ARTICLE

Recommended by Librarians

A Computational Citation Analysis Methodology for Identifying and Examining Books Promoted in LibGuides

Carmen Orth-Alfie and Erin Wolfe

ABSTRACT

To study library guides, as published on Springshare's LibGuides platform, new approaches are needed to expand the scope of the research, ensure comprehensiveness of data collection, and reduce bias for content analysis. Computational methods can be utilized to conduct a nuanced and thorough evaluation that critically assesses the resources promoted in library guides. Web-based library guides are curated by librarians to provide easy access to high-quality information and resources in a variety of formats to support the research needs of their users. Recent scholarship considers library guides as valuable resources and as de facto publications, highlighting the need for critical study. In this article, the authors present a novel model for comprehensively gathering data about a specific genre of books from individual LibGuide pages and applying computational methods to explore the resultant data. Beginning with a pre-selected list of 159 books, we programmatically queried the titles using the LibGuides Community search engine. After cleaning and filtering the resultant data, we compiled a list of 20,484 book references (of which 6,212 are unique) on 1,529 LibGuide pages. By testing against inclusion and exclusion criteria to ensure relevancy, we identified a total of 281 titles relevant to our topic. To gain insights for future study, citation analysis metrics are presented to reveal patterns of frequency, co-occurrence, and bibliographic coupling of books promoted in LibGuides. This proof-of-concept could be adopted for a variety of applications, including assessment of collections, public services, critical librarianship, and other complex questions to enable a richer and more thorough understanding of the information landscape of LibGuides.

INTRODUCTION

To advance the depth and breadth of research regarding LibGuides, the dominant platform for creating and hosting library guides, researchers Cody Hennesy and Annis Lee Adams, in their article "Measuring Actual Practices: A Computational Analysis of LibGuides in Academic Libraries," call upon others to use computational methods to gather and synthesize the collective wisdom of professional colleagues from around the world.[1] Following the definition proposed by Jennifer Emanuel and other authors, in this article we use "library guides" or "guides" as general terms to encompass course guides, lib guides, online library guides, research guides, general subject guides, subject guides, general topic guides, topical guides, pathfinders, and other similar terms. When referring to guides created using the Springshare platform, we will use "LibGuides" for both the platform and for multiple guides and "LibGuide" when a single guide is discussed.[2] While Hennesy and Adams used computational analysis methods to study how librarians structure content in

About the Authors

Carmen Orth-Alfie (corresponding author: corthalf@ku.edu) is Graduate Student Engagement and Business Librarian, University of Kansas. **Erin Wolfe** (edw@ku.edu) is Digital Initiatives Librarian, University of Kansas.

This peer-reviewed work was submitted 10 July 2023, accepted for publication 2 October 2023, and published 18 March 2024.

https://doi.org/10.5860/ital.v43i1.16687

LibGuides, there is also interest in the critical analysis of what specifically is presented.[3] To this end, a growing body of library research aggregates observations from multiple institutions to analyze the resources (book, databases, websites, etc.) and other content curated by librarians and promoted in online library guides, especially those created using Springshare's LibGuides platform. Thus far, few researchers report computational methods. Instead, researchers frequently set limits by narrowly focusing research questions and by applying sampling methods. Computational methods, such as web scraping and data analysis, could be used to advance research that critically assesses resources promoted in library guides.

To gain insights about a specific genre of books promoted by librarians in library guides, we use a novel computational methodology to comprehensively gather data about one resource type, books, from the LibGuides corpus and then applied computational citation analysis methods. The data collection starts by searching each title from a curated list of books within a specific genre and utilizes citation analysis methods to explore if and how these and related titles are promoted in LibGuides. Key goals for this computational methodology include the following:

- Comprehensively search the entire LibGuides corpus to collect data for citation analysis of books promoted in LibGuides.
- Identify relevant book titles meeting specific descriptive criteria.
- Collect data (institutional demographics, type of guide, etc.) about all LibGuides promoting a given book title.
- Collect data to identify co-occurring book titles.

LITERATURE REVIEW

LibGuides as Resource to Inform Practice

Exploring content promoted in the LibGuides corpus is one of several strategies librarians rely on to build their own skills, create library guides, and build collections, especially when time is limited or if they lack expertise in a subject or functional area.[4] Librarians explore libraries' websites specifically looking for salient resources, under the basic assumption that librarians with expertise curate the content.[5] As Melissa Fraser-Arnott noted in the midst of the COVID-19 pandemic, the practice of scanning guides can "reveal information resource patterns," especially for emerging topics.[6] Although librarians employ various tools to create web-based guides, the LibGuides platform from Springshare has been widely adopted for its ease of creating guides and pathfinders without the need for coding knowledge.[7] The platform also facilitates requests for permission to copy individual guides so that others can easily create derivatives with local branding.[8] LibGuides launched in 2007 and, as of April 2023, reports 913,616 total guides, created by 265,032 librarians, at 5,890 institutions, in 103 countries.[9] Despite commonly acknowledged shortcomings due to the ease of creation, Nester L. Osorio and Gabriel E. Osorio assert that "Today's subject guides are still produced by specialized librarians that use their expertise in the knowledge of a specific literature field to assist users in finding reliable sources of information."[10]

Research that considers LibGuides as a web creation tool has yielded insights around structure, design, usability, best practices, and standardization. Considering LibGuides "as distinct collections" conceptionally reframes how they are analyzed and discussed so that researchers "consider the unique content contained in a LibGuide as an independent information resource."[11] LibGuides has transformed into a publication genre created by librarians with expertise and domain knowledge.[12] In 2016, LibGuides was designated as a type of grey literature publication by the members of the Grey Net. Additionally, Marcus Vaska and Rosvita Vaska note that LibGuides

"are published online, easily accessible, provide up-to-date information, promote self-learning, and contain information that is often overlooked and neglected."[13] LibGuides have also been discussed as open educational resources (OER) publications, provided the content is clearly noted with a Creative Commons or similar license.[14]

Informal review of library guides to quickly inform practice is a relatively simple process that can be achieved through a targeted web search or by searching Springshare's LibGuides Community search engine. Researchers, however, have conducted more robust systematic analysis of resources promoted in library guides that support specific subjects, promote new services, address trending information needs, and critique the resources therein. Three recent articles illuminate this trend by critically assessing the resources and information presented in LibGuides on LGBTQ+ health,[15] anti-black racism,[16] and decolonizing indigenous studies.[17]

Books Promoted in Library Guides

The promotion of books (print and electronic) has been a common theme in LibGuides research literature. Noting the central role of books, Rosalinda Hernandez Linares and Anna Marie Johnson stress that books are "clearly an important piece of librarian's work as a tool for connecting users with collections."[18] When analyzing LibGuides' structure and usability, researchers often observe that "books" and "find books" are very common naming conventions for tabs/pages.[19] Recent studies, concerned more with the specific books included in the content of LibGuides, have reported positive correlations between patron views of book assets in LibGuides compared to the library's circulations statistics.[20] While not all cross-institutional studies focusing on content report on specific book resources, book citation variables are a significant focus of recent studies of LibGuides targeting sub-disciplinary topics (mechatronics;[21] non-engineering aviation;[22] and wine, enology, and viticulture[23]) and new service areas (3D printing in libraries[24] and data visualization research support services[25]).

Citation Analysis

Citation analysis methods can be used to assess the resources (books, databases, journals, websites, etc.) found in online library guides, especially the corpus of LibGuides, in order to synthesize the collective knowledge of librarians. Citation analysis is a branch of bibliometrics that analyzes the citations found in publication collections to determine patterns of use and to "identify information sources that are most critical to the success" of the authors.[26] Citation indicators commonly collected include "year, author, subject area, document type, source title, affiliation (or institution), country, source type, and language" that "reveal the distribution of literature" cited in the collection.[27] For example, when studying scholarly publications, academic librarians' application of citation analysis methods "typically involves recording the details of the reference lists of a number of publications to determine what materials are being consulted" by faculty and students.[28] Four citation analysis variables are commonly reported in this literature: "types of resources cited, age of cited resources, frequency of citation [to publication titles], and library holdings" as well as some "less commonly analyzed variables" that include "number of citations per publication" as well as the cited work's subject and language.[29]

Citation Analysis Variables of Resources Promoted in LibGuides

Several authors of cross-institutional studies of LibGuides explicitly report using content analysis methods to make observations about promoted resources.[30] Common metrics, clearly presented in Catherine V. Johnson and Scott Y. Johnson's assessment of resources (i.e., databases, books, e-book platform, journals, and video collections), report

- Total number of resources listed by type,
- Total number of unique titles by resource type,
- Total number of unique resources (by type) appearing on only one LibGuide,
- Average number of resource types listed on a LibGuide, and
- List of titles by the total number of LibGuides the title appears. [31]

This data collection can facilitate two important approaches to citation analysis (although not reported in the above study): "bibliographic coupling ... where the reference lists of two documents share one or more of the same cited documents" and "co-citation, [which] occurs when two citations are cited together" in the same publication.[32] This metric can help provide insight into how often specific resources are being promoted together, highlighting their perceived relevance to a particular topic or audience, or revealing potential relationships between resources that might not be immediately apparent.

