To Pre-filter, or Not to Pre-filter, That Is the Query:

A Multi-Campus Big Data Study

Heather L. Cribbs [a] & Gabriel J. Gardner [b*]

This is an electronic version of an article published as Cribbs, H. L., & Gardner, G. J. (2022). To pre-filter, or not to pre-filter, that is the query: A multi-campus big data study. *Journal of Librarianship and Information Science*, AHEAD-OF-PRINT. https://doi.org/10.1177/09610006221124609 JoLIS is available online at: https://journals.sagepub.com/home/LIS

Author Note

[a] Systems Librarian, Robert E. Kennedy Library, California Polytechnic State University, San Luis Obispo, United States of America; ORCiD https://orcid.org/0000-0001-7526-9037

hcribbs@calpoly.edu

[b] Associate Librarian & Discovery Coordinator, University Library, California State University Long Beach, Long Beach, United States of America; ORCiD https://orcid.org/0000-0002-9996-5587

gabriel.gardner@csulb.edu

Abstract

Library discovery platforms, which provide searchable user interfaces as their front-facing layer, aggregate tremendous amounts of metadata from multiple data streams describing a wide variety of print and electronic resources. Complicating the matter further, resources may differ in availability or delivery time depending not only on their media but also upon the source of the data stream describing them. How should libraries structure end users' options for searching discovery platforms in light of the many options available? This study used a nonexperimental design and quantitative methods to analyze users' revealed preferences for query type in twenty-four academic libraries in a data set containing metadata, sans queries, for over 64 million searches. Libraries studied were all located in California, used the same discovery layer software, and served similar user and faculty constituencies; however, the number of query types and pre-filtering options available differed between institutions. Results show that, when users were presented with the choice between search options, most conducted simple, more broad searches rather than complex and specific searches. When search options were highly constrained by the default choice architecture, but complex searches were possible, few users opted out of the default simple search. Implications for usability of discovery layers and the motivations of librarians in choice architecture are nontrivial and are discussed. The desires of librarians and "power user" faculty must be balanced with the fact that most users are novices and users of all abilities are largely habituated to commercial search products which emphasize post-search results filtering.

Keywords: discovery, Primo, user search behavior, UX, academic libraries, consortia

To Pre-filter, or Not to Pre-filter, That Is the Query: A Multi-Campus Big Data Study

Library discovery platforms are highly customizable interfaces for facilitating access to a library's unique offerings of collections and services. Marshall Breeding (Breeding, 2014) defined library discovery platforms as "tools that search seamlessly across a wide range of local and remote content and provide relevance-ranked results". Sometimes referred to as discovery layers, these front-facing, patron gateways are used in tandem with resource management systems to provide a searchable user environment that aggregates tremendous amounts of metadata from multiple data streams describing a wide variety of print and electronic resources. Complicating the matter further, resources may differ in availability or delivery time depending not only on their media but also upon the source of the data stream describing them. These technologies were meant to be customized according to specific local needs. One consideration faced by website designers and system librarians is how to structure end users' options for searching.

In this multi-campus big data study, researchers explored quantitative analytics to address essential considerations for reducing the complexity of options available to the end-user when creating a search environment. Libraries were all located in California, used the same discovery layer software, and served similar user and faculty constituencies; they differed importantly in the number of query types and pre-filtering options available. Every discovery layer had designers and every deviation from default settings implied a decision on behalf of the designer(s) about how they expected users to engage with the system. This research explores the mismatch between the behavior of our users versus designers' choice architecture expectations of how our users behave within these systems. Heterogeneity in the sample provided ample opportunity for variation under different choice architectures to emerge; yet, as shown in the results, little did.

Discovery layers are inherently distinct from web search engines, "crawler-based web search engines (e.g., Google Scholar), for example, function differently from bibliographic databases which have a curated catalogue of information (e.g., Scopus). Some of these search systems are large and multidisciplinary, while others have a narrower focus on a single or a few domains of research" (Gussenbauer and Haddaway, 2018). Research goals included data analysis to identify trends in usage of Ex Libris' Primo over 3 years and literature review to investigate relevant themes from similar research studies that explored distinct user searching behaviors. Researchers examined 3 years of search usage and browser session data collected from the California State University (CSU) libraries and compared user search-type behavior (i.e., use of pre-search options) across all libraries with the goal of understanding if users use the available search options, specifically the use of a pre-filter drop-down menu comprised of subdivisions of the available search indexes. As a consortia cooperative, the CSU brands Primo as "OneSearch" to users, meaning our patrons never learn the term Primo. However, the authors will use the technical terminology consistently when referencing this discovery layer throughout this study and analysis.

On the Present Situation

Relevant areas of this literature space include many theories and methods but are heavily tied to web usability and human-computer interaction studies. Important themes included: comparing discovery layers with Google, UX, system design, and the use of filters, facets, dropdowns, and tabs. While this literature review is highly focused on a specific user search behavior, there is an ever-expanding pool of UX research related to the online search experience including many studies exploring empirical methods for evaluation of information retrieval (IR) system design and the importance of end-user perspectives. Many authors mention the complexity of understanding human information-seeking behaviors. Targeted on system performance in terms of user's perceived relevance, including cognitive load and user satisfaction, (Hu et al., 1999) utilized frameworks adopted from the Model Human Processor to incorporate theoretical foundations in cognitive psychology within the development of IR systems and the role of interface design in system-user concept communication. Fidel et al (Fidel et al., 2004) explored user motivations as a construct of "information need" through a Cognitive Work Analysis framework to address assumptions associated with human actions being goal-driven, simultaneously examining the task the user is performing, the environment in which the task is carried out, and the perceptual, cognitive, and ergonomic capabilities of the users typically performing the task.

Zabed Ahmed et al. (Ahmed et al., 2009) advocated for user-centered design, summarizing relevant research on human-computer interfaces for library-based IR systems in terms of basic interface issues, cognitive engineering, and user interface engineering through user interface guidelines, usability evaluation, and interface engineering techniques. Lopatovska & Arapakis (Lopatovska and Arapakis, 2011) conducted a review of research studies that incorporate theories of emotions and their role in human information behavior to explore the relationships between library and information science, information retrieval, and human-computer interaction. Kelly & Sugimoto (Kelly and Sugimoto, 2013) performed a systematic review of interactive IR evaluations from 1967-2006 describing these types of studies as not having "prescribed experimental methods", but instead relying on a wide variety of methods and measures, perhaps due to the "complexity of evaluating user behavior and the system interfaces simultaneously" (p. 746). The authors of this review mention log analysis as the method of data collection to be "uniquely situated on the continuum because of their importance and uniqueness", defining A/B testing as a "term used to describe a live experiment when a slightly modified version of an interface, for instance, is distributed to a randomly selected number of users. The behaviors of these users are then compared to a set of users who function as a control group" (p. 748). Of relevance to our present study design rationale is that "while A/B and interleaving tests can also be considered as examples of experimental studies of information behavior, a distinction can be made between large-scale log studies and smaller scale experiments of search behavior of the type that often occur in laboratories, use controlled tasks and systems, and gather other data in addition to log data" (p. 748). More recently, Guo et al. (Guo et al., 2020) explore ranking models within the context of IR research and while query analysis is outside the scope of this investigation, it bears mentioning the importance of understanding the intent of the user when developing more efficient system interface designs. There is little doubt that our users' first experience with information retrieval is through Google. The similarities shared by Google and Google Scholar contribute to the latter's popularity. As early as 2013, Zhang employed search result assessment and feedback collection methods to examine users' perspectives on the usability and effectiveness of discovery layers when compared to Google Scholar (Zhang, 2013). Zhang mentioned an earlier study on users' information-seeking behavior where "participants mainly examined the first page of search results and relied heavily on the facets to distinguish between types of materials" (p. 314). Zhang challenged the expectation that discovery layers will simplify the users' workflow of searching for scholarly information (a unified index of pre-harvested metadata) finding that Primo received "significantly lower usability and preference ratings" (p. 313). This study emphasized users' preference for fewer clicks and lack of awareness of source types; users "did not fully understand the link resolver interface" that "forces participants to make an unnecessary choice they are not familiar with" (p. 320) concluding "participants did not understand this inconsistency caused by different types of materials". He recommended that "a discovery layer's interface should conform to common design practices in other search tools (e.g., Google Scholar) so that users are able to transfer their experiences of other systems to the discovery layer" (p. 321). Although focusing on Summon and not Primo, Namei and Young (Namei and Young, 2015) examined relevance algorithms and the impact of Google, highlighting the frustrations and ambivalence toward discovery layers in academic libraries. They noted that "most users tend to only look at the first page of results and many only click on the first item in the results list" (p. 522) and cited other studies where "students were heavily influenced by the position of items in the results list" (p. 522). Implicit in these findings is that users trust the library search engine's ability to retrieve and rank results.[i] These authors suggest that, rather than struggling to find the “best” information, searchers use results that are “good enough” or “satisfice” as a “coping mechanism for dealing with information abundance and overload”. In a later study, Mugulia and Namei (Muglia and Namei, 2017) explored the contradiction between not overwhelming users with choices and the supposed expectation of power users to have options by citing research findings where choice often led to less satisfaction, “choosing almost arbitrarily to get the process over with” and the need for filters to sort the subsequent abundance of information. These authors defined filters as any “mechanisms for narrowing, customizing, or even expanding option/content, depending on the parameters of need”, and further defined pre-filters as “any action taken with the goal of filtering the results before hitting the search button”. In relation to our present study, "over filtering” was directly responsible for several failed search attempts.

Rooted in user satisfaction methods including affinity theory, satisfaction-loyalty theory, information system success theory, and the Technology Acceptance Model (TAM), a recent study on satisfaction levels of university students in China (Xu and Du, 2018) highlights “perceived usefulness and perceived ease of use” of the system is highly correlated to users' acceptance of information technology (p. 65). This research study emphasized how user behaviors have transformed, specifically “the convenience and expedience of access to information resources and expecting to interact with information providers” (p. 64). Relevant findings and recommendations note the large variation between user types suggesting personalized services such as customized interfaces, “user differences including education level, age, gender, frequency of use, and user experience had significant effects on user satisfaction and user loyalty” (p. 72). Chapman et al. (Chapman et al., 2016) explored how to better support users’ interactions with the library’s website and digital resources. These authors presented strategies to reduce user frustrations with overly complex library websites based on cognitive science, human-computer interaction, and user experience studies to reduce users’ “cognitive load” (p. 48), highlighting the importance of adopting UX principles and using usage data in web page design. Authors recommend simplified decision making including establishing a clear visual path to direct users' attention where it needs to go, utilizing hierarchal organization and proximity as most users expect a pattern to exist, and “chunking” to address assumptions that objects close together tend to be related. This study lends credence to the “usability-aesthetic effect” which posits that the appearance or attractiveness of a site adds to its perceived value. Such an effect extends to the style and emphasis of text, which is highly related to users’ awareness, attention, and expected functionality (p. 49). Relevant findings surround the “burden of choice” and what these authors term “choice simplification” and “choice reduction”. Important advice included fighting the instinct/urge to provide an “exhaustive list of options and prominent access to advanced features just in case a user needs it” and the subsequent “disservice to most of our users who do not have preferences or advanced needs'' (p. 49). One perspective from this article centered on users' preference for Google despite exposure to library resources from direct instruction. When website usage was tracked, the study found that “two-thirds of visits were research related, with searching in Summon or EBSCO search boxes on the home page being the most popular activity” (p. 51). Chapman et al. emphasized the need to “move from a landscape of systems that were, to varying degrees, no longer meeting the needs and expectations of our users to a forward-looking environment where we could begin to implement best practices in user experience (UX) and discovery”, and noted “library professionals are accustomed to navigating within this patchwork environment, so it can be difficult to step back and realize that our users aren’t and don’t want to be” (p. 55).

There have been many usability studies performed on Primo and other discovery layers. Studies that specifically refer to user interactions with search scopes are highly relevant to our current findings. A recent study from another CSU library (Vargas Ochoa, 2020), highlighted the importance of limiting the number of decisions a user must make such as identifying themselves as faculty or student citing that “critics argue that this design forces users to identify themselves before searching for information, thus taking them out of their task mindset” (p. 2). Vargas Ochoa’s research centers on students lacking an understanding of library resources, maintaining that any sort of pre-filter is asking our users to make an informed decision without the necessary understanding of the search terminology or scope. While our research study did not examine the terminology (label/name) used for the various drop-down scopes at each CSU library, previous studies uphold that users are confused by “library terms” and “content should also not overwhelm the user with an abundance of information” (p. 3). Important findings included that “students scan... rather than [read] material” and that multiple search boxes are extremely confusing for users (p. 3). This study uncovered students' persistent expectation that any search box has “search all” functionality (library catalog and website together), specifically that “students approach library search boxes as if searching Google” (p. 3). This research reiterates the importance of familiarity with the search tool because use and adoption “took some getting used to” according to student feedback (p. 12). Dease et al. (Dease et al., 2020) focused specifically on web design standards that improve the UX of the library’s website. Relevant to our current study, “results indicated that the majority of the tabbed search box options were not being used” and “most users opted to start with the discovery service “Quick Start” option” (p. 415). Recommendations led to the implementation of a single search bar with “an approach that prioritizes a search experience that is simple and familiar to users a la Google” (p. 416). Adams & Hanson (Adams and Hanson, 2020) used similar metrics as our present study, investigating Primo Analytics sessions and searches in addition to usability testing to explore students’ experience with the library’s discovery layer on mobile devices. In a study focused on student search behavior, Hamlett & Georgas (Hamlett and Georgas, 2019) make recommendations for “modulating these ingrained habits” through improvements to discovery systems, citing multiple studies on users' misunderstanding of the scope (e.g., what is being searched) and source types such as format (e.g., articles, journals, library website, books). This research emphasizes distinct user behaviors covered in other recent studies such as students' difficulty with navigating Primo due to the overwhelming number of choices, finding that students did not use tools such as save, cite, or email. In this study, researchers observed that students often assumed the system knew what was most relevant and selected the top result. Galbreath et al. (Galbreath et al., 2018) presented a summary of usability studies to date and performed usability testing on the Primo New UI. These authors highlighted many issues with the interface, finding that “that test subjects had difficulty navigating and finding information in the Primo tabbed structure” as well as confusion among users in distinguishing between a journal and a journal article. Researchers also found that “participants preferred Basic Search to Advanced Search” (p. 14) referencing an earlier “2014 Ex Libris user study indicating that users are easily confused by too many interface options and thus tend to ignore them” (p. 19). In a similar study, Porat & Zinger (Porat and Zinger, 2018) found that “participants did not realize that Books & More included other materials such as journals, theses, video recordings and maps” and “mentioned that it was not clear that the title option in the advanced search of “Articles & More” tab (Primo Central Index) referred to the journal title and not the article title”. Extremely relevant to our present research is that “none of the participants used the search scope (the drop-down menu with the catalog that one chooses before the execution of a search)”. Gilmore et al. (Gilmore et al., 2017) tried to gain a full picture of user perceptions and experiences with Summon using surveys, focus groups, and usability testing that demonstrated users’ struggle with searching, misunderstanding of library terms and jargon, and preference for the familiar.

