Measuring Actual Practices: A Computational Analysis of LibGuides in Academic Libraries

Cody Hennesya and Annis Lee Adamsb

aUniversity of Minnesota-Twin Cities
bCalifornia State University-East Bay

This is an accepted manuscript version of an article accepted for publication in the Journal of Web Librarianship. https://doi.org/10.1080/19322909.2021.1964014

It is deposited under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.
This study investigates organizational patterns from 12,781 subject guides at 114 academic institutions, comparing practices evident on LibGuides with stated best practices from the literature of library and information science. Data from subject guide fields such as titles, tabs, and boxes were collected systematically from disciplinary guides at institutions in the United States that are described in Carnegie Classifications as "R1: Doctoral Universities – Very high research activity," and then analyzed according to two general methods. First, descriptive statistics were generated for quantitative aspects of the LibGuides data (e.g., the average number of boxes used per guide), to map both normal and atypical organizational practices. Second, word and phrase frequencies from text fields were compiled to explore differences in how fields such as guide titles and descriptions are commonly utilized. The findings suggest that the average range of guides tends to follow best practices, but that in the use of certain guide elements, especially tabs, most guides do not follow recommended guidelines. In addition, the data identify a number of guides that represent extreme outliers to typical practices: a single guide contained 144 tabs, for example, while another subject guide landing page was found to include 2,033 links.

Keywords: LibGuides, academic libraries, subject guides, quantitative analysis, information organization, term frequencies
Introduction

Academic libraries have long been in the business of providing subject directories of information resources for their users. Throughout the 20th century, academic librarians created in-depth reader’s aids and bibliographic essays describing books and reference volumes on the library shelves (and beyond), while shorter “pathfinders” to key subject resources began to be promoted as early as the 1970s (Smith, 2008). As information resources increasingly moved online in the 1990s and early 2000s, libraries migrated pathfinders online as well - initially as static HTML pages on library websites - adding links to the websites, databases, and ejournals most relevant to specific fields of inquiry. Since the launch of Springshare’s LibGuides platform in 2007, the number of online library guides at academic institutions has continued to grow. As of April 2021, the LibGuides Community page reported that there were over 817,000 guides, created by 211,238 librarians\(^1\) at 5,655 institutions, 2,631 of which were academic institutions (Springshare, 2021).

In the same year that Springshare launched LibGuides, the *Journal of Web Librarianship* published its inaugural issue, representing a growing research interest in the new library web environment. Over the subsequent 15 years, library and information science (LIS) scholars and practitioners have frequently studied the usability of LibGuides and other online subject guides, extrapolating best practices and recommendations for administrators and creators of library guides in various contexts. Alongside these studies, researchers have also turned to the growing corpus of LibGuides to undertake qualitative content analyses of guides from specific disciplines, tabulating common recommended resources, and sometimes noting a range of organizational practices across guides. This study builds upon these foundations to document actual practices - as opposed to best practices - related to the organization of online subject guides at academic libraries on a large scale. Specifically, the authors used computational methods to collect and analyze elements of 12,781 subject guides at 114 U.S. "R1: Doctoral Universities – Very high research activity" (R1) academic institutions, outlining both common and atypical practices found on these guides. The authors examine the use of guide titles, descriptions, tabs, boxes, profiles, subject tags, and links across a wide range of subject guides to uncover implicit organizational principles undergirding guides at academic institutions. Where organizational elements from the guides run counter to best practices - in the high number of tabs featured on the average guide, for example - the paper examines underlying assumptions about whom LibGuides are for. The findings also provide a macroscopic view of the disciplines most commonly supported by subject guides at R1 institutions in the United States.

To synthesize organizational trends across a large number of subject guides, the authors apply two primary methods. First, descriptive statistics are generated for quantitative data extracted from the guides (e.g., “What is the mean number of boxes/tabs per guide?”). Second, the frequencies of terms and phrases from key text fields such as guide titles, tab names, and guide descriptions are calculated across the entire dataset. These outputs establish a high-level summary of subject guides at academic libraries in

---

\(^1\) The term “librarian” is intended throughout to be inclusive of all library workers, regardless of job role. In practice, even as Springshare refers to LibGuide creators and editors as “librarians,” those responsible for guides at specific institutions range widely.
2020. The macroscopic view of subject-focused LibGuides also contributes to the LIS literature by focusing on how subject guides are actually organized, regardless of expectations or assumptions of how they should be organized. This viewpoint allows us to reconsider questions of the utility and audience of subject guides at academic libraries.

**Literature Review**

There is a substantial amount of literature about library subject guides and, specifically, about the use of LibGuides, the most-commonly-used platform for creating library guides. This literature review examines three aspects: best practices, guide structure and content analysis, and usability studies.

Regarding best practices, LibGuides allow for a lot of flexibility in how institutions and guide creators format and populate their guides. Librarians have recommended that libraries develop guidelines, best practices, or standards for guides so that users have an easier time understanding the information presented (Brown et al., 2018; Chen & Chen, 2015; Fritch & Pitts, 2016). Del Bosque & Morris (2021) surveyed LibGuide administrators and authors at institutions belonging to the Greater Western Library Alliance (GWLA) about the existence of standards or style guides. Only 14% said that there were no standards to follow. Despite a majority of respondents affirming the existence of standards, the respondents reported that enforcement was weak and only 50% stated they undertook any regular review of LibGuides. The primary reason respondents offered for not adhering to standards, updating their guides, or enforcing the standards was lack of time. Logan & Spence (2021) also surveyed academic libraries to learn whether libraries employed content guidelines to govern the creation and maintenance of their LibGuides, finding that 53% had content guidelines of some kind. Of those that had guidelines, 80% reflected page design, 73% naming conventions, another 73% navigation and information architecture, while 59% covered author profile boxes.

Several papers and many actual LibGuides highlight best practices for creating LibGuides, though there is not universal consensus about what those practices should be. Four studies suggest limiting the number of tabs, with Bergstrom-Lynch (2019) specifying a maximum of six, Lowe (2020) a maximum of seven, and Nicol (2020) a maximum of ten (Bergstrom-Lynch 2019; Goodsett, et al., 2020; Lowe, 2020; Nicol, 2020). Three sources recommend using concise, clear titles for the tabs, which Lowe (2020) indicates should be a maximum of three words long. DeJonghe (2020) and Nicol (2020) recommend that the first letter of each major word in tab titles should be capitalized, and another three sources suggest that guides should include a “best bets” or “top resources” box (Bergstrom-Lynch, 2019; Goodsett et al., 2020; Illinois Library, 2019). Some recommendations indicate that LibGuides should include a librarian profile box or librarian contact information and photograph (Goodsett et al., 2020; Nicol, 2020), and that friendly URLs be created for users to remember and find more easily (Illinois Library, 2019; Nicol, 2020). Finally, Bergstrom-Lynch (2019) and Fritch & Pitts (2016) outline several recommendations including standardizing navigation and vocabulary and conducting regular link-checking to prevent “link rot.” Fritch & Pitts (2016) further recommends standardizing how guides employ guide types, such as course, subject, and topic guide.