Manual Methods Currently Used for Data Collection

Most cross-institutional studies of LibGuides rely on manual data collection processes.[33] This is true for the research studies about subject guides reviewed by Amanda McCormick in which all described manual processes for data collection and data tabulation.[34] The LibGuides content analysis study "The Anatomy of Nursing LibGuides" by Tony Stankus and Martha A. Parker appears to be an early example of this type of research and a model for others.[35] The basic data collection process for subject guides research begins by considering at least two basic inclusion criteria: thematic subject of the guides to be observed and institution demographics (type, affiliation, and location). If the subject of the guides is the initial criteria, most researchers search by subject/keyword using Springshare's LibGuides Community search engine and then filter for the pre-determined list of libraries. When it is assumed that relevant webpages are not built using Springshare's platform, researchers conduct keyword searches using internet search engines. If the initial search criteria are by institutions, researchers manually search and browse library websites to identify relevant webpages and navigate to the library's guides. By narrowly defining the research inclusion criteria (thematic scope and institutional demographics) and limiting to LibGuides, it is possible to achieve a sense of comprehensiveness; however, some studies still use additional sampling techniques to restrict the quantity of data collected.

Researchers must pose limits, such as those noted above, to narrow the scope of their work, while also addressing biases introduced by the methods. For example, in their study to critically assess colonization biases perpetuated in LibGuides, Kristen J. Nyitray and Dana Reijerkerk intentionally limited data collection to Association of Research Libraries (ARL) members to shine a light on the practices of elite institutions in the United States and Canada while also noting that future research could examine other types of institutions.[36] Other researchers, attempting to mitigate sample biases, devised sampling methods to disperse the selection of institutions by geographic locations and/or type of institution.[37]

Computational Approaches to Library Guide Research

Considering the time and effort that is needed to select library guides from multiple institutions, code information, and then to tabulate the data, it is surprising that few authors report using computational methods. There are two distinct and notable exceptions. In their study synthesizing observations regarding the structural organization of LibGuides, Hennesy and Adams "used computational methods to collect and analyze elements of 12,781 subject guides at 114 US 'R1: Doctoral Universities—Very high research activity' (R1) academic institutions."[38] The researchers used an Application Programming Interface (API) provided by LibGuides, in combination with

manual review, to identify all LibGuides designated as subject guides and limited the data collection to only home/landing pages for each guide. A second study of note sought to find a specific resource (DataONE) in 78 Association of Research Libraries (ARL) LibGuides. Leah Cannon, Brianne Dorch, and Hannah Gunderman utilized a Python script to identify and test links to the DataONE.org domain, with the primary objective of validating the URLs.[39] The research did not report critical analysis of the LibGuide content.

Collaborative Research Teams

Research teams of two or more collaborators may be helpful to implement computational methods and to address complex research questions. From the minimal information about the researchers' credentials available in the articles (such as authors' job titles, functional areas, research interests and skills along with some discussion of peoples' roles to complete methodology), it was observed that research teams were necessary to complete processes using computational methodologies. The brief bios for Hennesy and Adams elucidate complementary credentials that range from Hennesy's functional responsibilities that include "text data mining research" while Adams's include "reference, instruction, and LibGuide work."[40] Cannon, Dorch and Gunderman specify that beyond the research team, they also had the assistance of their institution's Programming Research Group to replace manual processes with computational processes.[41] Many of the studies discussed in the literature review were completed by research teams rather that individuals, in particular the research teams conducting the critique of content promoted in LibGuides.[42]

METHODOLOGY

To assess how librarians promote particular book resources, we gathered extensive data about the books and LibGuides. Given the large amount of data expected, a computational approach to data collection and cleaning was applied using Python in a series of Jupyter notebooks.[43] Several widely used Python libraries and packages were identified and used throughout the processing, including Selenium WebDriver and BeautifulSoup (for querying web sites and scraping relevant data) and pandas (for data collection, clean-up, and analysis).[44] The data collection and cleaning was an iterative process, involving multiple approaches to ensure accuracy and reliability.

Gathering the Data

Initial Dataset

An initial dataset of book titles used for this case study on computational citation analysis methodology was compiled in 2022 through a manual selection process as part of a larger research project to identify "thesis and dissertation self-help book" titles. A larger goal is to create an extensive list of book titles in which the subject matter is (a) targeted toward graduate (masters or doctoral) students and (b) directly related to the process of writing the thesis or dissertation.[45] Thesis and dissertation self-help books held in library collections were identified by manually searching WorldCat using parameters including the descriptor, subjects, genre, and resource type.[46] Searches, conducted in July 2022, with publication date limited to 2012 through July 2022, yielded 679 results, which were downloaded into Zotero and organized by year of publication.[47] Duplicates and inaccurately described books were removed, and titles were manually screened for relevance to the study's scope. In cases where cataloguing metadata was insufficient, additional information about the content and intended audience was gathered through a review of online sources, including Google Books, Amazon, and publisher websites. Some books were topically related (such as literature reviews, research methods, and style manuals) but did not meet all the criteria for this study; these were classified as out-of-scope and

recorded separately to facilitate exclusion in later review. The process, while somewhat effective, was not exhaustive for compiling the corpus of the genre. Some additional titles, known to the authors but not included in the WorldCat results, were manually added to the final list. A total of 159 books met the above criteria and were classified as in-scope (In-scope List) for this case study.

Web-scraping LibGuides

Computational methods were used to scape data related to the "thesis and dissertation self-help books" recommended by librarians, as curated in LibGuides. Starting with the In-scope List as our foundation, we employed Selenium WebDriver in Python to automate a search of the LibGuides Community site to identify potentially relevant LibGuides. Searches were performed in two phases. The first phase used a combination of the book's main title (excluding leading articles, subtitles, or other extraneous content) enclosed in quotation marks, along with the primary author's last name (Search One). If this initial search yielded no results, we proceeded to the second phase, which consisted of the main title as a phrase in quotation marks (excluding leading articles, subtitles, or other extraneous content) and excluded the author's name (Search Two).

The search term formatting is intended to weed out false positive results that could otherwise clutter our dataset. For example, one title in our dataset is *Writing a Thesis: Substance and Style* by Keith Van Wagenen. The Search One query—"writing a thesis" wagenen—has higher precision and returns only one LibGuide page, whereas the Search Two query—"writing a thesis"—has greater recall but lower precision, returning well over 1,350 results. Although there may still be false positives from the Search One query due to the inability to ensure that the author is directly associated with the book (simply that both appear on the same page), the results are much easier to manage and have a greater precision than the Search Two query. Conversely, as will be discussed in a later section, the bibliographic metadata within LibGuides is highly uncontrolled, so a full title search without the author may miss a number of hits. In this example, the search query—"writing a thesis substance and style"—returns zero results, as the author of the sole LibGuide on which this title was found did not include the subtitle in the bibliographic metadata.

Of the 159 titles, 60.4% (n=96) were matched by Search One, an additional 9.4% (n=15) were matched by Search Two, and the remaining 30.2% (n=48) resulted in no matches from searching in LibGuides Community. Search strings with no results (n=48) were recorded in a separate CSV file for reference. The following metadata was extracted from the HTML of successful search results using the BeautifulSoup library (a process known as web scraping or scraping) and captured in a pandas dataframe, a tabular data format commonly used in data analysis, to facilitate systematic data tracking (see appendix A): LibGuide title, page title, author, institution, date updated, URL, and any subjects or tags applied to the page. The results comprised a list of 4,379 LibGuides pages (3,043 unique pages on 2,248 unique LibGuides published by 921 institutions).[48]

Duplicate URLs were removed, and some additional pages were excluded from the results: 296 pages that were not relevant to our search (e.g., LibGuides consisting of a list of the library's new acquisitions were excluded, as they were not part of a targeted thematic guide) and 51 pages that self-identified as non-English LibGuides in the page's HTML headers. This left 4,032 LibGuides to analyze. We used a combination of Python libraries (requests, Selenium WebDriver, and BeautifulSoup) to perform web scraping on each of the individual pages, looking specifically for items that were classified in the LibGuide as a book asset during the guide's creation.[49] Results that did not conform to this standard (e.g., if they were referenced as a different asset type) were excluded from the final count.

Once the books were identified, bibliographic metadata about each book on the page was harvested and tracked in a pandas dataframe: title, author, ISBN, publication date, and description (note that not each data point was available for each book—see table 1), along with the LibGuide's title, individual page title, publishing institution, as well as any subjects or tags that had been applied to the page. On the earliest attempts, this process resulted in a list of over 115,000 books, most of which were not related to the research topic. Subsequent data clean-up and filtering reduced this initial list to 64,812 book titles (All Titles).

Table 1. Summary count of bibliographic metadata elements for the All Titles dataset

Bibliographic element	Count	Percentage
Title	64,812	100%
Author	63,033	97.3%
ISBN	51,232	79.0%
Publication date	55,720	86.0%
Description	18,725	28.9%

An additional data point available is the LibGuide publishing institution's country of origin, which can enable analysis of where these books are being promoted. This data is not readily available within most LibGuides or in the LibGuides Community search results. However, it is accessible by searching specific institutions in LibGuides Community under the Find Institutions tab or by using an API provided by Springshare.[50] For this study, we used Selenium WebDriver to query each of the unique institutions in the dataset and extract the institution's country from the results and then map it back into the full dataset.

Processing the Data

After compiling this new dataset of book metadata, our next step was to begin the process of data cleaning and standardization. The final data review aimed to achieve three key objectives: (a) limit the dataset to include only relevant books that we wished to measure, (b) attempt to identify additional pertinent titles beyond the In-scope List, and (c) get an accurate count of each title's occurrence across all LibGuides. These metrics can be analyzed to identify patterns in the promotion of specific resources.

