A Longitudinal Study of Database Usage Within a General Audience Digital Library

Iris Xie & Dietmar Wolfram
School of Information Studies
University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA
Email: {hiris, dwolfram}@uwm.edu

Abstract

This study reports on a longitudinal investigation of database usage available through BadgerLink, a general audience digital library available to Wisconsin residents in the United States. The authors analyzed BadgerLink database usage, including EBSCO databases sampled every two years over a six-year period between 1999 and 2005 and four years of usage for ProQuest databases between 2002 and 2005. A quantitative analysis of the transaction log summaries was carried out. Available data included database usage, title requests, session usage by institution, format requests (full-text and abstracts), and search feature usage. The results reveal changes in usage patterns, with relative requests for resources in areas such as social sciences and education increasing, and requests for resources in business/finance and leisure/entertainment decreasing. More advanced search feature usage was also observed over time. Relative usage by searchers affiliated with academic institutions has grown dramatically. Longitudinal analysis of database usage presents a picture of dynamic change of resources usage and search interactions over time. The findings of this study are more in line with results from other online database and digital library environments than Web search engine and Web page environments.

Keywords: transaction log analysis, longitudinal study, digital libraries, information searching

1. Introduction and Literature Review

Transaction log analysis of user interactions with electronic retrieval systems have been undertaken in a variety of environments over the past few decades, and particularly over the past fifteen years with the popularity of Internet-based search services to support information retrieval (IR). Transaction logs provide an unobtrusive way to document and study user interactions with IR systems. In particular, longitudinal study provides an effective way for researchers to understand users' usage of electronic IR systems and change over time. The term "longitudinal" has been defined in different ways by researchers. After analyzing definitions of longitudinal analysis from different researchers, Davis (2006) concluded that longitudinal analysis needs to satisfy the following requirements: the research needs to cover a certain period of time to record the process or change; the research can be in prospective or retrospective direction; the analysis can be conducted qualitatively or quantitatively; multiple units need to be analyzed; data have to be collected more than once.

Researchers have conducted longitudinal studies in different digital environments. In order to compare research of a similar nature, studies that analyze data over a long period of time are included, even though the data are not collected and analyzed in interval times as part of the literature review. Most research to date on longitudinal analysis has focused on Web search engines. In particular, analysis of Excite log data (Jansen, Spink, & Saracevic, 2000; Spink & Jansen, 2004; Spink, Wolfram, Jansen, & Saracevic, 2001) revealed the following usage patterns across time: (1) there is not much change for number of queries per session with average 2.5 queries per session in 1997 and 2.3 queries in 2001); (2) Short queries were formulated by users in different times with average 2.4 terms per query in 1997 and 2.6 terms in 2001; (3) users' view habits do not change either with average 1.7 pages per query in both 1997 and 2001; (4) one of the major changes is that users have increasingly submitted Boolean queries over the study time frame with 5% of queries in 1997 and 10% in 2001; and 5) users' search topics range from entertainment, recreation, and sex to e-commerce.

Organizational Web page searching behavior has been another area for longitudinal study. Wang, Berry, and Yang (2003) analyzed longitudinal user queries submitted to an academic website over a four-year period. They found that the patterns of user queries between academic websites and search engines such as Excite and AltaVista were similar, e.g. most of the queries were unique and short. However, the authors of the study noted academic environments have their own unique query content characteristics, for example, search topics related to career, academics, finances, sports, etc. The longitudinal data presented similar patterns across time on search behaviors and topics even though there were seasonal changes in relation to topics and the number of queries entered. In addition, query length increased slightly over time. Thirty percent of queries consistently resulted in zero hits over the years. Lack of basic IR knowledge and misspelling contributed to a high number of zero hits. Unlike Wang, Berry and Yang's findings, Cothey (2002) did identify changes in users' Web search behavior over a ten month period. Users took a more passive or browsing approach in information searching when they gained experience. Their frequency of Web use also decreased over time. Users also became more diverse in selecting their Web hosts over the study period. Cockburn and McKenzie (2001) focused their research on URL access for users over a four-month period, in particular, users' revisitation rate for websites. They pointed out that even though users visited more sites compared to other studies, there were few common sites visited by users.