Gusenbauer & Haddaway (Gusenbauer and Haddaway, 2020) explored how technological advancement has impacted user research through the changes in search functionality and workflows as modern researchers struggle with this new abundance of information and multitude of search systems. While highly focused on evidence-synthesis and comprehensive, unbiased search capabilities of current popular search platforms, this study mentions that “users are perhaps guided by convenience rather than strategic consideration when choosing their search system” (p. 211). These authors stress the value of “search literacy”, exploring the tradeoff in performance as “a search system with a smaller size, covering only a single discipline, might bring more relevant search results than a large search system covering multiple disciplines” (p. 192). Expounding on the importance of Boolean logic to establish the perfect balance between precision and recall needed for the most effective search (p. 194), the authors categorized this capability as a *necessary* criterion since “Boolean queries retrieve the largest portion of relevant records”. However, the findings were in direct contrast to these recommendations as “half the search systems we examined have at least some issues with Boolean queries” (p. 209). Lowe et al. (Lowe et al., 2018) studied how first-year students search for information, mentioning Mann’s research on the “principle of least effort” including that students do not “venture past the first article and rarely mov[e] beyond the first page of results” (p. 518). These researchers discovered that Boolean is not “better” than natural language/phrase searching (p. 529). Subsequent work by Lowe et al. (Lowe et al., 2020) continued to explore researchers’ habits and behaviors related to searching, finding that “based on relevance, there is no compelling evidence that either search is superior” and that “simple search is likely to be much more realistic in anticipating a student approach” (p. 5). Those researchers supported their findings noting the congruity with other recent studies, citing students' difficulty in finding information such as defining search terms and using discipline-specific databases. Using a grounded theory lens and highly focused on library instruction, Pickard & Desilets (Pickard and Desilets, 2020) explored students' information-seeking behavior, specifically where students go to find sources to use in a research assignment. Of interest to our research study is that “students did not immediately turn to Google when doing independent research” instead they almost exclusively reused sources from databases introduced within their course. This suggests that instructor expectations and familiarity may have a larger impact on how students search, “in terms of searching for sources, multiple studies have found that students prefer what they perceive as ease-of-use over credibility” including “what students perceived as easy was relative to what they were accustomed to doing” and “search methods also ‘appear to be driven by familiarity and habit’”. This study highlights the importance of expanding students' understanding of the larger breadth of search possibilities as students seem unaware of individual search features within specific databases.

Huvilla et al. (Huvilla et al., 2022) investigated the disconnect between information behavior and practices literature and information system development, and a "need for a better and more holistic understanding of user needs and perspectives" (p. 1043). These authors mention the need for findings to be contextualized as "explanations and recommendations have the most value when they are connected to specific services and contexts of use – even if there are many general traits in how people interact with information" (p. 1052). Li & Liu (Li and Liu, 2019) recognize the difficulty in evaluating digital libraries and propose a model to explore users' perspectives on interactions as a multi-dimensional construct with 3 dimensions: information resource, interface, and tasks. The results support factors related to DL performance considerations in terms of appropriateness, rich and valid links, reasonable page layout, the salience of topics, search task difficulty, a well-organized website, learning curve, accessibility, usefulness, and familiarity with task procedure. This study mentions usability as the most investigated measure in DL evaluation including two methods: usability testing and Web statistics for data collection. "Web statistics are often used to analyze usage and search patterns based on a holistic view of users" (p. 707). Analyzing web analytics data to understand user actions and correlated user behaviors has been well explored. While controversial due to patron privacy concerns, Scarnò (Scarnò, 2012) used such methods to study user behavior by mining web server log files for distinct search sessions. This research cites the work of Swanson (Swanson, 1977) as the process of searching being trial and error, meaning that all users are starting with a guess and users better understand information retrieval systems as they use them; essentially learning and refining based on the results retrieved. Scarnò's major takeaway was that users prefer the simple search and that they tend to repeat the previously used search action (with some refinement). Greenberg & Bar-Ilan (Greenberg and Bar-Ilan, 2017) examined log files, reports, and publishers' counts to investigate library users' information retrieval behavior in the process of discovering scholarly information including six major expectations of users as they search for research materials. Authors mentioned that "libraries and library services are perceived as complicated, while other sources (such as Google) are easy to use" (p. 455). Their study used data from logs obtained from the open URL link resolver log files to count requests for full text of articles received from the system and Google Analytics data from the library's main homepage for visits and the number of hits of the discovery tool search box. Finding from the study reflects the use of the library website's main function as access to the discovery tool search box including "most users come to a library site wanting to do research, and the shorter their paths, the happier they are" (p. 465). Related to extracting information from available system data, Ndumbaro (Ndumbaro, 2018) studied user interactions with the library's Online Public Access Catalogue (OPAC) through transaction log analysis to explore potential causes for search failures. Results indicated low use of the integrated library system ADLIB and relevant to our present study, users were inclined toward the "default keyword search, author, title and subject terms being the most preferred access points". Findings included that users rarely used Boolean operators or advanced search features (p. 299) and addressed users' preconceived notions related to system performance as users assume that "library catalogues function like search engines" (p. 300). In a recent study, Fu et al (Fu et al., 2021) utilize Primo Analytics data and Google Analytics data and highlight log analysis as "one of the less overtly intrusive ways to study information seeking behavior online" through monitoring patterns in systems usage to explore user activities and actions. This research study demonstrates the use of log analysis as a method for understanding a particular cultural group's information seeking behaviors. Their study mentioned "the findings from the log analysis alone are hard to generalize as motivations and individual characteristics are hidden behind collective behavior" (p. 2) and citing literature "that information behavior is a complex event that 'involves changes in cognition, feeling, and/or events during the information seeking process'. Log analysis alone was not able to demonstrate all the relevant factors that impact on the information seeking behavior of different user groups, such as culture, usage context, personality, although it can highlight some user behaviors at scale" (p. 2). Primo offers many dimensions of useful analytics related to users' interaction with our discovery system. These analytics can shed light on questions asked and issues raised in the literature pertaining to the complexity of the search interface and how traffic arrives at a list of results from the discovery layer. It is with those questions in mind that this study explores how prior conclusions about discovery were reflected in our libraries.

A Study in Scope

This study used a nonexperimental design. The goal was primarily a descriptive one, to look at how a large sample of library users behaved in the real world across different discovery layer configurations and to answer the question: how many pre-search options should be presented? The study sample were the 23 libraries of the California State University (CSU) system and the library at Moss Landing Marine Laboratories (MLML), a multi-campus research consortium. The CSU is the largest four-year public university system in the United States. The campuses are remarkably similar, 17 of them are Carnegie Classification #18 (Master's Colleges & Universities: Larger Programs), all of them are classed as either “high undergraduate” or “very high undergraduate” in their Carnegie Enrollment Profile, 18 of them are located in large or midsize cities and suburbs according to the IPEDS degree of urbanization variable, and 14 of them are designated Hispanic-Serving Institutions. Regarding scholastic aptitude, twenty-one of the campuses were classified as either “inclusive” (i.e., not selective) or “selective” in their Carnegie ACT category with only California Polytechnic State University-San Luis Obispo and San Diego State University being “very selective”. Where there are notable differences between campuses, they are not demographic but pertain primarily to geography and size, and to a lesser extent academic aptitude. (The singular exception to this being Maritime Academy which was over 80% male during the entire study period.) To the extent that the students, faculty, and staff of the California State University system are similar to analogous users at other academic libraries, our large sample is informative. Notable differences between campuses are displayed in Table 1 along with measures of their variation.

Table 1: Selected Characteristics of California State University Campuses

CCIHE or IPEDS Variable	**Mean**	**Median**	**Average Absolute Deviation**	**Standard Deviation**
Bachelor's degrees conferred	4400.65	4045	2049.12	2461.76
Master's degrees conferred	881.74	764	610.79	750.30
SAT-Verbal 25th percentile score	365.65*	480	190.78	219.36
SAT-Math 25th percentile score	360.43*	470	188.05	216.64
ACT Composite Score, 25 percentile	13.83	18	7.21	8.50

Note: All data and calculations derived from: (Indiana University Center for Postsecondary Research, 2018)
* Readers may note these figures seem low, given that the minimum possible score is 200. This is because several campuses are de facto open admissions which is scored as zero, not null, in IPEDS data. Rather than remove those as outliers we have retained them as an indicator that there is a wide range of college readiness and cognitive ability in the CSU represented in the sample.

The data collection period was January 1, 2017, through December 31, 2019. This window of time was chosen deliberately as all the CSU campuses had experienced at least one complete academic year with Primo as their catalog by 2017, minimizing any effects that users learning a new system or staff testing might have had on the data. Despite the existence of 2020 calendar year data, 2019 was chosen as the cutoff point to rule out any effects from the transition to virtual learning driven by the SARS-CoV-2 pandemic. The sample, therefore, captured normal user behavior in Primo, under various configurations, during unexceptional circumstances.

Data Sources

Modeling each campus library as one ‘subject’, between-subjects (cross-institutional) comparisons were made across several variables derived from manual inspection of the 24 Primo instances and library website homepages as well as a large dataset of usage metrics extracted from Ex Libris’ Primo Analytics. This study, lacking an ex-ante hypothesis of how the data might look was exploratory and used nonexperimental methods. The primary task was cleaning the data, adding dichotomous grouping variables (e.g. Signed In/Not Signed In), adding supplemental synthetic variables, and preparing it for graphical analysis and simple statistical analysis. No query terms were extracted or analyzed, the data was devoid of any context for information-seeking behavior, presenting purely abstracted measures of usage. Because of this acontextuality, we refrained from considering “why” questions and restricted ourselves to descriptive “how” questions.

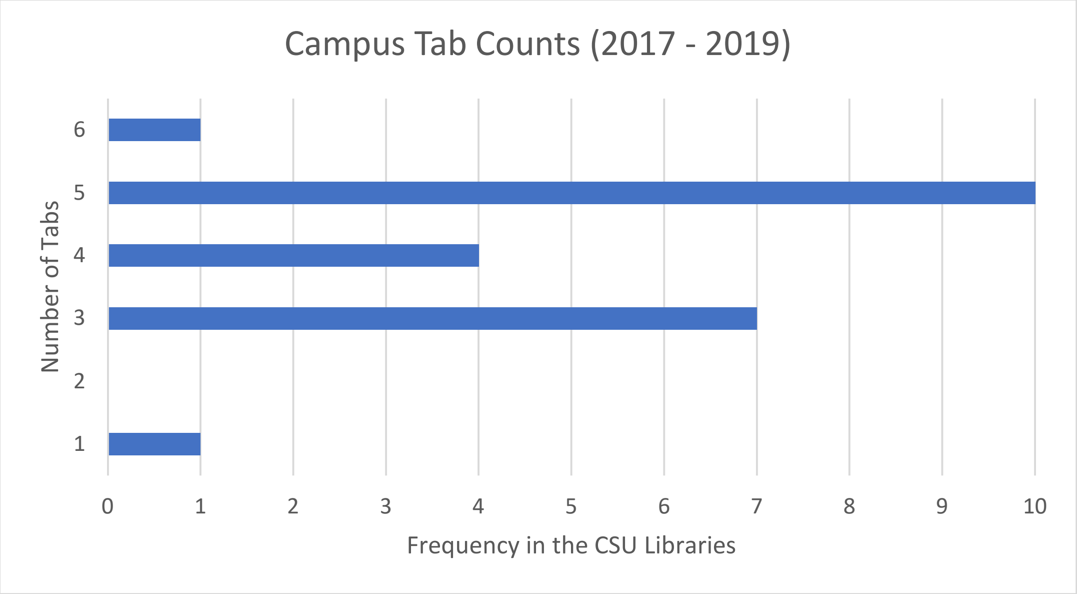
Primo Analytics

Data from Primo Analytics for all 23 CSU campus libraries plus the Moss Landing Marine Laboratories Primo instance was obtained via the CSU Chancellor’s Office. The variables queried were: Institution Name, Action, Action Sub Group, Search Scope Type, Active Tab, Referrer, User Group, Signed In, Actions, Sessions, and On Campus. Each variable took the standard form, definition, and operationalization assigned to it by Ex Libris in their implementation of Oracle Business Intelligence Enterprise Edition (Ex Libris, 2019). Usage was captured through two distinct variables and measures, the Actions variable and the Sessions variable. Actions was a numeric variable for a family of metrics in Primo Analytics that records the number of times a selected action took place; our dataset held tallies for all possible search actions, hereafter this data was labeled and called ‘Searches’. Sessions was numeric and recorded the number of web browser sessions in which a search action was taken. The Sessions variable tallies were necessarily smaller than the Searches as multiple search actions could be taken during one browser session. Other numeric variables were Signed In and On Campus which tracked the respective frequencies for authenticated searches and searcher location. The remaining variables were nominal and there was wide variation in terminology across the CSU system in Tab and User Group naming conventions. Reconciling user type codes, i.e. affiliation, and active tab names was a two-step process. First, staff (occasionally faculty librarians) were contacted at the libraries where a User Group code or Primo tab name was not clear. Second, after a clear picture of intent and meaning behind each ambiguous User Group and Tab value was available, a Python script was used to recode analogous values into new merged codes. One persistent problem with the Primo Analytics data was the presence of blank values for Active Tab, Referrer, and User Group. Blank values were not discarded; we address the implications and adjustments made due to these values in the limitations section of the discussion. Data associated with testing, done either on Primo Tabs or done by User Groups containing the name ‘test’ or some variant, was discarded. Once cleaned, the data from Primo Analytics was imported into MS Excel where dichotomous grouping variables were made for further analysis. These grouping variables were related to Yes/No questions and included: 1) Was the library homepage the HTTP Referrer into Primo? 2) Were the queries executed from on campus? 3) Did users authenticate? 4) Did the queries use Advanced Search or other options which required exploration? And 5) Did the queries use Primo Central Index data?

CSU Libraries Web Homepages

Library homepages, specifically the search options presented therein, were an essential piece of our puzzle because we knew anecdotally from our respective institutions that a library’s landing page was often a student’s entry point to library collections and services. To document the various search configurations available to users, researchers collected URLs from the CSU Chancellor’s Office website in January 2022. This website presents current links to each CSU library’s website (assumed to be the homepage or "default" landing page for most users). Since this was a retrospective study, researchers used Internet Archive's (IA) Wayback Machine to gain an approximate idea of how much variation occurred in the time under investigation. Each campus library website URL was verified by navigating to the current site and then copy/pasted into the search box provided by the Wayback Machine, where researchers could toggle between date options available. There is possible bias associated with the method to select available IA images as researchers had to rely on the limitations of this system, including trial and error, to discover which dates would load within a Google Chrome browser. Figure 1 is a nonscientific comparison of the number of search options presented on CSU Library websites. Captured screenshots of 2022, 2019, 2018, and 2017 for each campus where there was data available are on figshare. A dichotomous grouping variable was also created based on web page data; this variable grouped all Sessions and Searches which were run with Action, Action Sub Group, Search Scope Type, and Active Tab values that matched the default value of each homepage search box. This variable was combined with Referrer values to form a synthetic variable denoting which queries originated from a library's homepage using the default search configuration.

Figure 1: Primo Tab Frequency in the CSU Libraries 2017 - 2019



Cross-Institutional Analysis

Data analysis was primarily graphical. Bar graphs were created in MS Excel to view and understand the dataset. These included: clustered bar charts showing counts of sessions and searches for all of the Primo Analytics variables (excepting Sessions and Actions which were the basis of the counts), search to session ratios by institution, average searches per session by affiliation type, usage of Primo's Advanced Search by institution, percentage of referrer traffic originating from the library home page by institution, and percentage of queries originating from the homepage which used the default search settings by institution. Those latter two charts and associated analyses required the use of the synthetic variables. Behavioral patterns among campuses were sufficiently clear from graphical analysis in many cases. Due to the nonexperimental study design and lack of a priori hypothesis, statistical testing was used to supplement the graphical analysis and for creating reportable results. Graphical analysis revealed highly skewed distributions which were confirmed with a Kolmogorov-Smirnov test of normality for the Searches and Sessions variables. Both Searches (*SD* = 30,064.67, skewness = 37.14, kurtosis = 1874.09, test statistic = .473, *df* = 31031, *p* < .00) and Sessions (*SD* = 14,009.58, skewness = 32.88, kurtosis = 1376.53, test statistic = .472, *df* = 31031, *p* < .00) showed statistically significant deviations from normality. Searches and Sessions were also highly correlated (Spearman's ρ = .981). Kruskal-Wallis *H*-tests, a distribution-free alternative for an ANOVA, were run inter-institutionally to compare differences in search behavior on or off campus, signed in or not signed in (i.e. authenticated), and between affiliation types. These nonparametric statistics, which are calculated on rank order data and compare medians rather than means, were required due to the non-normality of the data. The search behaviors analyzed were: Action activity, Search Scope Type activity, and Active Tab activity. For each comparison, Searches and Sessions activity on-campus was compared with off-campus, and authenticated activity was compared with not authenticated activity. For the subset of data where users were signed in, the previous comparisons were supplemented by comparing User Group activity in querying Primo Central Index, traffic originating from the library homepage, and whether a user used the Advanced Search, Browse Search, Journal Search, or Newspapers Search. Statistical testing was done using IBM SPSS. The Kruskal-Wallis test can only determine that at least one of the groups analyzed differed from the others. It cannot tell which group that is; for that task, post hoc pairwise comparisons were needed and conducted in SPSS. Graphical comparison revealed the overall similarity of distributions examined using Kruskal-Wallis. Rather than report out lengthy post hoc comparison tables, we instead reported descriptive statistics for each distribution combined with graphs that clearly convey their similarities. Distributional statistics included skewness, a measure of asymmetry, and kurtosis, a measure of tailedness.