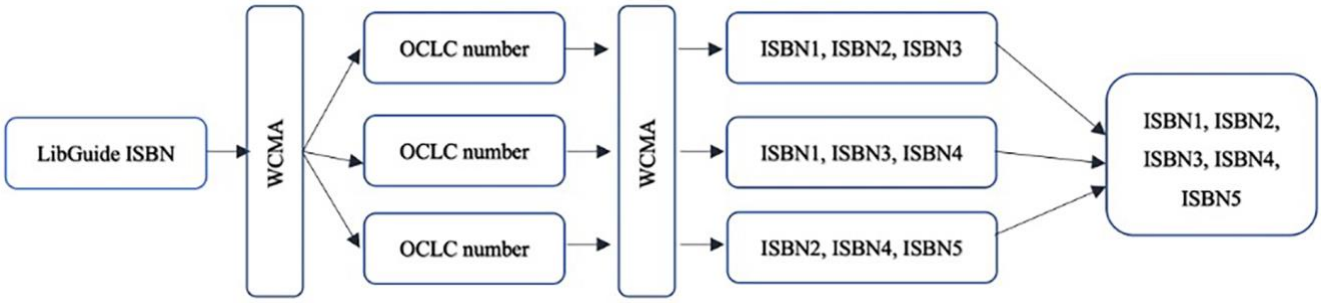
Identifying and Normalizing Variations in Bibliographic Metadata

Reliably identifying all the variations of a given book proved to be one of the most challenging steps in this process. Accounting for different editions, formats, and variants in data entry was crucial to accurately assess the popularity of individual titles. The LibGuides template for adding book assets includes fields for structured metadata (e.g., Title, Author/Editor, etc.); however, there is no data control imposed, and disparities in bibliographic metadata were significant. One example contained 19 recorded variations of the same title between the different guides, including differences in punctuation, inclusion/exclusion of the edition in the title, omission of the subtitle, among others. These inconsistencies made it difficult to group the entries accurately based on the Title field alone. Attempts to group by a simplified title string introduced a different issue: the data includes different books with the same title. For example, our dataset includes seven different books with the main title *Research Methodology*, so any grouping using just the main title will be inaccurate. Similar levels of variance were found in the author field, and accounting for all combinations of author and title was infeasible, despite several attempts using partial string or fuzzy matching.

Another approach was to match titles using the provided ISBN, which presented a new set of challenges. Many book assets did not include an ISBN, and some book asset entries used non-ISBN identifiers in the ISBN field, such as a local shelving number or an Amazon Standard Identification Number (ASIN).[51] After filtering out non-ISBN entries during the initial data collection, only 79.0% of the books (n=51,232) used a 10- or 13-digit ISBN in this field. Some LibGuides included an ISBN as part of a text string (n=1,907), and these were extracted using regular expressions. We were able to isolate additional missing ISBNs by using a regular expression in the URL for the book's cover image (n=7,553), bringing the total to 93.6% (n=60,746) of the entries with a valid ISBN. However, an additional major complication of grouping by ISBN is that a single book can have multiple ISBNs (e.g., different ISBNs for hardcover, softcover, e-book, different editions, etc.), so grouping by a single ISBN is as ineffective as earlier approaches.

To work around this problem, we attempted to gather all the related metadata for the ISBNs of a given title. After some testing using the open APIs of Google Books, Wikipedia, Open Library, and the Library of Congress' Voyager Z39.50 server (the best of which only mapped successfully to approximately 63% of the book data), we settled on using a multistep approach in the WorldCat Metadata API (WCMA).[52] For each book entry in the All Titles list, we queried the WCMA by the ISBN listed in the LibGuide. The API returns metadata in JSON format, including a list of OCLC numbers that reference that ISBN. Each OCLC number was then queried in WCMA to get a list of associated ISBNs, as well as author and title variations. Using the Python library thefuzz, we performed fuzzy matching on the original title and the OCLC recorded title to help reduce any false positives, such as typos or inaccurate catalog metadata. Duplicates were removed, and the results were sorted to facilitate matching in a later step (see fig. 1). Due to rate limits put in place by WorldCat, this process took several days. Then, we iterated through the All Titles dataset, mapping the single ISBN from the LibGuide to the list of ISBNs obtained through WorldCat. Now, all related books should have the same field to enable grouping.

Figure 1. Simulated workflow using the WorldCat Metadata API (WCMA) to gather multiple related ISBNs from a single ISBN.



For the final step, we again used the Python library thefuzz, attempting to match the remaining books (i.e., those that did not have an ISBN from the original LibGuide, n=4,120) against the other known groups in the dataset by using various matching patterns such as (a) a cleaned version of the book's title without punctuation, subtitles, edition number, or other extraneous content, along with (b) the primary author's last name. For matches with a confidence level of 85% or higher, the relevant WorldCat ISBNs were added to that item's metadata. This iterative approach allowed us to capture an additional 725 items, resulting in 94.8% (n=61,417) of our complete dataset having consistent ISBN metadata. With this metadata now available, we can more confidently identify the

number of unique titles regardless of other metadata inconsistencies and perform data analysis more effectively and with greater accuracy.

One challenge inherent in this approach is that it does rely on correct ISBNs to be entered in the LibGuide. While not common, there were instances noted during manual review in which some LibGuide authors had associated a book with another title's ISBN. While this did lead to some slightly skewed data, the process of manually verifying all included ISBN/Title matches was infeasible, and the decision was made to allow them to remain, acknowledging a certain amount of error in the data.

Filtering to Relevant Results

Once the above was completed, the next step was to narrow the dataset to include only titles relevant to the topic (in our case, books on writing dissertations and theses). Identification of relevant books was approached in three primary ways. The first way consisted of matching against the In-scope List, which also had ISBN matching completed in the same manner as above. The second approach was through keyword matching in the title or description, by looking for the inclusion of any of the following terms: dissertation, dissertations, thesis, theses, doctoral, doctorates, doctorate. If a book was found to match either of the above conditions, it was marked as relevant and moved to the next stage of review. The third method involved iteratively removing false positives through (a) manual reviewing books marked as relevant, (b) excluding books based on phrases found in the title (e.g., books with phrases such as "undergraduate dissertation" or "honors thesis" were excluded, as they are not pertinent to the topic of graduate and doctoral work), then repeating steps (a) and (b) until satisfied with the results. The resultant subset contains 3,973 book asset entries, which refer to 281 unique books, found on 1,529 individual LibGuide pages (Relevant Titles).[53]

Identifying Additional Titles

After compiling the Relevant Titles list, the next phase involved identifying additional books that were potentially related to our topic. Our approach was to utilize co-occurrence within LibGuide pages, following the co-citation analysis concept that the LibGuides remaining in our dataset should be targeted to a specific topic, and related books should appear on the same page. Going back to the original All Titles dataset, we first filtered to include only the Relevant Titles and created a new list that included all books that appeared on all those LibGuide pages. This new list comprised an additional 16,509 total books (5,931 unique) (Additional Titles).

RESULTS

With the processing complete, we developed a series of Jupyter notebooks to conduct various analytical tasks on the data. These tasks were designed to address our initial research question and also serve as proof-of-concept examples for exploring or visualizing the data that can enhance our understanding of how these books are promoted. Results were calculated via grouping in pandas and visualized using the Python library plotly.

Frequency and Co-occurrences of Titles

Information about the unique titles was recorded in a separate dataframe. This process involved grouping the Relevant Titles and Additional Titles by the merged WorldCat ISBN field and recording the number of occurrences, all title and author variations, along with the URLs of the source LibGuides. By sorting this resultant data by the number of occurrences, we can identify the most frequent titles in the datasets (see tables 2 and 3).

In addition to a simple frequency count of these titles, we wanted to be able to examine which titles (relevant or otherwise) would be included on the same page. This metric can help provide insight into how often specific book titles are being promoted together, highlighting their perceived relevance to a particular topic or audience, or revealing potential relationships between resources that might not be immediately apparent. Beginning with the most frequent titles, determining the co-occurrences involved creating subsets of data that included only URLs on which a given relevant book appeared and counting occurrences of other top titles on those pages. In this way, we could plot relative frequencies against relevant books (see fig. 2) and additional books (see fig. 3) using pandas. This co-occurrence analysis can facilitate the identification of common themes or topics related to the primary subject. In our case, this included books on literature reviews, research methods, style manuals, and other related themes. Understanding the resources that tend to be promoted together can help librarians create more balanced and tailored LibGuides that more effectively meet the needs of and support for their patrons.

To highlight one example, the most frequently referenced relevant book (A1, table 2) was *Doing a Systematic Review: A Student's Guide*, edited by Boland, Dickson, and Cherry. When evaluating usage alongside other frequently included relevant books, this title only co-occurs with four other titles (and only once each) (see fig. 2). However, when checking co-occurrence with the most frequent additional titles, A1 appeared on 194 LibGuides alongside seven of the top ten titles, with the highest co-occurrence (n=98) against *An Introduction to Systematic Reviews* (B6, table 3) (see fig. 3). The similarity between books A1 and B6 is apparent from the title, and the co-occurrence is not surprising.

