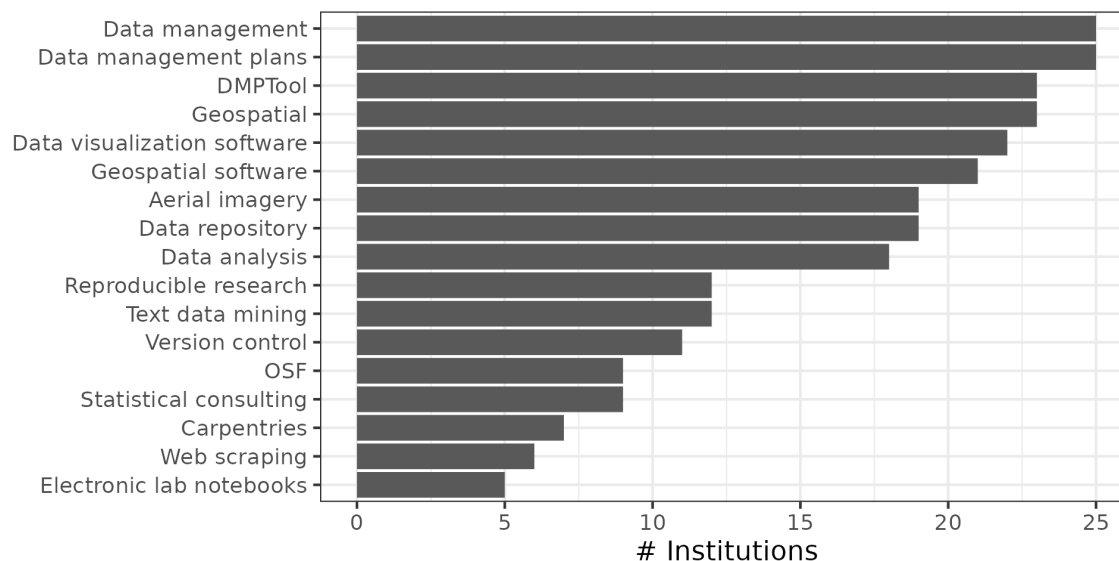


Figure 1: Frequency of data services supported across institutions.

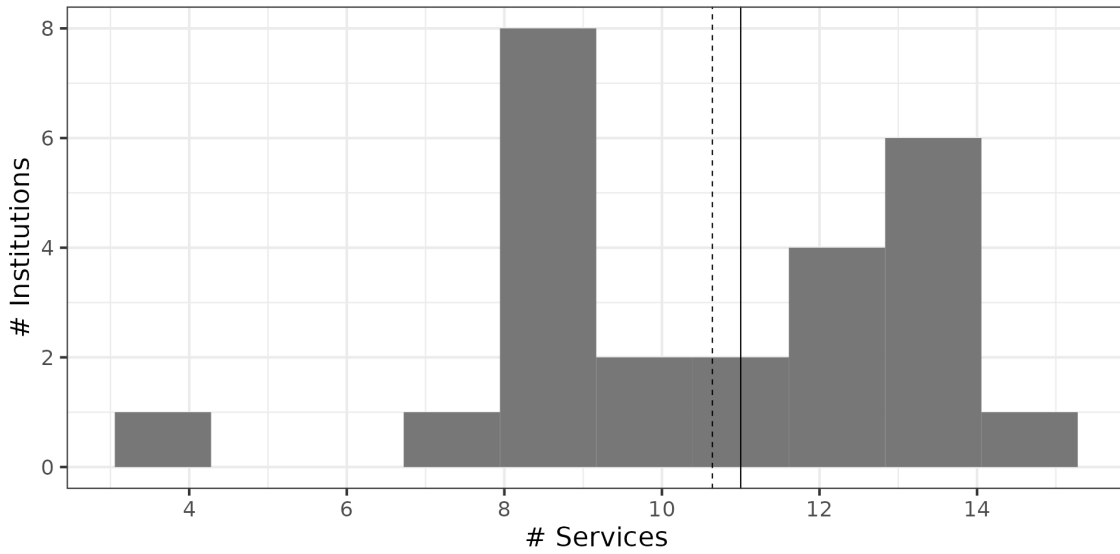


Figure 2: Number of data services across institutions; vertical lines are the mean (dotted) and median (solid) number of data services offered by libraries (10.6 and 11.0, respectively).

