Introduction
Library guides have a long history and prominent place in traditional library service. Librarians put forth tremendous effort in researching, compiling, and formatting research guides to help users navigate complex and vast library resources. Yet, despite their potential value, we know that library guides are generally underutilized by our patrons. In an effort to remedy this, many libraries conduct usability and/or user studies to improve the relevance and value of library guides to users. While many studies focus on findability, access, guide navigation and terminology for increasing usage, other studies suggest developing focused course guides rather than general subject guides.

Historically, library guides were primarily available as print bibliographies. Recently, studies have begun to enumerate the widespread integration of electronic guides on library websites as tools for facilitating the research process. Further evidence of this is found in a recent study by Ghaphery and White; this study documents that out of 99 ARL libraries, two thirds utilized the LibGuides platform (developed by Springshare) as a content creation and delivery tool for these guides.

At Cornell University, LibGuides have been enthusiastically adopted by librarians with over 1894 library guides created since its launch on campus in January 2008. Librarians heavily rely on LibGuides to promote electronic and physical library resources to user groups across campus. In an online environment, these guides are an essential conduit to extensive resources available in the library’s collection. However, our own usability testing reveals that users are generally unaware of library guides and their relevance to their research.

Intuitively, LibGuide authors may suspect that their library guides are not heavily utilized, yet, data provided by Springshare suggests otherwise. Individual LibGuide statistics summarize page views or hits, widget and API hits, link hits, file hits, event hits, and, even, book hits. Despite the wealth of data provided by Springshare, what these statistics reveal about guide usage remains ambiguous. Given this ambiguity, LibGuide authors interpret the numbers as significant patron usage and quantitative support for the creation of more LibGuides. In fact, the number of Cornell University Library (CUL) LibGuides rose from 637 guides in February 2012 to 829 guides in February 2013; an increase of 23% in 12 months.

In an effort to extract unfiltered user-behavior that might provide more meaning on the value of LibGuides, we obtained the raw log file from Springshare for our instance of the tool (guides.library.cornell.edu) and loaded the data into our locally developed Web analytics tool.

In this paper, we present the analysis and use of raw log files used to contextualize and understand user access as a novel approach that complements both traditional usability testing of the LibGuides tool.
and the Springshare statistics that LibGuide authors obtain. Our inclusion of the user’s location into the analysis contrasts, for example, with the usage analysis approach by Smith7 who looked at the total counts for all users, or Staley8 who surveyed members of particular campus communities to discern more about their use of library guides. We anticipate that revealing patterns derived directly from user locations will allow us to make compelling recommendations for our academic library community to enhance the use and value of library guides for our patrons.

Method
Our analysis for this study consisted of two complementary approaches. First, we analyzed log data (see below) for 637 LibGuides available to us at the start of our study to identify aggregate use by location of the user. This essential first step was based on raw Apache server log data furnished by Springshare and allowed us to determine not only the traffic each guide received, but also the location of users—internationally, nationally, as well as on and off-campus. In order to illustrate in more depth the value of user location and affiliation, as well as explore these in contrast to statistics available to authors via Springshare, we pared down our 637 guides to a sample of 20 LibGuides9 (each guide numbered consecutively from 1 through 20). Our second approach thus consisted of selecting two LibGuides—one with high use session counts (LG#4, 760 session counts) and one with low session counts (LG#16, 380 session counts) from the 20 selected LibGuides for further analysis using Springshare statistics and logs data.

Logs Data Approach
We were prompted to make our data request to Springshare after reading Dana Ouellette’s10 article, “Subject Guides in Academic Libraries: A User-Centered Study of Uses and Perceptions.” Ouellette’s interview-based qualitative research is a valuable contribution to the literature and provides a clearer picture of what students find valuable about LibGuides (see also Staley11). What is missing from that picture though is students’ actual use of LibGuides. At Cornell, we have a locally developed Web analytics tool that provides us with information about who is using our locally hosted content. The distinguishing feature of our system is its ability to aggregate use by location of the user, including detailed on-campus reports that tell us, for example, if the patron is using a public workstation in a physical library, or if the patron is located in a residential hall, using wireless on campus, etc.12 Given this capacity, we wondered what could we learn about how our patrons actually use the hundreds of LibGuides we offer if we acquired the raw usage logs from Springshare and loaded them into our local Web analytics system.