The variety of best practices and the different emphases on what is included in best practices is worth noting. Goodsett et al. (2020) acknowledge that “there is still not enough rigorous evidence about best usability design practices for research guides, and much of what does exist is specific to one
institution. We also found that some of the evidence was conflicting, so more research into those specific areas would be helpful” (p. 225).

The second aspect reviewed here is guide structure and content analysis. Most studies that have sought to summarize trends across guides have focused on particular disciplines or topics, such as engineering, theology, nursing, or geography. All of the content analysis studies included here employed manual qualitative methods so their datasets, ranging from 20 guides to 260 guides, are generally small samples. Many studies have examined various elements of LibGuides structures and content in different combinations. Two of the most common guide elements tabulated are the top resources (databases, journals, ebooks, websites, etc.) for specific disciplinary guides (Dougherty, 2013a; Dougherty 2013b; Furay, 2018; Horton, 2017; Johnson & Johnson, 2017; Osorio, 2014; Osorio, 2015; Stankus & Parker, 2012; Van Dyk, 2015) and the number of tabs (pages) that guides included (Dougherty, 2013a; Dougherty 2013b; Insua, 2018; Osorio, 2014; Stankus & Parker, 2012; Van Dyk, 2015). These studies reported that most guides had between five and eight tabs. There were outliers, however, ranging from zero tabs to up to 21. Some studies found that between 65% and 80% of the guides evaluated had a librarian profile box (Dougherty, 2013a; Dougherty 2013b; Stankus & Parker, 2012; Van Dyk, 2015), and others noted the presence of instructional videos or tutorials (Dougherty, 2013a; Insua, 2018; Stankus & Parker, 2012).

Researchers have analyzed column structure (Stankus & Parker, 2012; Van Dyk, 2015) and search box settings (Dougherty, 2013a; Dougherty 2013b). Osorio (2014) found that engineering guides included an average of 5.1 boxes. Jackson and Stacy-Bates (2016) found that the average number of links per guide at Association of Research Libraries (ARL) institutions varied across three disciplines - philosophy, journalism, and chemistry - ranging from 77 links per philosophy guide to 151 links per journalism guide. The same study noted a maximum of 1,117 links on a single guide, a minimum of seven links on another guide, and that the guides included a high number of links unique to their institution (i.e., they were not included on the other guides under analysis).

Stankus and Parker (2012) found that 68% of the LibGuides in their study included chat widgets. Another study examined the presence and length of resource descriptions, finding that 87% of respondents preferred using short descriptions, 11% included detailed descriptions, and 2% did not include resource descriptions at all (Chen & Chen, 2015). Furay (2018) examined language consistency across theater guides, noting that 73% of the guides were listed under the subject heading "theater/theatre", 3% were under “drama”, and 2% were under “performing arts”. Five percent of guides included the words “theater” or “drama”, and the remaining guides were filed under broader subject headings such as “fine arts”, “humanities”, or “arts”.

The third aspect to be reviewed here is usability studies. Usability studies have investigated how students interact with LibGuides with the intent of improving the guides based on student input. Griffin and Taylor (2018) utilized both LibGuides Statistics and Google Analytics to observe how users interacted with their guides. They found that the average LibGuide session duration was two minutes and eight seconds with an average of 1.86 pages viewed per session. They also noted differences in session duration and page views depending on where users came to LibGuides from. Most referrals to LibGuides came from their learning management system, and those users stayed for an average of 39 seconds, viewing 1.29 pages. Comparatively, when users came to LibGuides from the library’s website, the average length of time they stayed was three minutes, viewing an average of 2.79 pages.
Sonstey and DeJonghe (2013) found that the guides at their institution included between seven to 12 tabs, and their study participants expressed confusion about their utility, while others described the guides as “cluttered” and “busy” (p. 87), and others never noticed the tabs at all. Quintel (2016) found that 57% of their study participants did not explore the guide beyond the landing page. Conerton and Goldenstein (2017) found that some students never noticed guide tabs or took a bit of time before discovering and using them. They also found that students preferred shorter pages that required less scrolling.

Methods

To investigate how academic libraries construct and organize online guides to disciplinary resources, we first narrowed our scope to include the 131 U.S. institutions listed by the 2018 Carnegie Classification of Institutions of Higher Education as “R1: Doctoral Universities - Very high research activity.” The Carnegie classifications define R1 institutions as those that have “awarded at least 20 research/scholarship doctoral degrees and had at least $5 million in total research expenditures” (Center for Postsecondary Research, Indiana University, 2019). R1 institutions include academic libraries that are likely to serve researchers from a wide variety of backgrounds, from undergraduates to doctoral students, as well as faculty, staff, members of the public, and alumni. While practices associated with the organization of library resources at Master’s, Baccalaureate, Associate’s, Special Focus, and Tribal colleges are all of great interest, it was beneficial to limit the scope of this study to institutions with a similar variety of researcher needs.

A number of structural elements make LibGuides an excellent source of data about how academic libraries organize and present their resources to users. First, the platform provides a consistent infrastructure for web pages that allow data elements such as page titles, content links, and profile information to be programmatically downloaded from guides across institutions. Second, the market penetration of LibGuides at R1 institutions is phenomenal: of the list of 131 R1 institutions for which the authors sought subject guides, 124 of them (94.7%) had a public LibGuides site, a significant increase from a 2016 study in which 71% of 101 ARL member libraries were found to use the platform (Jackson & Stacy-Bates). LibGuides present a unique opportunity to programmatically collect data representing the organization of similar kinds of resources from a wide selection of academic libraries.

The “requests” and “Beautiful Soup” Python packages were used to scrape data from publicly accessible LibGuides HTML pages and from an Application Programming Interface (API) provided by LibGuides to enable machine-access (Reitz, 2015; Richardson, 2015). Before embarking on data collection, a fair use analysis was conducted to ensure that the research was legal and adhered to best practices related to the text mining of public web sites. Data were collected on October 15 and 16, 2020, with another round of collection for seven institutions on December 8 and 9, 2020, to correct for inaccuracies in the target subject guide URLs.