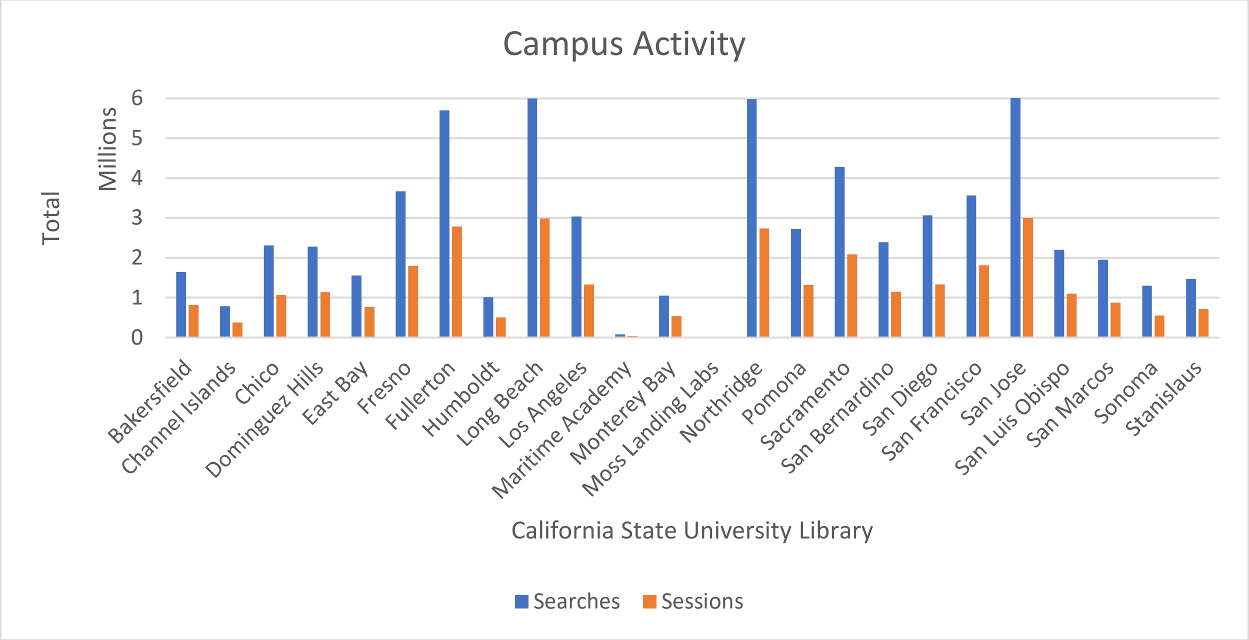
The Results of the Sixty-Four Million Searches

Table 2: Distribution Statistics for Searches and Sessions Variables

Usage	Total Activity	Mean		Standard Deviation	Skewness		Kurtosis	
		M	*SE*		Stat.	*SE*	Stat.	*SE*
Searches (*n* = 31,031 cases)	64,058,932	2064.35	170.67	30,064.67	37.14	.01	1,874.09	.03
Sessions (*n* = 31,031 cases)	30,822,052	993.27	79.53	14,009.58	32.88	.01	1,376.53	.03

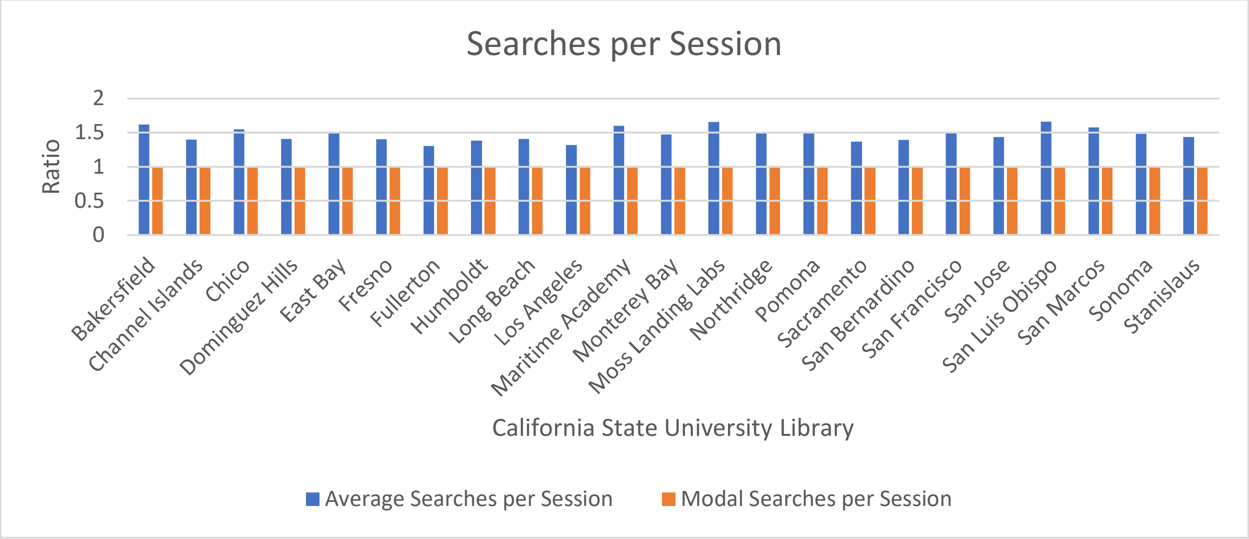
On a general note, we observed that search activity in Primo was a function of the size of the user base. This was clear based on a plot of Searches and Sessions activity by Institution combined with researchers' background knowledge of the sample; see Figure 2 for detail. Specifically, the correlation between total Searches 2017 – 2019 per campus and total degrees conferred in 2018 was *r* of 0.94 with an r^2 value of 0.89 showing that total degrees conferred explained 89% of the variation in search volume. Descriptive statistics about the shape of the Searches and Sessions variables overall, in particular their high variance, positive skew, and positive (as well as high) kurtosis are shown in Table 2.

Figure 2: Searches and Sessions for the CSU Libraries 2017 - 2019



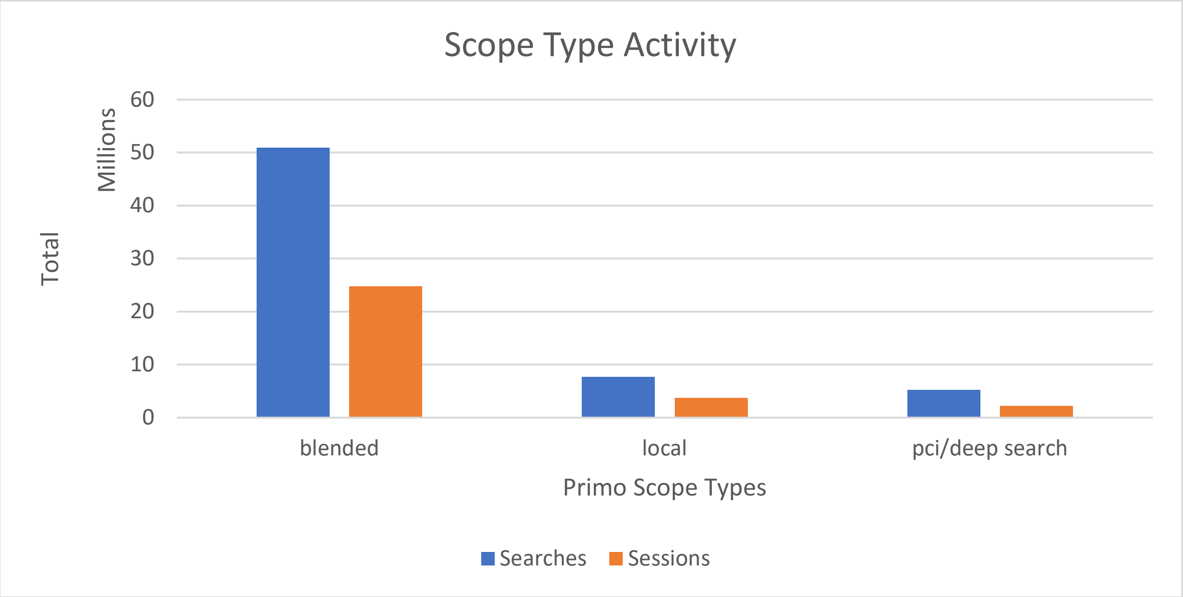
Search activity dwarfed all other ways of retrieving results from Primo including browse activity, the newspaper search, and the journal search (called the e-journal A to Z List and referred to as AZ List in Primo Analytics). Basic Search accounted for 86% of all search activity, while Advanced Search accounted for 11%. The remaining 3% of activity was all other query methods with usage of the journal search accounting for 2% of activity and the Newspaper search and Browse options making up the residual. The time investment in searching, represented by the Search to Session ratio, was very similar across institutions and not dependent upon the structure of search options and pre-filtering options available. The modal Search to Session ratio was 1 across all libraries and the average number of Searches to Sessions only varied between 1.29 (San Diego) and 1.66 (San Luis Obispo). Figure 3 shows the lack of variation in the number of searches conducted across libraries.

Figure 3: Ratio of Searches to Sessions for the CSU Libraries 2017 - 2019



When searching, there were three possible scope configurations drawing upon the data sources of local data from Alma (i.e., bibliographic records managed by the library), data from Ex Libris' Primo Central Index (i.e., bibliographic records managed by Ex Libris, now called Central Discovery Index), or a blend of the two (i.e., combined bibliographic records with priority/weighting determined by each library). The volume of Searches and Sessions using blended scopes was almost four times larger than the volume of usage on local and dedicated PCI scopes combined; see Figure 4 for detail. Seventy-nine percent of all Searches used a blended scope while 12% used a local scope and 8% used a PCI scope. This was likely a result of multiple causal factors. Most campuses had a blended scope as their default setting in Primo. One library, San Luis Obispo, only offered a blended scope. A small amount of Search_Scope_Type data was blank, expected behavior for miscellaneous search types in Primo Analytics, for both Searches (0.3%) and Sessions (0.4%) this accounted for less than zero percent of the data (Ex Libris, 2017).

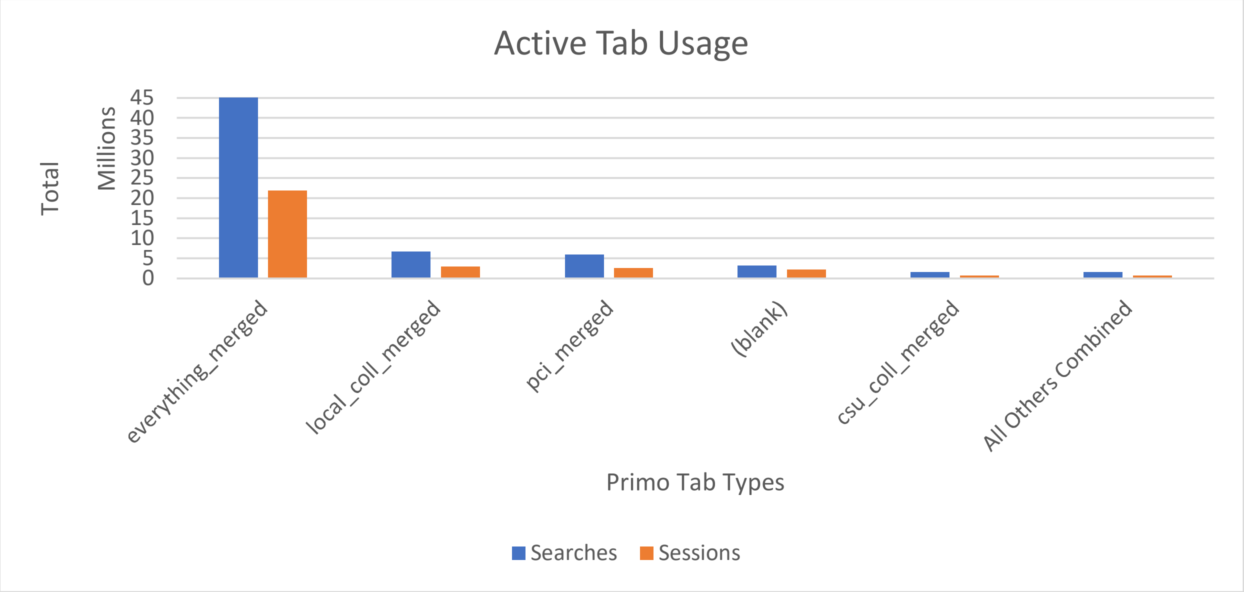
Figure 4: Primo Scope Type Usage Across CSU Libraries 2017 - 2019



The aspect of Primo that users actually saw displayed both prior and after a query, and which they could manipulate, was the Tab. While Scopes were obscured to users and implicit in Tab configuration, Tabs had public-facing labels. These labels, as noted above and below, had variations in their naming across institutions. When viewed for the entire sample Tab usage showed a highly skewed distribution. The overwhelming majority of searches, 71%, were conducted on tabs with variation on an "everything" name and function (e.g., Everything or All Collections). The second-most used tab function and naming convention was for exclusively local collections coming from Alma, which was 10% of searches. The third-most used, with 9% of searching, was a tab function exclusively querying Primo Central Index. These PCI tabs had variations just like the local tab names, but all emphasized the fact that article results could be retrieved (e.g., Articles & More, Articles+). Due to a product defect in Primo Analytics, blank data was present for the Tab value accounting for 5% of Searches but 7% of Sessions. With one exception, all other Tab functions and naming conventions received comparatively negligible Sessions traffic and Searches usage, functionally 0%. The exception being a tab querying Alma records shared across the California State University libraries (a service branded as CSU+), present at 11 libraries, which accounted for 2% of

Searches. Figure 5 shows the uneven usage pattern. Search to Session ratios did not show considerable variation and rounded to 2 with no decimal places for all Tabs with non-trivial usage.

Figure 5: Active Tab Usage Across CSU Libraries 2017 – 2019



Query Origination

From where online did users begin their Primo Sessions? HTTP Referrer data collected by Primo Analytics supplied an initial answer, which was that the majority of Referrer data (50%) was blank. Across all Institution values, blank Referrer was in the four most frequently occurring Referrer values. This raised a question about the presence of another possible software defect, but none was confirmed by Ex Libris, leading researchers to conclude that blanks were correct values that arose organically. Many possible user behaviors could result in a blank (i.e., empty) Referrer value being passed, we address this further below. After investigation of library homepages and the HTML and JavaScript code in their Primo query forms, we determined the blank values to likely be due to the way the forms were programmed. The blank Referrer data was then grouped together with Referrer values for each library homepage under a new dichotomous synthetic variable called ReferrerBlankORHomepage. There were statistically significant differences in Searches (*H* = 1821.13, *df* = 1, *p* < .00) and Sessions (*H* = 1594.32, *df* = 1, *p* < .00) values determined via Kruskal-Wallis tests grouping on the ReferrerBlankORHomepage variable. Descriptive statistics for the distributions, shown in Table 3 showed the four distributions to have positive skewness and kurtosis. Graphical comparison revealed no substantial differences between libraries. Under researcher assumptions, the percentage of Session traffic originating from each library's homepage was broadly similar; results can be seen in Figure 6. At every library, more than 52% of traffic into Primo came from the homepage and at 14 libraries, a majority, the percentage exceeded 70%.

