Library Services Contributing to Institutional Success at R1 Universities: An Exploratory Mixed-Source Quantitative Model

Elizabeth Szkirpan

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

abstract: As libraries are increasingly asked to demonstrate their value in a fluctuating academic landscape, new methods of evaluation and comparison must be explored to communicate the value of the library quantitatively to university stakeholders. This exploratory study seeks to measure the impact of library services on key institutional metrics by combining multiple data sources for Research 1 (R1) university libraries and analyzing the resulting data using machine learning methods. While data availability impacted the overall outcome of this study, early analysis reveals promising correlations between library services and key university success metrics.

Introduction

Libraries have been in a state of massive flux for decades. While libraries objectively play an important role on university campuses in provisioning space for activities and study, the traditional role of the library as a book repository has undeniably ended. Echoing the changing face of American higher education, library services have evolved drastically in recent years resulting in the expansion and transformation of the traditional role of the library and of the librarian, especially in terms of providing access to digital information and content.[1] However, as the library's role as a physical destination and traditionally in-person only resource declines, some institutions may also see parallel reductions in staffing and perceived importance on university campuses.[2] In addition to these challenges, libraries generally struggle to capture and relay the breadth and value of their services to campus administration.[3]

portal: Libraries and the Academy, Vol. 26, No. 1 (2026), pp. 47–84.
Copyright © 2026 by Johns Hopkins University Press, Baltimore, MD 21218.

There is no unified system in American libraries to measure services and their depth. While some central guidance to libraries on their role within the academy exists by way of the Association of College and Research Libraries (ACRL)'s Standards for Libraries in Higher Education, available resources are high-level and do not prescribe the services that libraries should provide to their constituents based on institutional size or goals.[4] Libraries seeking to communicate value to their institutions often cite geographical or aspirational peer practices to request additional funding, staffing, salaries, and resources, but these metrics may be disregarded if library goals and institutional goals misalign. Libraries generally plan to support institutional goals with the addition of new programs and services to support those aims, but reception of the contribution to the institution's overarching goals may not align with the library's intent. Moreover, the lack of a formally recognized library metrics system further disadvantages small academic libraries, especially those that may be struggling to provide library services even before rising to meet institutional growth plans.

The challenges facing libraries parallel those of the American academy, which has seen a rise in business-like practices in recent decades while simultaneously weathering the changing political, social, and technological landscape of the United States. Researchers attribute the rise in commercialization or commodification of universities to numerous possible causes, such as inclusion of successful business executives on university advisory boards or alignment of American materialism and consumerism with education.[5] Meanwhile, recent Supreme Court rulings centering on college admissions along with campus protests in the wake of the Israel-Hamas war have thrown institutions into the spotlight leading to increased scrutiny from lawmakers and the public alike.[6] As Americans' confidence in higher education trends downward and consumers question the return on investment (ROI) for a university education, higher education is also grappling with the impact of decreased enrollment numbers and decreased funding through both government and tuition sources.[7] As a piece of the academy's ecosystem, libraries inevitably experience the trickle-down effects from a changing academic landscape, which could impact staffing or prioritization of library needs in the broader university landscape. In recent years, universities have increasingly shifted their strategic practices in response to the aforementioned challenges by focusing on overarching institutional goals like boosting enrollment, developing competitive degree programs, securing more research funding, and achieving national accolades, such as a higher Carnegie Classification of Institutions of Higher Education status, to enhance competitiveness and prestige in a crowded academic landscape.

As a piece of the academy's ecosystem, libraries inevitably experience the trickle-down effects from a changing academic landscape, which could impact staffing or prioritization of library needs in the broader university landscape.

In particular, the Carnegie Classification was created to organize and classify the higher education sector, but it is often used to benchmark institutions, inform academic

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

policies, determine institutional peer groups, and more.[8] The Classification system iterates categorical tiers for various higher education types, such as doctoral universities or master's colleges and universities, with Research 1 (R1) encompassing doctoral universities with very high research activity and is the highest possible designation, while Research 2 (R2) encompasses doctoral universities with high research activity.[9] Very high versus high research activity is determined by several variables, including doctoral degrees awarded, doctoral programs offered, and funding spent on research activities.[10] Combined with publicly available datasets, current and aspirational Carnegie Classification statuses can easily be compared to provide valuable insight into successful peer performance.

Achieving R1 status indicates significant institutional investment in the research process while elevating institutional status to that of a select group of peers, thus potentially resulting in increased enrollment, improved faculty recruitment and retention, and grant funding. As such, achieving and maintaining R1 status is often synonymous with institutional definitions of success, and R1 metrics may inform overarching institutional goals.[11] Should institutions succeed in their R1 endeavors, increases in enrollment, awarded degrees, doctoral programming, and research activity will likely require institutional strategizing to absorb increased enrollment, additional faculty hires, and investment in new research areas. However, without library representation in this fluctuating landscape, necessary additional library support, including additional funds for qualified library staff and expansion of resources, may be entirely forgotten or overlooked during the institutional growth cycle.

The lack of a comprehensive, publicly available dataset promoting quantitative analysis of the library's impact on and relationship with overall institutional metrics represents a significant barrier to most institutions seeking a comparative quantitative approach to analyzing service impact and communicating value.

The lack of a national framework and a lack of reliable, open-source data robs libraries of a means to compare metrics among institutions, a way to determine scale of need for institutions aiming for R1 status, or quantitative data to advocate for the library's role in the academy. This study seeks to establish a reproducible method of library service and value communication utilizing predominantly public data sources for R1 institutions as a representative baseline to demonstrate the relationship between library services and metrics of institutional interest, such as enrollment and annual publication outputs.

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Review of the Literature

Assessment of library services, even at the quantitative level, is well documented in the literature with a myriad of books, articles, and web posts continually published on the topic. While researchers investigate library services and their community impact in a variety of ways, many assessments center on the individual library's impact on aspects

of the institution, such as student success and retention, rather than providing benchmark data between institutions. As well, assessments are typically localized, and while they are reproducible with surveys and questionnaires provided by the literature, there is no required contribution to publicly available benchmark data outside of academic library data supplied to the Integrated Postsecondary Education Data System (IPEDS).[12]

Megan Oakleaf's 2010 report is perhaps the most comprehensive in compiling and comparing various avenues to academic library assessment.[13] While there are many proposed ways to assess the value and impact of library services, there is no consensus about the best or universally accepted approach. Many assessment studies point to the creation of institution-specific assessment tools to bolster library research, which ultimately requires a significant investment of staff time and may lead to impossible barriers for libraries with limited staffing. Additionally, many of these studies result in qualitative data, which requires additional processing and can be less valued by institutional administrators compared to other data types.[14]

Judy Luther's model appears to be the most relevant to the discussion of quantifying library services to justify institutional investment. Luther sought to determine the library's impact on university income saying, "Academic libraries are increasingly being challenged to demonstrate their value to their institution in compelling quantitative terms. There is a growing need to provide a response based on sound methodology to questions about the value of the university's investment in the library."[15] Luther builds on earlier explorations of return on investment (ROI) models and their possible application to libraries to connect financial investment in library resources to metrics and trend data that might represent value to university administrators, such as grant awards and citations. However, Luther's model incorporates metrics that may be difficult for individual libraries to obtain and does not include the full range of services that libraries may provide to aid in faculty success, such as library staffing or programming.

The inability to meaningfully connect the work of the library to the overarching success of the institution potentially impacts the library in several negative ways, including and especially from the stakeholder perspective. Available studies find a range of stakeholder perspectives of the academic library, with some administrators generally trusting the work of their libraries while others question the relevancy of library services in today's university model. Deborah Jeanne Grimes tackled the disconnect between the library's view of its own centrality and organizational reality in a 1993 dissertation that investigated the history and accuracy of "the library is the heart of the university" metaphor.[16] Grimes's study found that while the library held an important symbolic role on many university campuses, faculty and academic officers did not view the library as an important contributor to the university's educational goals. They instead pointed to students, faculty, or even technology as a more apt metaphor for university centrism. Grimes attributed this disconnect to many possible causes and emphasized that decades of discourse pointed to the disconnect between the library's approach to theory, research, instruction, and integration into the organizational structure as being at odds with organizational reality and often placing libraries into a poorly defined role within the academy. Leigh S. Estabrook echoes many of these sentiments and succinctly captures the challenges of a changing academic landscape and communicating value to university administrators in a commodified academy using a chief academic officer's

This mss. is peer reviewed, copy edited, and accepted for publication. portal 26.1.

quote, reporting "what we see today in top academic leadership is people who are not academics themselves or whose priority is not the academic mission, but the business model of advancing the institution. They are going to invest in things that they can point out, not the number of people who appreciate we've doubled the size of the collection."[17]

The conflicting perceptions of the library's role and purpose continue to be evident in multiple recent studies. Jennifer K. Frederick and Christine Wolff-Eisenberg's 2020 survey found that "library directors perceive the value of their roles—and the roles of their libraries—to be declining in the eyes of their supervisors and other higher education leaders. Continuing a trend found in our 2016 survey, directors at all institution types feel less valued by, involved with, and aligned strategically with their supervisors and other senior academic leadership."[18] Adam Murray and Ashley Ireland's 2019 study mirrors these findings from the provosts' perspective, saying "the overwhelming response by provosts that their campuses overall do not recognize the role the library can play in retention initiatives is likewise evidence of the continued work librarians and library directors need to do in making these connections in a way that is visible to faculty and administrators."[19]

The lack of a comprehensive, publicly available dataset promoting quantitative analysis of the library's impact on and relationship with overall institutional metrics represents a significant barrier to most institutions seeking a comparative quantitative approach to analyzing service impact and communicating value. One avenue to an open quantitative approach is the academic libraries data collected by the Integrated Postsecondary Education Data System (IPEDS). Integrated into the IPEDS survey in various ways from 1988 to 2014 before being permanently integrated in 2014, IPEDS captures standardized library variables "from academic libraries serving degree-granting, Title IV postsecondary institutions in the 50 states, the District of Columbia, and the outlying areas."[20] IPEDS is unique compared to other surveys capturing library data, as "[t]he completion of all IPEDS surveys is mandatory for institutions that participate in or are applicants for participation in any federal student financial aid program" and IPEDS results are freely available.[21]

Library studies have utilized IPEDS data in many ways, including to analyze the relationship between library services and student retention or graduation rates, to analyze long-range trends in academic library staffing and expenditures, and to estimate numbers of solo librarians.[22] While these studies look at overall trends, they do not rise to John Cocklin's 2008 assertion that IPEDS can potentially play a valuable role in communicating the library's role in its institution, saying, "As assessment grows in importance to academic libraries, librarians are looking for information that will complement their user surveys and place the library within the larger context of the people and institutions they support. IPEDS not only provides consistent data over time for the college or university that a library directly supports, it also provides consistent data for peer institutions as well."[23]

One library-oriented application of IPEDS data is carried out annually by the Association of College & Research Libraries (ACRL) by way of their annual Academic Library Trends and Statistics Survey. ACRL combines IPEDS library data with a supplemental survey to analyze overarching academic library trends resulting in subsequent reports, academic publications, and an enhanced dataset that compliments each year's IPEDS