First, a base URL for each institution’s LibGuides site was collected by searching the LibGuides Community Site by institution name (Springshare, 2021). In some cases, multiple LibGuides instances for a single institution were retrieved, most often when a law or medical library had a LibGuides site separate from the institution’s main libraries. In those cases, only the LibGuides URL for an institution’s primary
library instance was retrieved, and data from the more specialized LibGuides pages (e.g., law or medical) were not collected. After querying the LibGuides Community site, it was determined that seven institutions did not use LibGuides, which narrowed the initial targets to 124 institutional guides.

The second step involved identifying and collecting the URLs for all of the subject guides at each of the 124 libraries. Springshare provides a “Guide Types” metadata field that guide creators can use to assign each guide any one of six possible values: General Purpose, Course Guide, Subject Guide, Topic Guide, Internal Guide, or Template Guide. How a particular library defines and utilizes these guide-type categories varies, but most libraries assign guides with Subject and/or Course guide-type designators. Each institution’s unique site ID was captured from its base URL, then the LibGuides API was queried by site ID to download a structured list of every guide where the guide_types filter was equal to “3” - the default for Subject Guides - at each of the 124 institutions.

Looking over the initial results it became evident that 10 sites either did not use the default Subject Guide type of 3 or used it too inconsistently to include their guides in the final analysis. Every institution with fewer than 50 subject guides was checked manually to determine whether or not those guides were representative of the actual subject guides available on their sites. Ten libraries were found to have no accessible listing of disciplinary/subject guides and were therefore dropped from the dataset. An additional seven LibGuides sites were found not to use the default guide_types value of 3 to represent their subject guides but had other accessible lists of disciplinary guides, either on a public webpage or by way of a different guide_types value set for disciplinary guides. The data collection for these seven institutions was updated on December 8 and 9, 2020, using the more accurate subject guide lists. After removing the 10 libraries with no viable subject guide listings, the dataset included 114 institutional LibGuides sites - 87.0% of the R1 institutions on the initial list - from which 12,837 subject guide URLs were gathered.

Another round of data verification involved manually examining guides that appeared to have no titles (LibGuides are required to have a title field), guides on which zero or very few links were found on the page, or guides with abnormally small page sizes (measured by the number of characters in the source code). Of those, 56 guides were determined to point to URLs that were either no longer publicly available or that redirected to a non-LibGuides site. Those were removed from the dataset, leading to a final dataset of 12,781 guides.

The final step of data collection involved scraping every subject guide page to collect key elements (see Table 1). While specific subject guides often include more than one webpage, also often referred to as “tabs” in LibGuides jargon, only content elements from the landing page for each guide were captured. In other words, if an Economics subject guide had four pages/tabs - Home, Books, Articles, and Data - only content from the Home page was collected. This approach suits the focus on practices related to subject resource organization on LibGuides but does not shed much light on the content that is

2 In cases where the institution was a medical school, such as University of Colorado Anschutz, the primary LibGuides instance captured may only reflect medical library content.
shared on these guides. The specific article databases, books, or data resources that are most common to all Economics guides, for example, are not necessarily present on the guide’s landing page.

Table 1: Data collected from each subject guide page.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Title of the guide.</td>
</tr>
<tr>
<td>Description</td>
<td>Description of guide contents.</td>
</tr>
<tr>
<td>Updated</td>
<td>Date the guide was last updated.</td>
</tr>
<tr>
<td>Tabs</td>
<td>Title and URL for every tab/page listed on the guide. Both “horizontal” and “vertical tabs” are included.</td>
</tr>
<tr>
<td>Boxes</td>
<td>Titles for sections of modular guide content that can contain various content types (e.g., database listings, free-text, RSS feeds).</td>
</tr>
<tr>
<td>Profiles</td>
<td>Names of library workers and their subject expertise. Listed in guide “profile” boxes, usually in the left or right hand columns of the guide.</td>
</tr>
<tr>
<td>Subjects</td>
<td>Controlled vocabulary “subject” tags that can be assigned by LibGuides editors to individual guides to provide subject navigation across the system.</td>
</tr>
<tr>
<td>Number of links</td>
<td>A count of the total number of links on each subject guide. Includes navigational links, headers, footers, etc.</td>
</tr>
</tbody>
</table>

The data elements reflected in Table 1 were collected from 12,781 subject guides and compiled into a single tabular “dataframe” using the Python Pandas package, which associated each guide with its parent institution, as well as its guide URL (Pandas Development Team, 2020). To examine practices across libraries, the authors undertook two analytic approaches. First, descriptive statistics were generated for quantitative aspects of specific data elements (e.g., the average number of tabs per guide). Second, lists of word and phrase frequencies from specific fields, such as guide titles, were compiled and compared.

The number of tabs, boxes, profiles, subjects, content links, and total links per guide were each assessed by generating the mean, standard deviation, maximum, minimum, and overall count for each using built-in Pandas methods. The same statistics were also aggregated by institution to examine how
practices differed between institutions. Statistics related to the character lengths of text strings in guide, tab, and box titles, as well as the description fields were also calculated. To interpret these descriptive statistics, small subsets of guides showing an atypically large or small number of links, tabs, and boxes, were viewed in a web browser. Several data elements from guides required different methods of analysis. The “last updated” dates for guides were converted using the Pandas `to_datetime` function, and then tabulated by year.

Term and field frequencies for text from the title, subject, description, tab and box name fields were generated using two different approaches. First, entire fields were cleaned and counted to enable the creation of lists of the most-frequent full guide titles, subject tags, tab and box names. Second, cleaned terms were extracted from the title, subject, description, tab and box fields, after which they were stemmed to allow for token frequencies to be generated separately. In both cases, the following cleaning steps were taken:

1. Terms were converted to all lower-case letters.
2. Punctuation and special characters were removed from the text strings.
3. Numerical digits in the range 0-9 were removed.
4. Stop words were removed from all term lists, using the English language stop word package from the Natural Language Toolkit (NLTK) corpus library (Bird et al., 2009).

The Python `Counter` module from the `collections` package was used to generate lists of the most-common full title, subject tag, tab name, and box name phrases. To create more meaningful counts of similar titles, additional words that were common to the title field - `home`, `guide`, and `library`, for example - were added to the stop word list and removed from those phrases. Guides named `Psychology: Home` and `Psychology Resources` in this case were both counted under the same title of `psychology`.