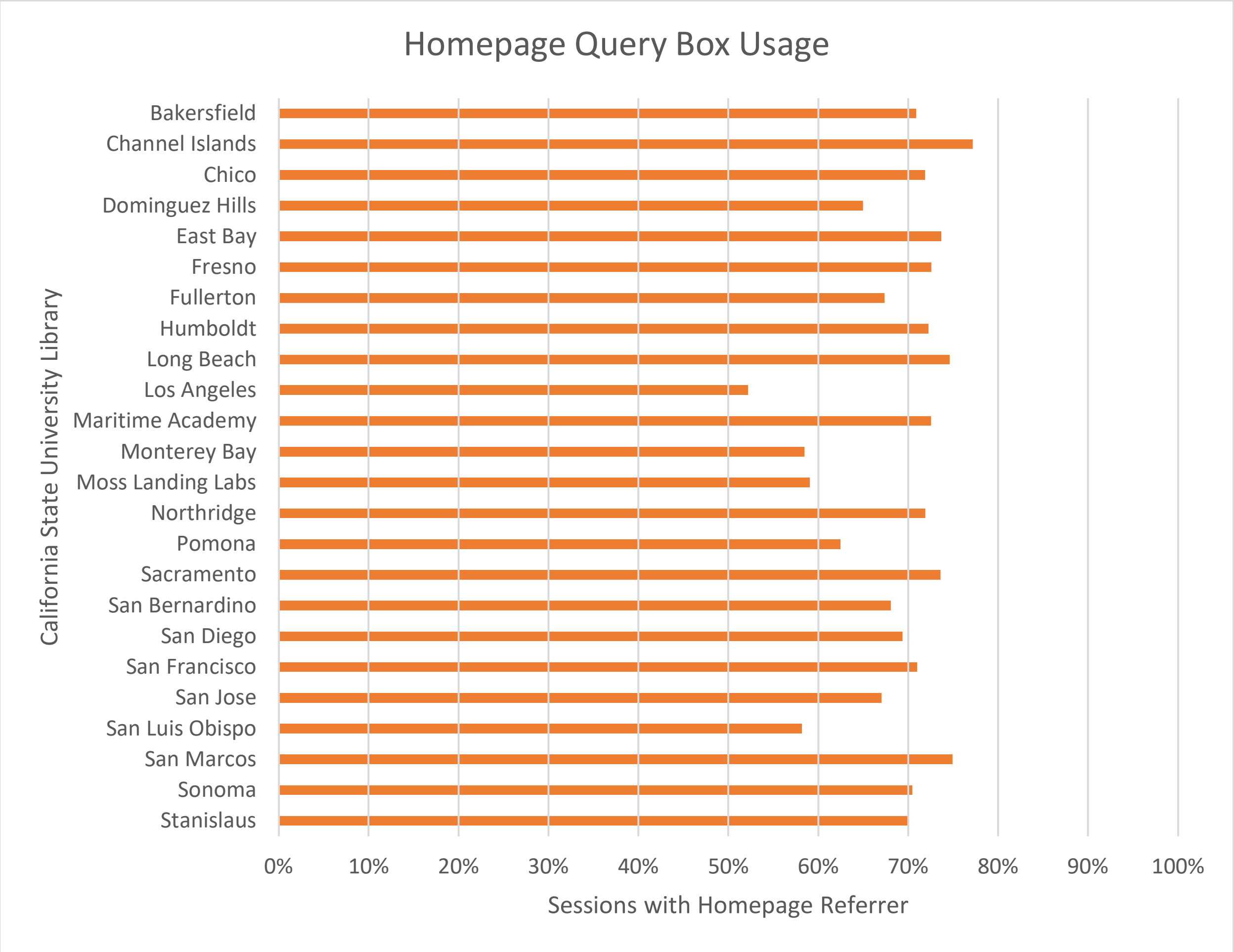
For non-blank Referrer values, they followed a general pattern in Session volume. First, with the most traffic being for an institution's library homepage (e.g., lib.calpoly.edu), the next most frequently appearing were the authentication URLs for the institutions (e.g., idp.calpoly.edu and shibboleth.csuchico.edu), followed by institutional dashboards (e.g., my.calstatela.edu), then by URLs for a library's research guides or finding aides (e.g., guides.lib.calpoly.edu and libguides.csuchico.edu). In total there were 2,453 unique Referrer values recorded over the sample period with the modal Session value for any given Referrer being 1; this showed the incredible variety of places across the internet from which users could find a link into a CSU Primo instance.

Table 3: Distribution Statistics for Searches and Sessions by Referrer

Usage Behavior	Total Activity	Mean		Standard Deviation	Skewness		Kurtosis	
		M	*SE*		Stat.	*SE*	Stat.	*SE*
Searches Referrer Homepage (*n* = 9,850 cases)	42,649,395	4329.89	492.83	48,912.37	25.38	.03	818.52	.05
Searches Referrer Not Homepage (*n* = 21,181 cases)	21,409,537	1010.79	99.15	14,430.53	30.25	.02	1126.89	.03

Sessions Referrer Homepage (*n* = 9,850 cases)	21,267,618	2159.15	233.74	23,198.42	21.51	.03	559.04	.05
Sessions Referrer Not Homepage (*n* = 21,181 cases)	9,554,434	451.09	41.43	6,030.25	30.93	.02	1223.59	.03

Figure 6: Percentage of Sessions with Homepage Referrer Across CSU Libraries 2017 – 2019



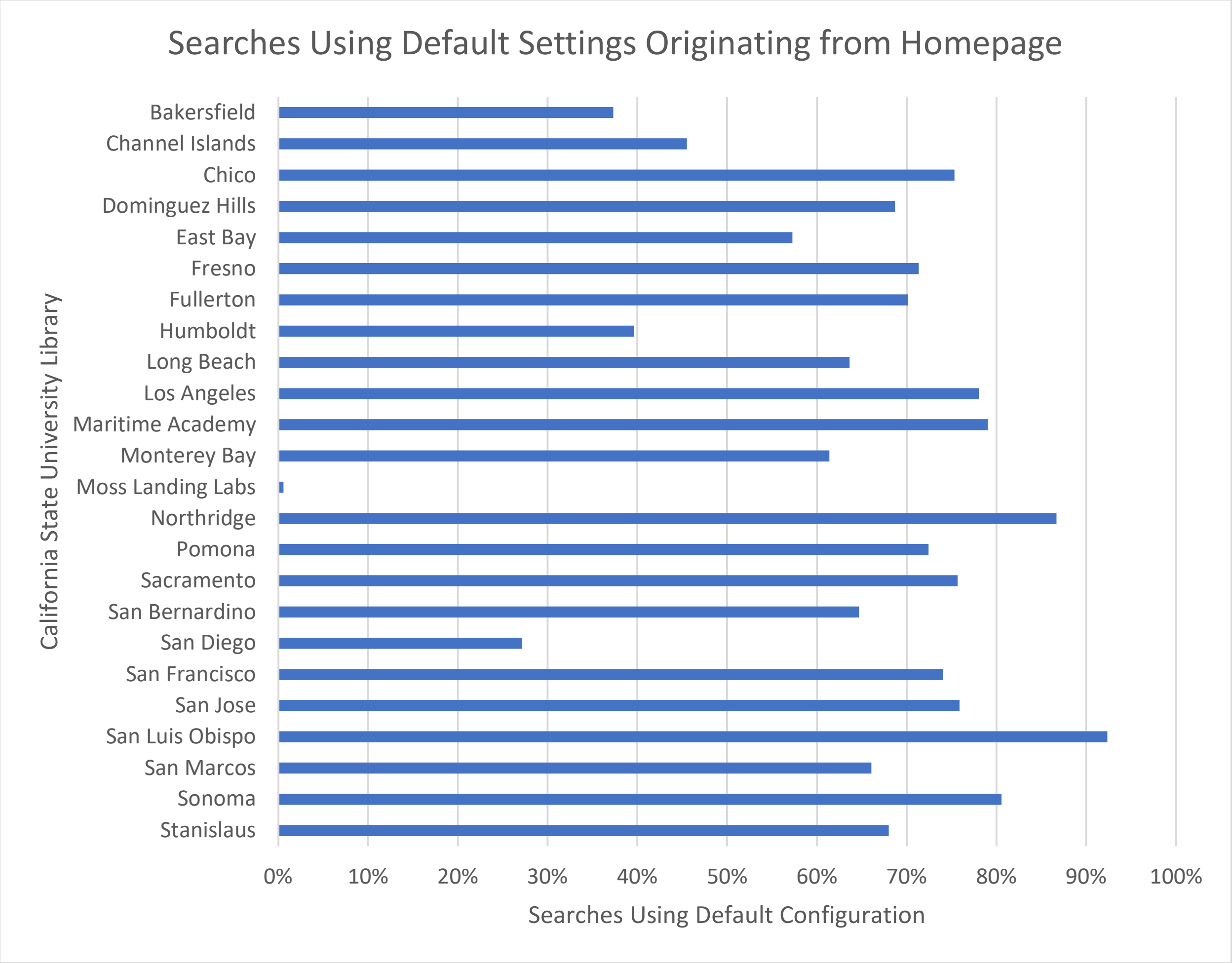
* The majority of Sessions (77%) were generated by the User Group values ‘guest’, an Ex Libris code indicating users who do not sign in, and ‘undefined’, the researcher-assigned variable to account for blank values (Ex Libris, 2019). The remaining data can be seen disaggregated in Figure 15.

By using the information on library homepages from the Wayback Machine we were able to also measure the amount of Primo usage that simply originated from a homepage and accepted the preset default search settings; results shown in Table 4. In other words, for what percentage of traffic originating on the homepage was the default pre-search option acceptable? This was determined through a synthetic variable corresponding to the Scope, Tab, and Mode (i.e., Basic or Advanced Search) values that matched the default from each library's homepage in a majority of their archived pages in the Wayback Machine. Total queries for the synthetic variable indicating default settings were then divided by the total queries for the synthetic variable indicating a homepage referrer; results are shown in Figure 7.

Table 4: Distribution Statistics for Searches and Sessions by Homepage Default Search

Usage Behavior	Total Activity	Mean		Standard Deviation	Skewness		Kurtosis	
		M	*SE*		Stat.	*SE*	Stat.	*SE*
Searches Used Homepage Default (*n* = 883 cases)	29,458,030	33361.30	5227.41	155,334.01	8.10	.08	80.54	.16
Searches Did Not Use Homepage Default (*n* = 30,148 cases)	34,600,902	1147.70	80.41	13,960.88	31.56	.01	1312.91	.03
Sessions Used Homepage Default (*n* = 883 cases)	14,935,252	16914.22	2495.44	74,152.92	6.67	.08	51.59	.16
Sessions Did Not Use Homepage Default (*n* = 30,148 cases)	15,886,800	526.96	33.34	5,788.18	30.65	.01	1271.46	.03

Figure 7: Percentage of Searches Using the Homepage Defaults Across CSU Libraries 2017 - 2019



Query Location (on/off campus)

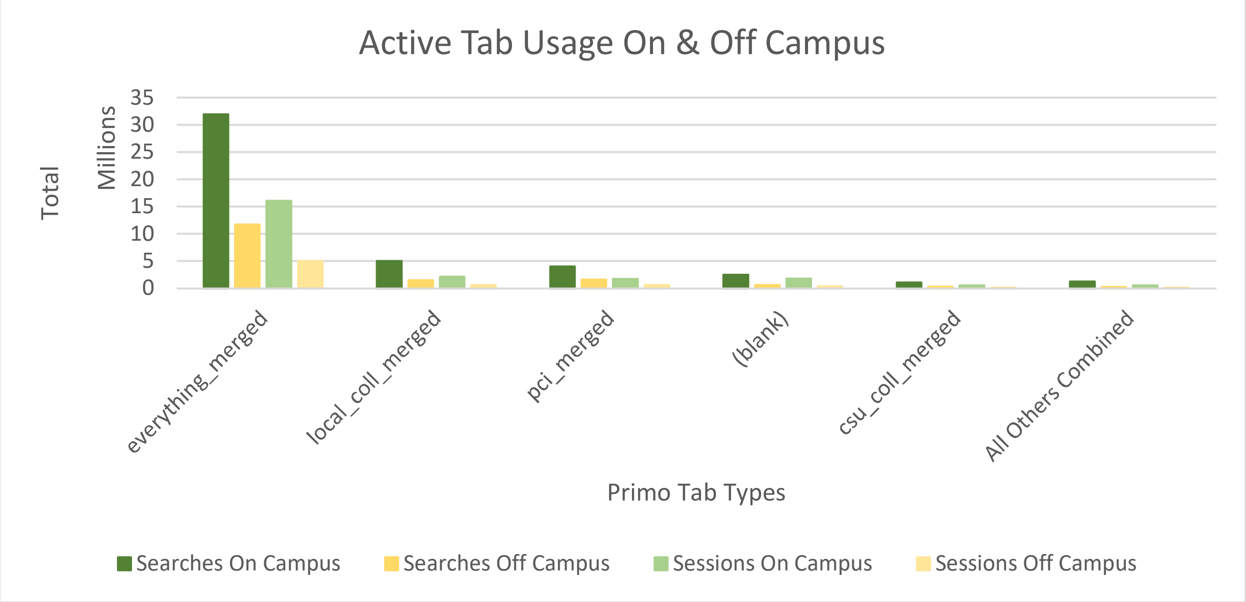
To determine if there were differences in query behavior between on-campus and off-campus groups, Kruskal-Wallis tests were run. Results showed statistically significant differences across both Searches (H = 2984.77, df = 1, p < .00) and Sessions (H = 3028.49, df = 1, p < .00). Descriptive statistics for the distributions, below in Table 5, showed the four distributions to have positive skewness and kurtosis. Graphical comparison revealed no substantial differences across the various search descriptors; for brevity, we include here only the data for Active_Tab in Figure 8 as that is most relevant to our final recommendations. Users exhibited similar query behavior in terms of Action type (e.g., Basic or Advanced Search), Scope type (blended or local), and Active Tab on-campus as they did off.

Table 5: Distribution Statistics for Searches and Sessions by Location

Usage Behavior	Total Activity*	Mean		Standard Deviation	Skewness		Kurtosis	
		M	*SE*		Stat.	*SE*	Stat.	*SE*
Searches On Campus (*n* = 4,864 cases)	45,783,970	9412.82	1038.76	72,445.56	16.31	.04	344.28	.07
Searches Off Campus (*n* = 24,888 cases)	15,998,576	642.82	52.97	8,356.64	29.68	.02	1,117.22	.03
Sessions On Campus (*n* = 4,864 cases)	22,804,874	4688.50	489.19	34,117.78	14.01	.04	240.50	.07
Sessions Off Campus (*n* = 24,888 cases)	6,878,030	276.36	19.53	3,080.95	24.63	.02	770.36	.03

* Calculations exclude CSU Dominguez Hills which had errors in their Alma/Primo configuration causing all activity to appear as if coming from off campus.