In February 2012, we approached Springshare about acquiring the raw Apache server log data for our instance of LibGuides, guides.library.cornell.edu. To our knowledge, we are the first library to make such a request. Springshare graciously complied and generated a custom log data extract for us. Our study’s data set included four months of LibGuides use, November 2011—February 2012. The log file Springshare sent us includes the following data for each request: IP address, timestamp, requested page, server response code, referrer, user agent (browser type). The raw log data set we received also includes the terms (if there are any) typed by users to locate our LibGuides in search engines. The defining feature of our “bibliomining” tool, however, are its geo-coded usage reports based on IP address—within libraries on campus, across campus, and outside of campus.13

The next step in making use of the files was to create a list of Cornell LibGuides. We wrote a script to extract all the librarian page URLs from within that list and wrote another script to harvest each librarian’s LibGuide page, and recursively extract the guide URLs associated with the librarian page. Using this method, we compiled a list of 637 Cornell library guides for the study.

With the list in hand, and the Springshare file loaded into our Cornell bibliomining tool, our next task was to automate the extraction of a large number of usage reports. Given the data available to us and the way in which we extracted the list of Cornell LibGuides, we had to make the assumption that the primary URL that corresponds to the home page of the guide’s home page tab (e.g., http://guides.library.cornell.edu/nameofguide) is representative of the use of the whole guide. Given that these are the URLs that the librarian’s themselves use to link to the pages when they are being shared, we felt this was a reasonable assumption.

Springshare Data Approach
In order to more fully understand LibGuides at CUL,
we felt it was necessary to examine the logs data in context with the actual Springshare statistics provided to each LibGuide author. Using the sample of 20 guides (based on the log data analysis), we contacted the guide authors and requested Springshare statistics for their guides for November 2011 through February 2012. Compiling this data into a spreadsheet, we compared page views for the guides. Page view data were broken down by total page views, home page views, and subsequent tab and subtab page views. Once the Springshare statistics were compiled into a spreadsheet, the data could then be compared with the logs data from the raw Apache files.

Findings—Logs Reports and Springshare Data

Log Session Counts and Affiliation—The Broad View

Our analysis is based around the concept of sessions. In our system, a session starts the first time a unique IP address makes a request to one of the pages in the LibGuide. The session is closed when 30 minutes elapses without any clicks from the user. Basic session statistics across all 637 Cornell LibGuides studied are presented in table 1.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Session counts</th>
<th>No. of guides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sessions analyzed</td>
<td>41,814</td>
<td>637</td>
</tr>
<tr>
<td>Lowest use LibGuide</td>
<td>0</td>
<td>155</td>
</tr>
<tr>
<td>Highest use LibGuide</td>
<td>3,114</td>
<td>1</td>
</tr>
<tr>
<td>Average</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

The assumption here, based on the work of Nicholson and our own intuition is that we can make an inference between the location of the person using the system and their affiliation. Table 2 shows aggregate usage of our LibGuides across nine broad location types. One drawback in the logs data is our inability to distinguish between author editing of LibGuides and reference staff usage of LibGuides for the category of CU Lib Staff, but what we do know for sure is all of the sessions in this category were generated by Cornell affiliated users. All of the numbers are for sessions, not page hits.

Looking at sessions by inferred affiliation illustrates the distribution by percentage between these two categories. Some 69.9% (n=29,211 sessions) out of a total of 41,814 sessions were generated by users not affiliated with the university (figure 1). As an example, one of our most popular guides was accessed over 3000 times in a four month span, but less than 1% of that use was by on-campus patrons.
Another way of looking at the aggregate session location data is to categorize each individual LibGuide according to the percentage of campus (and Ithaca) activity. Out of 637 LibGuides, 155 had zero activity recorded during the four months of data collected for this study. The campus and Ithaca activity recorded for the remaining 482 LibGuides, and represented in figure 2, broken down by percent usage, illustrates that 244 LibGuides (that is, over 50% of those with activity) only enjoyed up to 20% of use. At the other end of the spectrum, the 31 LibGuides that garnered high traffic (80 -100%), represent less than 7% of LibGuides with activity.

Logs Geographic Location—The Close View

The aggregate location type data summarized above is for broad location type or affiliation categories. Taking a closer look at location reports for the two LibGuides selected from our sample—one of high affiliated session counts and the other low—shows striking differences in terms of user location. Starting with the low guide (LG #16), table 3 lists where the heaviest users are located. At the top of list is the Russian Federation. Ithaca, NY and Cornell campus show up as Cornell affiliated, but the rest of the traffic comes from places such as Atlanta, GA (USA), the Ukraine, and China.