To address some of the shortcomings of not being able to accurately cluster similar titles, subject tags, box or tab names, which had slight variations in exact phrasing, a separate list of the most common terms in those fields were created, along with the terms from the description field. These terms were stemmed using the NLTK Snowball Stemmer to reduce words to their morphological roots and the original English NLTK stop word list was used (leaving in words like `home`, `guide`, and `library`). Given the titles `Health Sciences`, `Health`, and `Biological Science`, for example, `health` would be counted twice, `scienc` would be counted twice, and `biolog` counted once.

**Findings**

Of the 12,781 guides collected from 114 institutions, the mean number of subject guides per institution was 112.1, with a median of 92.0 (see Figure 1). The number of subject guides within one standard deviation from the mean ranged from 33 to 192. On the high end, 15 institutions had more than 192 subject guides, with a maximum of 555 subject guides found at a single institution. On the low end, only two institutions had fewer than 33 subject guides. The fewest guides from a single institution was two, though one of the guides utilized separate tabs to capture the subject-specific content for dozens of disciplines. To put this into context, as of April 2021, Springshare reported the two institutions with the highest number of guides (regardless of guide type) were the Digital Theological Library with 5,667
guides and Walden University with 5,637. The 100 institutions with the highest number of guides overall, had on average 1,334 guides each (Springshare, 2021).

Figure 1: Distribution of the number of subject guides collected per institution

Several institutions customized the HTML templates for LibGuides such that the authors were unable to retrieve title fields for 255 guides (1.9% of the dataset). Of the remaining 12,526 titles, the mean length was 30.6 characters, with a maximum title length of 152 (Japanese Studies Databases for Researchers and Graduate Students @ Pitt: Resources for Japan Studies While You Are Teaching and Learning from Home: Home) and a minimum of three (Art).

The most-common subject guide titles, after cleaning the text strings as described in the Methods section, were psychology and sociology (84 each, representing 73.7% of institutions), followed by philosophy (81, 71.1%), music (78, 68.4%), and anthropology (76, 66.7%) (see Table 2). Because the counts are based on exact string matches on the cleaned titles, they tend to favor disciplines with fewer natural subdisciplines: engineering guides, for example, are usually not titled simply Engineering, but more frequently Mechanical Engineering, Chemical Engineering, or Biomedical Engineering. This manner of counting also likely favors single words or other short titles. Still, this list highlights how commonly R1 institutions create and provide guides to many of the same subjects. All of the subject guide titles listed in Table 2 appear at over 50% of the 114 institutions under study.
Table 2: The most common subject guide titles

<table>
<thead>
<tr>
<th>Guide titles</th>
<th>Count</th>
<th>Percentage of institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>psychology</td>
<td>84</td>
<td>73.7%</td>
</tr>
<tr>
<td>sociology</td>
<td>84</td>
<td>73.7%</td>
</tr>
<tr>
<td>philosophy</td>
<td>81</td>
<td>71.1%</td>
</tr>
<tr>
<td>music</td>
<td>78</td>
<td>68.4%</td>
</tr>
<tr>
<td>anthropology</td>
<td>76</td>
<td>66.7%</td>
</tr>
<tr>
<td>chemistry</td>
<td>73</td>
<td>64.0%</td>
</tr>
<tr>
<td>economics</td>
<td>71</td>
<td>62.3%</td>
</tr>
<tr>
<td>education</td>
<td>70</td>
<td>61.4%</td>
</tr>
<tr>
<td>history</td>
<td>67</td>
<td>58.8%</td>
</tr>
<tr>
<td>political science</td>
<td>64</td>
<td>56.1%</td>
</tr>
<tr>
<td>mathematics</td>
<td>60</td>
<td>52.6%</td>
</tr>
</tbody>
</table>

Table 2 percentages refer to the proportion of each guide title relative to the 114 institutional sites under review.

The most-common stemmed title terms (see Table 3) highlight a number of terms unrelated to subject disciplines. The top term, home (2,832, representing 6.5% of 43,417 words collected from the title field), is an artifact of the inclusion of the name of the landing page in the title of guides (e.g., Psychology: Home). Other common terms refer to the context and focus of the guides themselves: research (1,320, 3.0%), guid (1,290, 3.0%), resourc (1,201, 2.8%), and librari (530, 1.2%). Guide titles appear to highlight the library as an institution or place (librari) less frequently than they point to the activity of the users (research), or the general items described on the guide (resourc). A few key active verbs also show up here - start (494, 1.1%), get (311, 0.7%), and find (286, 0.7%) - suggesting that these guides are commonly framed as introductory materials (start) aimed at assisting users to connect to (get and find) research materials.
LibGuides include an optional description field that is displayed directly under a guide title. A total of 6,550 guides (51.2%) used the description field, while 6,231 guides (48.8%) left the field empty. The LibGuides description field is limited to 255 characters, and the shortest description included was only three characters long ("ESL"). The mean number of characters in the description field for all guides was 97.3 characters.

The most-common stemmed terms from the description field (see Table 3) largely echo the context terms found in guide titles - guid (3,531, representing 5.8% of the 60,773 words from the description field), resourc (3,303, 5.4%), research (2,883, 4.7%), and librari (1,402, 2.3%) - but also include a longer list of related contextual terminology such as inform (839, 1.4%), sourc (536, 0.9%), univers (494, 0.8%), topic (401, 0.7%), and collect (394, 0.6%). Of all of the textual fields analyzed, the description field provides the only instance in which a potential guide user, student (764, 1.3%), appears to be frequently mentioned.

Each guide on the LibGuides platform can include multiple pages, which are often also referred to as “tabs.” Horizontal or vertical tabs are the default navigational options that allow users to browse from one page to the next in a LibGuide and no differentiation was made in the data collection between horizontal or vertical tabs. Some tab links redirect users to other guides or websites, and so do not necessarily point to pages that belong to the guide on which they appear. While the name and URL were collected for every tab on a subject guide, no determination was made regarding whether a webpage referred to in a tab belonged to the guide or not. This study uses the term tab quite literally, then, and it should not be understood to be synonymous with a guide page.