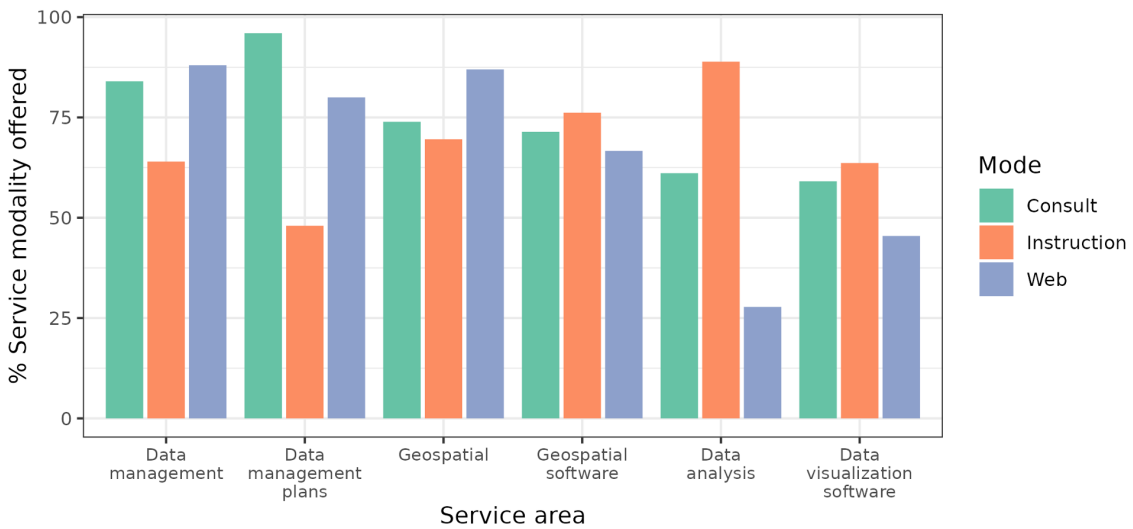


Figure 3: Modes of delivery for the different data service areas: data management, geospatial support, and data science (data analysis and data visualization software).

The amount of resourcing, as measured by a library's total salary and wage expenditures, influenced the total number of data services offered. In a regression analysis, total salary and wage expenditures were significantly correlated with the total number of services a library offered ($p = 0.046$; Figure 4).



Figure 4: Total number of data services offered as a function of library's total annual salaries and wages. Line shows predicted relationship from generalized linear regression model (GLM).

Discussion

While our categorization of data services was different from Radecki and Springer (2020), we can make useful comparisons for those data services that overlapped between this study and prior work. In general, we found geospatial services and statistical support to be more widespread in the academic libraries included here than in the libraries at R1 universities surveyed by Radecki and Springer (2020). In our survey, 92% (23 of 25) of libraries provided some form of geospatial support, compared with 48% (19 of 40) in Radecki and Springer (2020). Similarly, we found higher levels of statistics support at academic libraries (current work: 36%, 9 of 25; Radecki and Springer (2020): 18%, 7 of 40). Methodological differences between the two studies likely explain the differences in percentages: Radecki and Springer (2020) included consulting and training events as modes of research data service delivery; the current work included those two modes as well as web resources, such as online guides and static learning resources. A fine-scale comparison of the seven universities included in both studies illustrates minor discrepancies between the two studies: both studies indicated six of the seven libraries offered geospatial support, but the one library that did not offer geospatial services differed between the current work (Washington State University) and Radecki and Springer (2020) (University of California, Los Angeles). Statistical support in the library was found in only one of the overlapping seven universities in the current work, while Radecki and Springer (2020) reported that none of the same seven universities offered statistical support through the academic library. These differences are not a critique of the web-based survey methodology or of the conclusions offered in Radecki and Springer (2020), but highlight how differences in criteria can manifest in results.

A finer-scale comparison among certain data services further highlights variation in the way services are offered and may reflect nuanced differences in the demands for support in each area. Data management services are offered more frequently through consultations and web resources than through instruction (Figure 3), potentially reflecting the “just-in-time” need for data management support. This contrasts with data science services, which are more likely to take the form of instruction over web resources and are best categorized as “just-in-case” support. That is, the demand for data management support is often tied to grant or journal submission deadlines, while data science skills development are less likely to be offered in an on-demand fashion. The relatively low amount of library-created web resources for data science (28% data analysis; 46% data visualization) may also reflect an abundance of reference and training resources that already exist online.

Variation in the number of research data services among academic libraries is likely due, in part, to the level of library resourcing. Radecki and Springer (2020) provide a similar conclusion in comparing the number of research data services offered at R1, R2, and small, liberal arts colleges. While our measurement of resourcing, a library’s total combined salary and wage expenditures, is a coarse proxy for resourcing dedicated to research data services, it does reinforce the idea that better-provisioned libraries will be able to offer a greater range of research data services. It warrants mentioning that many of the services included in our survey are often also provided by other, non-library campus units (Murray et al. 2019; Radecki and Springer 2020). While beyond the scope of the current work, future studies could investigate the factors determining not just if a data service is provided on an academic campus, but in which campus unit (library, information technology, discipline-specific center) such a service is offered. A similar cost analysis using estimates of salary expenditures for those other units could help paint a more complete picture of what it would take to offer services in collaboration with other units. Finally, finer scale quantifications of resources, through targeted surveys of individual library units providing data services (i.e. labor costs for data service units) would further refine the nuanced relationship between financial support and the quantity and quality of data services offered.