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

data for libraries. Resulting reports help contextualize trends in libraries over multiple years and summarize each year's data in terms of impact to libraries.[24] While ACRL's survey undeniably captures important insights into industry trends, participation in the survey is entirely voluntary with 45.2 percent (n=1,414) of US academic institutions with libraries responding to the 2023 survey.[25] Although this response rate represents a significant sample size, the voluntary nature of the survey also introduces the possibility of sampling bias, as it cannot fully capture the experiences of non-participating libraries. Additionally, access to current and previous survey results requires a paid subscription, which limits accessibility, restricts further exploration of the data, and marginalizes libraries or professionals unable to afford the subscription. [26]

Given the perception of library impact on academic success by key stakeholders, it is clear that libraries must improve upon communication of value using avenues that speak to stakeholder information needs. However, the lack of centralized library data that mesh with existing institutional data restricts use among institutions of various sizes, as well as the creation of national standards that can be used by both libraries and their overarching institutions to guide investment into the library and its services. This fact is even more stark given the proposed dissolution of IPEDS academic library data collection for the 2025-2026 cycle because of Department of Education budgetary and workload concerns.[27] Ultimately, doing so would eliminate the only required and standardized data collection for academic libraries. Librarians have objected to the removal of IPEDS academic library data citing that "the survey data gives academic libraries visibility as an integral component of higher education institutions that supports information literacy, faculty research and student success."[28]

The case for utilization of machine learning, big data, and open data library initiatives is also well documented.[29] There are ample opportunities for libraries to explore and integrate these tools into library services to better support users and embrace ever-changing technologies, as well as to use these emerging ideas to analyze library data. Despite existing research into the separate areas of library assessment, campus stakeholder perceptions of libraries, the use of IPEDS data in library research, R1 academic libraries, and the opportunities for use of machine learning or data in libraries, there is no consistent intersection of these research areas nor consistent implementation of tactics when conveying the library's critical contributions to campus in support of institutional goals.

Methodology and Data Collection

While not fully encompassing all research library experiences and capabilities, this exploratory study focused on Research 1 (R1) libraries, as R1 institutions are often the aspirational peers for universities aiming to increase research output, and because the services provided by libraries positively impact metrics of importance to R1 institutions. Hypothetically, institutions prioritizing research output will invest in library services to support high research output. In turn, this may be correlated to other institutional impact factors, like retention, as libraries typically provide services that benefit researchers at every skill level and provide access to auxiliary resources like study space and technology to support student learning. The variables and resulting data for this project

This mss is peer reviewed, copy edited, and accepted for publication, portal 26.1.

evolved numerous times to increase transparency and reproducibility of the dataset and its conclusions. In particular, the investigator aimed to use openly available or readily accessible data sources to ensure broad possible use of the data and data sources. This study builds upon the researcher's prior investigation into services provided by R1 and R2 libraries to support online learners and utilized a similar method of manual compilation of data from library websites.[30] However, the investigation into the role of library service impact on institutional success differs significantly in its sole focus on R1 libraries. It revises the areas of service impact to eliminate emphasis on online services, while expanding to include additional in-person services, and incorporates numerous data sources to bolster overall data and study reliability. Using lessons learned from the previous study, library services investigated for this study were devised by canvasing library websites of the first twenty institutions alphabetically from a list of 2023 R1 universities. Library services were then compiled from the library websites of the selected twenty institutions to identify the various features, services, and resources highlighted on library front pages, about pages, database pages, service pages, FAQs, library maps, and help center articles as services or benefits to their scholarly community that were standardly or commonly offered among institutions. Lastly, these variables were cross-referenced with variables from the prior study, which centered around online library services, to identify any services that may have been less prevalent in R1 libraries, such as microcredentialing services, but of potential interest to non-R1 libraries aiming to benchmark against the compiled data.

The services that libraries provide to their communities are of particular interest to this study because of the ever-evolving role of librarians as they serve the evolving academic landscape. There are no definitive lists of the services that libraries provide or should provide to their communities, so a library's resulting services may be devised based on a library's response to its environment, by patron or administrative request, in response to another library's services, to support specific curriculum, to address recurring requests, and so on. While some library services, like reference help and interlibrary loan, are likely available at all R1 libraries, exploration and inclusion into recurring services provides additional contrast to quantitative data and allows for the comparison of the impact of specific library services against specific institutional metrics. Also, while this specific study explores available data and data analysis for R1 libraries, identification of recurring library services in these institutions expands possible application to non-R1 libraries that may aim to benchmark their own services and metrics against any collected data.

Data for this study were gathered in two phases between June, 2023 and May, 2024. Phase one data were manually collected from each of the 146 R1 library websites from June to July, 2023, to identify explicit evidence of library services, such as interlibrary loan, document delivery, library instructional space, and tutorials within a spreadsheet. The investigator utilized library front pages, about pages, database pages, and other public pages to record evidence or lack of evidence for these services in a spreadsheet. Evidence of a provided library service was recorded as a "1" in the dataset while lack of evidence of a library service provided was recorded as a "0." Additional quantitative and categorical data, such as the number of databases, number and type of staffing, and library discovery system results were also recorded at the time of this first data collec-

This mss. is peer reviewed, copy edited and accepted for publication, portal 26.1.

tion pass. It is worth noting, however, that only explicit evidence of these services was recorded and without direct input from these libraries, the services and their depth are difficult to fully gauge. Additionally, library websites and systems vary greatly from one institution to another, potentially skewing data if a library website does not index or clearly communicate certain resources, staff members, et cetera. While estimating the depth of each library service is not possible using manual data collection techniques using library websites, the listed availability of these services reflects a similar experience to that which library users, stakeholders, and peer institutions experience when researching a library's services and reflects the availability of open data given these limitations.

Manually scraped data were then combined with US News & World Report college and university rankings data and Scopus research output data for each institution to help contrast library services data with other key institutional metrics, such as enrollment numbers. Data gathered from US News & World Report included Fall 2022 enrollment data while Scopus, a citation and indexing tool available by subscription through Elsevier, provided research output statistics and the top research output subject areas for the 2022 calendar year.

However, to improve data standardization where possible and ensure data use compliance, raw library services data were revised in the second iteration of this study's dataset created in May, 2024 to combine IPEDS academic library data and institutional metrics in place of US News & World Report data. Relying on IPEDS data instead ensured transparent data collection practices and standardized variables, such as staffing and library databases, across all institutions, while providing new data that could not previously be captured from library websites alone in terms of library expenditures, circulation statistics, and iterated staffing information. IPEDS data for graduation numbers, degrees awarded, enrollment, and research expenditures were extracted and combined with the library services data from the same time to provide insight into the relationship between libraries and key institutional metrics. While alternative sources for research output quantitative data were explored, the investigator selected Scopus for its clear variable explanations and ultimately received permission from Elsevier to utilize and publish data extracted from the platform. Both phases of data collection were combined with Scopus metrics for institutional research documents published in 2022 and top-level Scopus data concerning the largest subject area published by the institution to date.

Lastly, columns of data were also added for each institution to indicate whether the library was a member of the Association of Research Libraries (ARL) or not and the total number of records housed in WorldCat for each institution. ARL data were compiled in the June, 2023 collection phase by comparing institution names to the ARL membership list and recording ones for presence or zeros for absence within the dataset. Meanwhile, WorldCat data were added to the dataset during the May, 2024 revision by searching each R1 institution's name, identifying the main library entry within WorldCat's database, and then recording the number of items reported as cataloged in WorldCat.

While library data reported to IPEDS or collected by Scopus allow for point-in-time comparisons and data from the 2022 school year or 2022 calendar year are utilized where possible in this dataset's collection, data scraped in the May, 2024 collection phase from library websites or WorldCat cannot accurately be tied to academic or calendar year 2022 search results. Neither library websites nor WorldCat provide static data for certain points

This mss. is peer reviewed, copy edited, and accepted for publication in portal 26.1.

in time, meaning that these variables can fluctuate day-to-day. However, quantitative data pulled from these searches is unlikely to significantly change in one year (for example, library collection sizes and system search result numbers are fairly static year-to-year given decreased emphasis on print material or static library budgets) and there is not a reliable method for collecting historic library data using web resources. As a result, data available at the time of collection were recorded. Given the lack of reliable website time-series data for each library, data in these areas may be slightly skewed, reflect changes in library indexing and collections, or represent services that were not provided by the library during the corresponding academic or calendar 2022 year. Also, data collected from Scopus in June, 2023 was accurate at the time of collection, but periodic updates to resources indexed by Scopus can lead to fluctuations in citation counts, which could be further impacted by ongoing changes to the Scopus platform in 2024.[31]

The resulting CSV datafile utilized for subsequent analysis and conclusions contains 82 variables for 146 R1 libraries. Final variables used capture library staffing, expenditures, collection sizes and circulation, services provided, professional memberships by way of ARL, and additional institutional level data, such as enrollment, research output, and primary research areas. Resulting data exclusively represents R1 university libraries and does not include other Carnegie status designations. All final variables captured and their provenance, Python code created for analysis, and publicly accessible versions of the data are available in a dedicated open access Dataverse collection.[32] Generative AI in the form of ChatGPT was utilized in this project to help write, edit, and troubleshoot the Python code but was not utilized to collect or interpret data. Transcripts of the Chat GPT conversations regarding the code are also available within the Dataverse collection.

Data Analysis

Data were analyzed using both Excel and Python via Jupyter Notebook to create visualizations, calculate data points, and explore data relationships using machine learning methods. General data summaries were first explored using Excel to summarize the overall dataset, revealing that R1 institutions represent a diverse arrangement of identities and research foci with 107 public institutions, 39 private institutions, 103 ARL members, and 43 non-ARL members as represented in Figure 1.

As represented in Figure 2, the largest research output areas for R1 institutions generally fall into the broad science, technology, engineering, and mathematics (STEM) family with 41.8 percent (n=61) of R1 institutions outputting most research in medicine subjects and approximately 19.9 percent (n=29) of institutions producing their largest quantity of research in engineering.

Elsevier's Scopus database, which indexes abstracts and citations, was utilized to determine the largest research output area for each institution. Within Scopus, each R1 institution was individually searched to reveal the institution's document count for all time. Then, the subject area with the highest number was recorded as the largest research output area. Summarized from Figure 2, the largest research output areas for R1 institutions are as follows: agricultural and biological sciences (n=15); biochemistry, genetics, and molecular biology (n=2); computer science (n=2); earth and planetary sciences (n=2); engineering (n=29); medicine (n=61); physics and astronomy (n=16); and social sciences (n=19).