The mean number of tabs per guide across institutions was 8.4, with the standard deviation ranging from 1.5 to 15.2. The mean in this case is higher than the maximum number of six tabs suggested by Bergstrom-Lynch (2019) and seven tabs by Lowe (2020), with 15.2 at the high end of the standard deviation being even higher than the maximum of 10 recommended by Nicol (2020). Eight hundred and sixty-three subject guides, 6.8% of those evaluated, had no tabs at all, while the vast majority, 93.2%, included at least one (in addition to the landing page itself). Thirty-five subject guides featured more than 50 tabs, and the highest number of tabs on a single guide was 144. This is a far higher number of tabs than previous research had identified, when a guide was found with a maximum of 21 tabs (Osorio, 2014), and is over six standard deviations from the mean. In almost all cases of subject guides that had more than 50 tabs, a LibGuides feature that allows for “sub-pages” to display as dropdown menus from a tab title was utilized. A Music guide with a primary tab for Piano, for example, provided sub-page tabs for material types such as Bibliographies, Biographies, Discographies, and Encyclopedias/Dictionaries. In another common use of these hierarchical dropdown tabs, a Government Documents guide was organized alphabetically by location on the top-level (e.g., Places A-D, Places E-H...) and then broke down pages by country name in tab dropdown menus (e.g., Afghanistan, Albania, Algeria...).

The mean number of characters per tab was 24.2, with a mean of 3.4 words per tab, which is above the maximum word length of three words recommended by Lowe (2020). In fact, 17,982 tabs included more than three words, with a high count of 19 words in a single tab: Featured Resource:
Virtual Tour of Becoming Jane: The Evolution of Dr. Jane Goodall exhibit at the National Geographic Museum.

The most-common tab name (see Table 4), home (6,043, representing 5.7% of the 106,817 tabs collected), is the default title for a primary page on all guides, so it comes as little surprise. Seven of the 10 most-common tab names, however, refer to resources by format or type: (find) books, (find) articles, databases, websites, and journals. Along similar lines, the most-common stemmed terms found in the names of tabs (see Table 3) relate to formats, both specifically, as in book (6,963, 3.0% of the 235,749 words in tab names), article (5,998, 2.5%), databases (3,805, 1.6%), journal (3,339, 1.4%), data (2,605, 1.1%), refer (1,970, 0.8%), and statist (1,862, 0.8%), and generically, as in resource (6,642, 2.8%), source (5,195, 2.2%), inform (1,982, 0.8%), and collect (1,862, 0.8%). While format or media type terms also appear in the titles of boxes, these are both more common and appear more granularly in tab names than in any other text field under scrutiny.

[Insert Table 4 – see Appendix]

LibGuides boxes are sub-units of a page that contain various kinds of content. They are most-commonly represented visually as rectangular boxes on the page with their own headings. Within a box, an editor can add text, links, databases, book covers, RSS feeds, and many other kinds of content. A box is essentially a generic unit of content on a LibGuides page, one which can be organizationally useful since it can easily be moved, duplicated on other guides, or deleted. Each box, by default, includes a title as well as a section for content.

The mean number of boxes per subject guide landing page was 4.6, with a standard deviation range of 1.1 to 8.0 boxes per guide. Many of the subject guides with the most links also appear in the list of those with the highest number of boxes, perhaps unsurprisingly. The most boxes found on a single guide page was 143, which is 41 standard deviations from the mean, and seven other guides included more than 50 boxes each. Broken down by institution, no single site had a box-per-guide average higher than 8.2, or lower than 1.4, suggesting that institutions generally did not significantly deviate from the norm in terms of the number of boxes typically present, but that individual guide creators did.

Many of the same common terms from titles, tabs, and descriptions appear in the names of boxes (see Table 3). The frequent occurrence of contextual terminology such as guid (6,003, representing 4.0% of the 150,664 terms appearing in box titles), librarian (5,140, 3.4%), resource (3,923, 2.6%), research (3,230, 2.1%), and databases (2,404, 1.6%), suggests that boxes are used to organize many of the same kinds of content seen in tabs and titles. One notable exception is the single most-common stemmed term from box titles, librarian (7,507, 5.0%), which does not appear in the 20 most-common terms from other text fields. Key terms such as help (2,093, 1.4%) and ask (1,384, 0.9%) are also present here, which, together with librarian, point to the common presence of profile boxes, seen in the list of most-common full box titles in Table 4 as librarian (1,500, representing 2.6% of 58,380 boxes), your librarian (662, 1.1%), ask a librarian (524, 0.9%), and subject librarian (467, 0.8%). Where guide titles focus on organization by subject and tabs by format types, boxes are often used to highlight the availability of expert help. Among the top overall box titles, quick links appeared 380 times (0.7%), best bets 236 (0.4%), and key resources 249 (0.4%), signaling some level of adherence to the recommended best practice of including a box to highlight top resources (Goodsett et al., 2020; Illinois Library, 2019).
LibGuides editors are able to assign each guide with any number of controlled-vocabulary terms representing the “subjects” of the guide, commonly referred to as subject tags. A guide with a title *Chinese History*, for example, might be assigned multiple subject tags: *Chinese*, *History*, and *East Asian*. When a user selects a subject tag link on a specific guide, they will be taken to a listing of other guides (and sometimes “experts” and databases) on the platform with the same subject tag. Institutions are able to define the subject tags that are available for their editors to apply to guides.

A total of 9,998 guides (78.2%) included at least one subject tag, while 2,783 guides (21.8%) had none. The overall mean of subject tags per guide was 1.5, with a standard deviation ranging from 0.0 to 3.2. On the high end, 10 guides included 20 or more subject tags each, with the highest subject tag count being 44 for a single guide. The highest mean number of subject tags per guide by institution was 7.0 tags per guide. This was from an institution that only had two subject guides in the dataset, one of which was a “General” guide with 13 subject tags, while the other guide used only one tag. Excluding this outlier, institutional means were comparable to the standard deviation for guides overall, as the highest mean number of subject tags per guide by institution otherwise was 4.5.

The list of the most-common subject tags (see Table 4) overlaps in some cases with the most-common guide titles: *education*, *music*, *history*, and *political science* all appear on both lists. But *business* (531 tags, representing 2.8% of the 18,642 subject tags collected), *engineering* (387, 2.1%), *social sciences* (291, 1.6%), *law* (252, 1.4%), *medicine* (227, 1.2%), and *health sciences* (213, 1.1%) are all unique to the subject tags list. In each of these cases, the subjects named are broader formulations of disciplines than many of those found in the most-common subject guide titles. *Social sciences* appears as a common subject tag, for example, while specific disciplines such as *sociology*, *anthropology*, and *economics* appear in the top titles. This suggests that institutions are using the subject tag feature to categorize guides with more-specific subject content under broader headings.