The variation in the number of services offered prompts the oft-asked question of how can academic libraries with fewer resources still provide important research data services to their respective constituencies? That is, in the absence of an influx of resources to hire additional staff or afford significant professional development opportunities, how can libraries “do more with less”? Prior work has called out the utility of partnering with other campus units, including service cores and academic departments (Murray et al. 2019; Radecki and Springer 2020). We propose two additional solutions based on the collective expertise across academic libraries. First, formal collaborations among libraries at different institutions can provide mutual benefit to all parties involved. Such arrangements can provide clear expectations and reliable resources to audiences larger than those at a single institution (e.g. interlibrary loan, the Greater Western Library Alliance (<https://www.gwla.org>), the Data Curation Network (Johnston et al. 2018)). Through shared responsibilities and distributed workload, such efforts provide a means of data service support beyond the capacity of what

might be possible by a single institution, especially those with fewer resources.

A second option for increasing data services at current resourcing levels is through informal sharing of resources among institutions. Specifically, the recent pandemic has highlighted the utility of synchronous online meetings for both networking and instructional purposes. The online modality allows (1) gatherings of the library data services community of practice across institutions and geography and (2) participation in instructional sessions offered at other institutions. Such networking and sharing of practices has long been a benefit of regional, national, and international society meetings, but we argue that the pandemic has highlighted how such sharing online can occur at a higher frequency, at considerably lower cost to participants. Opening instructional sessions to participants at other institutions allows campus constituents to benefit from the collective data services expertise, rather than solely relying on resources offered at their own institutions. Such sharing is not without its challenges, including how information is disseminated among institutions and how libraries can still serve their primary audiences at their own respective institutions while instructing audiences at other institutions. Given these challenges and the rapidly evolving data landscape, we assert that informal agreements offer the flexibility best suited at this point in data services. A current example of this model is exemplified by the Data Science in Libraries quarterly meetings organized by personnel at Arizona State University. Born of an online national meeting in June 2022, this community of practice meets online to share best practices and invite participation in online instructional opportunities. Through this sharing network, campus constituents have been able to take advantage of training opportunities that are unavailable to them at their present institution (J. Oliver, pers. obs.). Such collaborations, both formal and informal, provide a means of supporting the collective data service needs in the absence of significant increases in academic library funding.

While we did not perform a systematic comparison of our direct survey methodology to alternative means of data collection (i.e. volunteer survey), it is worth mentioning differences between services identified and actual services provided at our institution, the University of Arizona Libraries (UAL). For example, two services categorized as “unsupported” at the UAL, statistical consulting and support for version control, are both provided through weekly R programming sessions, but are not explicitly mentioned on library websites. This is largely due to the limited capacity: advertising support for these services, especially statistical consulting, would be disingenuous, given the high demand for such services on campus and limited number of personnel who can support such services. We recommend future work to compare the difference between services that are advertised through library websites and communication channels to those services that are actually offered by an academic library.

Conclusion

Data-intensive research is growing across disciplines, and academic libraries are evolving to support the corresponding needs across campuses. Data services include not only the long-supported areas of data management and curation, but also geospatial services, data analyses, and data visualization. Compared to previous work by others, we provided a finer view of data service offerings, specifically at R1 academic

libraries in the U.S. Unlike prior work, we also provided an evaluation of how resourcing in the form of salaries and wages relates to services offered. While the current work found differences among services in the modalities in which they were offered, future comparisons could highlight changes brought about by the COVID-19 pandemic and shifts to online learning. Furthermore, recent emphasis on open research tied to federal funding, such as the NIH Policy on Data Management and Sharing (Collins 2020) and the “Nelson Memo” (Nelson 2022), have placed further demands for support from the academic library. The current work, performed before these two policy changes were announced, would serve as a useful point for comparison to determine how the data services at academic libraries have changed (or not) in response to the new federal mandates. There is also potential for a more precise investigation of the influence of resourcing on the amount and type of data services offered by an academic library: increased sampling and finer-scale resourcing information (e.g. total expenditures on data services personnel as opposed to all library personnel) could better measure the impact of resourcing on the data services provided. The influence of limited resources on data services provides an opportunity for inter-institutional collaborative efforts to meet data services needs.

Data Availability

All data and analytical code are available at <https://github.com/UAL-RE/data-service-landscape-scan> and archived at <https://doi.org/10.25422/azu.data.22297177>.