Table 2. Top ten most frequent titles and authors from the Relevant Titles list

Most frequent relevant titles	**Total occurrences**
A1 - *Doing a Systematic Review* (Boland)	207
A2 - *How to Write a Thesis* (Murray)	178
A3 - *How to Write a Thesis* (Eco)	127
A4 - *The Dissertation Journey* (Roberts)	97
A5 - *Succeeding with Your Master's Dissertation* (Biggam)	83
A6 - *Surviving Your Dissertation* (Rudestam)	80
A7 - *Writing a Graduate Thesis or Dissertation* (Blair)	75
A8 - *Proposals That Work* (Silverman)	74
A9 - *A Nurse's Step-By-Step Guide to Writing a Dissertation or Scholarly Project* (Roush)	74
A10 - *How to Write a Master's Thesis* (Bui)	71

Table 3. Top ten most frequent titles and author from the Additional Titles list

Most frequent additional titles	**Total occurrences**
B1 - *The Sage Handbook of Qualitative Research* (editor)	175
B2 - *Publication Manual of the American Psychological Association* (APA)	127
B3 - *A Manual for Writers of Research Papers, Theses, and Dissertations* (Turabian)	117
B4 – *Research Design* (Creswell)	111
B5 - *The Literature Review* (Machi)	103
B6 - *An Introduction to Systematic Reviews* (editor)	101
B7 - *Systematic Approaches to a Successful Literature Review* (Booth)	92
B8 - *The Elements of Style* (Strunk)	89
B9 - *MLA Handbook* (editors)	83
B10 - *The Craft of Research* (Booth)	83

Figure 2. Charting the most frequent relevant titles co-occurring within the same LibGuide page. Darker shading indicates a higher level of co-occurrence. See table 2 for full item titles.

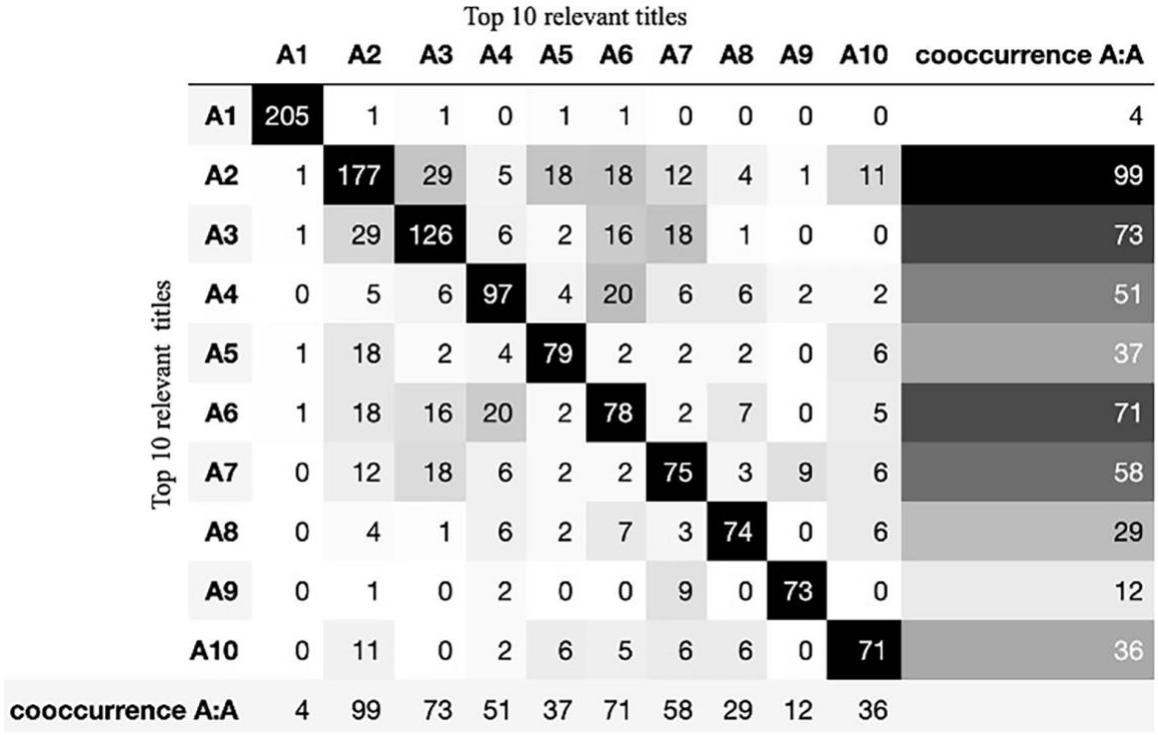


Figure 3. Charting the most frequent (A) relevant and (B) additional titles co-occurring within the same LibGuide page. Darker shading indicates a higher level of co-occurrence. See tables 2 and 3 for full item titles.

Top 10 relevant titles

Top 10 additional titles	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	cooccurrence B:A
B1	2	0	0	1	5	0	0	0	0	0	8
B2	6	5	3	9	19	5	3	2	7	7	66
B3	0	50	20	3	4	13	11	6	7	30	144
B4	5	0	0	9	6	1	0	8	0	2	31
B5	9	9	0	12	3	15	13	10	11	10	92
B6	98	1	1	0	2	1	0	0	0	0	103
B7	72	3	1	0	2	1	0	0	0	0	79
B8	0	30	31	7	0	2	14	4	3	6	97
B9	0	20	17	0	1	14	1	2	0	4	59
B10	2	6	4	6	0	6	3	7	2	6	42
cooccurrence A:B	194	124	77	47	42	58	45	39	30	65	

Distribution of Relevant Books

Identifying clusters of relevant resources provides an opportunity to review LibGuides with higher numbers of relevant titles, which can allow for a more efficient review process in at least two ways. First, distribution of these books among LibGuides can be examined through bibliographic coupling, a method to measure the number of relevant books that appear on a single page. As noted, the Relevant Titles dataset shows 3,975 relevant books appearing on 1,529 LibGuide pages. 51.6% (n=789) of these pages contain only one relevant title, 15.6% (n=238) contain two relevant titles, 10.1% (n=154) list three relevant titles, with the remaining 22.7% (n=348) pages having between 4 and 22 relevant titles (see fig. 4). Second, we can review the relative proportion of the LibGuides devoted to relevant books. When taken as a percentage of the total number of books represented on a given LibGuide page, 23.8% (n=364) pages were comprised of less than 10% relevant books, an additional 21.9% (n=334) included 10–20% relevant books, and 24.1% (n=368) of the pages were comprised of 20–40% relevant books. On the upper end of the data, 9.2% (n=141) of the pages were comprised of over 90% relevant books (see fig. 5).

Figure 4. Frequency distribution of the number of LibGuides containing a given number of relevant titles.

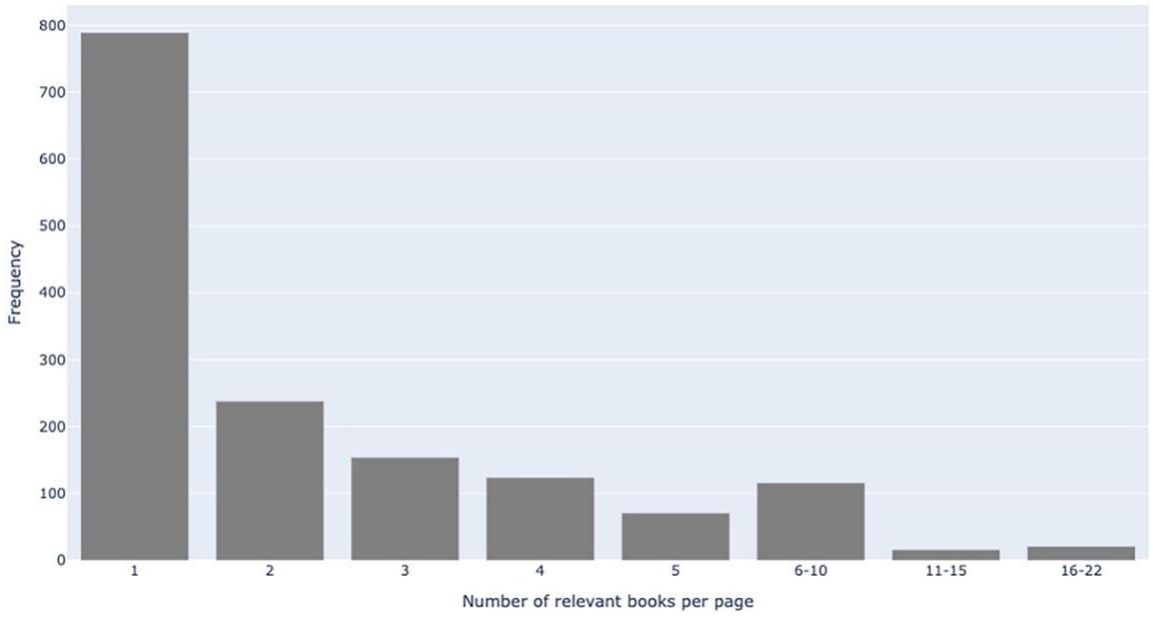
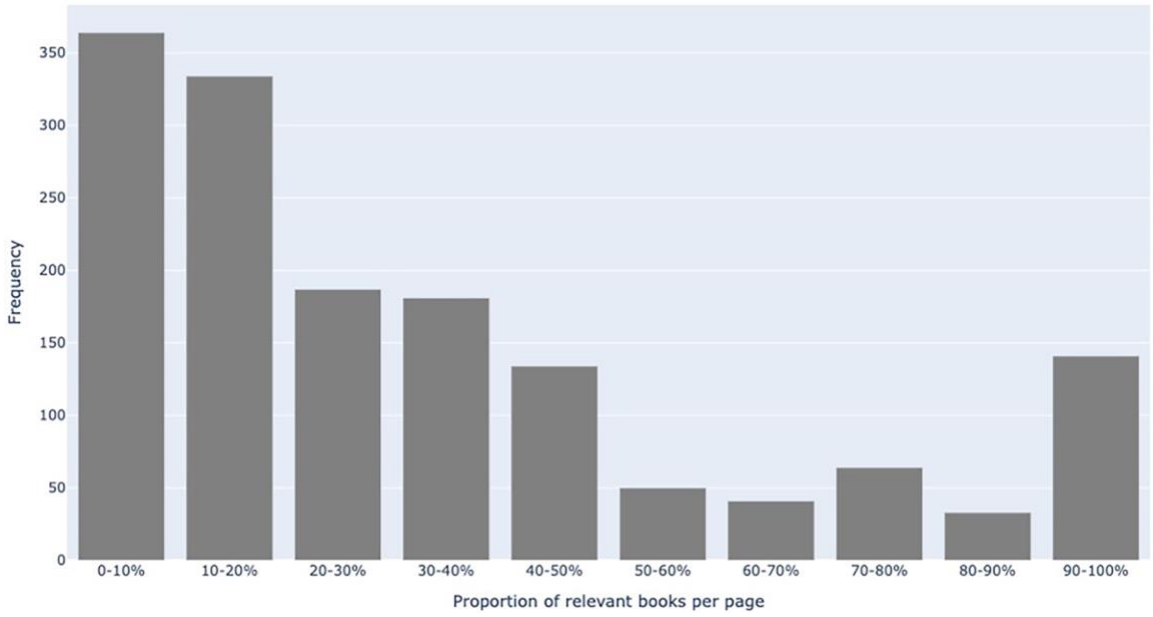