The list of most-common stemmed terms from the subject tag field found in Table 3 shows significant overlap with the most-common subject tag phrase list in Table 4. The stemmed terms generally highlight disciplinary terms that are common across fields: *studi* (2,505, representing 7.0% of the 36,028 terms in subject tags), *social* (580, 1.6%), and *public* (545, 1.5%), for example. The relative frequency of broader subject fields is better represented by this list of stemmed terms, however, as top terms could be read as reflections of clusters of disciplinary activity: *scient* (2,468, 6.9%), *histori* (1,172, 3.3%), *health* (942, 2.6%), *art* (906, 2.5%), *engin* (864, 2.4%), *busi* (832, 2.3%), *educ* (726, 2.0%) and *literatur* (628, 1.7%).

The overall number of links per subject guide page - including links from the site navigation, headers, and footers, as well as links within the subject guide content itself - were counted. The mean number of links per guide landing page was 77.0, with a range of 24 to 130 falling within one standard deviation of the mean. The higher end of the range, however, includes guides with a phenomenally high number of links: the maximum for a single guide was 2,033 links - nearly 37 standard deviations from the mean - while another guide included 1,123 links, and seven other guides had more than 500. Of the 20 guides with the highest number of links, eight were from a single institution, though it appears that only two LibGuide creators at the institution were responsible for the guides with the highest link counts. Accordingly, when calculating the mean number of links by institution, the same institution averaged only 72.1 links per guide, suggesting that the guides with especially high link counts were not
representative of institution-wide practices. There seems to be little disciplinary alignment among the guides that displayed a particularly high number of links: other than Legal Studies - for which there were six guides in the top 20, all of which were from the same institution - only two other subject areas appeared more than once, Archives and Health/Medicine. Otherwise, specific guides from a smattering of disciplinary and topic areas including Anthropology, Chemistry, East Asian Studies, Economics, Engineering, Government Documents, and the History of Math all appeared on the list of guides with the highest number of links. The highest mean number of links, when broken down by institution, was 268.3 links per guide, while the low was an average of 34.8 links.

A Profile box is a type of box that displays information such as the name, photo, and email address of a LibGuides editor, most commonly the owner of the subject guide on which it is displayed. Profiles are derived from the accounts of those creating LibGuides at a specific institution and are often included as a contact point for users who would like to get help. A total of 9,075 guides (71.0%) included at least one profile, while 3,706 (29.0%) included none, in line with previous research reporting a range of 65% to 80% of guides having a profile box (Dougherty, 2013a; Dougherty 2013b; Stankus & Parker, 2012; Van Dyk, 2015). The mean number of profile boxes per guide was 0.8, with a narrow standard deviation range from 0.2 to 1.3. Of the guides that included a profile, 8,686 guides (95.7%) included one profile box, while 412 (4.5%) included two profiles, 29 (0.3%) included three, three (0.03%) guides included four, and a single guide included six profile boxes. Subject guides are most commonly associated with a single person, and it seems to be quite uncommon for more than one individual to be associated with any single subject guide.

Some profiles turn up frequently at a single institution: a shared profile for one institution’s “Law Librarians” appeared 61 times while the most-common profile for an individual showed up on 39 guides. On the other hand, 781 profiles appeared on only a single guide. It is important to note that guide authors need not include their profile, and so these figures do not reflect the number of guides created by specific individuals.3

Most subject guides include a listing of when a guide was last updated, though LibGuides administrators can remove this from a guide’s template. This study captured dates for 11,702 of the guides in the dataset (91.6%), a higher percentage than previous research in which 79% of guides from a smaller sample included last-updated dates (Jackson & Stacy-Bates, 2016). These dates reflect any change that is saved to the guide, whether it is a minor edit of a typo or a major overhaul of the guide content. Broken down by year, 90.5% of the guides for which dates were available (10,591) were updated in 2020 - the year that the data was collected - while 5.8% (682 guides) were updated in 2019, and 2.1% (240 guides) were updated in 2018. Only 1.6% of guides were shown to have been updated between 2015

---

3 The LibGuides Community site tracks the number of guides, of any type, created by each author across the entire platform. The highest number of guides created by a single LibGuides user as of April 8, 2021, was 4,875 guides, the equivalent of creating one new guide every day for over 13 years. Fewer than 200 of these guides are available publicly, however, which makes it difficult to ascertain why or how the vast majority of the guides were created.
Discussion

Ninety-five percent of R1 libraries use the LibGuides platform, and on average they have published 112 subject guides each. Considering that in 2012 63% of ARL members used LibGuides (Gaphery & White), and in 2016 71% of ARL member libraries used the platform (Jackson & Stacy-Bates), the data in this study show a tremendous uptake of LibGuides at academic libraries in the United States in the last decade. This suggests both a striking market dominance by Springshare, as well as a remarkable amount of labor expended by library workers in the creation of subject guides. The LibGuides platform is clearly an important tool for academic libraries in organizing their disciplinary resources online. And while previous research demonstrates that the prevalence of dead links is a non-trivial issue for library guides (Jackson & Stacy-Bates, 2016), 90.5% of the guides in this study were updated within the current year, suggesting that they are usually maintained at least on some minimal level.

Looking at the practices evident across these nearly 13,000 subject guides, the “normal” range of behavior often aligns with best practices, but a significant number of guides across institutions represent outliers to those norms. Guide pages included a mean of 77.0 links and 4.6 boxes, while outlier guides were found with 2,033 links and 143 boxes, respectively. Also, the outliers were not without company: nine guides each had more than 500 links, and eight guides each included more than 50 boxes.

The use of tabs represents one area that deviates significantly from the recommended best practices, which might help to unpack a possible cause of other outliers. The mean number of 8.4 tabs per guide is above some recommended maximums (Bergstrom-Lynch, 2019; Lowe, 2020), though lower than Nicol’s (2020) recommended maximum of 10, while a substantial number of guides include over 50 tabs. The heavy reliance on tabs seems to suggest that librarians are either unaware of, or simply ignoring, best practices. Usability literature shows that users are confused by high numbers of tabs, finding them “cluttered” and overwhelming (Sonstey & DeJonghe, 2013), and that the majority of users do not navigate beyond a guide’s landing page (Quintel, 2016). Why then do LibGuides still include so many tabs? One possibility is that there are simply so many relevant subject resources available at R1 institutions that librarians have little alternative but to organize them into subcategories. A subject specialist could create one extremely long guide page instead - lacking the navigation and organizational advantages of tabs - or they could leave potentially relevant resources out of their subject guides. But neither of these alternatives is particularly appealing, especially considering that leaving relevant resources off of subject guides could leave users with no other way to discover niche subscription databases and websites associated with a discipline. The idea that librarians are attempting to wrangle vast numbers of subject resources is supported by the high number of links - 77.0 on average - per guide page. Extrapolating from the mean, it can be estimated that an average LibGuide, with 8.4 tabs and 77.0 links per page, has 646.8 links. That is an enormous amount of content for a library subject specialist to track. While a user is unlikely to be able to browse such a substantial directory effectively, it is also unclear how else a subject specialist can keep track of so many resources. Indeed, librarians have been reported to observe “I feel like I’m creating LibGuides for me. I feel like my students don’t look at them” (Brown et al., 2018), and "after years of experience, I’m convinced that only librarians use LibGuides" (Del Bosque & Morris, 2021, p. 16). These sentiments align with Griffin and Taylor’s (2018) findings that users
landing on a subject guide from the library website stayed considerably longer on the subject guide than did users navigating from elsewhere, such as a learning management system or a web search. The users navigating from the library website may be librarians rather than students.