Figure 8: Active Tab Usage by Location 2017 - 2019



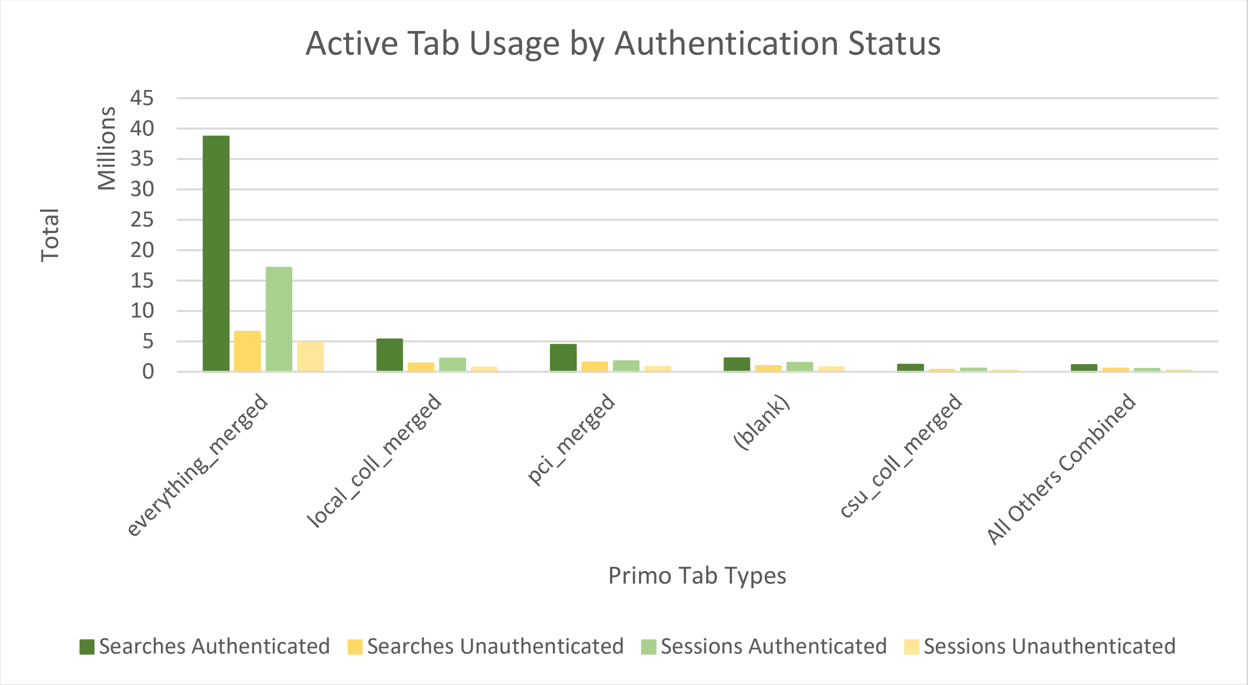
Authenticated Queries

To determine if there were differences in query behavior between signed-in and non-signed-in usage, Kruskal-Wallis tests were run. Results showed statistically significant differences across both Searches (*H* = 1785.20, *df* = 1, *p* < .00) and Sessions (*H* = 1344.42, *df* = 1, *p* < .00). Descriptive statistics for the distributions, below in Table 6, showed the four distributions to have positive skewness and kurtosis as well as large variance. Graphical comparison revealed no substantial differences, as seen in Figure 9. Users exhibited similar query behavior in terms of Action type, Scope type, and Active Tab when there was an authentication event as when there was not.

Table 6: Distribution Statistics for Searches and Sessions by Authentication

Usage Behavior	Total Activity	Mean		Standard Deviation	Skewness		Kurtosis	
		M	*SE*		Stat.	*SE*	Stat.	*SE*
Searches Authenticated (*n* = 19,251 cases)	52,859,189	2745.79	266.02	36,909.59	31.64	.02	1,318.85	.04
Searches Not Authenticated (*n* = 11,780 cases)	11,199,743	950.74	113.88	12,359.62	29.14	.02	1,043.67	.05
Sessions Authenticated (*n* = 19,251 cases)	23,433,566	1217.26	114.89	15,941.41	30.15	.02	1,157.99	.04
Sessions Not Authenticated (*n* = 11,780 cases)	7,388,486	627.21	92.83	10,075.22	37.56	.02	1,681.21	.05

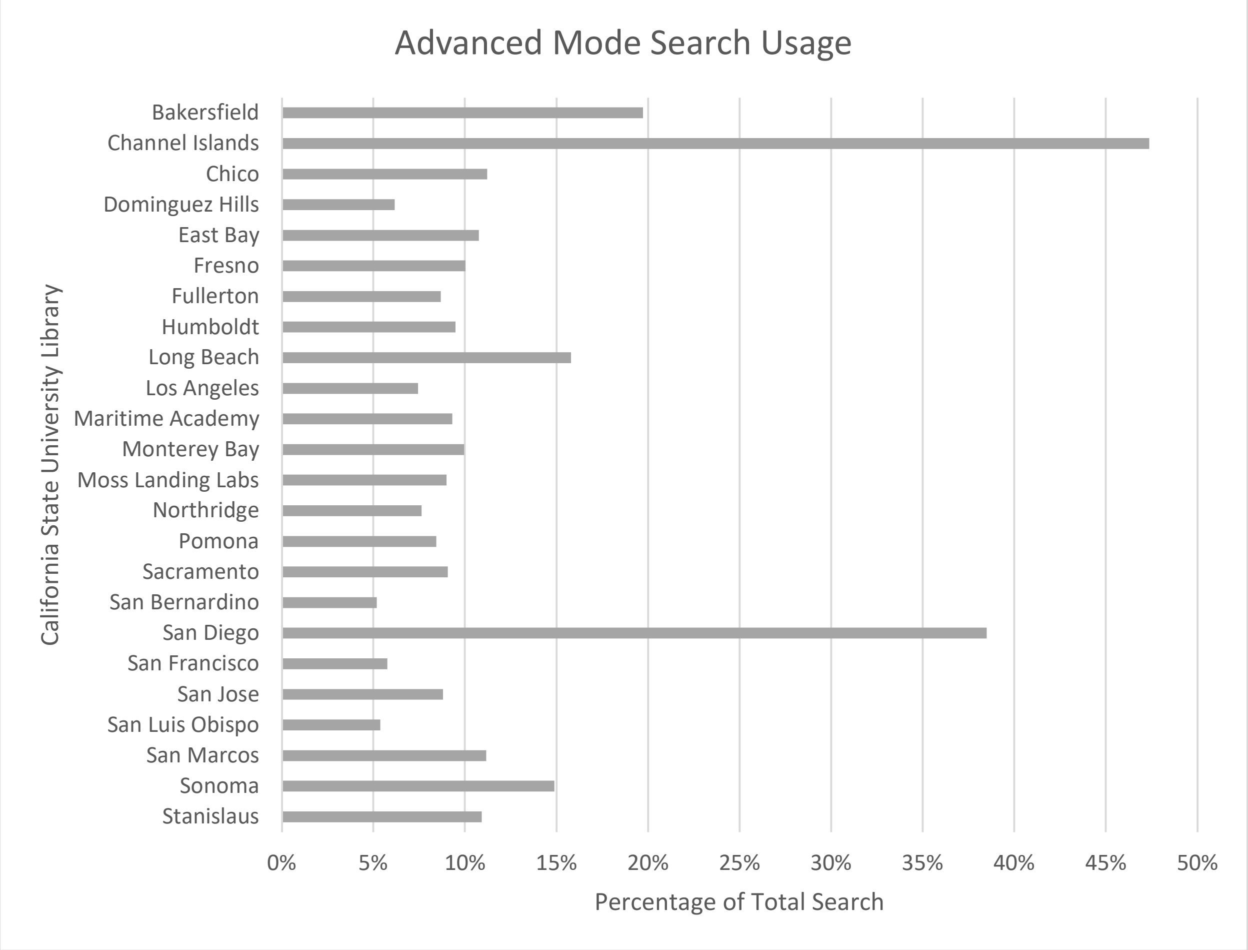
Figure 9: Active Tab Usage by Authentication 2017 - 2019



Discovery Layer Exploration

Usage of the Advanced Search feature was broadly very similar in percentage terms across 22 of the 24 libraries. Excluding the two outliers, the percentage of Advanced Search usage was between 5% and 19% with the average and median usage patterns both being 9% of all search activity. Including outliers, the ceiling of Advanced Search usage rose to 47% while the average was 13% and the median was 9%. The two outliers, San Diego State University and CSU Channel Islands are both partially explainable. Results are displayed in Figure 10.

Figure 10: Percentage of Advanced Search Usage Across CSU Libraries 2017 - 2019



The query box webform on SDSU Library's homepage offered the ability to search books, articles, journals, databases, archives, their website, and all the above. The default setting was all the above which served up a bento box display of the results page. From that page, up to 3 results were displayed for each query type with a link to "See All *n* Results" with all such links that pointed to Primo directing users to the Advanced Search. Put another way, when starting a search from the SDSU Library's homepage into Primo, Advanced Search was the default. (It is therefore striking that even when a design choice was made to proactively give users the Advanced Search options, that 62% of their total Primo queries were using the Basic Search option with one query box.)

The situation at CSU Channel Islands was different in specifics but otherwise broadly similar. Analysis of the library's homepage revealed several small and some dramatic changes over the study period. Researchers reached out to colleagues from the institution for clarification and details. From "Spring 2018" through the end of data collection in 2019, the homepage query box was set to default to Advanced Search. Another factor may have been user education. Colleen Harris, Head of Instruction, Engagement, & Assessment, explained how they market the benefits of using an advanced search feature in not only Primo, but also when approaching database searching as this search option "better curates results and leads to less overwhelm at the results list". She stated that "we very strongly encourage students we encounter at the desk, in research appointments, and during IL sessions (we have a popular program) to use advanced search" (Ruiz et al., 2022). Yet with most Primo sessions originating at the library homepage where Advanced Search was the default for approximately 2/3rds of the study period, coupled with user education to reinforce advanced searching in general, 53% of their total Primo queries were in Basic Search mode.

There were statistically significant differences across both Searches ($H = 108.38$, $df = 1$, $p < .00$) and Sessions ($H = 60.40$, $df = 1$, $p < .00$) for user exploration. There were also statistically significant differences across both Searches ($H = 103.62$, $df = 1$, $p < .00$) and Sessions ($H = 50.21$, $df = 1$, $p < .00$) for whether queries used Primo Central Index or not. The four distributions about the question of 'exploration' are presented in Table 7 below. Also below in Table 8 are data about the four distributions for whether queries sought Primo Central Index data or not. As with the original distributions of Searches and Sessions, all the distributions show large variance with positive skewness and kurtosis (Figures 11 and 12).

Table 7: Distribution Statistics for Searches and Sessions by Exploration

Usage Behavior	Total Activity	Mean		Standard Deviation	Skewness		Kurtosis	
		M	*SE*		Stat.	*SE*	Stat.	*SE*
Searches Explored (*n* = 14,454 cases)	9,079,898	628.19	45.05	5,415.73	29.47	.02	1,339.70	.04
Searches Did Not Explore (*n* = 16,577 cases)	54,979,034	3316.59	316.74	40,781.33	27.71	.02	1,031.34	.04
Sessions Explored (*n* = 14,454 cases)	4,261,592	294.84	18.80	2,260.41	26.05	.02	1,022.56	.04
Sessions Did Not Explore (*n* = 16,577 cases)	26,560,460	1602.25	147.81	19,030.49	24.43	.02	752.28	.04

Figure 11: Active Tab Usage by Exploration 2017 - 2019

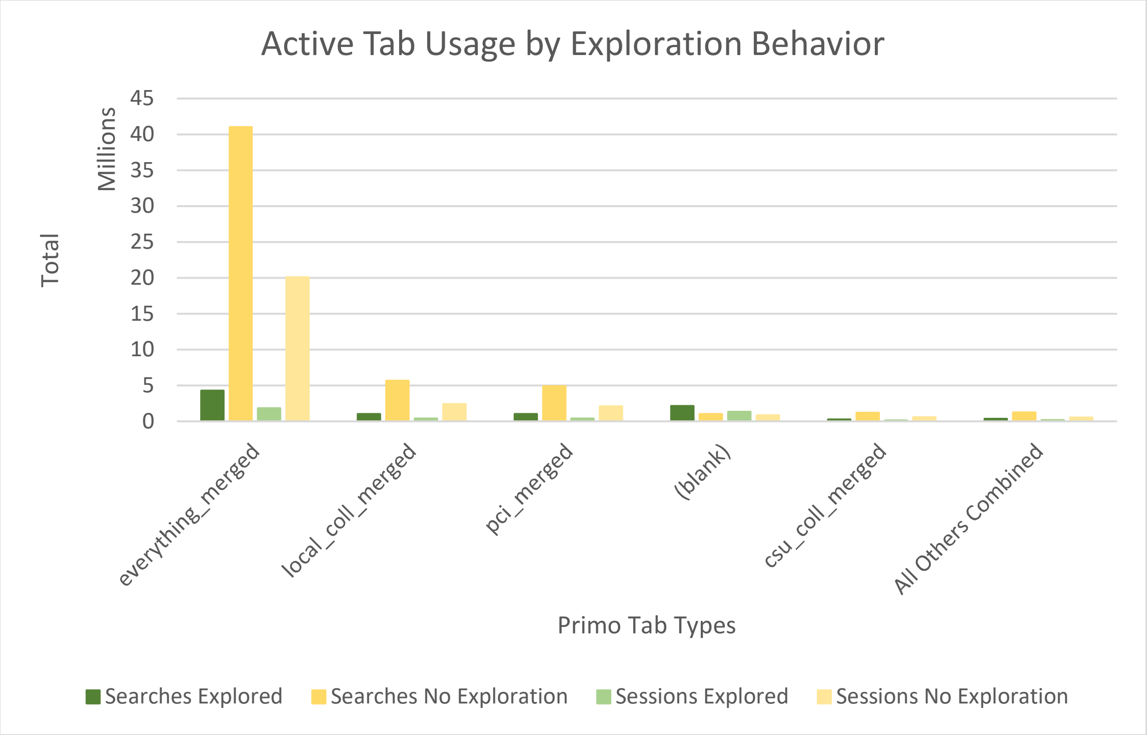
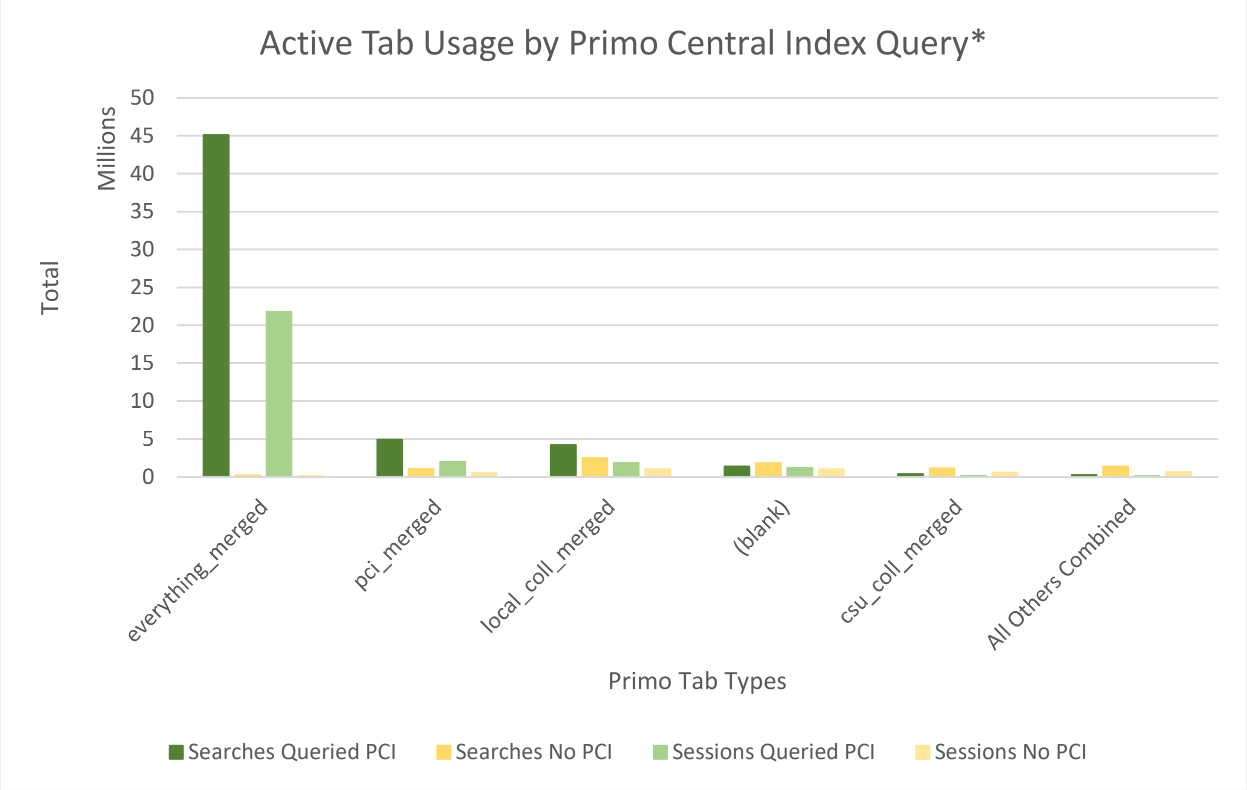