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

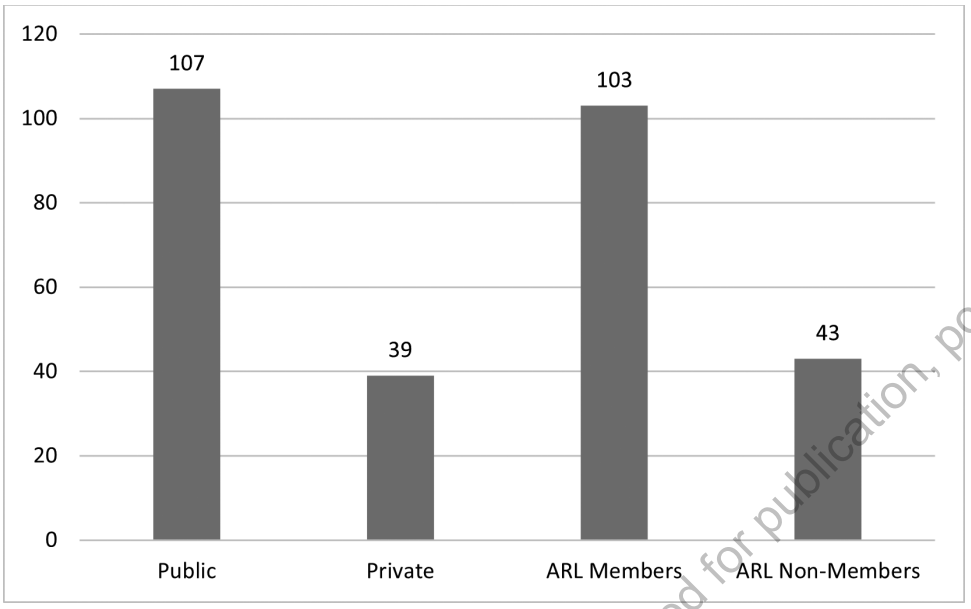

Figure 1. Research 1 institutions: Public versus privately funded universities and Association of Research Libraries (ARL) members versus non-members.

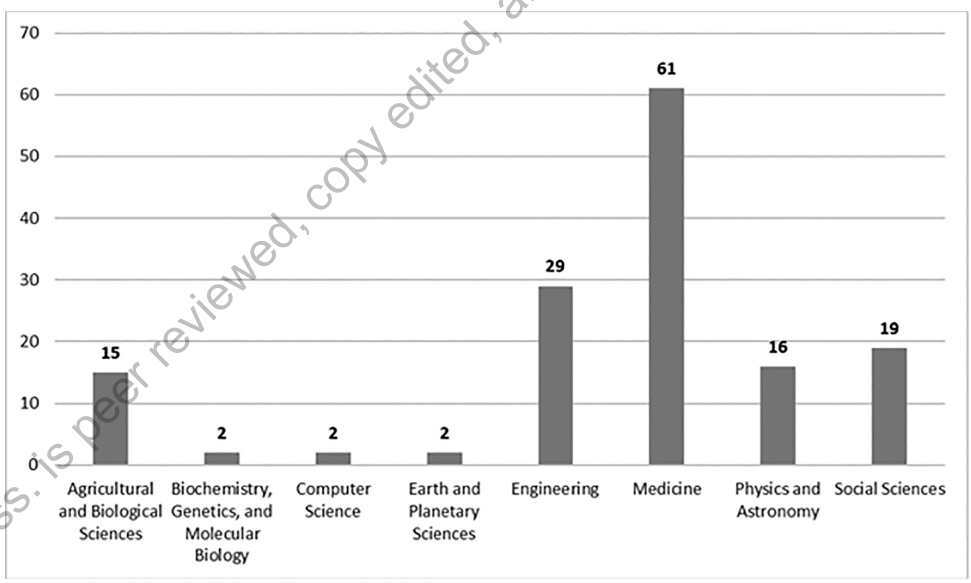

Figure 2. Research 1 institutions: Largest research output area based on Scopus publication materials.

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Table 1.

Calculated average, minimum, and maximum values for selected variables from the dataset for R1 libraries.

Metric	Average	Minimum	Maximum
Research Output	5,222.4	585.0	32,447.0
Total Enrollment	29,590.3	2,401.0	80,065.0
Total Library Staff	187.9	16.8	811.5
Total Physical Items	3,153,045.7	220,439.0	26,767,405.0
Total Electronic Items	2,393,384.2	350,134.0	10,540,505.0
Total Salaries	$10,577,128.50	$848,067.00	$56,138,193.00
Total Expenses	$28,515,045.20	$2,997,735.00	$162,603,770.00

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Analyzing the average, minimum, and maximum of data variables for the overarching institution similarly reveals a range of institutional experiences as summarized in Table 1.

Data summaries reveal significant outliers, especially at institutions with markedly larger collections, staff sizes, and research output. These outliers were retained in the subsequent data analysis but skew resulting averages and equations. Separating the data by private and public institutional control also revealed higher average, minimum, and maximum values for private institutions in nearly all variables, also likely impacting overall data analysis.

While R1 institutions represent a wide range of experiences, R1 libraries are less diverse, offering similar services and benefits regardless of overall institution size or metrics, as outlined in Figure 3.

R1 library services and experiences look homogeneous in the graphs produced using Excel, such as in Figure 4 comparing library space services. However, when combined with other collected library datapoints, despite similar levels of offered services, datapoints such as library staffing or collection size vary drastically within the dataset. Depth of these services certainly plays a role in interpreting how these library services specifically vary from one library to another. For example, if a library offers copyright support, is that through a research guide, a dedicated librarian, or a department within the library? Nevertheless, the data also indicate that there is something of an expectation for academic libraries to provide most or all the identified library services regardless of the actual library's resources or bandwidth. Alternatively, aspiring R1 libraries may look to the homogeneity of other R1 library services as a roadmap to determine the types of services they may be called upon to provide should their institution achieve future R1 status.

After initial exploration using Excel, institution names were removed from a dedicated copy of the data before importing it into Jupyter Notebook for further analysis. The

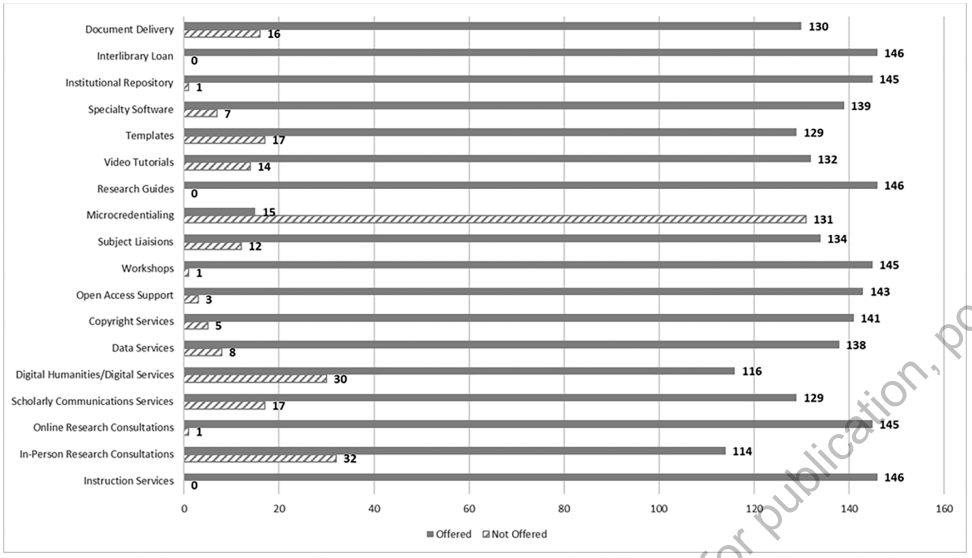

Figure 3. Research 1 library-offered research support services.

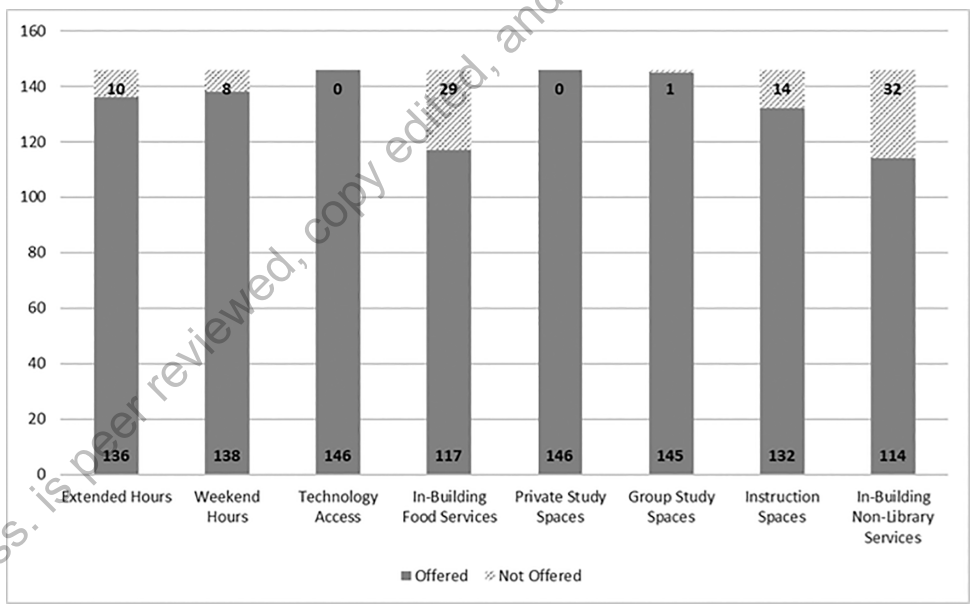

Figure 4. Research 1 library-offered spaces.

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

categorical variables for public or private institutional control and research areas were converted to binary variables, a series of 0s and 1s to represent whether an institution does (1) or does not (0) match a categorical variable, resulting in 87 total variables for 146 institutions. Data were then explored visually again to review simple one-to-one relationships amongst key variables, such as total library staffing, institutional research output, institutional enrollment, and library system results. Additional information about the variables represented in subsequent analysis, including definitions and data source, is available in Appendix A.

Total library staffing and institutional research output was first explored with the researcher hypothesizing that institutions with high research output would need more library staffing to support research. As depicted in Figure 5, the comparison of total library staffing and institutional research output reveals a positive skew within the dataset, with institutions generally reporting a higher number of library staffers as institutional research output increases.

Along the same lines, hypothetically, institutions with large enrollment numbers should have more library staffers for scholarly support. Similar to total library staffing against research output, Figure 6 revealed that plotting total library staff against total enrollment reveals a relationship of moderate positive skew with institutions typically reporting a higher number of library staffers as total enrollment increases. While this data indicates positive correlation between library staffing and key institutional metrics, the data cannot be utilized in its current form to determine whether there is a direct correlation between these variables or if the relationship is the result of other factors, such as institutions with higher research output and enrollment potentially distributing more

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

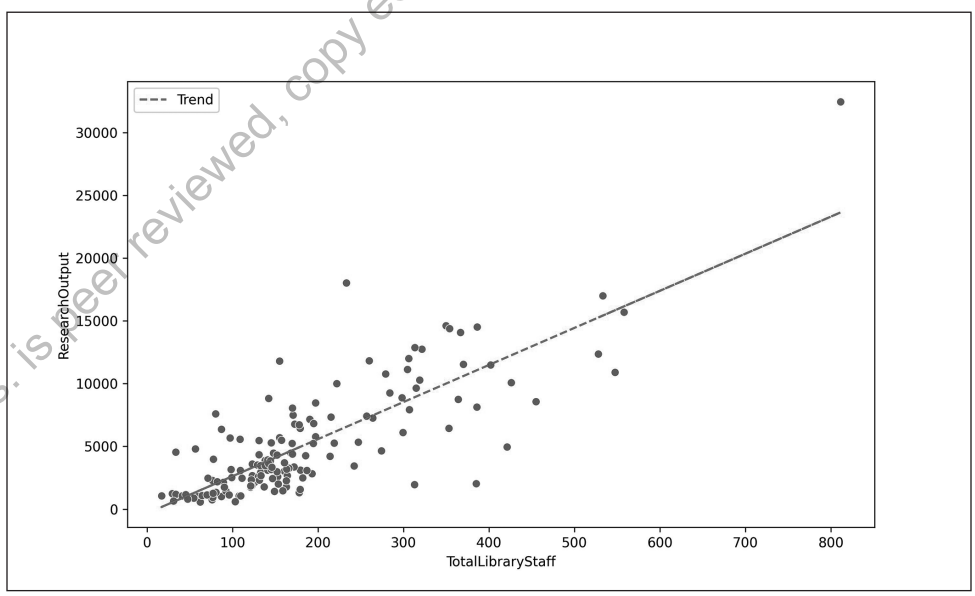

Figure 5. Scatter plot showing the relationship between total library staff and institutional research output.