Another way to approach the question of why high numbers of tabs are so common is to look at how they are used. First, tabs are clearly preferred for organizational purposes to boxes, which appear less frequently than tabs: on average, only 4.6 boxes per guide landing page. Based on the high prevalence of terms such as books, articles, databases and journals in tab names, it is clear that tabs are commonly used to organize resources by format/type. This maps to how vendors often package and sell resources to libraries, and to how many librarians conceptualize collections (e.g., ebook packages, scholarly article databases). But there is little evidence that this organization by type is helpful to library users, especially those with less research experience. Still, the format of resources appears to provide a fundamental organizing principle for subject guides at R1 libraries. One could read this as further evidence that a primary audience for subject guides is often subject guide creators and their library colleagues, and not, as is more commonly assumed, beginning researchers.

Unsurprisingly, guide titles are used primarily to name guides according to discipline. Looking at those titles, there does not appear to be a preponderance of guides for any specific subject or broader subject area such as the sciences or arts. Of the 11 guide titles present at over 50% of the institutions, only two, chemistry and mathematics, were from the sciences, but the list of most common terms from titles included science (934), engin (529), and health (407), along with other science-related terms. It is also evident that there is significant overlap in subject content covered from institution to institution. Eleven of the exact same guide titles were present at over 50% of the institutions. Eighty-four of 114 institutions (73.7%) included guides named Psychology and Sociology, for example, suggesting there is substantial overlap in broad subject coverage. While measuring the overlap of the actual content included in each guide was beyond the scope of this study - and an area ripe for future research - it appears that there is significant redundancy in subject guides from institution to institution, and likely opportunities for libraries to more closely collaborate and share guide content.

Limitations and future directions

Using a computational and quantitative approach to look from a distance at so many subject guides prioritizes the general over the specific and, by design, flattens the landscape of a rich and varied corpus of unique guides into a series of statistics. While those statistics provide a fresh perspective from which to consider the organization of subject resources at R1 institutions, they are not intended to replace close and careful readings of disciplinary guides. There are other limitations to the study. First, the data provides a somewhat skewed representation of subjects across U.S. institutions, underrepresenting law and medical library subject guides, many of which were not collected for institutions with multiple LibGuides instances. Second, the application of the subject guide type is inconsistent from institution to institution and provides a less than ideal hook for collecting disciplinary guides. It is likely the case that there are both many relevant subject guides from these institutions which were not included, and that some small number of guides (course or topic guides) that were included are not actually subject guides. Third, the authors did not collect sub-pages from each guide, but only the guide’s landing page. It is worth noting that in a study of user behavior on LibGuides at the University of South Florida, Griffin and Taylor (2018) found that “the vast majority of users… visit only one page, almost always the guide’s home page”
(p. 152). But in terms of organizational practices, there may also be significant differences observable by a deeper dive into the full content of these subject guides.

There is a dearth of research on the intentions, assumptions, and strategies of librarians and others who are using LibGuides to organize subject resources online. There is an often-explicit assumption in the literature on LibGuides that the primary intended audiences for LibGuides are the students, staff and faculty at the institutions where they are created. The high number of resources shared on those guides and the difficulty that novice users have in using them effectively, however, suggests that guides are more often by and for expert users. More research is needed to understand why this might be the case. Are librarians designing LibGuides for themselves? Are they doing so because they need a tool to help them keep track of resources in an environment of information overload? Rather than continuing to study how only students use LibGuides, there is a need for usability studies focused on how librarians and library workers engage with their institutional guides.

Further, the methods employed in this paper could be applied to greatly expand the scope of future content analyses of LibGuides. Rather than sampling dozens or hundreds of disciplinary library guides to track the inclusion of specific kinds of content such as academic databases, web scraping and text cleaning approaches could be employed to compile and analyze the content that is shared across thousands of institutions and hundreds of thousands of guides.

**Conclusion**

There is a popular meme based on a cartoon by illustrator K.C. Green that depicts a dog in a burning house, sometimes accompanied by a second panel in which the dog states “This is fine” (Cavna, 2021). The first panel, of the dog smiling with a cup of coffee while flames engulf the room, has circulated widely on social media among librarians with the text “Why don’t we make a LibGuide?” The joke - that librarians are so enamored of LibGuides that even as the proverbial world burns, they can think of no better action to take than to create a LibGuide about it - reflects a frustrating puzzle that many in libraries sense. Librarians spend a lot of time creating LibGuides, but for whom? The data collected here certainly confirm that U.S. academic librarians create a massive number of subject guides, and they do so with a striking amount of redundancy across institutions. How different, really, are the 84 subject guides from R1 institutions titled Psychology? And are the differences significant enough for each of these institutions to continue to invest the time and resources necessary to develop and maintain guides to the same disciplinary topic, ostensibly for similar audiences of students, staff, and faculty researchers? Perhaps, however, the organizational practices that this study observed across these guides - the high number of resources shared overall and their reliance on expert understandings of resources (such as format type) for navigation - suggest that LibGuides’ raison d’être is not, in fact, to help novice researchers but instead to help experts organize a complex universe of information resources for themselves. If the act of creating and maintaining a subject guide is the primary means by which a librarian can come to understand the resources for a field and serves as an intermediary tool for helping others, a certain amount of redundancy across institutions may be a necessary cost. However, librarians should consider the actual audiences for their LibGuides and develop best practices more intentionally with those audiences in mind. Whatever the case may be, the sheer volume of LibGuides suggests that it remains an indispensable platform for academic libraries to organize knowledge.
Acknowledgements

The authors would like to thank Kate McManus and Tiffany Carlson at the University of Minnesota-Twin Cities, for providing administrator access to the platform to assist with construction of LibGuides API queries.

Data availability

Derived data for this paper are available in the Data Repository for the University of Minnesota, https://doi.org/10.13020/AF1W-G563.