Our contrasting example shows a guide with high Cornell activity (LG#4; table 4). The top three location groups are Cornell campus, Ithaca, NY, and library staff working in Cornell’s Olin library (this would either be reference staff or perhaps the LibGuide author(s)). The rest of the top ten are from outside Ithaca, NY, but interestingly, compared to the previous guide, all of this activity comes from users within the USA.

Comparing the location use for these two guides demonstrates that the audience for LG#16 (low affiliated use) is very different from the audience for LG#4. Clearly this data begs the question why and propelled us to find out the unique characteristics of LibGuides based on session counts by conferring with LibGuide authors. The results for that study will be forthcoming, but the importance of the point remains. Location data would benefit authors by helping them examine their

<table>
<thead>
<tr>
<th>Location</th>
<th>Session Counts</th>
<th>Affiliation Assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russian Federation</td>
<td>21</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>Ithaca, NY</td>
<td>15</td>
<td>Cornell</td>
</tr>
<tr>
<td>Atlanta, GA</td>
<td>15</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>Cornell Campus Wifi</td>
<td>13</td>
<td>Cornell</td>
</tr>
<tr>
<td>Germany</td>
<td>12</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>Ukraine</td>
<td>10</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>8</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>Los Angeles, CA</td>
<td>8</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>Palo Alto, CA</td>
<td>8</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>China</td>
<td>8</td>
<td>Non-Cornell</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location</th>
<th>Session Counts</th>
<th>Affiliation Assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cornell Campus Wifi</td>
<td>698</td>
<td>Cornell</td>
</tr>
<tr>
<td>Ithaca, NY</td>
<td>61</td>
<td>Cornell</td>
</tr>
<tr>
<td>Cornell Olin Library Staff</td>
<td>36</td>
<td>Cornell</td>
</tr>
<tr>
<td>New York, NY</td>
<td>23</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>Laurel, MD</td>
<td>20</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>Congers, NY</td>
<td>16</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>Elmsford, NY</td>
<td>9</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>Westborough, MA</td>
<td>6</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>Nashville, TN</td>
<td>4</td>
<td>Non-Cornell</td>
</tr>
<tr>
<td>Morganton, NC</td>
<td>4</td>
<td>Non-Cornell</td>
</tr>
</tbody>
</table>
guides closely and tailor their efforts to cater to their constituencies in a more effective way. This type of information is unavailable to individual guide authors and is impossible to extract from traditional usability studies or to infer from Springshare statistics.

**Springshare Data**

Despite statistics provided by Springshare, some authors may be uncertain about what the data implies about usage and value of their guides. The Springshare data seems to indicate, via quantified page views, that library guides are being accessed and, thereby, utilized by the intended audience, Cornell users. The goal in analyzing the Springshare data was to look for patterns in the statistics available to guide authors. In the absence of our more substantive logs data regarding user groups and location, the Springshare statistics are the only measure of access available to authors. We hope our interpretation of these figures might provide strategies for authors to utilize their guide statistics in a more meaningful way.

The Springshare data appears straightforward; it provides a quantifiable number of page views for each tab on a LibGuide. Table 5 outlines tab page views for our two example guides for our sample period of four months.

According to Springshare’s FAQ, page view statistics are collected each time the page is loaded with a full page load equaling one view; page views are not based on requests for files to the server. In addition, page views are broken down according to the following distinctions: *all* (the sum of view locations), *standard* (views by a standard desktop browser), *mobile* (views by a mobile browser), *API* (views that originate from clicking on a LibGuide API call) and *widget* (views that originate from clicks on a LibGuide widget). Springshare data provides page view data for all pages on a guide, pages being the individual tabs or subtabs. The home page is the default name for the first page or tab on a guide and is also the landing page for users when they access the guide. Although their data is varied, what does it really reveal about the use of an individual LibGuide? We contend that for the typical LibGuide author, not much.

In order for guide authors to clearly interpret the data, Springshare needs to further define page views or page hits. If page views are not a request to a web server for a file, then what actually constitutes page views? Is a user’s IP address considered in calculating the number of page views for a tab? Are multiple clicks by one IP address counted once or each time they are executed? Are bot hits—bots being software applications that run automated tasks over the Internet—excluded from the page views calculation? The ambiguity in the definition of page views translate into little understanding of the numbers provided by Springshare.

Taking page views into consideration, Springshare data does indicate that the home tab will experience the highest number of page views for a guide. In our sample of 20 guides, the total page views across all guides were 12,365 views. Of those 12,635 views, 6,508 views were to the home pages on these guides. These calculations reveal that 51% of the total page views were on the home page of a guide. As an example, figures 3 and 4 outline the breakdown of page views across our two sample guides.