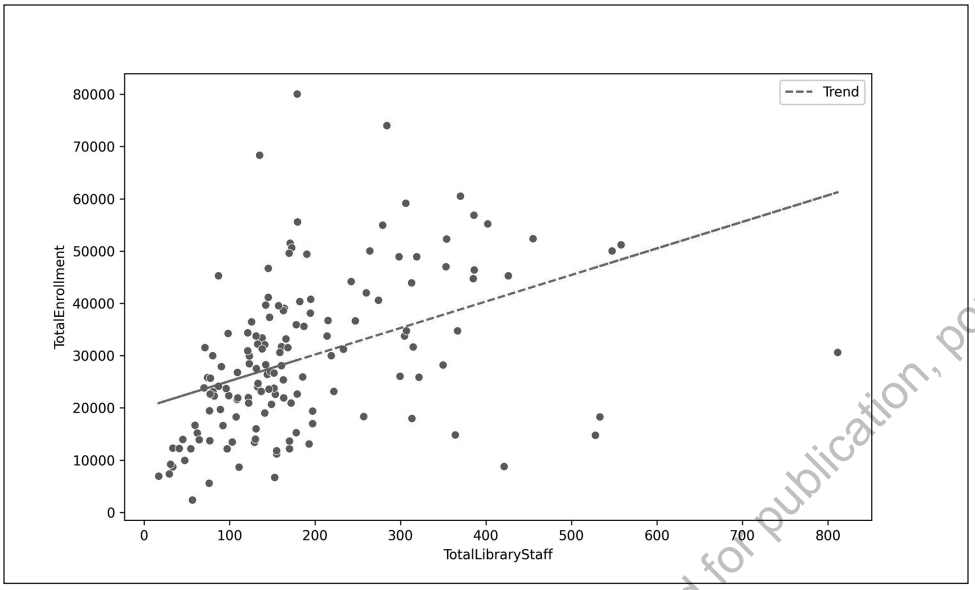

Figure 6. Scatter plot showing the relationship between total library staff and total enrollment.

funds to libraries as a result of higher institutional income, rather than preparation or reward for library contributions.

While the graphs for total library staffing in relation to institutional research output and enrollment are promising as a tool to demonstrate measurable library impact, the correlations between other institutional metrics do not always reveal positive relationships. For example, in Figure 7, plotting library system results against research output reveals little positive correlation. Hypothetically, library resources support institutional research output and libraries with more resources would see higher research output. Instead, the resulting graph reveals a myriad of experiences with some institutions producing large amounts of research output with smaller system results while other libraries have higher than average system results but average or low research output.

System results, a variable designed to emulate the broad library searches that researchers may conduct as they undertake a given project—crafting a literature review, obtaining access to current and historic journals, accessing research data—are not necessarily indicative of resources available to library researchers, especially as library systems are highly customizable and vary amongst institutions. As a result, rather than seeing a positive relationship between system results and research output indicating that more available resources equate to more research output, there is little obvious correlation in the data.

Data were analyzed using Python in Jupyter Notebook with the Pandas, Seaborn, Statsmodels, Scikit-learn, and NumPy libraries to analyze, visualize, manipulate, and conduct machine learning by way of linear regressions on the data. In particular, machine learning methods were employed to explain relationships within the data mathematically with the goal of creating a dependable model that allows libraries to input their own

This ms. is peer reviewed, copyedited, and accepted for publication, portal 26.1.

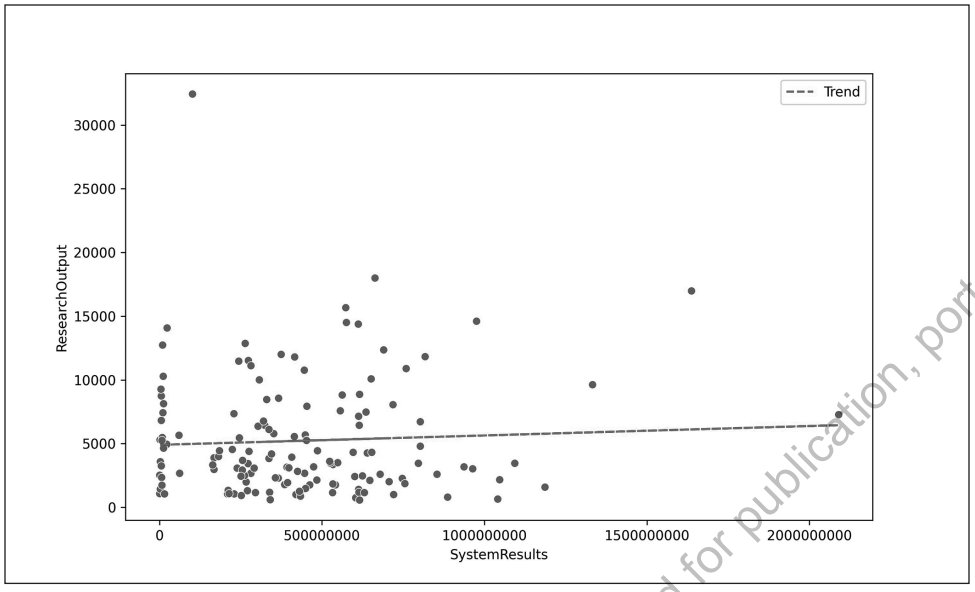

Figure 7. Scatter plot showing the relationship between total library staff and total enrollment.

known variables to calculate impact to the institution or to quantify needs. While data could be analyzed and explored indefinitely using machine learning approaches, linear regression equations, or equations that predict variables based on one or many other variables, were selected as an easy-to-use and easy-to-explain methodology for libraries.

Ultimately, three simple linear regression equations representing one-to-one relationships in the data, and three multilinear regression equations representing one-to-many relationships in the data were created using machine learning methods within Python and then analyzed for goodness of fit by reviewing their associated regression output summaries provided by the code. Multiple machine learning methods could be applied to this dataset but linear regression equations were selected to experiment with a "plug and play" solution for libraries to predict their needs for achieving certain institutional metrics. Subsequent equations were trained on the compiled R1 library data so equations best represent the experiences of R1 libraries and will skew towards those metrics. However, they could still be used for experimentation and estimation purposes by non-R1 libraries to benchmark against aspirational peers.

The relationship between research output and librarians Full Time Equivalency (FTE) was explored first to mathematically explain the previously generated scatterplot, resulting in:

$$\text{ResearchOutput} = 758.0567 + 73.4540 \times \text{LibrariansFTE}$$

In this equation, LibrariansFTE acts as the independent variable to help predict ResearchOutput, the dependent variable. Based on the analyzed data, institutional research output can be estimated based on librarians FTE by adding 758.0567 as the equation's coefficient to 73.4540 for each librarian FTE. For example, if there is only one (1) librarian FTE, the equation can be completed by calculating:

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

$$\text{ResearchOutput} = 758.0567 + 73.4540 \times (1)$$
$$\text{ResearchOutput} = 831.5107$$

Based on this example, an institution's research output can be estimated as approximately 831.5 documents annually if only one librarian FTE is present. As the number of librarian FTE increases at the institution, research output also increases. For all equations, the adjusted R-squared value, which measures the percentage of variance in the dependent variable that can be explained by the independent variable, was selected as the regression output to measure the goodness of fit of the equation. Simple linear regressions are typically measured using the R-squared value because the adjusted R-squared value incorporates the number of predictor variables into its output calculation. However, the R-squared and adjusted R-squared values are similar for simple linear regressions and can provide a simple comparison among all equations. Review of the adjusted R-squared value in this equation's regression output estimates that approximately 61.4 percent (n=0.614) of the variance in institutional research output can be explained by librarian staffers. While the threshold for adjusted R-squared values as an indication of a strong statistical relationship varies between industries, an adjusted R-squared above 50 percent (n=0.5) is generally considered an indication of good model fit.

The relationship between research output and total expenses was also explored to investigate whether well-funded libraries resulted in higher institutional research output, resulting in:

$$\text{ResearchOutput} = -88.2565 + 0.0002 \times \text{TotalExpenses}$$

In this equation, TotalExpenses acts as the independent variable to help predict ResearchOutput, the dependent variable. Based on the analyzed data, institutional ResearchOutput can be estimated based on TotalExpenses by subtracting 88.2565 as the equation's coefficient to 0.0002 for each dollar spent in TotalExpenses. For example, if $100,000 is spent annually by the library, the equation can be completed by calculating:

$$\text{ResearchOutput} = -88.2565 + 0.0002 \text{ x } (100{,}000)$$
$$\text{ResearchOutput} = -68.2565$$

Based this example, the institution's research output can be estimated as decreasing by approximately 68 documents annually for every $100,000 spent by the library. However, as the expenditures of the library increase, research output will also increase given the positive impact of expenses on the equation. Review of the adjusted R-squared value in this equation's regression output estimates that approximately 77 percent (n=0.771) of the variance in institutional research output can be explained by the library's total expenses.

The relationship between total institutional enrollment and total library staff was explored as the last of the simple linear regression equations to investigate whether higher enrollment resulted in proportionate library staffing, resulting in:

$$\text{TotalLibraryStaff} = 81.8161 + 0.0036 \times \text{TotalEnrollment}$$

In this equation, TotalEnrollment acts as the independent variable to help predict TotalLibraryStaff, the dependent variable. Based on the analyzed data, TotalLibraryStaff can be estimated based on adding 81.8161 as the equation's coefficient to 0.0036 for

This mss. is peer reviewed, copy edited, and accepted for publication portal 26.1.

each enrolled student. For example, if 1,000 students are enrolled, the equation can be completed by calculating:

$$\text{TotalLibraryStaff} = 81.8161 + 0.0036 \times (1{,}000)$$
$$\text{TotalLibraryStaff} = 85.4161$$

Based on this example, the institution's total library staffing can be estimated as amounting to approximately 85 staff members for 1,000 enrolled students. Review of the adjusted R-squared value in this equation's regression output estimates that only 17.6 percent (n=0.176) of the variance in total library staffing can be explained by the institution's total enrollment.

After the creation of these three simple linear regressions, the research area categories in the data were also converted to binary variables so they could be taken into consideration in multiple linear regression equations. Multiple linear regression equations allow multiple variables to be taken into consideration as independent variables to help predict a single dependent variable. Institutional research output was selected as the first dependent variable to be predicted with all 86 remaining variables:

Multiple linear regression equations allow multiple variables to be taken into consideration as independent variables to help predict a single dependent variable.