Notes on contributors

Cody Hennesy, MLIS, is the Journalism & Digital Media Librarian at the University of Minnesota-Twin Cities, where he focuses on library services and support for text data mining research and other emerging literacies in the digital humanities and computational social sciences.

Annis Lee Adams, MLIS, MA, is the E-Resources Librarian at California State University-East Bay. In addition to e-resources, her responsibilities include reference, instruction, liaison, and Libguide work, with special emphasis on user experience.

ORCID

Cody Hennesy, chennesy@umn.edu, https://orcid.org/0000-0002-9410-9810.

Annis Lee Adams, lee.adams@csueastbay.edu, https://orcid.org/0000-0003-0428-7793.

References


Cavna, M. (2021, May 18). The ‘This Is Fine’ dog is back. And his creator wants to show he’s more than a meme. The Washington Post. https://www.washingtonpost.com/arts-entertainment/2021/05/18/this-is-fine-dog-question-hound-kc-green/


Illinois Library. (2019, May 7). *Getting started with LibGuides: Best practices: Layout & content*. [https://guides.library.illinois.edu/libguides_intro/layout_content_design](https://guides.library.illinois.edu/libguides_intro/layout_content_design)


Richardson, L. (2015). *Beautiful Soup* (4.8.1) [Python].


<table>
<thead>
<tr>
<th>Guide titles</th>
<th>Subject tags</th>
<th>Tab names</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
<td>Count</td>
<td>%</td>
<td>Term</td>
</tr>
<tr>
<td>home 2832 6.5%</td>
<td>studi 2505 7.0%</td>
<td>librarian 7507 5.0%</td>
<td>find 7075 3.0%</td>
</tr>
<tr>
<td>studi 1343 3.1%</td>
<td>scienc 2468 6.9%</td>
<td>guid 6003 4.0%</td>
<td>book 6963 3.0%</td>
</tr>
<tr>
<td>research 1320 3.0%</td>
<td>histori 1172 3.3%</td>
<td>librari 5140 3.4%</td>
<td>resourc 6642 2.8%</td>
</tr>
<tr>
<td>guid 1290 3.0%</td>
<td>health 942 2.6%</td>
<td>resourc 3923 2.6%</td>
<td>home 6375 2.7%</td>
</tr>
<tr>
<td>resourc 1201 2.8%</td>
<td>art 906 2.5%</td>
<td>research 3230 2.1%</td>
<td>articl 5998 2.5%</td>
</tr>
<tr>
<td>scienc 934 2.2%</td>
<td>engin 864 2.4%</td>
<td>databas 2404 1.6%</td>
<td>sourc 5195 2.2%</td>
</tr>
<tr>
<td>histori 551 1.3%</td>
<td>busi 832 2.3%</td>
<td>subject 2326 1.5%</td>
<td>research 4405 1.9%</td>
</tr>
<tr>
<td>librari 530 1.2%</td>
<td>educ 726 2.0%</td>
<td>welcom 2170 1.4%</td>
<td>databas 3805 1.6%</td>
</tr>
<tr>
<td>engin 529 1.2%</td>
<td>literatur 628 1.7%</td>
<td>help 2093 1.4%</td>
<td>journal 3339 1.4%</td>
</tr>
<tr>
<td>start 494 1.1%</td>
<td>social 580 1.6%</td>
<td>relat 2052 1.4%</td>
<td>librari 2628 1.1%</td>
</tr>
<tr>
<td>health 407 0.9%</td>
<td>languag 579 1.6%</td>
<td>book 1787 1.2%</td>
<td>data 2605 1.1%</td>
</tr>
<tr>
<td>educ 393 0.9%</td>
<td>public 545 1.5%</td>
<td>search 1611 1.1%</td>
<td>cite 2564 1.1%</td>
</tr>
<tr>
<td>art 371 0.9%</td>
<td>law 482 1.3%</td>
<td>inform 1545 1.0%</td>
<td>guid 2541 1.1%</td>
</tr>
<tr>
<td>american 351 0.8%</td>
<td>medicin 472 1.3%</td>
<td>find 1479 1.0%</td>
<td>get 2424 1.0%</td>
</tr>
<tr>
<td>literatur 350 0.8%</td>
<td>govern 444 1.2%</td>
<td>get 1438 1.0%</td>
<td>start 2194 0.9%</td>
</tr>
<tr>
<td>inform 317 0.7%</td>
<td>american 443 1.2%</td>
<td>articl 1437 1.0%</td>
<td>inform 1982 0.8%</td>
</tr>
<tr>
<td>get 311 0.7%</td>
<td>human 438 1.2%</td>
<td>start 1424 0.9%</td>
<td>refer 1970 0.8%</td>
</tr>
<tr>
<td>music 290 0.7%</td>
<td>biolog 396 1.1%</td>
<td>scienc 1423 0.9%</td>
<td>statist 1862 0.8%</td>
</tr>
</tbody>
</table>
Table 3 percentages represent the proportion of each stemmed term in relation to the total number of non-unique words from each field, after the removal of stop words. The total number of terms in each field are: 43,417 guide title terms, 36,028 subject tag terms, 150,664 box title terms, 235,749 tab name terms, and 60,773 description words.

Table 4: The most common full phrases from LibGuides box names, tab names, and subject tags

<table>
<thead>
<tr>
<th>Tab names</th>
<th>Box names</th>
<th>Subject tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrase</td>
<td>Count</td>
<td>Percentage</td>
</tr>
<tr>
<td>home</td>
<td>6,043</td>
<td>5.7%</td>
</tr>
<tr>
<td>books</td>
<td>1,853</td>
<td>1.7%</td>
</tr>
<tr>
<td>articles</td>
<td>1,167</td>
<td>1.1%</td>
</tr>
<tr>
<td>find articles</td>
<td>1,099</td>
<td>1.0%</td>
</tr>
<tr>
<td>getting started</td>
<td>1,057</td>
<td>1.0%</td>
</tr>
<tr>
<td>find books</td>
<td>837</td>
<td>0.8%</td>
</tr>
<tr>
<td>databases</td>
<td>688</td>
<td>0.6%</td>
</tr>
<tr>
<td>websites</td>
<td>661</td>
<td>0.6%</td>
</tr>
<tr>
<td>citing sources</td>
<td>610</td>
<td>0.6%</td>
</tr>
<tr>
<td>journals</td>
<td>594</td>
<td>0.6%</td>
</tr>
</tbody>
</table>

Table 4 percentages represent the proportion of each field cleaned phrase relative to the overall number of tab titles (106,817), box titles (58,380), and subject tags (18,642).