Figure 5. Proportion of relevant books on a given LibGuide page.



Comparing Country Data

The Relevant Titles dataset consisted of 1,529 LibGuide pages published by 510 unique institutions from 33 countries. The highest proportion was from the United States, with 60.6% of the entries (n=927), followed by the United Kingdom, representing 13.5% (n=207) of unique pages, then Canada with 6% (n=91), with other countries following in rapidly decreasing proportions. Although our interest for this phase was simply in the raw count of occurrences, this additional data point could be used to weight the data to determine the top books by country, the relative proportion of a given book from a given country, and a variety of other research questions. Reviewing the most frequent titles along with an additional dimension denoting the

LibGuide's country of origin can provide valuable context when considering if the institution's location might play a role in the specific resources or types of resources promoted (see fig. 6).

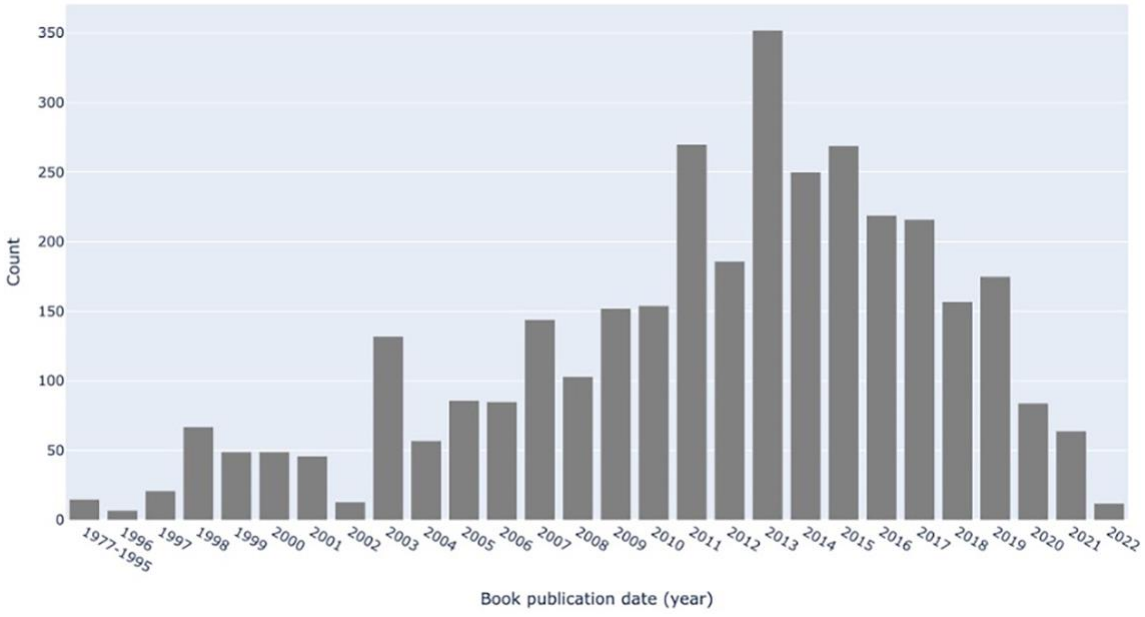
Figure 6. Top ten relevant titles illustrating the proportion of US LibGuides (dark grey) and non-US LibGuides (light grey) on which a given title appears.

Title/Author	Total occurrences	Proportion of US vs. non-US occurences
A1 - Doing a systematic review (Boland)	207	39% / 61%
A2 - How to write a thesis (Murray)	178	57% / 43%
A3 - How to write a thesis (Eco)	127	77% / 23%
A4 - The dissertation journey (Roberts)	97	63% / 37%
A5 - Succeeding with your master's dissertation (Biggam)	83	39% / 61%
A6 - Surviving your dissertation (Rudestam)	80	57% / 42%
A7 - Writing a graduate thesis or dissertation (Blair)	75	69% / 31%
A8 - Proposals that work (Silverman)	74	73% / 27%
A9 - A nurse's step-by-step guide to writing a dissertation or scholarly project (Roush)	74	89% / 11%
A10 - How to write a master's thesis (Bui)	71	79% / 21%

Reviewing Dates

Although our initial list used to gather data was scoped to include primarily books published since 2012, relevant book titles in use on LibGuides showed a much higher variance. Of the entries in the Relevant Titles dataset (n=3,973), 86.4% (n=3,434) had a valid publication date included in the metadata. Of these, 57.8% (n=1,983) had a publication date between 2012 and 2022, with the remaining 42.1% (n=1,450) published earlier—some as early as 1977 (see fig. 7).

Figure 7. Distribution of publication years of Relevant Titles. Note that the earliest dates (n=15) are grouped into a single category of 1977–1995, inclusive.



Subjects and Tags

LibGuide creators have the option to add page-specific subjects and tags to their LibGuides. These can provide context for the user and are also available as a filter for search results within LibGuides Community. Springshare characterizes the distinction as follows:

> Subject categories provide a way to organize guides by topic, discipline, or subject area. On the default LibGuides homepage, these will appear grouped under the By Subject tab. Tags provide a way to organize guides by keywords. Primarily used to improve search results, you can also create widgets that allow users to view guides assigned to your tags.[54]

Examining the application of and commonalities among these can further help us understand the ways that librarians are attempting to make these resources discoverable. Of the 1,529 relevant LibGuides, 68% (n=1,037) had one or more subjects assigned, and 48% (n=728) had one or more tags assigned. These were captured during the initial web scraping process.

For LibGuides that included subjects (n=1,037), 58.8% (n=609) had only a single subject assigned, 20.6% (n=214) had two subjects, 9.6% (n=100) had three subjects, with the remaining 11% (n=114) LibGuides containing between 4 and 22 subjects, with two outliers at 45 subjects each. By grouping all of these subjects together and counting frequencies, we can reveal the most commonly used subjects overall, as well as narrowing in to only LibGuides with only a single subject and comparing subject usage between these applications (see table 4).

Table 4. Top ten subjects assigned to 1,037 LibGuide pages containing at least one relevant title

All LibGuides (n=1,037)		**LibGuides with a single subject (n=609)**	
Term	**Count**	**Term**	**Count**
education	110	education	52
nursing	73	nursing	27
psychology	59	psychology	21
business	45	business	14
research	28	design	10
health sciences	28	academic writing	10
public health	25	entertainment	9
engineering	24	health sciences	8
social work	21	english	8
medicine	21	sociology	7

For LibGuides that included tags (n=728), 19.8% (n=144) had only a single tag assigned, 16.8% (n=122) had two tags, 13.5% (n=98) had three tags, with the remaining 50% (n=364) LibGuides containing between 4 and 29 tags, with one outlier which had 97 individual tags applied. As with subjects, we can also perform a comparison between tag application on all guides and on guides with only a single tag (see table 5).