ResearchOutput = -201.2896(LibraryStaff) + 6.5854(TotalLibraryStaff) +
10.8654(LibrariansFTE) + 2.0791(ProfessionalFTE) + -0.2582 (OtherFTE) +
-6.0999(StudentFTE) + -44.8469(NumberofLibraries) + 0.00000000383(SystemResults) +
-0.0000000224(ArticleResults) + 0.0002(WorldCatResults) + 0.00000250(PhysicalBooks)
+ -0.0002 (PhysicalMedia) + 0.0006(PhysicalSerials) + 0.0004(TotalPhysical) +
-0.00000413(PhysicalCirculation) + 0.0008(ElectronicBooks) + -0.0017(Databases)
+ 0.0007(ElectronicMedia) + -0.0003(ElectronicSerials) + -.0005(TotalElectronic) +
0.000000174(ElectronicCirculation) +
-0.0001(TotalCollections) + 0.00000612(TotalCirculations) +
-201.2896(InstructionServices) + -50.3082(InPersonConsultations) +
4657.8984(OnlineResearchConsultations) + 100.954(ScholarlyCommunications) +
-266.8486(DigitalServices) + 83.1289(DataServices) + -1618.5788(CopyrightServices) +
328.5681(OpenAccess) + 1153.8698(Workshops) + 126.0091(SubjectLiaisions) +
-619.1735(Microcredentialing) + -761.8584(ExtendedHours) +
247.4272(WeekendHours) + -201.2896(ResearchGuides) + 369.1752(VideoTutorials) +
47.7161(Templates) + 11.36(SpecialtySoftware) + -227.5878(InstitutionalRepository) +
-201.2896(ILL) + -0.0015(ILLProvided) + -0.0005(ILLReceived) +
-1091.0539(DocumentDelivery) +
-201.2896(TechnologyAccess) + 420.3605(FoodServices) + -201.2896(PrivateStudy) +
-2393.191(GroupStudy) + 139.2356(InstructionSpace) + -775.1881(NonLibraryServices)
+ -711.1792(ARLMembership) + -201.2896(ExpendituresOver100) +
0.0002(TotalSalaries) + -651.8139(FringeBenefits) + 0.0002(BenefitsAmount) +
-0.00000245(OTPurchases) + 0.00000419(Subscriptions) +
0.0001(OtherMaterialsExpenses) + 0.0001(TotalMaterialsExpenses) +

This mss. is peer reviewed, copy edited, and accepted for publication; portal 26.1.

-0.0017(PreservationExpenses) + 0.001(OtherOperationsExpenses) +
-0.0006(TotalOperationsandMaintenanceExpenses) + 0.00000547(TotalExpenses) +
-0.0003(TotalExpensesLessBenefits) + -0.007(BachelorsDegrees) +
-0.2624(MastersDegrees) + 5.7985(ResearchDoctoralDegrees) +
0.4879(ProfessionalDoctoralDegrees) + 1.1706(OtherDoctoralDegrees) +
0.0619(TotalEnrollment) + 0.0067(FTEnrollment) + 0.0552(PTEnrollment) +
-0.0708(UndergradEnrollment) + 0.1328(GraduateEnrollment) +
0.046(ResearchExpenses) + 165.8867(Control_Private) + -367.1764(Control_Public) +
-329.7073(ResearchArea_AgriculturalandBiologicalSciences) +
1166.8485(ResearchArea_BiochemistryGeneticsandMolecularBiology) +
-37.0143(ResearchArea_ComputerScience) + 732.6788(ResearchArea_
EarthandPlanetarySciences) +
-460.3653(ResearchArea_Engineering) + 392.3095(ResearchArea_Medicine) +
-689.0558(ResearchArea_PhysicsandAstronomy) +
-976.9837(ResearchArea_SocialSciences)

As with prior examples, libraries could utilize their own data to fill in the equation above to help predict their institutional research output based on a myriad of library and institutional variables. Some variables, such as LibraryStaffFTE and collection counts, could be completed using available IPEDS data, while other variables could be filled in directly with library collected data, such as System Results and Article Counts. Categorical variables representing library services, Control, and Research Area would be completed using ones (1) to represent presence or zeros (0) to represent absence in the equation. The regression output for this equation appears to perform well with an adjusted R-squared of 91.2 percent (n=0.912). However, p-values, a regression output measuring statistical significance of included variables, perform poorly with only research expenses and social sciences as a research area, indicating statistical significance to the model. These conflicting measurements of equation goodness are likely due to multicollinearity issues within the dataset as many included variables are heavily related or calculated from one another. Reduction of included variables would likely provide a more accurate equation.

A multiple linear regression model was also created to predict Total Enrollment as the dependent variable with all remaining 86 variables as independent variables or predictors.

Total Enrollment= -1.55E-15(ResearchOutput) + 1.25E-12(LibraryStaff) +
-7.34E-13(TotalLibraryStaff) + 7.71E-13(LibrariansFTE) + 9.20E-13(ProfessionalFTE) +
9.00E-13(OtherFTE) + 7.03E-13(StudentFTE) + 4.87E-12(NumberofLibraries) +
-6.06E-17(SystemResults) + 2.28E-17(ArticleResults) + 3.11E-17(WorldCatResults) +
2.62E-14(PhysicalBooks) + 2.63E-14(PhysicalMedia) + 2.61E-14(PhysicalSerials) +
-4.97E-14(TotalPhysical) + 1.31E-14(PhysicalCirculation) + 4.00E-15(ElectronicBooks) +
3.78E-15(Databases) + 4.11E-15(ElectronicMedia) + 4.03E-15(ElectronicSerials) +
-2.77E-14(TotalElectronic) + 1.26E-14(ElectronicCirculation) +
2.36E - 14(TotalCollections) + -1.25E-14(TotalCirculations) +
6.82E-13(InstructionServices) + 2.25E-11(InPersonConsultations) +
2.56E-11(OnlineResearchConsultations) + 8.98E-12(ScholarlyCommunications) +
9.21E-12(DigitalServices) + -2.41E-11(DataServices) +
-3.75E-11(CopyrightServices) + -3.64E-12(OpenAccess) + 2.05E-11(Workshops) +
9.32E-12 (SubjectLiaisions) + 2.16E-12(Microcredentialing) +

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

-7.73E-12(ExtendedHours) + -2.18E-11(WeekendHours) + -5.00E-12(ResearchGuides) +
3.32E-11(VideoTutorials) + 9.10E-12(Templates) + -4.09E-12(SpecialtySoftware) +
-1.73E-11(InstitutionalRepository) + -7.96E-13(ILL) + -3.47E-16(ILLProvided) +
-9.71E-17(ILLReceived) + -2.50E-11(DocumentDelivery) +
2.50E-12(TechnologyAccess) + -9.66E-12(FoodServices) + -2.39E-12(PrivateStudy) +
2.44E-11(GroupStudy) + -2.73E-11(InstructionSpace) + 1.30E-11(NonLibraryServices)
+ 8.64E-12(ARLMembership) + -2.39E-12(ExpendituresOver100) +
-1.06E-13(TotalSalaries) +
-2.05E-12(FringeBenefits) + 3.57E-13(BenefitsAmount) + -2.99E-14(OTPurchases) +
-2.99E-14(Subscriptions) + -2.98E-14(OtherMaterialsExpenses) +
-7.56E-14(TotalMaterialsExpenses) + -5.11E-14(PreservationExpenses) +
-5.04E-14(OtherOperationsExpenses) +
-5.52E-14(TotalOperationalandMaintenanceExpenses) + -3.57E-13(TotalExpenses) +
4.62E-13(TotalExpensesLessBenefits) + 2.93E-15(BachelorsDegrees) +
-5.99E-15(MastersDegrees) + -4.49E-14(ResearchDoctoralDegrees) +
-2.96E-14(ProfessionalDoctoralDegrees) + -1.99E-13(OtherDoctoralDegrees)
+ 0.5(FTEnrollment) + 0.5(PTEnrollment) + 0.5(UndergradEnrollment) +
0.5(GraduateEnrollment) + -1.65E-17(ResearchExpenses) +
-1.19E-11(Control_Private) + 5.68E-12(Control_Public) +
2.53E-12(ResearchArea_AgriculturalandBiologicalSciences) +
2.91E-11(ResearchArea_BiochemistryGeneticsandMolecularBiology) +
-2.64E-11(ResearchArea_ComputerScience) +
-1.00E-11(ResearchArea_EarthandPlanetarySciences) +
9.49E-12(ResearchArea_Engineering) + 1.23E-11(ResearchArea_Medicine) +
8.58E-12(ResearchArea_PhysicsandAstronomy) +
6.47E-13(ResearchArea_SocialSciences)

Similar to the previous multiple linear regression equation, this equation performed well from an adjusted R-squared perspective with 100 percent (n=1.00) of the variance in Total Enrollment explained by the 86 predictor variables. However, the p-values for this equation do not list a single included variable as being statistically significant to the equation and a perfect adjusted R-square value generally indicates serious equation issues, both likely due to multicollinearity issues within the dataset.

Lastly, while the multicollinearity within the dataset was identified early into data analysis, a multiple linear regression equation was still created using forward selection, a machine learning technique where variables are added to a multiple linear regression equation one at a time to ensure statistical significance of the included variables. This approach was used to create a model again predicting research output:

ResearchOutput = 0.0002(TotalExpenses) + 0.0367(ResearchExpenses) +
-0.0051(PreservationExpenses) + -944.463(NonLibraryServices) +
4.9058(ResearchDoctoralDegrees) + -0.0002(TotalMaterialsExpenses) +
0.1461(GraduateEnrollment) + 0.0002(PhysicalBooks) +
-975.9134(ResearchArea_SocialSciences) + -0.1331(BachelorsDegrees) +
0.0348(ILLReceived)

Similar to prior models, the adjusted R-squared for this model indicated that 96.4 percent (n=0.964) of the model's variance can be explained by the included independent variables of total expenses, research expenses, preservation expenses, non-library services,

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

research doctoral degrees, total materials expenses, graduate enrollment, physical books, social sciences as a research area, bachelor's degrees awarded, and interlibrary loan (ILL) requests received. Unlike other models, though, p-values for this equation are all less than or equal to 0.05, indicating statistical significance.

Discussion of Findings

This study originated as an effort to determine which specific library services best served the overall goal of the institution and to represent value in a quantifiable way. However, inequities in openly available data, variation caused by data collection methods or dates, and challenges caused by piecing together data from various sources morphed this study into a greater exploration of what is available in terms of library data, as well as what can be expressed using this imperfect data. While this study points to statistically significant relationships between library variables and select metrics indicative of institutional success, such as research output, limitations in available data and data reliability also limit the efficacy and trustworthiness of results.