Further, in our sample, the median for home tab page views as a percentage of total page view is 45% with a minimum of 27% and a maximum of 100%. This suggests that, for most guides, the majority of page views are to the home tab, thus leading us to infer that the most important content should be presented on a guide’s home tab.

In addition, we examined page views in relation to the number of tabs on a guide. For our sample of 20 LibGuides, the mean number of tabs per guide is 7.2; the median number of tabs per guide is 6 with a minimum of 1 tab and a maximum of 19 tabs.

It seems obvious that the number of tabs on a guide might influence the number of page views for each guide. To test if there was, indeed, a relation-

<table>
<thead>
<tr>
<th>TABLE 5</th>
<th>Springshare Total Page Views for Two Sample Guides with High (Lg4) and Low (Lg16) Session Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Page Views</td>
</tr>
<tr>
<td>LG4</td>
<td>1231</td>
</tr>
<tr>
<td>LG16</td>
<td>355</td>
</tr>
</tbody>
</table>
ship between page views and the number of tabs on a guide, we calculated the average page views per tab for each guide. Once we calculated this figure, we compared the mean for guides with tabs fewer than 6 (six is the median number of tabs on a guide) with the mean for guides with six or more tabs.

In this case, the mean for a total of seven LibGuides with fewer than 6 tabs garnered 217 average page views per tab. The mean for guides with 6 tabs or greater \( (n=13) \) was 78 average page views per tab. The results suggest that using fewer tabs on LibGuides might increase the page views for all tabs; a key point that bears further investigation with a larger sample size. If true, fewer tabs might increase the likelihood that users will access all of the content on a guide.

Discussion

The strength of bibliomining is evident when we look at the geo-coded sessions based on IP address and extract the location from which users access LibGuides. Assuming that user location is highly correlated with affiliation, the logs data indicates a high percentage of CUL LibGuide users are not located on campus or even within its vicinity, but rather, are unaffiliated users from around the globe. One of the most useful and surprising findings, is that our users are, in many cases, not the intended student or faculty audience for which the LibGuides were developed (see figure 1).

In our sample of 20 LibGuides, ten were course-related guides developed to support student research within a specific course. Unfortunately, because LibGuide authors do not have access to this type of location-specific user data, they can only infer that the Springshare data available to them represents usage by their intended audience. McMullin and Hutton provide an example of how LibGuide authors often project who is using their library guide based on who they think should be using the library guide. In a discussion on the efficacy of instruction, librarians are quoted stating, “While the numbers did drop in the following months [after the instruction session] (108 hits for October; 52 for November) they did remain higher than before the instruction sessions, indicating that students were perhaps returning to the guide for help with other assignments.”17 In other words, because the authors have only aggregate usage counts, they make the assumption that their intended audience for the LibGuide is the actual audience. In this case, they might have been correct that the instruction sessions increased use, but in reality they cannot be sure. We are certainly not implying that Springshare is deliberately misleading authors, but rather, suggesting that location data is useful in helping authors focus their efforts more effectively.

One could argue that Cornell affiliation is a narrow definition of intended audience. Clearly, most universities support more than just their on-campus or local students, faculty and staff. Distance education students, alumni, visiting scholars, and local, state and national visitors are all components of a library’s user base. Our logs data does not begin to identify our nuanced user groups, yet, our study does begin to address the need of understanding who is accessing our online resources. In our current economic climate, articulating the use of our services is essential.18 The
point is not to discourage access from outside of campus or internationally, but to enrich our content and target our energies to foster the use of tools which really serve our constituencies.

Another interesting finding from our study is that the logs home page session counts and the Springshare home page views are remarkably consistent across our sample of 20 LibGuides. While drawing from the same raw data, we use a very different metric in defining sessions than Springshare uses in defining their page views. Yet, despite this difference, the values for tracking activity closely parallel each other. Figure 5 illustrates this across our 20 LibGuides.

Figure 5 also illustrates that, while the total logs session views and the Springshare home page views are closely aligned, neither are a sufficient metric in measuring user affiliation. Only by looking at the logs Cornell affiliate session counts do we get a meaningful gauge of LibGuide usage by our intended audience. The top five affiliated-use LibGuides from our sample (LG#1-5), across the X axis, show a close relationship between total sessions, home page views, and Cornell affiliated session counts. In contrast, LibGuides on the opposite end of X axis show little relationship between total sessions or home page views and Cornell affiliated use.