Table 5. Top ten tags assigned to 728 LibGuide pages containing at least one relevant title

All LibGuides (n=728)		LibGuides with a single tag (n=144)	
Term	**Count**	**Term**	**Count**
research	67	education	8
writing	57	reading & resources list	7
thesis	51	health & social care	6
nursing	38	systematic reviews	6
education	38	psychology	6
literature review	37	disciplines	4
dissertation	37	nursing	4
health	30	english	3
dissertations	29	research	3
reading & resources list	27	course materials	3

LibGuide Titles

When considering how certain materials are promoted to users by librarians in this venue, it can be helpful to examine how the LibGuide itself is characterized by looking at the title applied to both the LibGuide and the individual page. Although there are a few ways this could be evaluated, we found it useful to simply look at (a) the most common titles used, and (b) the most frequent n-grams (a term used in Natural Language Processing (NLP) to indicate a sequence of *n* words) across all relevant LibGuides to help determine commonalities among the guides. To obtain these values, we utilized the Natural Language Toolkit's (NLTK) FreqDist function, which can extract and analyze individual terms from longer texts.[55] By evaluating these results, we can gain some insight into how these materials are being promoted (see tables 6 and 7). For this case study, the authors are most interested in the raw data and are evaluating the terms as used, rather than combining similar terms (e.g., "dissertation" and "dissertations"). Future research could involve various NLP methods to combine similar terms or concepts, such as stemming or lemmatizing terms.

Table 6. Most frequent titles among relevant LibGuides (n=1,529) and individual pages (n=1,527)

LibGuides		Individual pages	
Title	**Count**	**Title**	**Count**
Systematic Reviews	33	Home	78
Nursing	26	Books	49
Psychology	16	Writing Help	30
Literature Reviews	16	Master's Thesis Projects	26
Education	11	Writing Resources	24
Academic Writing	11	Writing	21
Research Support	7	Research and Writing	21
Sociology	6	Dissertations & Theses	19
Writing Resources	6	Dissertations	19
Resources for Dissertation Writing	6	Writing & Citing	6

Table 7. Most frequent n-grams (1 and 2) for relevant LibGuide titles (n=1,529) and individual pages (n=1,527)

Common n-grams (LibGuide title)				Common n-grams (page title)			
Term (n=1)	**Count**	**Term (n=2)**	**Count**	**Term (n=1)**	**Count**	**Term (n=2)**	**Count**
research	235	systematic reviews	63	writing	337	and writing	37
guide	165	research guide	44	research	198	literature review	37
resources	122	resources for	29	resources	165	writing resources	36
writing	110	literature reviews	27	books	162	systematic reviews	36
reviews	93	research methods	26	thesis	112	writing a	34
systematic	80	library resources	19	dissertation	90	research and	33
education	78	public health	19	home	80	writing help	33
graduate	78	graduate students	18	dissertations	79	study skills	28
health	72	for graduate	16	literature	78	a thesis	27
nursing	69	academic writing	16	reviews	63	master 's	27

DISCUSSION

Through the methods described here, we have presented a model for programmatically gathering data from the LibGuides corpus and computationally applying citation analysis methods to gain insight about books that are recommended by librarians in these guides. These computational methods in addition to manual processes allowed for a more comprehensive analysis than reliance only on manual methods. This project represents what can be accomplished through collaboration and would have not been completed by either author individually. Our two-person research team with various complementary skills from differing functional responsibilities and research interests created synergy throughout the project and enabled us to envision a means to explore increasingly complex and critical research questions.

The goal of this case study, which focused on promotion of book resources aimed at assisting students in the production of their master's theses or doctoral dissertations, was to collect and evaluate data published on LibGuides to inform future decision-making. Thus far, the preliminary results informed recommendations for collection development. Forty-two of the most frequently promoted titles, which were listed on at least 29 different LibGuides, were compared to local paper and electronic holdings. From a gap list, which included 17 titles recently published or new editions, four titles have been added to the library's electronic books collections. Additionally, the research yielded an additional 170 for further analysis, which expanded our understanding of the corpus self-help book to include over 300 titles. Future research will subcategorize the genre by

audience and analyze bibliographic coupling/co-citation of titles to inform decisions as to which of these self-help books to highlight and promote in various LibGuides.

Limitations

Focusing on book resources promoted on the LibGuides platform does limit the data to a subset of *all* the resources (e.g., websites, videos, etc.) that libraries may publish on any number of platforms (e.g., libraries' web pages or LibGuide alternatives). As noted above, if the LibGuide author did not use the book asset during the creation of the page, these resources were likewise not included in the results. Similarly, false positives were sometimes found when non-book resources, such as databases, were added to the LibGuide using the book asset type. Furthermore, the LibGuide metadata does not create a distinction between physical or electronic resources, so evaluating any differences between these formats would require additional steps.

In some ways, this study was necessarily limited by our initial list. Although it was systematically gathered with the intention of achieving saturation in the search results, item selection was limited by keyword inclusion with library catalog metadata. While the 159 titles on the In-scope List were all specifically relevant to the topic, 30.1% (n=48) were not found in LibGuide search results. This lack of results contributes to our understanding of which materials are being promoted by specifically providing a list of certain books that are not being promoted. In addition, having identified 170 additional in-scope titles through this process, we could repeat the initial search process with these new titles included, resulting in additional titles and LibGuides for further analysis and exploration.

The relevancy determination itself was likewise limited by the inclusion of specific keywords in the title or (in some few cases) description. Although this may be a good indicator of relevance for a particular book, it is hard to know how many titles were omitted because they did not contain one of these keywords. For example, one of the titles on our In-scope List was *Surviving and Thriving in Postgraduate Research* by Ray W. Cooksey. This title was included in eight LibGuides, but it could easily have been omitted from the results due to the absence of our selected keywords in the title. Further study could include checking for subject terms using the Library of Congress's Voyager Z39.50 API or the WorldCat Classify API, either of which could serve to capture additional relevant books, although the level of false positives that could be introduced through this approach is unknown.

Possibilities for Future Research

While we were able to collect some institutional demographics to look at the data by country, future study is needed to associate the data by library type and affiliation. Some libraries self-identify with LibGuides Communities as academic institution, public library, special library, etc. However, not all libraries do this, and the metadata is not presented in a way that can be easily accessed by the user, making this data point hard to track. Institutions could be mapped to lists of research libraries, such as R1/R2 research institutions and ARL (Association of Research Libraries) member libraries, allowing for potential review of a given library's peer institutions. There may be a lack of consistency in institution name across lists that would require manual review (e.g., University of Connecticut [R1 listing] vs. UConn Library [ARL listing]), but this sort of analysis is well within reach.

Exploration of the LibGuide's country of origin can be taken further to compare the inclusion of resources from specific countries and the relative frequency of occurrence. As one example, the most frequently included relevant book (A1, table 2) was *Doing a Systematic Review: A Student's*

Guide, edited by Boland, Dickson, and Cherry. Figure 6 shows that 39.1% (n=81) of these were found in US-based guides, with the remaining 60.9% (n=126) originating outside of the US. This can easily be taken further to review the count from each country that includes this title, as well as calculate a weighted proportion relative to a given country's LibGuides in the dataset. This more detailed analysis could provide a more nuanced understanding of the usage of a particular resource across different countries.

Future research could measure the effectiveness of this methodology. While the methodology presented here begins with a pre-selected list of book titles, similar results could be achieved by initiating the search using keywords in LibGuides Community, which mimics the earlier research using manual methods. For example, a general search (e.g., "science") would likely be too broad to be applicable, but a combination of more specific search keywords (e.g., "mycology", "fungi", "mycological", etc.) could produce a useful list of book references and could easily be accomplished and analysed using this methodology.

Applications of citation analysis methodology can reveal patterns of frequency by various measures but do not measure quality.[56] Citation analysis methods uncover only part of what can be studied; thus, additional methods are needed to assess quality and usages of the resources promoted by librarians via LibGuides. In our case, identifying the books promoted by librarians is a preliminary step within a larger research agenda. Future research will examine the patterns of what librarians choose to promote as well as critically assess the content of the books promoted in these guides.

CONCLUSION

Presented here is a proof-of-concept computational methodology to identify books promoted on LibGuides for a specific user group and to report citation analysis variables. While some researchers aim to identify core books and other resources by assessing what is promoted at multiple institutions, the manual methods necessitate limits for identifying a small sample set and/or the type of analysis. These limits can introduce or perpetuate unintentional biases. For example, limiting the data to only ARL member libraries' LibGuides neglects the knowledge and expertise of colleagues from other types of libraries and from libraries in other parts of the world. Turning a critical lens on affiliated libraries can help address biases of content within the guides; however, the effort necessary for completing critical content analysis using various qualitative methods might limit reporting of citation analysis variables.[57]

While the citation analysis results shown in the case study are similar to earlier research, our methodology is novel. Computational methods for searching and initial data collection of LibGuide pages avoids sampling by narrowly defined types of guides or by institution demographics. Our case study (self-help books for graduate students researching and writing theses and dissertations) is centered on a specific user group rather than a particular set of guides (type and/or subject) or set of institutions. To minimize inherent biases, this computational methodology searches the entire corpus of LibGuides to identify guides that include relevant books, which can be cross tabulated by type of guides and types of institutions to look for patterns that reflect the expertise of librarians from around the world. By describing our process in detail and providing access to our research data, we invite others to adapt this computational methodology for citation analysis to increase comprehensiveness of the LibGuides data observed for decision-making.