In particular, while the resulting linear regression equations from this study represent a new approach to communicating library value and assessing library impact, it is important to note that they are rudimentary and exist as explorations into a possible return on investment (ROI)-type approach to library metrics, rather than a perfected technique. Inputting library data or even the R1 data into the linear regressions will likely return skewed or inaccurate calculations at this phase of exploration. Additionally, given the limited scope of this study and focus on R1 academic libraries, only the three resulting multilinear regression equations evaluated all dataset variables against the investigated dependent variables, meaning that additional relationships between variables cannot be analyzed outside the scope of corresponding equations to predict other library metrics (such as using the Association of Research Libraries (ARL) coefficient from an equation predicting enrollment to determine whether ARL membership has an impact on library budgets). At this time, only research output, total library staff, and total enrollment were explored via regression equations but each variable could be further explored as an independent candidate to investigate other interdependent relationships in the dataset. More work must be done in the future to achieve broadly applicable and reliable "plug and play" equations that further investigate the many relationships between library services and institutional metrics.

Notwithstanding the limitations of openly available data, this study and the materials resulting from it do indicate strong correlations between library services and institutional success. This early attempt at quantifying variables that are often distinctly qualitative in nature is promising and points to infinite applications or adaptations. As seen with prior attempts at quantifying the value of library services, a lack of centrally gathered and defined library data often limits studies to local environments while broad attempts at creating industry standards will ultimately require the creation and maintenance of central data or measurement mechanisms. Large scale solutions must be carefully constructed to adequately capture the quantitative and qualitative nature of library services, as well as their depth. As well, eventual solutions must be openly available to serve the greatest benefit and ensure industry transparency.

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Ideally, IPEDS's existing survey of academic libraries will continue and even expand under library advisement or partnership to include detailed and updated questions that describe library services. IPEDS data represent a critical and primary source of academic library comparison, much like the various other metrics IPEDS collects for peer-to-peer and year-to-year benchmarking, making its planned elimination alarming. While library associations or organizations may take up the library data collection cause, participation will no longer be mandatory, data will exist outside the context of greater academic data aggregation, and perhaps most importantly, may become locked behind a paywall thus limiting use and transparency of results. Moreover, exclusion of academic library data from the broader academic data collection process points to a devaluation of the library as a key campus service, which reflects the repetitive and increasing concerns of libraries in the academy.

As seen with prior attempts at quantifying the value of library services, a lack of centrally gathered and defined library data often limits studies to local environments while broad attempts at creating industry standards will ultimately require the creation and maintenance of central data or measurement mechanisms.

In the context of the study as it stands, there appears to be a high correlation between library services, collection sizes, staffing, and metrics like research output. These results echo the belief that libraries provide added value and service to their communities, including toward metrics of value to the broader institution. High-level exploration points to correlation between some variables, which is likely skewing overall results while also introducing a paradox: generally, institutions with more funding, staffing, and resources have higher research output, but it is impossible to say whether these institutions experience these results as a direct result of library support, if high library support is a result of the abundance and high achievement at successful institutions, or if there is not an intentional relationship between library support and institutional success. Other unexplored relationships, such as research areas, may have further impact on this quandary as some fields, such as medicine, require significant investment in both resources and staff to support medical programs. Expanding this study into R2 institutions could help clarify these relationships.

Future Study Considerations

As an exploratory study, this research could be expanded in numerous directions. For one, a return to the original question of library services contributing to institutional success using self-reported data on institutional services and their depth could resolve data collection challenges while enhancing the validity of library services data. Directly collaborating with libraries to investigate services offered and their depth would provide new contrast to this existing study's data by segmenting R1 services more meaningfully while providing more concrete investigations into library service depth for aspiring R1

This mss. is peer reviewed, copy edited and accepted for publication, Portal 26.1.

institutions. An expansion into the behaviors and data for R2 libraries would also provide clarity surrounding library support, services, and institutional success, while bolstering the reliability of the study's methodology overall. Were the study expanded under its current design to include R2 libraries, study results would reflect a significantly wider range of library experiences and services and provide additional training data for any created linear regression equations to help improve accuracy. Using the existing dataset, as is, also provides opportunities for the identification and removal of heavily correlated variables and further machine learning-based approaches to better understand the statistical relationships between libraries and their broader institutions. Revisiting the existing dataset could result in more reliable equations post-data cleanup, investigate new interdependent relationships not explored in this study's scope, or explore entirely new ways of parsing the data for comparison, such as by public versus private control or location to determine how factors outside of a library's control impact their overall support to institutional success metrics. Lastly, revisitation to IPEDS's wealth of academic data and academic library data could provide more insight into historic library patterns and correlation between the library and the overall success of its institution. Were this study to be scaled back to just include IPEDS data, concerns about gauging depth of library services, multicollinearity within the data, and data reproducibility could be eliminated while presenting new opportunities to expand investigation into R2 libraries and yet unexplored relationships within the available dataset.

Conclusion

The relationship between library services and institutional metrics of success closely parallels the somewhat confusing and disjointed relationships between libraries and their academies, with no one-size-fits-all solution. While libraries are capable of providing valuable services to their campus communities, decades of misaligned services between libraries and their universities have devalued their role in some administrators' eyes at a time when libraries are also experiencing decreased status and funding on campus. Despite challenges in library data compilation and reliability, early exploration into the world of library service value modeling to explain contribution to institutional goals or even to predict library resources to achieve desired impact is both warranted and promising. Improvements to data availability will improve data modeling and provide more applicable solutions for libraries of all sizes, which may ultimately provide avenues to better library-university goal integration or quantitative-based communication on library needs.

While libraries are capable of providing valuable services to their campus communities, decades of misaligned services between libraries and their universities have devalued their role in some administrators' eyes at a time when libraries are also experiencing decreased status and funding on campus.

This mss. is peer reviewed, copy edited, and accepted for publication. portal 26.1.

Disclosures

The researcher began work on this project while employed at the University of Tulsa and completed it at her current institution. Neither institution required institutional review board (IRB) approval or review to undertake this study, and the study was not funded via grant or research funding. Data utilized in the study was collected from publicly available sources apart from Elsevier's Scopus data. The researcher received approval from Elsevier to include extracted Scopus data in this project and to share extracted data for study reproducibility. The researcher utilized ChatGPT to help create, edit, and troubleshoot Python code within Jupyter Notebook but ChatGPT was not utilized to analyze data, draw conclusions, or craft this article.

Acknowledgements

The researcher thanks her family and colleagues for their support and encouragement during this multi-year project.

Elizabeth Szkirpan is the collections and discovery specialist at Harvard Business School's Baker Library, email: eszkirp@gmail.com, ORCID: 0000-0003-3811-8840.

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Appendix A

Variable	Definition	Purpose	Source
Institution	Research 1 (R1) institution name.	Ensure unique variables in data collection process.	Carnegie Classification of Institutions of Higher Education
Control	Delineates public or private control of the institution information.	Adds categorical variable to the dataset that can provide additional insight into variables that ensure R1 success.	Carnegie Classification of Institutions of Higher Education
State	Delineates state where institution is located.	Delineates state where institution is located.	Carnegie Classification of Institutions of Higher Education
ResearchOutput	Research output for calendar year 2022 (January-December 2022) research documents published by the institution.	Provides metric to help gauge library impact towards overall institutional outputs.	Elsevier's Scopus
ResearchArea	Largest research areas by number of document subjects indexed in Scopus.	Helps identify categorical correlations that may exist between institutional priorities, funding, et cetera and major research focuses.	Elsevier's Scopus
LibraryStaff	Categorical variable: Does the institution have library staff, yes (1) or no (0)?	Identifies whether the institution's academic library has staff or is unstaffed.	Integrated Postsecondary Education Data System (IPEDS), FY2022
TotalLibraryStaff	Summed from other staffing numbers, this variable tallies the total number of full-time equivalency (FTE) library staff supported by the library's budget.	Identifies the number of full-time equivalency library staffers.	Integrated Postsecondary Education Data System (IPEDS), FY2022
LibrariansFTE	Professional staff requiring education in library studies or equivalent studies.	Identifies the number of professional librarian staffers.	Integrated Postsecondary Education Data System (IPEDS), FY2022

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Appendix A, cont.

Variable	Definition	Purpose	Source
ProfessionalFTE	Professional staff that may have equivalent education or training in fields related to librarianship but are not librarians as defined by NISO.	Identifies the number of professional non-librarian staffers.	Integrated Postsecondary Education Data System (IPEDS) , FY2022
OtherFTE	Library staffers without formal qualifications.	Identifies the number of paraprofessional library staffers.	Integrated Postsecondary Education Data System (IPEDS), FY2022
StudentFTE	Student assistants, including undergraduate and graduate students, employed on an hourly basis using library funds or accounts.	Identifies the number of student library staffers.	Integrated Postsecondary Education Data System (IPEDS), FY2022
NumberofLibraries	Auxiliary library branches outside of the main campus branch that are open most or all or the fiscal year.	Identifies the total number of libraries in multi-library systems.	Integrated Postsecondary Education Data System (IPEDS), FY2022
SystemResults	Conducted in each library's discovery system, this variable used "~" where possible and "and" elsewhere to broadly estimate the number of library search results researchers might expect to find. No filters are utilized to refine these searches and indexing can vary widely among institutions.	Broadly returns the number of results immediately available to researchers to mimic the top-level results that many institutional researchers will encounter when beginning an academic search. Some institutions drastically expand or contract index results based on their interpretation of researcher needs.	Compiled by the researcher from library discovery systems

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Appendix A, cont.

Variable	Definition	Purpose	Source
ArticleResults	Conducted in the library's discovery system, this variable used "*" where possible and "and" elsewhere with an article or peer-reviewed article filter as relevant to broadly estimate the number of library search article results researchers might expect to find. No filters are utilized to refine these searches and indexing can vary widely among institutions.	Broadly returns the number of article results immediately available to researchers to mimic the top-level results that many institutional researchers will encounter when beginning an academic search. Some institutions drastically expand or contract index results based on their interpretation of researcher needs.	Compiled by the researcher from library discovery systems
WorldCatResults	Conducted by searching for each institution's name in WorldCat, this variable is the number of items indexed in WorldCat by each institution. Main campus libraries were used for the number of reported results.	Given the extreme variance in library system indexing, this variable sought to standardize discoverable items attributed to each institution by WorldCat. Note that not all libraries catalog records into WorldCat and some libraries may forget to delete old data out of WorldCat.	Compiled by the researcher from WorldCat.org
ElectronicCollection	Categorical variable: Is the library's collection entirely electronic, yes (1) or no (0)?	Identifies whether the institution's library is fully electronic/digital or not.	Integrated Postsecondary Education Data System (IPEDS), FY2022
PhysicalBooks	Number of physical books reported by the library to IPEDS.	Identifies and standardizes the number of physical books reported by each library.	Integrated Postsecondary Education Data System (IPEDS), FY2022
PhysicalMedia	Number of physical media items reported by the library to IPEDS.	Identifies and standardizes the number of physical media reported by each library.	Integrated Postsecondary Education Data System (IPEDS), FY2022
PhysicalSerials	Number of physical serials reported by the library to IPEDS.	Identifies and standardizes the number of physical serials reported by each library.	Integrated Postsecondary Education Data System (IPEDS), FY2022

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Appendix A, cont.