Table 8: Distribution Statistics for Searches and Sessions by Primo Central Index Query

Usage Behavior	Total Activity	Mean		Standard Deviation	Skewness		Kurtosis	
		M	*SE*		Stat.	*SE*	Stat.	*SE*
Searches Queried PCI (*n* = 17,403 cases)	56,100,636	3223.62	302.59	39,917.66	28.19	.02	1,070.86	.04
Searches No PCI Query (*n* = 13,628 cases)	7,958,296	583.97	37.82	4,414.82	19.72	.02	554.96	.04
Sessions Queried PCI (*n* = 17,403 cases)	27,016,401	1552.40	141.06	18,608.35	24.92	.02	784.36	.04

Sessions No PCI Query (*n* = 13,628 cases)	3,805,651	279.25	16.74	1,953.74	18.87	.02	511.89	.04

Figure 12: Active Tab Usage by Primo Central Index Query 2017 - 2019



* Readers will note that some Searches and Sessions marked as 'No PCI' occurred for Active_Tab values for pci_merged. This might seem impossible but the dichotomous grouping variable to separate the two distributions was dependent upon the Active_Tab's Search_Scope_Type being 'blended' or 'pci/deep search', meaning a mix of Primo Central Index content or only that content, respectively. Through post-search filtering (such as using the Institution filter or the availability filter to limit to a user's local holdings) it is technically possible to execute searches that do not query PCI from Tabs that are 'blended'.

User Group Query Behavior

As noted above, our dataset held two usage metrics pulled from Primo Analytics without any researcher intervention, Searches and Sessions. User Group categories queried Primo at different rates. This was learned through a Kruskal-Wallis test between groups on the Searches (H = 3448.79, df = 50, p < .00) and

Sessions (*H* = 3422.57, *df* = 50, *p* < .00) variables. Though the magnitudes of Searches and Sessions differed, the general shapes of their distributions were similar as can be seen in Figures 13, 14, 15, & 16. The User Groups with the most usage in descending order were: undergraduate students, graduate students, faculty, then staff. The usage distribution for both Searches and Sessions was skewed towards usage by those groups with most other User Group categories making up comparatively little of total Primo usage. Results are visible in multiple figures below; the full long tail is available in the raw data online.

Given the differences in background knowledge and potential goals, it might be expected that User Groups would query Primo differently. When comparing User Group behavior, as measured in Searches and Sessions, across the dichotomous grouping variables and synthetic variables, the unsurprising finding that User Groups queried Primo at different rates continued to hold. Kruskal-Wallis tests were run on Sessions and Searches counts comparing User Group behavior for: pre-search filters querying Primo Central Index (or not), queries originating from the homepage, whether the default search from the homepage was used, and a synthetic grouping variable that captured ‘exploration’. That last variable captured any usage of the Browse feature, Journal search, Newspapers search, or Advanced Search. However, graphical analysis did not reveal any substantial differences in User Group behavior along the four grouping dimensions. Table 9 displays Kruskal-Wallis results for the tests on pre-search filters querying Primo Central Index, queries originating from the homepage, those using the homepage default, and the user ‘exploration’ variable. Our data cannot say whether there were differences in post-search behavior, such as facet/filter usage, between User Groups.

Table 9: Kruskal-Wallis Results for User Group Query Behavior Across Comparative Distributions

Behavior	Status & Cases	Searches			Sessions		
		H Stat.*	*df*	*p*	*H* Stat.*	*df*	*p*
User Group explored	Yes (*n* = 14,454 cases)	1717.81	40	< .00	1919.83	40	< .00
	No (*n* = 16,577 cases)	1845.15	43	< .00	1707.37	43	< .00
User Group queried PCI	Yes (*n* = 17,403 cases)	2079.52	43	< .00	1952.78	43	< .00
	No (*n* = 13,628 cases)	1539.99	45	< .00	1706.36	45	< .00
User Group referrer was Homepage	Yes (*n* = 9,850 cases)	1976.59	44	< .00	2343.39	44	< .00
	No (*n* = 21,181 cases)	1878.53	48	< .00	1601.64	48	< .00
User Group chose default on Homepage	Yes (*n* = 883 cases)	138.98	38	< .00	164.51	38	< .00
	No (*n* = 30,148 cases)	3449.12	49	< .00	3450.37	49	< .00

* All *H* values are adjusted for ties in pairwise comparisons.

Figure 13: User Group Searches and Sessions by Exploration 2017 - 2019

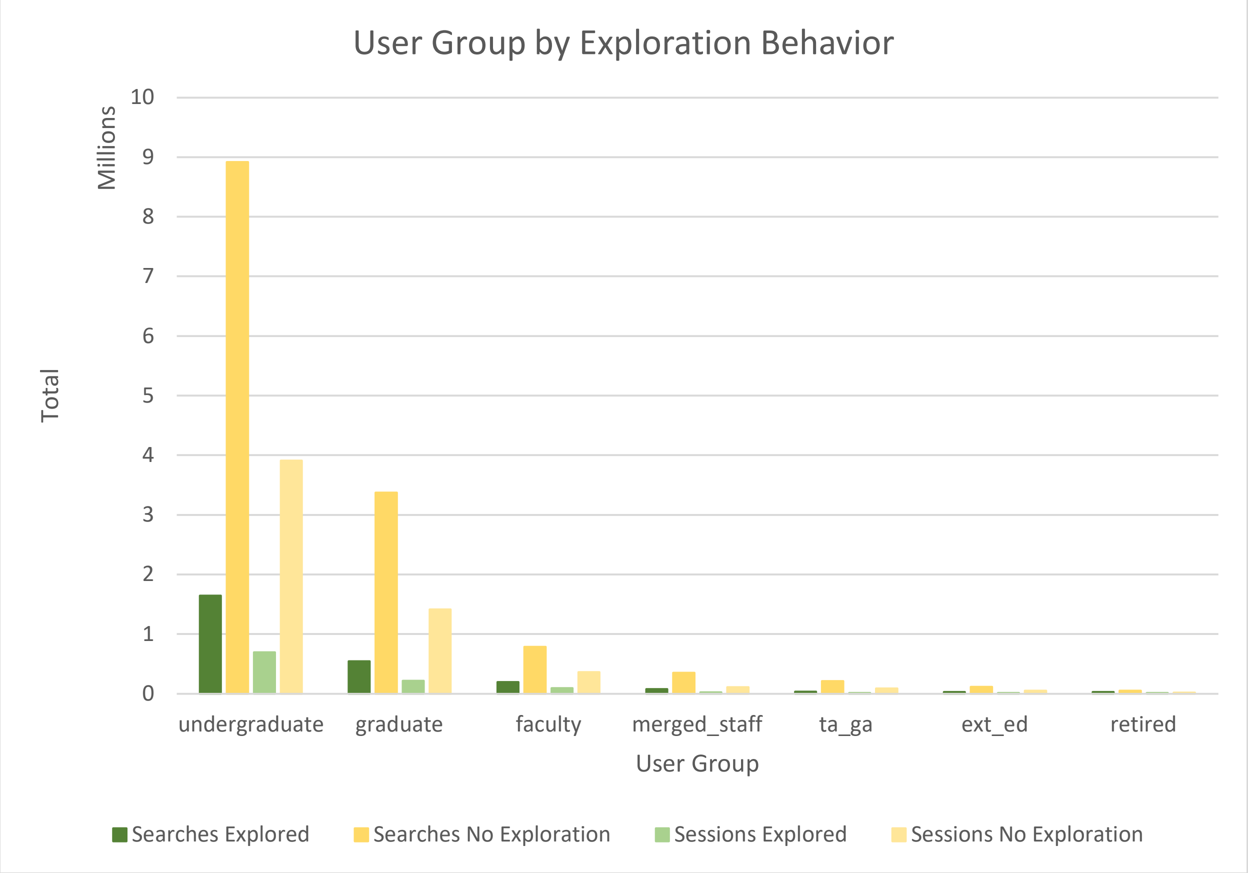


Figure 14: User Group Searches and Sessions by Primo Central Index Query 2017 - 2019

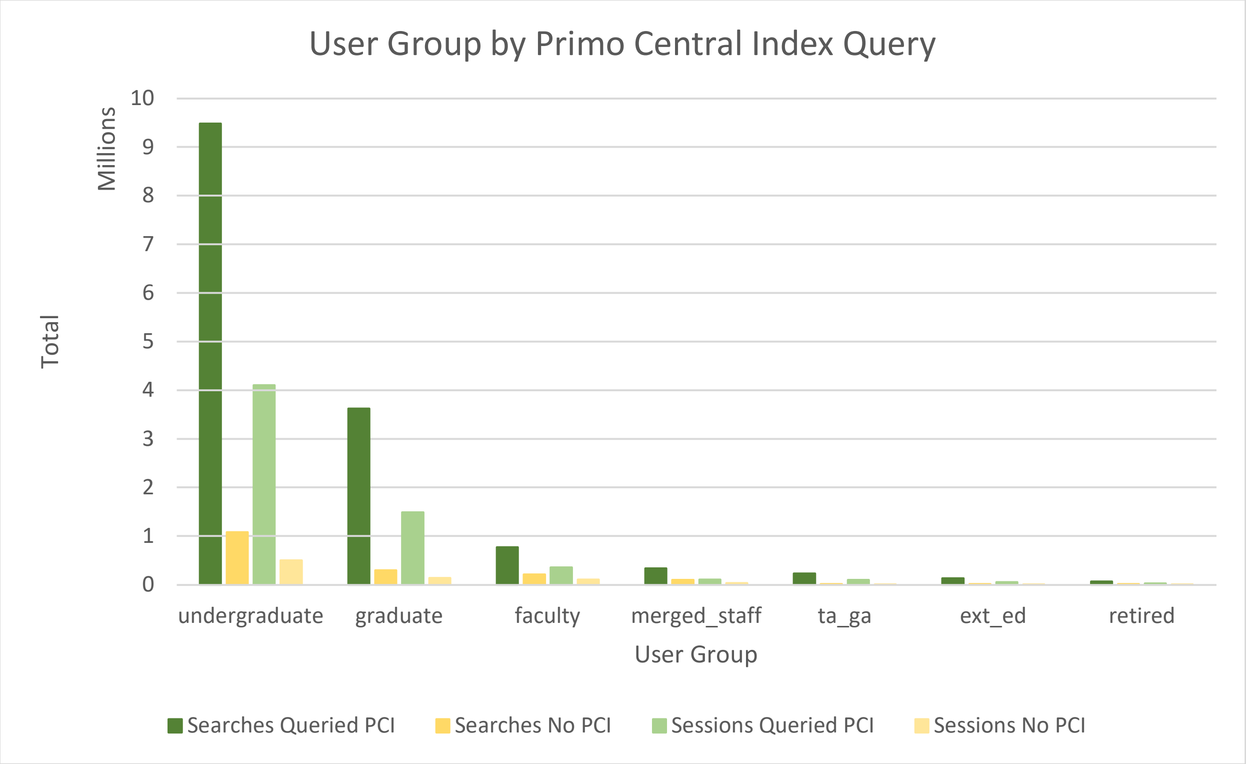
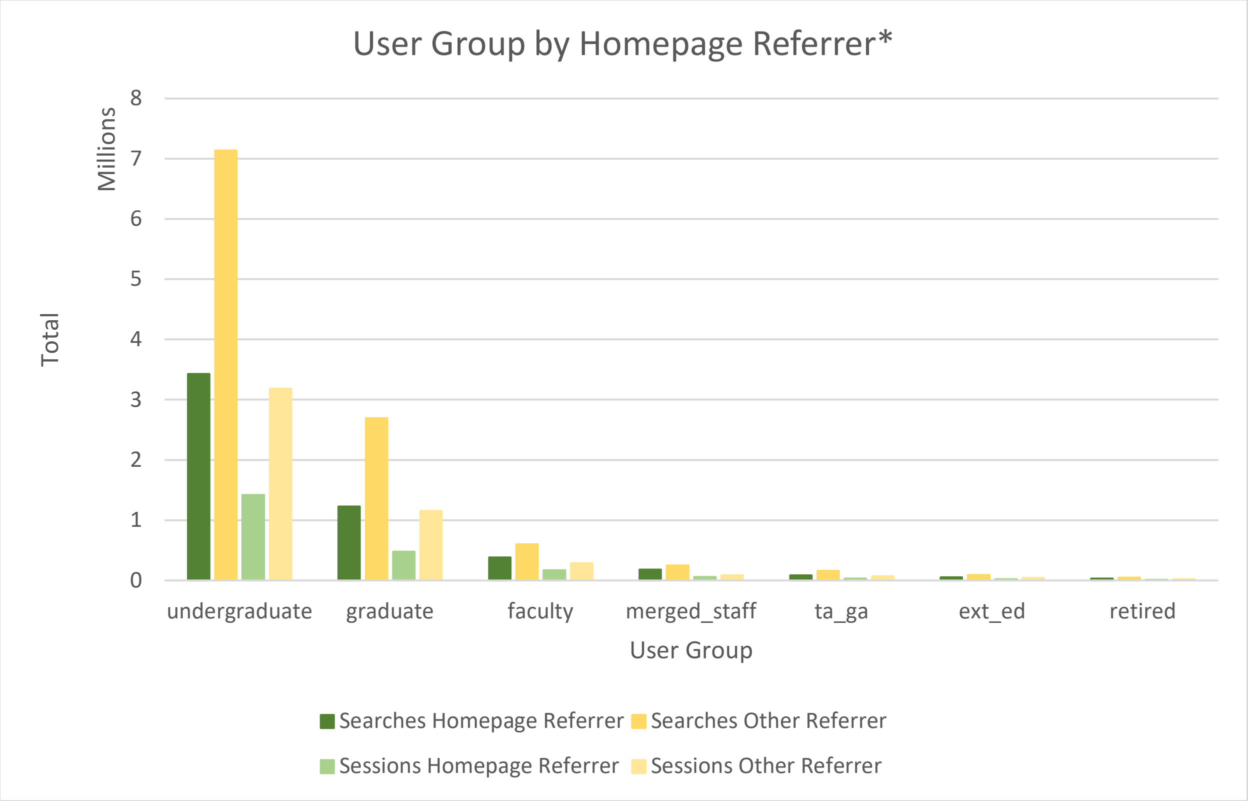
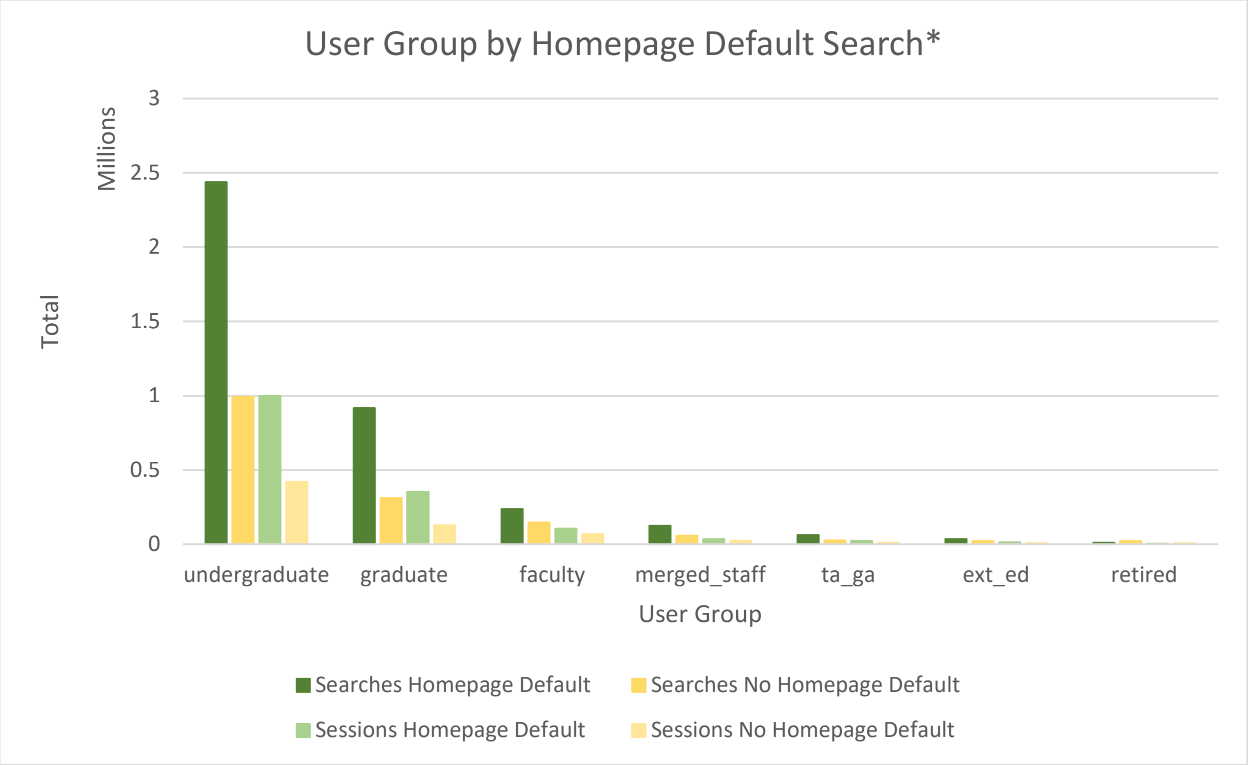