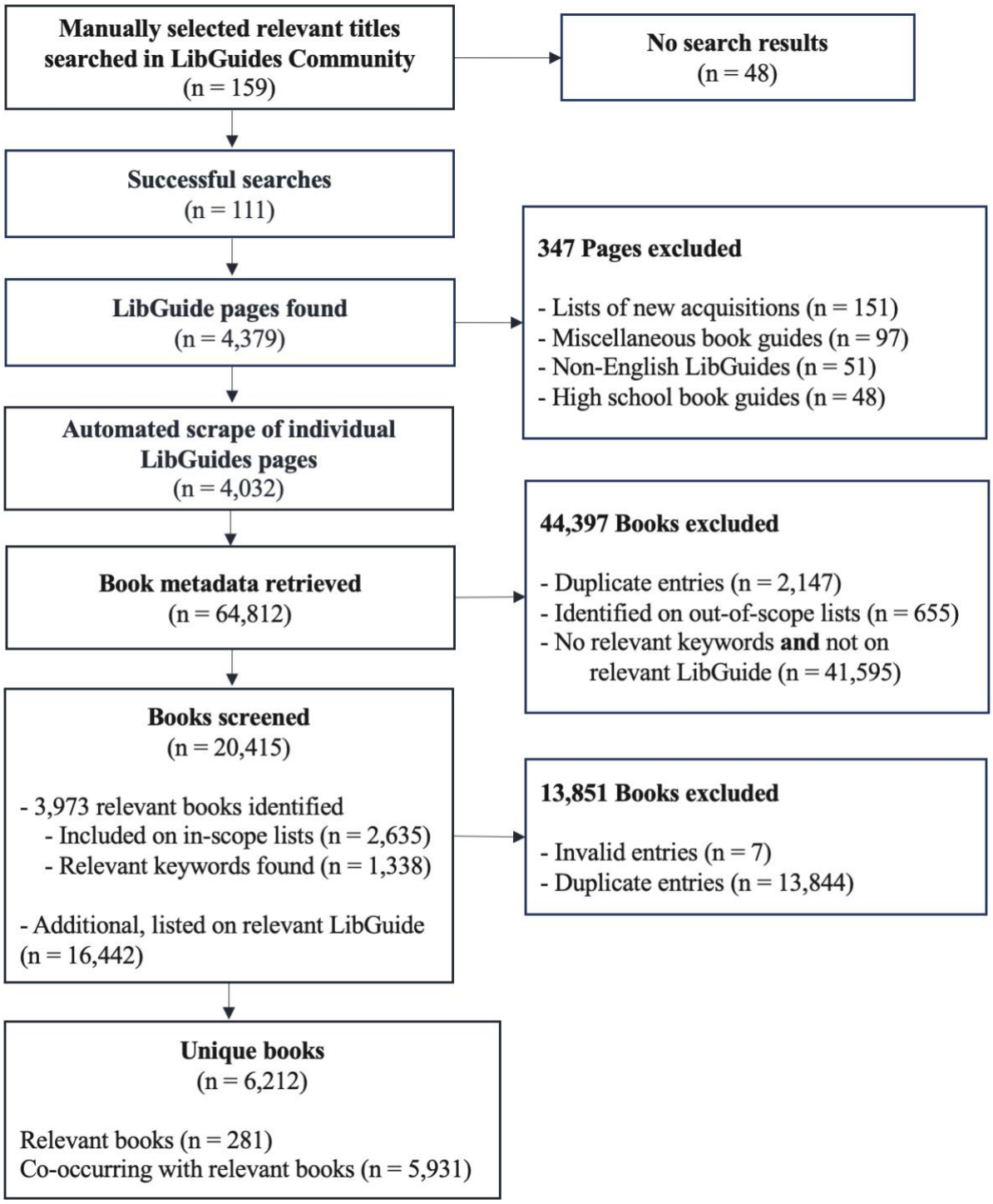
DATA AVAILABILITY

As with any web-based content, LibGuides may be added, removed, or modified at any given time, making replicability of our data processing a challenge. All the data collected, manipulated, and enhanced throughout this process has been saved to csv files to facilitate review. In addition, we used the Internet Archive's Save Page Now (SPN) Public API to submit each of our final set of LibGuide pages (n=1,529) to the Internet Archive's Wayback Machine to create a contemporaneous snapshot.[58] The data reviewed for this article was harvested on March 2, 2023, and the pages were submitted to SPN on March 21, 2023. All data and notebooks used in this paper will be made available in the University of Kansas institutional repository: https://doi.org/10.17161/1808.34184.

Disclosure statement: The authors report there are no competing interests to declare.

Acknowledgements: We would like to thank our colleagues for their thoughtful feedback and encouragement.

APPENDIX A – PRISMA FLOW-STYLE DIAGRAM OF THE DATA COLLECTION AND SCREENING PROCESS



ENDNOTES

[1] Cody Hennesy and Annis Lee Adams, "Measuring Actual Practices: A Computational Analysis of LibGuides in Academic Libraries," *Journal of Web Librarianship* 15, no. 4 (2021): 219–42, https://doi.org/10.1080/19322909.2021.1964014.

[2] Jennifer Emanuel, "A Short History of Library Guides and Their Usefulness to Librarians and Patrons," in *Using LibGuides to Enhance Library Services* (Chicago: ALA TechSource, 2013), 3–20.

[3] Hennesy and Adams, "Measuring Actual Practices."

[4] Anna Bagshaw and Phil Yorke-Barber, "Guiding Librarians: Rethinking Library Guides as a Staff Development Tool," *Journal of the Australian Library and Information Association* 67, no. 1 (January 2, 2018): 31–41, https://doi.org/10.1080/24750158.2017.1410629; Amanda McCormick, "Collection Development for Librarians in a Hurry: A Survey of the Physics Resources of the Libraries of the Association of American Universities," *Issues in Science and Technology Librarianship*, no. 96 (2020), https://doi.org/10.29173/istl68.

[5] Bagshaw and Yorke-Barber, "Guiding Librarians"; Robert M. Britton and Jie Li, "Using Subject Guides to Support Curriculum and Patient Care: A Multipronged Approach," *Medical Reference Services Quarterly* 38, no. 1 (January 2019): 31–40, https://doi.org/10.1080/02763869.2018.1547048; Judith Logan and Michelle Spence, "Content Strategy in LibGuides: An Exploratory Study," *The Journal of Academic Librarianship* 47, no. 1 (January 1, 2021): 102282, https://doi.org/10.1016/j.acalib.2020.102282; Nestor L. Osorio and Gabriel E. Osorio, "An Analysis of Technical Information for Mechatronics Research," *Collection & Curation* 39, no. 4 (October 2020): 117–29, https://doi.org/10.1108/CC-09-2019-0030; Marcus Vaska and Rosvita Vaska, "Looking for Information That Is Not Easy to Find: An Inventory of LibGuides in Canadian Post-Secondary Institutions Devoted to Grey Literature," in *Eighteenth International Conference on Grey Literature* (2017), 59, http://www.textrelease.com/gl18conference.html.

[6] Melissa Fraser-Arnott, "Academic Library COVID-19 Subject Guides," *The Reference Librarian* 61, no. 3–4 (2020): 165–84, https://doi.org/10.1080/02763877.2020.1862021.

[7] Chris Neuhaus et al., "Ubiquitous LibGuides: Variations in Presence, Production, Application, and Convention," *Journal of Web Librarianship* 15, no. 3 (July 2021): 107–27, https://doi.org/10.1080/19322909.2021.1946457.

[8] "Guides: Create or Copy a Guide," Springshare, accessed January 29, 2024, https://ask.springshare.com/libguides/faq/770.

[9] "LibGuides Community," Springshare, accessed April 28, 2023, https://community.libguides.com/#.

[10] Osorio and Osorio, "An Analysis of Technical Information," 118.

[11] Michael Chee, "Weeding the Web: Applying a Collections Management Mentality to Editing Academic Library LibGuides," *Emerging Library & Information Perspectives* 3, no. 1 (October 30, 2020): 152, https://doi.org/10.5206/elip.v3i1.8713.

[12] Chee, "Weeding the Web"; "Document Types in Grey Literature," GreyNet International, accessed April 19, 2023, http://www.greynet.org/greysourceindex/documenttypes.html; Vaska and Vaska, "Looking for Information That Is Not Easy to Find."

[13] Vaska and Vaska, "Looking for Information That Is Not Easy to Find," Abstract.

[14] Deborah L. Lauseng et al., "Assessing Online Library Guide Use and Open Educational Resource (OER) Potential: An Evidence-Based Decision-Making Approach," *Journal of Web Librarianship* 15, no. 3 (July 2021): 128–53, https://doi.org/10.1080/19322909.2021.1935396.

[15] Gregg A. Stevens and Francisco J. Fajardo, "LGBTQ+ Health Research Guides at North American Health Sciences Libraries: A Survey and Content Analysis," *Journal of the Medical Library Association: JMLA* 109, no. 3 (2021): 406–13, https://doi.org/10.5195/jmla.2021.1189.

[16] Gemmicka Piper, Mahasin Ameen, and M. Sara Lowe, "An Investigation of Anti-Black Racism LibGuides at ARL Member Institutions," *Communications in Information Literacy* 15, no. 2 (2021): 188–207, https://doi.org/10.1016/j.acalib.2021.102397.

[17] Kristen J. Nyitray and Dana Reijerkerk, "From Pathfinder to Indigenized: An Assessment of LibGuides for Indigenous Studies by ARL Member Institutions," *College & Research Libraries* 83, no. 6 (2022): 1014–31, https://doi.org/10.5860/crl.83.6.1014.

[18] Rosalinda Hernandez Linares and Anna Marie Johnson, "Comparing Apples to Apples Oranges: An Exploration of the Use of LibGuides in ARL Libraries," *The Southeastern Librarian* 64, no. 1 (2016): 5, https://digitalcommons.kennesaw.edu/seln/vol64/iss1/16/.