Variable	Definition	Purpose	Source
TotalPhysical	Total number of physical materials reported by the library to IPEDS. This variable is summed from other physical material numbers.	Identifies and standardizes the number of physical materials reported by each library.	Integrated Postsecondary Education Data System (IPEDS), FY2022
PhysicalCirculation	Total number of materials to physically circulate during FY2022 reported by the library to IPEDS.	Identifies the number of physical circulations for each institution.	Integrated Postsecondary Education Data System (IPEDS), FY2022
ElectronicBooks	Total number of electronic or digital books reported by the library to IPEDS.	Identifies and standardizes the number of electronic books reported by each library.	Integrated Postsecondary Education Data System (IPEDS), FY2022
Databases	Total number of electronic or digital databases reported by the library to IPEDS.	Identifies and standardizes the number of databases reported by each library.	Integrated Postsecondary Education Data System (IPEDS), FY2022
ElectronicMedia	Total number of electronic or digital media reported by the library to IPEDS.	Identifies and standardizes the number of electronic media reported by each library.	Integrated Postsecondary Education Data System (IPEDS), FY2022
ElectronicSerials	Total number of electronic or digital serials reported by the library to IPEDS.	Identifies and standardizes the number of electronic serials reported by each library.	Integrated Postsecondary Education Data System (IPEDS), FY2022
TotalElectronic	Total number of electronic materials reported by the library to IPEDS. This variable is summed from other electronic material numbers.	Identifies and standardizes the number of electronic materials reported by each library.	Integrated Postsecondary Education Data System (IPEDS), FY2022
ElectronicCirculation	Total number of materials to electronically or digitally circulate during FY2022 reported by the library to IPEDS.	Identifies the number of electronic or digital circulations for each institution.	Integrated Postsecondary Education Data System (IPEDS), FY2022
TotalCollections	Total physical and electronic collections reported by the library to IPEDS. Summed from TotalPhysical and TotalElectronic variables.	Identifies and standardizes the number of collection materials reported by each library.	Integrated Postsecondary Education Data System (IPEDS), FY2022

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Appendix A, cont.

Variable	Definition	Purpose	Source
TotalCirculations	Total physical and electronic circulations reported by the library to IPEDS. Summed from PhysicalCirculations and ElectronicCirculations variables.	Identifies and standardizes the number of circulations reported by each library.	Integrated Postsecondary Education Data System (IPEDS), FY2022
InstructionServices	Does the library's website explicitly outline availability of instruction services in any format, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
InPersonConsultations	Does the library's website explicitly outline availability of in-person research consultation services, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
OnlineConsultations	Does the library's website explicitly outline availability of online research consultation services, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
ScholarlyComms	Does the library's website explicitly outline availability of scholarly communications services in any format, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
DigitalServices	Does the library's website explicitly outline availability of digital humanities and/or digital services in any format, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
DataServices	Does the library's website explicitly outline availability of data services in any format, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Appendix A, cont.

Variable	Definition	Purpose	Source
CopyrightServices	Does the library's website explicitly outline availability of copyright services in any format, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
OpenAccess	Does the library's website explicitly outline availability of open access support (e.g. publishing or indexing OA materials) in any format, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
Workshops	Does the library's website explicitly outline availability of library-hosted workshops in any format, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
SubjectLiaisions	Does the library's website explicitly outline availability of library subject area experts, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
Microcredentialing	Does the library's website explicitly outline availability of library-backed microcredential programs, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
ExtendedHours	Does the library's website explicitly outline building hours that are earlier than 8 AM and/or later than 10 PM during Monday-Friday hours during the typical Fall/Spring semester, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
WeekendHours	Does the library's website explicitly outline building hours include both Saturdays and Sundays during the typical Fall/Spring semester, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Appendix A, cont.

Variable	Definition	Purpose	Source
ResearchGuides	Does the library's website explicitly outline availability of research guides, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
VideoTutorials	Does the library's website explicitly outline availability of library video tutorials, yes (1) or no (0)?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
Templates	Does the library's website explicitly outline availability of templates, yes (1) or no (0)? May include research data management, citation, research templates, or other.	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
SpecialtySoftware	Does the library's website explicitly outline availability of specialty software, either online or in-person, yes (1) or no (0)? May include online software such as Zotero, makerspace-specific software, or library facilitation of other software.	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
InstitutionalRepository	Does the library's website explicitly outline availability of an institutional repository, yes (1) or no (0)? The library may or may not manage the repository itself but they cite it for researchers on their site.	Identifies availability of library services that may support research and scholarship at their corresponding university.	Compiled by the researcher from each library's website
ILL	Does the library report supplying interlibrary loan (ILL) services, yes (1) or no (0), to IPEDS in their FY2022 data?	Identifies availability of library services that may support research and scholarship at their corresponding university.	Integrated Postsecondary Education Data System (IPEDS), FY2022

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Appendix A, cont.

Variable	Definition	Purpose	Source
ILLProvided	Number of interlibrary loan (ILL) documents and materials provided to other libraries.	Materials provided outside of the immediate university community. Provides contrast as to whether collection size or prestige impacts loaning of materials.	Integrated Postsecondary Education Data System (IPEDS), FY2022
ILLReceived	Number of interlibrary loan (ILL) documents and materials received from other libraries.	Materials required to support on-going research and study efforts that expand past the university's current collections.	Integrated Postsecondary Education Data System (IPEDS), FY2022
DocumentDelivery	Does the library's website explicitly outline Document Delivery as a service provided by or facilitated by the library, yes (1) or no (0)?	Aims to understand whether the library is able to provide delivery of library materials to facilitate on-going research and study efforts or is otherwise involved in that process.	Compiled by the researcher from each library's website
TechnologyAccess	Does the library's website explicitly outline technology access (e.g. computers, copiers, scanners, makerspaces) as a service provided by or facilitated by the library, yes (1) or no (0)?	Aims to understand services provided by the library outside of the traditional book and media accessibility role that may support research and scholarship or encourage use of the library as space.	Compiled by the researcher from each library's website
FoodServices	Does the library's website explicitly outline food services (e.g. On-site café, restaurant, or vending machine) available within the library, yes (1) or no (0)?	Aims to understand services provided by the library outside of the traditional book and media accessibility role that may support research and scholarship or encourage use of the library as space.	Compiled by the researcher from each library's website
PrivateStudy	Does the library's website explicitly outline private study spaces that are reservable or first-come-first-serve available within the library, yes (1) or no (0)?	Aims to understand services provided by the library outside of the traditional book and media accessibility role that may support research and scholarship or encourage use of the library as space.	Compiled by the researcher from each library's website

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Appendix A, cont.

Variable	Definition	Purpose	Source
GroupStudy	Does the library's website explicitly outline group study spaces that are reservable or first-come-first-serve available within the library, yes (1) or no (0)?	Aims to understand services provided by the library outside of the traditional book and media accessibility role that may support research and scholarship or encourage use of the library as space.	Compiled by the researcher from each library's website
InstructionSpace	Does the library's website explicitly outline instruction spaces for library instruction or other classroom instruction as available within the library, yes (1) or no (0)?	Aims to understand services provided by the library outside of the traditional book and media accessibility role that may support research and scholarship or encourage use of the library as space.	Compiled by the researcher from each library's website
NonLibServices	Does the library's website explicitly outline non-library services (e.g. Writing center, student services, Information Technology) as being physically housed within the library, yes (1) or no (0)?	Aims to understand services provided by the library outside of the traditional book and media accessibility role that may support research and scholarship or encourage use of the library as space.	Compiled by the researcher from each library's website
ARLMembership	Is the university affiliated with the library listed as a current Association of Research Libraries (ARL) member, yes (1) or no (0)?	Aims to understand how formal organizational membership in the Association of Research Libraries (ARL), which provides community and training for ARL members, impacts library services and university metrics.	Compiled by the researcher from ARL's membership section on their website
ExpendituresOver100	Are the library's FY2022 expenditures greater than or equal to $100,000, yes (1) or no (0)?	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022
TotalSalaries	Salaries and wages for full- and part-time staff paid for from the library's FY2022 budget.	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022
FringeBenefits	Are fringe benefits for staff paid for from the library's FY2022 budget, yes (1) or no (0)?	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Appendix A, cont.

Variable	Definition	Purpose	Source
BenefitsAmount	Fringe benefits paid for from the library's FY2022 budget.	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022
OTPurchases	One-time library material expenses recorded for each library's FY2022 budget.	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022
Subscriptions	On-going subscription expenses recorded for reach library's FY2022 budget.	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022
OtherMaterialsExpenses	Other material and services costs/fees not recorded elsewhere in the IPEDS collection of academic library financial data for each library's FY2022 budget.	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022
TotalMaterialsExpenses	A summed column of one-time costs, subscriptions, and other material/service fees from each library's FY2022 budget.	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022
PreservationExpenses	Preservation expenses from each library's FY2022 budget.	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022
OtherOperationsExpenses	All other operations and maintenance expenses not recorded elsewhere in the IPEDS collection of academic library financial data for each library's FY2022 budget.	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022
TotalOpMaintExpenses	A summed column of preservation expenses and other operations/maintenance expenditures from each library's FY2022 budget.	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022

This mss. is peer reviewed, copy edited, and accepted for publication portal 26.1.

Appendix A, cont.

Variable	Definition	Purpose	Source
TotalExpenses	A summed column of wages, fringe benefits, total materials and services, and operations and maintenance fees reported by each academic library for FY2022.	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022
TotalExpensesLessBen	A summed column of wages, total materials and services, and operations and maintenance fees but not fringe benefit costs reported by each academic library for FY2022.	Financial metric that is not available elsewhere for academic libraries.	Integrated Postsecondary Education Data System (IPEDS), FY2022
BachelorsDegrees	The number of bachelor's degrees awarded by each university in their FY2022 data.	Identifies and standardizes the number of degrees awarded reported by each university.	Integrated Postsecondary Education Data System (IPEDS), FY2022
MastersDegrees	The number of master's degrees awarded by each university in their FY2022 data.	Identifies and standardizes the number of degrees awarded reported by each university.	Integrated Postsecondary Education Data System (IPEDS), FY2022
ResearchDocDegrees	The number of research doctoral degrees awarded by each university in their FY2022 data.	Identifies and standardizes the number of degrees awarded reported by each university.	Integrated Postsecondary Education Data System (IPEDS), FY2022
ProfessionalDocDegrees	The number of professional doctor's degrees awarded by each university in their FY2022 data.	Identifies and standardizes the number of degrees awarded reported by each university.	Integrated Postsecondary Education Data System (IPEDS), FY2022
OtherDocDegrees	The number of other doctoral degrees awarded by each university in their FY2022 data.	Identifies and standardizes the number of degrees awarded reported by each university.	Integrated Postsecondary Education Data System (IPEDS), FY2022
TotalEnrollment	The total number of enrolled students reported by each university in their FY2022 data.	Identifies and standardizes the number of enrolled students reported by each university.	Integrated Postsecondary Education Data System (IPEDS), FY2022
FTEnrollment	The full-time (FT) enrollment reported by each university in their FY2022 data.	Identifies and standardizes the number of enrolled students reported by each university.	Integrated Postsecondary Education Data System (IPEDS), FY2022

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.

Appendix A, cont.