Figure 15: User Group Searches and Sessions by Homepage Referrer 2017 - 2019



* The majority of Sessions (77%) were generated by the User Group values ‘guest’, an Ex Libris code indicating users who do not sign in, and ‘undefined’, the researcher-assigned variable to account for blank values (Ex Libris, 2019). A full picture of the data aggregated to include those values and broken down by campus is in Figure 6.

Figure 16: User Group Searches and Sessions by Homepage Default Search 2017 - 2019



*Shown for cases with the synthetic variable value of ReferrerBlankORHomepage based on HTTP Referrer data. See also notes for Figure 6 and Figure 15.

The Discussion

Discussion topics are drawn from current themes introduced within the literature review to understand our findings under the context of other impacting factors. Broadly, there are three themes. First, the dominance of Basic Search over Advanced Search and other types of searching as well as browsing features. Second, the lack of substantial differences in search type and pre-filtering among User Groups. Third, the importance of a library homepage as an entry point into the discovery layer with the associated finding that default settings were used in most circumstances. While data alone cannot fully explain users' rationale and motivations, many recent studies have reported similar findings related to users searching behaviors. The present study, the largest of its kind, shows the power of defaults and that most users do not choose to take advantage of more precise pre-search options when available. When our results are synthesized with other studies, it lends support to efforts aimed at simplifying search interfaces and reducing the number of decisions a user must make before searching. That overarching conclusion is a prominent theme related to web design best practices and congruent with our three broad findings. Regarding search pre-filters, users seem to perform the search process almost exclusively based off previous search experiences in Google, perhaps even being unaware of this ingrained search literacy. Our findings cannot speak to what happens post-search on a discovery layer search results page, which is an avenue for future research. Many of the studies referenced within the literature review explore searching as a process, including the difficulty in honing the ability to correctly identify terms to combine into a short phrase that will return the most relevant information. Lowe et al. (2018) note that “asking users to translate a complex question into keywords is an oversimplification of their information need” (p. 519). Under the context that users come to the library’s website looking for information, most users seem to want to type in a few keywords relating to their inquiry and receive a list of relevant results. Search box configurations are an essential aspect of meeting users at their point of need. Al-Qallah & Ridha (Al-Qallaf and Ridha, 2019) developed a Library Website Evaluation Checklist (L-WEC) through a content analysis of 110 academic library websites, mentioning “one significant finding was that students strongly desired digital content but preferred to use Google and other search engines over library websites. Other common themes identified were: (1) an increasing demand to download e-journal articles; (2) a preference for discovery services; and (3) the desired ability to search and retrieve information in the fewest possible steps.” While there is no way to predict what users are specifically seeking, understanding *how* they search leads to designing system interfaces that better respond to users’ expectations.

The diverse scope configurations at each campus supplied multiple variable scenarios for users to encounter the Primo system; this comparable data demonstrates the first theme of the dominance of simple Basic Search and relative lack of use of other search scopes and advanced modes. Although this study was nonexperimental, the examples of CPSU San Luis Obispo, San Diego State University, and CSU Channel Islands when compared with the other CSU libraries are instructive; they are exceptions that prove a general rule. In the case of CPSU San Luis Obispo, when placed in a search environment permitting no pre-filtering (unless via Advanced Search) and given the option on every basic results page of toggling over to gain query features by using Advanced Search, few (5%) users do so. In the case of SDSU and CSUCI, even though all Primo queries originating on their homepage defaulted to the Advanced Search, Basic Search queries still accounted for the majority of all search activity. Therefore, the majority of all queries performed by users were of the Basic Search type and this fact was invariant across all pre-filtering configurations observed in this study; most users have a clearly demonstrated preference for a simple query box that imitates commercial web search providers such as Google. This conclusion confirms results obtained by less robust studies using smaller samples (Galbreath et al., 2018; Kliewer et al., 2016; Wells, 2016; Zavalina and Vassilieva, 2014).

The second broad theme was a lack of difference in User Group behavior in how they approached Primo, whether they queried Primo Central Index, and whether they engaged in ‘exploratory’ selection of search options. The distributions of Searches and Sessions across the various permutations of these questions all showed the same general patterns of having a high variance, positive skewness indicating a right-tailed distribution, with positive and generally high kurtosis indicating substantial outliers present in the tails. Graphical analysis confirmed the similarity between User Group values. This can be generalized to a claim about super- or power-users; whatever differences there are otherwise between an experienced professor and an undergraduate, they use search pre-filter options similarly. Certainly, there are differences in post-search behavior and interpretation of search results pages, but those issues were not captured in our dataset. What about the unique needs of a “power user” or “advanced researcher” such as an experienced Faculty member or graduate student? All readers of this article know that such a population exists (as we are a part of it) and requires more out of a discovery layer than what undergraduate students expect out of commercial web search. A 2018 study found in the specific domains of legal research, healthcare, and patent analysis the need to execute complex multi-line Boolean queries is particularly acute (Russell-Rose et al., 2018). There is a clear use case and need for such features as Primo's Advanced Search, offering multiple query boxes and explicit Boolean logic, and the ability to conduct Browse searches over indices such as Library of Congress Subject Headings or call numbers. While most users did not use them, this does not mean that advanced features should be hidden or unavailable.

Librarians often prefer Advanced Search with the goal of helping students to understand how to build a better search for more relevant results and be better equipped for database-style searching. But are we working under a false assumption? Studies done on student understanding and usage of Boolean logic and operators can answer that question and inform design. Comparing unfiltered Boolean, filtered Boolean, unfiltered natural language, and filtered natural language queries across 8 important databases, Lowe et al (Lowe et al., 2018) found remarkable similarities between Boolean and natural language results using a relevancy rubric. Furthermore, when looking at the subset of results with high relevancy, the overlap in results between Boolean and natural language queries was high and only in one instance was below 50%. In other words, users will tend to find the same highly relevant results whether they use Boolean or natural language searching. Primo was not included in Lowe et al.'s 2018 study but assuming their results hold for Primo, this further supports the simplification of an initial encounter with a search interface. In a study comparing relevance and overlap of simple Boolean and advanced Boolean Lowe et al. (Lowe et al., 2020) found that simple Boolean actually outperformed advanced Boolean overall. When one considers the fact that the Basic Search mode in Primo returns the exact same results for explicitly typed Boolean queries using nested parentheses as it does when those same queries are executed using the Advanced Search mode, the case against having a complex search interface on a homepage is closed.

Users who wish to query with explicitly typed simple or advanced Boolean may do so, leaving many users to query a simple search box with natural language strings.

Every academic librarian knows that there are differences in keyword production and query formation between novices and seasoned searchers. This anecdotal experience has been confirmed empirically across many studies and is a multi-causal phenomenon involving intelligence/cognitive ability (Hsieh-Yee, 2001; Naghib et al., 2020; Tella et al., 2017) and background domain knowledge (Hsieh-Yee, 1993; Sanchiz et al., 2017) at a minimum. However, little research exists on faculty or power-user use of pre-filtering and general query approach (as opposed to specific query string analysis). This present study is the largest of its kind to compare faculty (i.e., experienced researcher) to undergraduate (i.e., novice) use of search pre-filters. As noted above there were no substantive differences in how the various types of users approached their search, either in terms of arriving at the search results page or in their use of pre-filters. Further null findings included the hypotheses that users might query differently when on or off campus and when authenticated or not.

Our third finding regarding query origination points to the power of choice architecture and has implications for library homepage design outside of a discovery layer such as Primo. It is important to realize that people can enter Primo from many other places: a bookmark, a link on a research guide, the Learning Management System, or via the link resolver in a subscription database. For fourteen of the 24 instances, all other methods of entry combined account for less than 30% of traffic into Primo while the library homepage provides the majority. In no campus instance was the homepage's share of traffic below 52% (Los Angeles). (Data for Moss Landing Marine Laboratories should be disregarded because for the majority of the study period they lacked a query box form on their library homepage; instead, a link to Primo was present.) This shows how incredibly important a homepage is for directing traffic into the catalog or discovery layer. Given the sample size here, there is good reason to believe this general finding holds for all academic libraries that allow for a query of their catalog from the homepage. Furthermore, the configuration of the homepage query box exerts strong effects on user behavior. In this study, there were only five Primo instances where the majority of all Searches were not run according to the default configuration of the query box on a library's homepage. The well-known truths of the power of habit and default adoption explain the highly skewed distributions of usage for Primo Tabs (and their component Scopes) observed in Figures 4 & 5. The fact that at 18 campuses over 60% of all search activity from the homepage just used the default settings shows the power of choice architecture and should make library UX designers think carefully about just what the default settings are. A library's homepage and the design choices composing it are incredibly important.

Three Limitations

Relying on Primo Analytics for the majority of our data necessitated a high-level aggregate view; this limited the range of questions we could ask. Also, as noted above, the context of information-seeking search behavior was abstracted away in the absence of data about query terms. This study could have been improved by including qualitative feedback from users via a usability study. However, researchers worked at different campuses and wanted to forgo any complications related to needing IRB approval for such a research study. The lack of direct contact with users restricted the questions we could answer. However, the questions we examined required researcher intervention to pose and answer given the data quality. The nonparametric statistical tests conducted cannot account for demonstrated correlation between variables; future work on this subject should develop a richer model and examine more independent variables with more rigorous statistical testing. We consider three subsets of the issues we regard as limitations around data quality, and ideas for future research, below.

Inconsistent Public Terminology

Despite the standardization of several aspects of Primo and Alma across the CSU Libraries, the fact remains that different libraries used terminology inconsistently in their Primo instances. Most relevant for this study is that Primo Tabs and their associated Scopes carried different names across campuses. The Tab querying local collections of metadata (excluding Primo Central Index) had idiosyncratic names. These local collections Tabs produced inconsistent results as well; some included holdings from an institutional repository or their course reserves software while other libraries segregated such content into either separate Tabs, excluded it from Primo altogether, or bundled it into their “Everything” Tab. Many CSU libraries offered a search named “Everything”, always a literal misnomer given that no Scope could truly query every thing and which produced additionally inconsistent behavior depending on whether the “Expand My Results” toggle was activated by default or not. The “Expand My Results” feature widens the search scope to include Ex Libris’ Central Discovery Index (still called Primo Central Index during the study period) which contains items not directly owned or licensed by the library, but which can be requested through interlibrary loan (ILL) requests (Ex Libris, 2016). When approaching search scope configuration, what is a true *everything* search? Is it everything the library owns and licenses or everything the library can eventually get? While the policy in all CSU libraries is ILL requests are free to users, document delivery and interlibrary loan services are not costless. Libraries pay the cost to borrow and ship other libraries’ materials as well as staffing overhead. The CSU libraries have established their own lending network, marketed as CSU+, and share resources between campus libraries at no charge. This specialized lending network is highlighted at most campuses with multiple scope configurations as its own sub-collection of books and resources from all CSU libraries. While this study is solely focused on one distinct user behavior, specifically using a dropdown pre-filter, future research using Primo Analytics data included in the Facet Usage subject area could also include post-search filters with facet use and the Primo “Expand My Results” checkbox. Such a study would also suffer from the fact that post-search filters and facets also lack standard terminology across institutions. Because of the richness in motivations for post-search filtering and faceting, research on them would need to include qualitative and quantitative feedback from users and not rely only on Primo Analytics.

When pre-search options are limited, the user experience with post-search options becomes more important.

Primo Analytics as a Data Source

Quite apart from the self-inflicted data quality issue of inconsistent terminology across institutions in the same network, our Primo Analytics data was marred by two software defects. Before the August 2021 release of Primo (including Primo Analytics), two issues, bugs, affected data entry and resulted in blank values being recorded. Specifically, the variables affected were User Group and Active Tab. Our Referrer data also had many blanks, which not being due to a bug we address separately below.

Before August 2021, Primo Analytics allowed for a blank User Group value to be recorded. We first noticed this in cases where blank User Group values were present for a substantial portion of the On Campus data. Off Campus data with blank User Groups only accounted for 3,477 Sessions in which 5,373 Searches were performed. In all, 41% of our Search activity data was recorded as User Group = blank. That gave researchers pause; we subdivided these into two groups: 'no_sign_in' and 'undefined'. User Group 'no_sign_in' was assigned to all blank entries where the Signed In variable was 0, indicating the user did not authenticate and Primo Analytics recorded a blank. User Group 'undefined' was assigned to all blank entries where the Signed In variable was non-zero, indicating users authenticated but Primo Analytics failed to capture their Alma User Group code. As readers may wonder about the justifiability of these synthetic research-created user groups, we note that Kruskal-Wallis *H*-testing was conducted a second time on all the variables and comparisons noted above but with the 'no_sign_in' and 'undefined' cases excluded; while *H* statistics differed, statistical significance did not. Our labeling intervention in the data was thus tested and did not affect the study outcome; for brevity, we reported only 1 set of testing above.