The corroboration between logs data and Springshare statistics highlights another important finding in our data—the importance of the home tab (Tab 1) for our users. Between 45% and 84% of the traffic in 20 LibGuides is driven by the home tab, an impressive number that quickly dwindles with more tabs per guide. Staley’s 2007 study found that students generally favor the articles and databases pages in subject guides, an indication that users generally know what they are looking for and zero in on their need without exploring other pages. She also suggested that for over 60% of Nursing, Journalism and Mass Communication students, the home page was recognized as a starting point for research. The content that students recognize and expect could be combined or added to the first tab as a means of ensuring that they find and access to what they need promptly.

In contrasting the log data and the Springshare statistics, we would have liked to be able to contrast the tab and subtab page views with the tab and subtab log sessions. Unfortunately, the raw Apache data only included data for the home page views and we were, therefore, only able to draw comparisons between the Springshare home page views and the logs data session views for the home page. Ultimately, we do feel that there is great value in analyzing the relationship of page views to guide tabs.

Study Limitations
One limitation of our study is that we lacked a full semester’s data. Our findings may be somewhat skewed since the raw data we utilized included the months of

![FIGURE 5](image-url)
December and January, Cornell’s winter intersession period. A full semester or a year’s worth of logs data would provide a more realistic picture of the life-cycle of LibGuide access.

In addition, our analysis largely ignores the ephemeral nature of LibGuides. Springshare designed LibGuides to be easily edited and updated. This makes drawing comparisons between numbers difficult because it is impossible to determine if spikes in usage are due to marketing, content updates, or another unknown. For our study, we could only assume that the guide content was static when we evaluated guide traffic via page views and sessions, although we recognized that updates do occur.

Finally, our study only measures access of LibGuides; access clearly does not imply usage of library resources and services. Access is only a first measure in articulating the value of LibGuides among other resources within the Cornell University Library system.

Springshare’s willingness to share their data with us speaks of goodwill and a real openness to work with their clients. This is also evident in their newsletters and the added bells and whistles that LibGuides seem to be updated with often. Nonetheless, there is information that is not as easy to locate as we think, which could be more effectively tagged or available.

Conclusions

We started this analysis with two objectives in mind: to investigate LibGuide usage based on geographic location and to use that data to complement and contrast the statistics available to authors directly via Springshare. Both of these objectives sprung from a desire to assess the visibility and relevance of library guides to users and provide meaningful recommendations to the library staff whose efforts have created hundreds of guides.

In the process of analyzing geo-coded data on LibGuide usage it became apparent that we had more questions than answers. Knowing that over two thirds of the usage on approximately 640 LibGuides comes from non-affiliated users, is an eye-opening result. At one level, curiosity drives the analysis—where are these users coming from? What countries are represented? Why are there so many from international locations? At another level, the motivation to find out what is unique about guides with fantastic numerical activity and those with less activity, drives the research. We find ourselves at the second level, attempting to build on this initial eye-opening result to extract meaningful data by looking closely at content location, type of content, format of content and intended audience. Clearly, this study is an important first step in understanding our audiences better, and one where Springshare can make a difference. Librarians need more robust open statistics reports on the guides they create, building from IP address/location.

In our initial attempts at applying usability methods to understand LibGuides use, we found that most students had little knowledge of them or had used them but only minimally. Thus we embarked on this study aiming to complement and provide another facet to that approach. From that perspective the logs data lays the groundwork for future studies that can then grant attention to the nuances that speak to users. With that knowledge in mind, and the hard data from this study, we are beginning to understand that only a fraction of our LibGuides are benefiting Cornell users. And while interesting to see the particular location of groups using a particular guide, we did not find any intrinsic patterns, based on the log data alone that helped us to understand why some guides receive more campus use than others. Thus our future work is centered on detailing the unique characteristics of a subset of LibGuides, including information gained from the authors themselves. With the creation of LibGuides rising yearly, at least at CUL, we hope that this provides a meaningful foundation for furthering the discussion on services that reach and may impact our users.

Notes

6. Thanks to Springshare for sharing the data set we used for this study.
7. Charles Smith, “Meta-Assessment of Online Research


9. To do this, we sorted log data results (campus and Ithaca) by session counts from high to low and used that to identify natural breaks in the data, resulting in four sample sets of five LibGuides each (20 in total), from high session counts to low session counts. An in-depth analysis of the characteristics of each sample set including authors’ perspectives, forms part of a forthcoming analysis.


11. Staley, "Academic Subject Guides."


Bibliography


---


---


---


---