Variable	Definition	Purpose	Source
PTEnrollment	The part-time (PT) enrollment reported by each university in their FY2022 data.	Identifies and standardizes the number of enrolled students reported by each university.	Integrated Postsecondary Education Data System (IPEDS), FY2022
UndergradEnrollment	The number of enrolled undergraduates reported by each university in their FY2022 data.	Identifies and standardizes the number of enrolled students reported by each university.	Integrated Postsecondary Education Data System (IPEDS), FY2022
GraduateEnrollment	The number of enrolled undergraduates reported by each university in their FY2022 data.	Identifies and standardizes the number of enrolled students reported by each university.	Integrated Postsecondary Education Data System (IPEDS), FY2022
ResearchExpenses	Research expenses per Full-Time Equivalency (FTE) reported by each university in their FY2022 data. Combines GASB and FASB column data from IPEDS so each university has a single line of research expenses in the dataset.	Identifies and standardizes research expenses reported by each university.	Integrated Postsecondary Education Data System (IPEDS), FY2022

This mss. is peer reviewed, copy edited, and accepted for publication portal 26.1.

Notes

1. Inka Tappenbeck and Wilfried Sühl-Strohmenger, "The Evolving Role of the Academic Librarian – From Traditional to Digital," *De Gruyter Conversations,* August 15, 2023, https://blog.degruyter.com/the-evolving-role-of-the-academic-librarian-from-traditional-to-digital/; Graham Matthews and Graham Walton, *University Libraries and Space in the Digital World* (Routledge, 2013).
2. Beverly P. Lynch, Catherine Murray-Rust, Susan E. Parker, Deborah Turner, Diane Parr Walker, Frances C. Wilkinson, and Julia Zimmerman, "Attitudes of Presidents and Provosts on the University Library," *College & Research Libraries* 68, no. 3 (May 1, 2007): 213–28, https://doi.org/10.5860/crl.68.3.213; Joshua Kim, "Have Academic Library Staffing Numbers Really Declined That Much over the Past Decade?" *Inside Higher Ed*, February 12, 2023, https://www.insidehighered.com/blogs/learning-innovation/have-academic-library-staffing-numbers-really-declined-much-over-past; Adam Murry and Ashley Ireland, "Provosts' Perceptions of Academic Library Value and Preferences for Communication: A National Study," *College & Research Libraries* 79, no. 3 (2021): 336–65, https://doi.org/10.5860/crl.79.3.336.
3. Murray and Ireland, "Provosts' Perceptions of Academic Library Value and Preferences."
4. "Standards for Libraries in Higher Education," Guidelines, Standards, and Frameworks, Association of College and Research Libraries, February 2018, https://www.ala.org/acrl/standards/standardslibraries.
5. Frank W. Lutz and Robert W. Field, "Business Valuing in Academia." *Higher Education* 36, no. 4 (December 1998): 383–420, https://doi.org/10.1023/a:1003473919179; Gary A. Miller, "The Two-Way Street of Higher Education Commodification," in *Schools or Markets? Commercialism, Privatization, and School-Business Partnerships* (Routledge, 2004), 149–69.
6. Cole Clark, Megan Cluver, Tiffany Fishman, and Danylle Kunkel, "2024 Higher Education Trends: A Look at the Challenges and Opportunities Shaping America's Higher Education Sector," Deloitte Center for Government Insights, 2024, https://www.deloitte.com/us/en/insights/industry/articles-on-higher-education/latest-trends-in-higher-education.html.
7. Megan Brenan, "Americans' Confidence in Higher Education Down Sharply," Gallup, Inc., July 11, 2023, https://news.gallup.com/poll/508352/americans-confidence-higher-education-down-sharply.aspx; Emily Wielk and Tristan Stein, "Measuring the Return on Investment of Higher Education: Breaking Down the Complexity," Bipartisan Policy Center, April 4, 2024, https://bipartisanpolicy.org/explainer/measuring-the-return-on-investment-of-higher-education-breaking-down-the-complexity/; Matt Zalaznick, "Higher Ed Funding: A Look at 2 Surprising Trends." University Business, May 17, 2024, https://universitybusiness.com/higher-ed-funding-a-look-at-2-surprising-trends/.
8. Mushtaq Gunja and Sara Gast, "How Carnegie Classification Updates Could Affect State Higher Education Policy," Carnegie Classification of Institutions of Higher Education, American Council on Education, February 1, 2024, https://carnegieclassifications.acenet.edu/how-carnegie-classification-updates-could-affect-state-higher-education-policy/#:~:text=Beyond%20funding%2C%20accrediting%20agencies%20often,sits%20within%20a%20national%20context.
9. "Basic Classification," Carnegie Classification of Institutions of Higher Education, archived August 23, 2024 at https://web.archive.org/web/20240823044812/https://carnegieclassifications.acenet.edu/carnegie-classification/classification-methodology/basic-classification/.
10. Ibid.
11. Khaleel Seecharan, "Exploring the Impetus of R2 Universities that Attain R1 Status" *(PhD diss.,* University of Pennsylvania, 2020), ProQuest (2447055542).
12. "About IPEDS," Integrated Postsecondary Education Data System (IPEDS), accessed August, 2024, https://nces.ed.gov/ipeds/about-ipeds#:~:text=The%20completion%20of%20all%20IPEDS.

This mss. is peer reviewed copy edited and accepted for publication, portal 26.1.

13. Megan Oakleaf, *Value of Academic Libraries: A Comprehensive Research Review and Report* (American Library Association, 2010), https://www.ala.org/sites/default/files/2024-06/val_report.pdf.
14. Murray and Ireland, "Provosts' Perceptions of Academic Library Value and Preferences."
15. Judy Luther, "University Investment in the Library: What's the Return? A Case Study at the University of Illinois at Urbana-Champaign," (white paper, Elsevier, 2008), http://www.informedstrategies.com/wp-content/uploads/2015/10/lcwp0101_1_.pdf.
16. Deborah Jeanne Grimes, "Centrality and the academic library" (*PhD diss,* University of Alabama, 1993), ProQuest (304044825).
17. Leigh S. Estabrook, "What Chief Academic Officers Want from Their Libraries: Findings from Interviews with Provosts and Chief Academic Officers," (white paper, 2007), https://www.ala.org/sites/default/files/acrl/content/publications/whitepapers/Finalreport-ACRLCAOs.pdf.
18. Jennifer K. Frederick and Christine Wolff-Eisenberg, "Leading the Library by Looking Beyond the Library," *Ithaka S+R Blog* (blog), *Ithaka S+R*, April 12, 2020, https://sr.ithaka.org/blog/leading-the-library-by-looking-beyond-the-library/.
19. Murray and Ireland, "Provosts' Perceptions of Academic Library Value and Preferences."
20. William Miller, "Changes Coming to the Collection of Federal Library Data," *Library Issues* 35, no. 1 (September 2014); "Academic Libraries Resources Center," Integrated Postsecondary Education Data System (IPEDS), accessed August, 2024, https://nces.ed.gov/ipeds/report-your-data/resource-center-academic-libraries.
21. "About IPEDS."
22. Seth Allen, "Towards a Conceptual Map of Academic Libraries' Role in Student Retention," *The Christian Librarian* 57, no. 1 (January 1, 2014): 7–19, https://doi.org/10.55221/2572-7478.1434; Gregory A. Crawford, "The Academic Library and Student Retention and Graduation: An Exploratory Study," *portal: Libraries and the Academy* 15, no. 1 (January 2015): 41–57, https://doi.org/10.1353/pla.2015.0003; Starr Hoffman and Samantha Godbey, "US Academic Libraries' Staffing and Expenditure Trends (1996–2016)," *Library Management* 41, no. 4/5 (April 14, 2020): 247–68, https://doi.org/10.1108/lm-12-2019-0093; Clare Keating, "Quantification of Solo Librarianship within American Academic Libraries between Fiscal Years 2016 and 2019 Using IPEDS Data," *College & Undergraduate Libraries*, April 16, 2024, 1–19, https://doi.org/10.1080/10691316.2024.2341332.
23. John Cocklin, "Adding Context to Academic Library Assessment: Using the Integrated Postsecondary Education Data System (IPEDS) for Institutional and Comparative Statistics," in *Proceedings of the 2008 Library Assessment Conference*, (Association of Research Libraries, 2009), 253–56.
24. "ACRL Academic Library Trends and Statistics: 2024 Survey Information," Association of College & Research Libraries (ACRL), December 2024, https://acrl.libguides.com/stats/surveyhelp; Association of College & Research Libraries, *The State of U.S. Academic Libraries: Findings from the ACRL 2023 Annual Survey* (Association of Collection & Research Libraries, 2024). https://www.ala.org/sites/default/files/2024-10/2023%20State%20of%20Academic%20Libraries%20Report.pdf; Devin Savage and Steve Borrelli, "2022 ACRL Academic Library Trends and Statistics Survey: Highlights and Key Findings," *College & Research Libraries News* 85, no. 5 (2024); "Academic Library Trends and Statistics Survey," Association of College & Research Libraries (ACRL), 2024, https://www.ala.org/acrl/proftools/benchmark/survey.
25. ACRL, *The State of U.S. Academic Libraries.*
26. "Benchmark: Library Metrics and Trends," Association of College & Research Libraries (ACRL), 2024, https://www.ala.org/acrl/benchmark-library-metrics-and-trends.
27. Kathryn Palmer, "Academic Librarians Oppose Plan to Eliminate Key Federal Data," *Inside Higher Ed*, May 10, 2024, https://www.insidehighered.com/news/tech-innovation/libraries/2024/05/10/academic-librarians-oppose-plan-eliminate-key-federal#:~:text=The%20National%20Center%20for%20Education,stem%20from%20concerns%20about%20cost.

This mss. is under review; could change and accepted for publication, portal 26.1.

28. Ibid.
29. Ryan Cordell, *Machine Learning + Libraries: A Report on the State of the Field* (LC Labs, Library of Congress, July 14, 2020) https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf; Emmanouel Garoufallou and Panorea Gaitanou, "Big Data: Opportunities and Challenges in Libraries, a Systematic Literature Review," *College & Research Libraries* 82, no. 3 (March 2021): 410, https://doi.org/10.5860/crl.82.3.410; "Libraries and Open Data in 2021: Access, Skills, Engagement," International Federation of Library Associations and Institutions (IFLA), March 6, 2021, https://www.ifla.org/news/libraries-and-open-data-in-2021-access-skills-engagement/#:~:text=As%20the%20Open%20Data%20movement,of%20access%20to%20such%20information.
30. Elizabeth Szkirpan, "Data Services and the Librarian: Supporting Early Career Researchers and Online Learners via the Library," *Advances in Online Education: A Peer-Reviewed Journal* 1, no. 4 (June 1, 2023): 326–45, https://doi.org/10.69554/lrwr4673.
31. Mohamed Parvees, "How Many Times Scopus Changes It's Indexing (Journals) per Year?" ResearchGate, November 12, 2021, https://www.researchgate.net/post/How_many_times_Scopus_changes_its_indexing_Journals_per_year; D. Feldner, "Enhancements to Citation Overview on Scopus," *Scopus blog* (blog), April 29, 2024, https://blog.scopus.com/posts/enhancements-to-citation-overview-on-scopus.
32. Elizabeth Szkirpan, "Library Services Contributing to Institutional Success Dataverse," Harvard Dataverse, August 2024, https://dataverse.harvard.edu/dataverse/LibraryServicesInstiutionalSuccess.

This mss. is peer reviewed, copy edited, and accepted for publication, portal 26.1.