The Active_Tab value was blank for 2,168,405 Sessions in which 3,163,253 Searches were performed. These are large numbers but represent a small part of the overall sample, 7% of all Sessions accounting for 5% of all Searches. Blank data appeared seemingly at random; Active_Tab might have been blank while all other numeric and descriptive string variables in a row might have values. Similarly, much Referrer data was blank as noted above; these blank values were often the only ones in their respective case row as exported from Primo Analytics. Since there was no discernable pattern in blank values, we did not discard any instances with blank data. In one sense this inflated our sample size but there was no one single non-numeric variable other than Active_Tab where blank values would change the findings about pre-filter search usage. As just noted, those blank values only account for 5% of all Searches and we consider the results robust despite the absence of a full picture due to software defects in Primo Analytics.

Researcher Assumptions

Faced with much blank Referrer data, researchers resorted to an ad hoc hypothesis which then sent us to the Internet Archive's Wayback Machine in search of old website layouts and HTML code inside <form> elements. Some readers may have found the logic behind those assumptions as detailed above unpersuasive. There is no unambiguous evidence for or against the theory that blank Referrer data in Primo Analytics arises when a query originates from a form where the form tag had an action attribute pointing to Primo and a target attribute with the value _self. Perhaps users prefer opening Primo in a new tab or from a bookmark? Perhaps there were other Primo query forms not on the homepages that contained rel=noreferrer code? We cannot rule out these possibilities. Given that a full 50% of Sessions contained blank Referrer values it is certainly possible that there was heterogeneity in the blank Referrers and it was not all created through the same data-generating process. To the extent that readers doubt our comingling of blank Referrer and verified homepage URL Referrer data, that doubt must lower the credence readers have in our findings about query origination.

Although some organizations' landing pages saw little change, many institutions altered some aspects of their homepage in some way from 2017-2019. Due to the inherent nature of digital information and recent CSU systemwide migration to Primo around this period, it is impossible to fully grasp the impact these refinements may have caused on the results of this study. Also, the present research study is exclusively based on quantitative data and does not directly incorporate user perspectives or conceptual frameworks for the evaluation of information retrieval systems. These findings could be contextualized by incorporating qualitative measures such as usability studies, focus groups, or surveys. Allowing users to articulate the reasons behind their preference for a single search box would add insight towards developing a better understanding of these results.

What has Happened?

Though this study was nonexperimental, through the power of a large sample size, a clear and convincing picture of user behavior has been presented. The data from intra-institutional and inter-institutional comparisons when combined with the extant literature on usability testing suggest reducing cognitive overload improves the user experience for students, faculty, and librarians across a consortium. When search options were highly constrained by the default choice architecture, but complex searches were possible, few users opted out of the default simple search. A straightforward interpretation of the data would be that the needs of librarians and power user faculty must be balanced with the fact that most users are novices. Users of all abilities are largely habituated to commercial search products which emphasize post-search results filtering. Can we reconsider and articulate our purpose within this technological arena to rebrand discovery layers as distinct in their search of exclusively curated resources and move away from what is easily discovered on Google? One solution could be embracing the clear differences and marketing Primo as a *research tool* that is supplemental to Google and Google Scholar, potentially having a more reputable sub-collection with the ability to request other authoritative and credible resources from other libraries and academic institutions. Much of the literature supports users' preference for easy-to-find over credible sources, but ideally, these would be one and the same within a library's discovery layer. Our present findings are robust and congruent with similar studies within this field; other academic libraries may use this growing body of evidence to inform decisions related to layout, design, and instructional practices related to these search tools. Libraries should make an active investment to improve the UX of library discovery layers, portals, and gateways, specifically central landing pages like homepages. Designing for all users involves considering a diverse userbase with a wide range of abilities, keeping edge cases in mind, and building from proven web design principles.

Acknowledgements

We thank the members of the California State University Unified Library Management System Discovery Committee (academic year 2021 – 2022) for encouraging this research and for offering productive criticism of the preliminary results; in alphabetical order: Ian Chan (Vice-Chair), Christine Holmes, Hannah Lee, Cathy Outten, Julie Shen, David Walker, and Christian Ward (Chair). Brandon Dudley at the California State University Office of the Chancellor is a steady hand at Alma and Primo Analytics in Oracle Business Intelligence Enterprise Edition and supplied our raw data from each campus library. Timothy T. Strawn and Brandon Dudley improved this article with their comments; any errors remaining are our own. Lastly, we appreciate the two anonymous reviewers assigned by Editor Anne Goulding; their pointed criticism made the present iteration simultaneously much stronger and more humble than our original submission to *JOLIS*.

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Data Availability Statement

The data that support the findings of this study are openly available in figshare at 10.6084/m9.figshare.19071578

References

Adams AL and Hanson M (2020) Primo on the Go: A Usability Study of the Primo Mobile Interface. *Journal of Web Librarianship* 14(1–2). Routledge: 1–27. DOI: 10.1080/19322909.2020.1784820.

Ahmed SMZ, McKnight C and Oppenheim C (2009) A review of research on human-computer interfaces for online information retrieval systems. *The Electronic Library* 27(1): 96–116. DOI: 10.1108/02640470910934623.

Al-Qallaf CL and Ridha A (2019) A Comprehensive Analysis of Academic Library Websites: Design, Navigation, Content, Services, and Web 2.0 Tools. *International Information & Library Review* 51(2): 93–106. DOI: 10.1080/10572317.2018.1467166.

Breeding M (2014) Web-Scale Discovery Services. *American Libraries*, 1 February. https://americanlibrariesmagazine.org/2014/01/14/web-scale-discovery-services/

Chapman S, Fry A, Deschenes A, et al. (2016) Strategies to Improve the User Experience. *Serials Review* 42(1): 47–58. DOI: 10.1080/00987913.2016.1140614.

Dease N, Villaespesa E and MacDonald CM (2020) Working together: Using student-driven UX projects to improve library websites. *College & Undergraduate Libraries* 27(2–4): 397–419. DOI: 10.1080/10691316.2021.1888838.

Ex Libris (2016) Performing Basic Searches. Available at: https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Primo/End_User_Help_-_New_UI/010Performing_Basic_Searches (accessed 1 March 2022).

Ex Libris (2017) Blank Search Scope Type in Primo Analytics. Available at: https://knowledge.exlibrisgroup.com/Primo/Knowledge_Articles/Blank_Search_Scope_Type_in_Primo_Analytics (accessed 5 October 2021).

Ex Libris (2019) Primo Action Usage. Available at: https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Primo/Analytics/Primo_Analytics_Subject_Areas/Primo_Action_Usage (accessed 5 October 2021).

Fidel R, Pejtersen AM, Cleal B, et al. (2004) A Multidimensional Approach to the Study of Human Information Interaction: A Case Study of Collaborative Information Retrieval. *Journal of the American Society for Information Science and Technology* 55(11): 939–953. DOI: 10.1002/asi.20041.

Fu Y, Lomas E and Inskip C (2021) Library log analysis and its implications for studying online information seeking behavior of cultural groups. *The Journal of Academic Librarianship* 47(5): 1–11. DOI: 10.1016/j.acalib.2021.102421.

Galbreath BL, Johnson C and Hvizdak E (2018) Primo New User Interface: Usability Testing and Local Customizations Implemented in Response. *Information Technology & Libraries* 37(2): 10–35. DOI: 10.6017/ital.v37i2.10191.

Gilmore T, Metko S and Gilbert C (2017) Sailing the Wide-Open Seas of Discovery: Assessing Students' Use and Perceptions of Summon for Conducting Research. In: *At the Helm: Leading Transformation: The Proceedings of the ACRL 2017 Conference*, Baltimore, MD, 22 March 2017, pp. 544–557. Association of College and Research Libraries. Available at: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2017/SailingtheWide-OpenSeasofDiscovery.pdf (accessed 21 March 2017).

Greenberg R and Bar-Ilan J (2017) Library metrics – studying academic users' information retrieval behavior: A case study of an Israeli university library. *Journal of Librarianship and Information Science* 49(4): 454–467. DOI: 10.1177/0961000616640031.

Guo J, Fan Y, Pang L, et al. (2020) A Deep Look into neural ranking models for information retrieval. *Information Processing & Management* 57(6): 1–20. DOI: 10.1016/j.ipm.2019.102067.

Gusenbauer M and Haddaway NR (2020) Which Academic Search Systems Are Suitable for Systematic Reviews or Meta-analyses? Evaluating Retrieval Qualities of Google Scholar, PubMed, and 26 Other Resources. *Research Synthesis Methods* 11(2): 181–217. DOI: 10.1002/jrsm.1378.

Hamlett A and Georgas H (2019) In the Wake of Discovery: Student Perceptions, Integration, and Instructional Design. *Journal of Web Librarianship* 13(3): 230–245. DOI: 10.1080/19322909.2019.1598919.

Hsieh-Yee I (1993) Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science* 44(3): 161–174. DOI: 10.1002/(SICI)1097-4571(199304)44:3<161::AID-ASI5>3.0.CO;2-8.

Hsieh-Yee I (2001) Research on Web search behavior. *Library & Information Science Research* 23(2): 167–185. DOI: 10.1016/S0740-8188(01)00069-X.

Hu PJ-H, Ma P-C and Chau PYK (1999) Evaluation of user interface designs for information retrieval systems: a computer-based experiment. *Discovery Support Systems* 27(1–2): 125–143. DOI: 10.1016/S0167-9236(99)00040-8.

Huvilla I, Enwald H, Eriksson-Backa K, et al. (2022) Information behavior and practices research informing information systems design. *The Journal of the Association for Information Science and Technology* 73(7): 1043–1057. DOI: 10.1002/asi.24611.

Indiana University Center for Postsecondary Research (2018) Carnegie Classifications 2018 public data file. Indiana University Center for Postsecondary Research. Available at: http://carnegieclassifications.iu.edu/downloads/CCIHE2018-PublicDataFile.xlsx (accessed 22 January 2022).

Kelly D and Sugimoto CR (2013) A Systematic Review of Interactive Information Retrieval Evaluation Studies, 1967–2006. *Journal of the American Society for Information Science and Technology* 64(4): 745–770. DOI: 10.1002/asi.22799.

Kliewer G, Monroe-Gulick A, Gamble S, et al. (2016) Using Primo for undergraduate research: a usability study. *Library Hi Tech* 34(4): 566–584. DOI: 10.1108/LHT-05-2016-0052.

Li Y and Liu C (2019) Information Resource, Interface, and Tasks as User Interaction Components for Digital Library Evaluation. *Information Processing & Management* 56(3): 704–720. DOI: 10.1016/j.ipm.2018.10.012.

Lopatovska I and Arapakis I (2011) Theories, methods and current research on emotions in library and information science, information retrieval and human–computer interaction. *Information Processing & Management* 47(4): 575–592. DOI: 10.1016/j.ipm.2010.09.001.

Lowe MS, Maxson BK, Stone SM, et al. (2018) The Boolean is Dead, Long Live the Boolean! Natural Language versus Boolean Searching in Introductory Undergraduate Instruction. *College & Research Libraries* 79(4): 517–534. DOI: 10.5860/crl.79.4.517.

Lowe MS, Stone SM, Maxson BK, et al. (2020) Boolean redux: Performance of advanced versus simple boolean searches and implications for upper-level instruction. *The Journal of Academic Librarianship* 46(6): 102234. DOI: 10.1016/j.acalib.2020.102234.

Muglia C and Namei ES (2017) Academic Libraries, Filtering, and the “Tyranny of Choice”. In: *At the Helm: Leading Transformation: The Proceedings of the ACRL 2017 Conference*, Baltimore, MD, 22 March 2017, pp. 1–14. Association of College and Research Libraries. Available at: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2017/AcademicLibrariesFilteringandtheTyrannyofChoice.pdf (accessed 28 March 2017).

Naghib F, Mirzabeigi M and Alborzi M (2020) The role of spatial intelligence in predicting web information searching behavior and performance of high school students. *Library Hi Tech* 39(1): 48–63. DOI: 10.1108/LHT-07-2019-0139.

Namei ES and Young CA (2015) Measuring Our Relevancy: Comparing Results in a Web-Scale Discovery Tool, Google & Google Scholar. In: *Creating sustainable community: The Proceedings of the ACRL 2015 Conference*, Portland, OR, 25 March 2015, pp. 522–535. Association of College and Research Libraries. Available at: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2015/Namei_Young.pdf (accessed 28 March 2017).

Ndumbaro F (2018) Understanding user-system interactions: An analysis of OPAC users’ digital footprints. *Information Development* 34(3): 297–308. DOI: 10.1177/0266666917693885.

Pickard E and Desilets MR (2020) Accidental Information Literacy Instruction: The Work a Link Landing Page Can Do. *Scholarship of Teaching and Learning, Innovative Pedagogy* 2(1): 16–27. https://digitalcommons.humboldt.edu/sotl_ip/vol2/iss1/2/

Porat L and Zinger N (2018) Primo New User Interface—Not Just for Undergrads: A Usability Study. *Journal of Library User Experience* 1(9): 1–17. DOI: 10.3998/weave.12535642.0001.904.

Reidsma M (2019) *Masked by Trust: Bias in Library Discovery*. Sacramento, CA: Litwin Books.

Ruiz M, Harris C and Worden L (2022) Re: Curiously High Advanced Search Usage in Primo at CI. [Personal Communication].

Russell-Rose T, Chamberlain J and Azzopardi L (2018) Information retrieval in the workplace: A comparison of professional search practices. *Information Processing & Management* 54(6): 1042–1057. DOI: 10.1016/j.ipm.2018.07.003.

Sanchiz M, Chin J, Chevalier A, et al. (2017) Searching for information on the web: Impact of cognitive aging, prior domain knowledge and complexity of the search problems. *Information Processing & Management* 53(1): 281–294. DOI: 10.1016/j.ipm.2016.09.003.

Scarnò M (2012) User's behaviour inside a digital library. In: Zaraté P (ed.) *Integrated and Strategic Advancements in Decision Making Support Systems*. Hershey, PA: IGI Global, pp. 138–146. DOI: 10.4018/978-1-4666-1746-9.ch011.

Swanson DR (1977) Information Retrieval as a Trial-And-Error Process. *The Library Quarterly: Information, Community, Policy* 47(2): 128–148. DOI: 10.1086/620653.

Tella A, Anyim O, Memudu SA, et al. (2017) Predictors of Information Retrieval Effectiveness among Library and Information Science Undergraduates in Kwara State Universities. *Library Philosophy and Practice*: 1–21. https://digitalcommons.unl.edu/libphilprac/1626/

Vargas Ochoa I (2020) Navigation Design and Library Terminology: Findings from a User-Centered Usability Study on a Library Website. *Information Technology & Libraries* 39(4): 1–15. DOI: 10.6017/ital.v39i4.12123.

Wells D (2016) Library Discovery Systems and their Users: a Case Study from Curtin University Library. *Australian Academic & Research Libraries* 47(2): 92–105. DOI: 10.1080/00048623.2016.1187249.

Xu F and Du JT (2018) Factors influencing users' satisfaction and loyalty to digital libraries in Chinese universities. *Computers in Human Behavior* 83: 64–72. DOI: 10.1016/j.chb.2018.01.029.

Zavalina O and Vassilieva EV (2014) Understanding the Information Needs of Large-Scale Digital Library Users. *Library Resources & Technical Services* 58(2): 84–99. DOI: 10.5860/lrts.58n2.84.

Zhang T (2013) User-Centered Evaluation of a Discovery Layer System with Google Scholar. In: Marcus A (ed.) *Design, User Experience, and Usability. Web, Mobile, and Product Design*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 313–322. DOI: 10.1007/978-3-642-39253-5_34.

[i] Though beyond the scope of this present study, there is convincing evidence that such trust is misplaced. See: (Reidsma, 2019) *Masked by Trust: Bias in Library Discovery*. Sacramento, CA: Litwin Books.