[19] Hennesy and Adams, "Measuring Actual Practices."

[20] Siviwe Bangani et al., "LibGuides as a Collection Development Tool: The Case of North-West University, South Africa," *Library Philosophy and Practice* (2019): 1–21, https://digitalcommons.unl.edu/libphilprac/2944/; Sarah Hicks, Kristy White, and Rob Behary, "The Correlation of LibGuides to Print and Electronic Book Usage: A Method for Assessing LibGuide Usage," *Journal of Web Librarianship* 15, no. 1 (January 2, 2021): 1–17, https://doi.org/10.1080/19322909.2021.1884927.

[21] Osorio and Osorio, "An Analysis of Technical Information."

[22] Jennifer Horton, "Survey of Aviation Online Library Research Guides," *Science & Technology Libraries* 41, no. 4 (2021): 411–21, https://doi.org/10.1080/0194262X.2021.1999883.

[23] Jennifer Horton, "An Analysis of Online Research Guides for Wine, Enology, and Viticulture," *Journal of Agricultural & Food Information* (2022): 1–12, https://doi.org/10.1080/10496505.2022.2127736.

[24] Jennifer Horton, "An Analysis of Academic Library 3D Printing LibGuides," *Internet Reference Services Quarterly* 22, no. 2–3 (2017): 123–31, https://doi.org/10.1080/10875301.2017.1375059.

[25] Mahmoud Sherif Zakaria, "Data Visualization as a Research Support Service in Academic Libraries: An Investigation of World-Class Universities," *The Journal of Academic Librarianship* 47, no. 5 (2021): 102397, https://doi.org/10.1016/j.acalib.2021.102397.

[26] Philip B. White, "Using Data Mining for Citation Analysis," *College & Research Libraries* 80, no. 1 (January 2019): 76–77, https://doi.org/10.5860/crl.80.1.76.

[27] Aidi Ahmi, *Bibliometric Analysis for Beginners: A Starter Guide to Begin with a Bibliometric Study Using Scopus Dataset and Tools Such as Microsoft Excel, Harzing's Publish or Perish and VOSviewer Software* (Kedah Darul Aman, Malaysia: UUM Press, Universiti Utara Malaysia, 2021), 5–6.

[28] Kristin Hoffmann and Lise Doucette, "A Review of Citation Analysis Methodologies for Collection Management," *College & Research Libraries* 74, no. 4 (2012): 321, https://doi.org/10.5860/crl-254.

[29] Hoffmann and Doucette, "A Review," 325–26.

[30] Bangani et al., "LibGuides as a Collection Development Tool"; Catherine V. Johnson and Scott Y. Johnson, "An Analysis of Physician Assistant LibGuides: A Tool for Collection Development," *Medical Reference Services Quarterly* 36, no. 4 (October 2017): 323–33, https://doi.org/10.1080/02763869.2017.1369241; Neuhaus et al., "Ubiquitous LibGuides"; Nyitray and Reijerkerk, "From Pathfinder to Indigenized"; Hennesy and Adams, "Measuring Actual Practices."

[31] Johnson and Johnson, "An Analysis of Physician Assistant LibGuides."

[32] Ronald R. Powell and Lynn Silipigni Connaway, *Basic Research Methods for Librarians*, 4th ed, Library and Information Science Text Series (Westport, Conn: Libraries Unlimited, 2004), 63.

[33] Hennesy and Adams, "Measuring Actual Practices."

[34] McCormick, "Collection Development for Librarians in a Hurry."

[35] Tony Stankus and Martha A. Parker, "The Anatomy of Nursing LibGuides," *Science & Technology Libraries* 31, no. 2 (April 1, 2012): 242–55, https://doi.org/10.1080/0194262X.2012.678222.

[36] Nyitray and Reijerkerk, "From Pathfinder to Indigenized"; Piper, Ameen, and Lowe, "An Investigation."

[37] Kathleen Carlson, Joachim Schopfel, and Marcus Vaska, "Grey Literature LibGuides or LibGuides about Grey Literature: A Two-Continent Environmental Scan of Common Themes & Trends," 2017, https://hal.univ-lille.fr/hal-03433941; Kate Dougherty, "Getting to the Core of Geology LibGuides," *Science & Technology Libraries* 32, no. 2 (April 1, 2013): 145–59, https://doi.org/10.1080/0194262X.2013.777233; Fraser-Arnott, "Academic Library COVID-19 Subject Guides."

[38] Hennesy and Adams, "Measuring Actual Practices," 220.

[39] Leah Cannon, Brianne Dosch, and Hannah Gunderman, "Preparing, Not Repairing," *Qualitative & Quantitative Methods in Libraries* 9, no. 1 (March 2020): 9–18.

[40] Hennesy and Adams, "Measuring Actual Practices," 241.

[41] Cannon, Dosch, and Gunderman, "Preparing, Not Repairing," 11–12.

[42] Vaska and Vaska, "Looking for Information That Is Not Easy to Find," 59; Deborah L. Lauseng et al., "Assessing Online Library Guide Use and Open Educational Resource (OER) Potential: An Evidence-Based Decision-Making Approach," *Journal of Web Librarianship* 15, no. 3 (July 2021): 128–53, https://doi.org/10.1080/19322909.2021.1935396; Gregg A. Stevens and Francisco J. Fajardo, "LGBTQ+ Health Research Guides at North American Health Sciences Libraries: A Survey and Content Analysis," *Journal of the Medical Library Association: JMLA* 109, no. 3 (2021): 406–13, https://doi.org/10.5195/jmla.2021.1189.

[43] Brian E. Granger and Fernando Pérez, "Jupyter: Thinking and Storytelling with Code and Data," *Computing in Science & Engineering* 23, no. 2 (2021): 7–14, https://doi.org/10.1109/MCSE.2021.3059263.

[44] Rakesh Vidya Chandra and Bala Subrahmanyam Varanasi, *Python Requests Essentials* (Packt Publishing Birmingham, UK, 2015); Wes McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference*, vol. 445 (Austin, TX, 2010), 51–56, https://10.0.97.248/Majora-92bf1922-00a; "WebDriver," Selenium, 2021, https://web.archive.org/web/20230505182152/https://www.selenium.dev/documentation/webdriver/; Leonard Richardson, "Beautiful Soup Documentation — Beautiful Soup 4.4.0 Documentation," 2015, https://web.archive.org/web/20220709220201/https://beautiful-soup.readthedocs.io/en/latest/.

[45] Tim Anderson and Tomoyo Okuda, "Writing a Manuscript-Style Dissertation in TESOL/Applied Linguistics," *BC TEAL Journal* 4, no. 1 (November 9, 2019): 33–52, https://doi.org/10.14288/bctj.v4i1.334. While Anderson and Okuda note that this "self-help" genre "come[s] in many shapes and sizes, and span[s] disciplinary allegiances, but most often take[s] the form of book-length monographs or chapters in edited volumes," this research is narrowly focused on the book-length monographs.

[46] WorldCat search query: (de: dissertation* and (su: authorship OR su: handbook* OR su: guide* OR su: manual*) not ge: dissertation and yr: 20XX and la= "eng" and dt= "bks").

[47] To query Springshare's LibGuides Community, book publications were normalized to title and authors to represent all possible editions and formats regardless of original copyright date. As a possible measure for currency of the information, we looked at the publication date as reported on the LibGuides asset metadata.

[48] Note that a single LibGuide page can reference more than one relevant book, and a LibGuide is usually comprised of multiple pages.

[49] Aaron W. Dobbs, Ryan L. Sittler, and Douglas Cook, *Using LibGuides to Enhance Library Services: A LITA Guide* (American Library Association, 2013). Springshare provides several types of

assets that can be used when adding content to a LibGuide, including books, databases, files, and others. A book asset is added as "Book from the Catalog."

[50] Hennesy and Adams, "Measuring Actual Practices."

[51] "How to Start Selling on Amazon: Product Listing Details," Amazon, March 29, 2023, https://web.archive.org/web/20230329221118/https://sell.amazon.com/sell.

[52] "WorldCat Metadata API," OCLC Developer Network, March 14, 2023, https://web.archive.org/web/20230314213658/https://www.oclc.org/developer/api/oclc-apis/worldcat-metadata-api.en.html.

[53] The 281 unique books identified by this method consists of a mix of some titles from (a) our In-scope List (111 titles that were found in the LibGuides Community search) and (b) new titles appearing on the same pages that are directly relevant to our topic (281 – 111 = 170). Consequently, this approach yielded 170 newly discovered titles for further analysis.

[54] "Guides: Assign Subjects and Tags to a Guide – Help Center," Springshare, accessed March 21, 2023, https://ask.springshare.com/libguides/faq/791.

[55] Steven Bird, Ewan Klein, and Edward Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit* (O'Reilly Media, Inc., 2009).

[56] Powell and Connaway, *Basic Research Methods for Librarians*.

[57] Nyitray and Reijerkerk, "From Pathfinder to Indigenized."

[58] Vengelis Banos, "SPN2 Public API Page Docs," February 22, 2023, https://web.archive.org/web/20230331011842/https://docs.google.com/document/d/1Nsv52MvSjbLb2PCpHlat0gkzw0EvtSgpKHu4mk0MnrA/edit.