Google form questions and criteria matrix are available under the article [Supplementary Files](#):

Appendix 1: Text of the Google form for web-based inventory

Appendix 2: Criteria matrix for determining whether or not a service is offered by an academic library

Acknowledgements

The authors would like to thank Anna Cirell, Claudio Garcia, Scott Jones, and Christine Kollen for their support in data collection efforts. We also thank the editor Regina Fisher Raboin and two anonymous reviewers for providing constructive feedback on prior versions of this work.

Competing Interests

The authors declare that they have no competing interests.

References

Belsley, David A., Edwin Kuh, and Roy E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

Burton, Matt, Liz Lyon, Chris Erdmann, and Bonnie Tijerina. 2018. “Shifting to Data Savvy: The Future of Data Science In Libraries.” Pittsburgh, PA: University of Pittsburgh. <https://d-scholarship.pitt.edu/33891>.

- Collins, Francis. 2020. "Final NIH Policy for Data Management and Sharing." <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>.
- Cook, R. Dennis. 1977. "Detection of Influential Observation in Linear Regression." *Technometrics* 19 (1): 15. <https://doi.org/10.2307/1268249>.
- Cox, Andrew M., Mary Anne Kennan, Liz Lyon, Stephen Pinfield, and Laura Sbaffi. 2019. "Maturing Research Data Services and the Transformation of Academic Libraries." *Journal of Documentation* 75 (6): 1432–1462. <https://doi.org/10.1108/JD-12-2018-0211>.
- Hamad, Faten, Maha Al-Fadel, and Hussam Fakhouri. 2022. "The Provision of Smart Service at Academic Libraries and Associated Challenges." *Journal of Librarianship and Information Science* July: 096100062211141. <https://doi.org/10.1177/09610006221114173>.
- Hooper, Rachel. 2023. "Big Data: What Is It and How Can Academic Libraries Use It?" *Alabama Libraries* 60 (2): 1–18. https://jagworks.southalabama.edu/alabamalibraries_journal/vol60/iss2/3.
- Hothorn, Torsten, Frank Bretz, and Peter Westfall. 2008. "Simultaneous Inference in General Parametric Models." *Biometrical Journal* 50 (3): 346–363. <https://doi.org/10.1002/bimj.200810425>.
- Johnston, Lisa R, Jake Carlson, Cynthia Hudson-Vitale, Heidi Imker, Wendy Kozlowski, Robert Olendorf, Claire Stewart, et al. 2018. "Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data." *International Journal of Digital Curation* 13 (1): 125–140. <https://doi.org/10.2218/ijdc.v13i1.616>.
- Murray, Matthew, Megan O'Donnell, Mark Laufersweiler, John Novak, Betty Rozum, and Santi Thompson. 2019. "A Survey of the State of Research Data Services in 35 U.S. Academic Libraries, or 'Wow, What a Sweeping Question.'" *Research Ideas and Outcomes* 5 (December): e48809. <https://doi.org/10.3897/rio.5.e48809>.
- Nelson, Alondra. 2022. "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research." August 25, 2022. <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.
- Radecki, Jane, and Rebecca Springer. 2020. "Research Data Services in US Higher Education: A Web-Based Inventory." *Ithaka S+R*. <https://doi.org/10.18665/sr.314397>.
- Schatteman, Alicia, and Li-Yin Liu. 2022. "Measuring What Matters: Comparing Costs and Performance of Municipal Libraries and Library Districts." *Public Library Quarterly* August: 1–23. <https://doi.org/10.1080/01616846.2022.2110631>.
- Si, Li, Wenming Xing, Xiaozhe Zhuang, Xiaoqin Hua, and Limei Zhou. 2015. "Investigation and Analysis of Research Data Services in University Libraries." *The Electronic Library* 33 (3): 417–449. <https://doi.org/10.1108/EL-07-2013-0130>.
- Tenopir, Carol, Ben Birch, and Suzie Allard. 2012. "Academic Libraries and Research Data Services: Current Practices and Plans for the Future." Report. Association of College and Research Libraries. <https://alair.ala.org/handle/11213/17190>.
- Tenopir, Carol, Jordan Kaufman, Robert Sandusky, and Danielle Pollock. 2019. "Research Data Services in Academic Libraries: Where Are Today?" <https://www.choice360.org/research/research-data-services-in-academic-libraries-where-are-we-today>.

U.S. Department of Education. 2021. "Integrated Postsecondary Education Data System." 2021. <https://nces.ed.gov/ipeds/use-the-data>.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.

Wickham, Hadley. 2019. *Stringr: Simple, Consistent Wrappers for Common String Operations*. <https://CRAN.R-project.org/package=stringr>.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.

Wickham, Hadley, and Maximilian Girlich. 2022. *tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr>.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.