Longitudinal analysis has also been applied to online public access catalog (OPAC) environments. Based on transaction log analysis, Blecic, Dorsch, Koeing, and Bangalore (1999) investigated the effects of OPAC screen changes on searching behaviors of users over a four-year period. The findings illustrated search behaviors at four different times including the total number of transactions, search statements, correct syntax searches, keyword searches, etc. Interestingly, they concluded that not all the positive improvement on interfaces for successful searches could be sustained. This indicates the importance of conducting longitudinal studies to examine which types of screen changes can lead to changes of long-term behaviors. Analysis of a Web-based library catalog log covering 479 days was conducted by Cooper (2001). This study differentiated three different types of users: individual users, tourists, and Web spiders. The findings demonstrated that while the length of sessions increased from 6 to 10

minutes, user errors and help requests decreased. Simultaneously, there was no change in terms of the number of searches per session and numbers of display actions per session after users became familiar with the system. However, the also results revealed that there were major differences in search behaviors in different databases, from number of searches per sessions, display behaviors, etc

In a study of online bibliographic databases, Ke, Kwakkelaar, Tai, and Chen (2002) analyzed log data of user behavior in ScienceDirect in Taiwan for approximately eight and a half months. They reported usage patterns over time, query behaviors (query length, query modes, query operators, query fields, query refinement and term occurrences), as well as download behaviors. Yi et al. (2006) analyzed and compared log data from one psychology and two history databases for one month to identify user search behaviors in domain-specific IR systems. Echoing Cooper's (2001) results on different searching behaviors in different online databases, they found that while users applied more conceptual terms in searching PsycINFO, they employed more specific terms for events, people and regions in searching ABC-Clio. However, the average query lengths were similar for online databases in different domains (3.16 for PsycINFO and 3.42 for ABC-Clio).

Transaction log analysis of digital library usage has also been studied. Employing deep log methods of analysis of one million users' requests to a digital library generated between February-March 2003, Nicholas, Huntington, Monopoli, and Watkinson (2006), and Nicholas, Huntington, and Watkinson (2005) explored users' viewing behaviors. To understand their degree of penetration of a system the authors were particularly interested in what a user viewed in terms of time spent, subjects viewed, types of items viewed, number of items viewed, users' geographic locations, etc. Similarly, based on the analysis of the transaction logs of a Korean digital library over two years, Zhang, Lee, and You (2001) found that the usage of the digital library increased across time, and a frequent user group was identified in the second year. They concluded that the search function was the most frequently used system function according to usage statistics.

Usage patterns can also be analyzed to identify research trends and research impact in digital libraries. Bollen, Luce, Vemulapalli, and Xu (2003) analyzed the usage patterns derived from the log analysis of an institution's digital library usage from 1998 to 2001, and compared the usage pattern to Scientific Indexing (ISI) Impact Factor values during the same years to identify the local research trends in the institution. Based on an analysis of a 12-month time series of transaction logs derived from the Alexandria Digital Library, Buttenfield and Reitsma (2002) developed a three-dimensional, origin-destination-time flow/transaction matrix to model transactions in terms of their time, origin, and destination components in order to detect the patterns of navigation through an Internet-based digital library. The findings showed that user training instead of changes in the user interface affected transaction patterns significantly.

Researchers have also compared usage patterns in different types of IR environments. Jones, Cunningham, McNab, and Boddie (2000) conducted a transaction log analysis of user interactions with the Computer Science Technical Reports (CSTR) collection of the New Zealand Digital Library, and further compared the similarities and differences between their study and studies of Web search engines and OPACs. Several parallel usage patterns were identified when compared with studies of Web search engines including short queries, short sessions, few results viewed, and a relatively low

percentage of unique queries. The main difference between their study and studies of Web search engines was in the use of Boolean operators. The authors noted greater usage of Boolean operators in the CSTR query data, which was attributed to the more specialized audiences of CSTR who are more knowledgeable about Boolean logic. A comparison with transaction analyses of OPAC studies also yielded similar short queries. However, the length of each OPAC search session varies from 7 minutes to 30 minutes. Because OPAC studies have reported large proportions of search errors, the level of search errors observed in CSTR was not unexpected. Wolfram (2008) compared and analyzed search characteristics in different Web-based environments (OPAC, bibliographic database, general search engine, specialized search service dealing with health information). He reported that the number of terms used per query was shorter for the bibliographic database than for the OPAC and search engine environments.

Most research to date has focused on the identification of users' search behaviors, in particular query formulation and display actions. Few studies have focused on an in-depth analysis of information sources and subject areas of search topics. In addition, many of the longitudinal studies do not cover log analysis over a long period of time. The limitation of the existing literature prevents researchers from obtaining a more complete understanding of users' information-seeking and retrieval behaviors in digital environments.

The current study presents a longitudinal analysis of summary transaction log data for a state-supported general audience digital library in the United States called BadgerLink. BadgerLink provides access to a range of databases from EBSCO (EBSCOhost) and Bell & Howell (ProQuest) to residents of Wisconsin via the Internet as a Web-based service. EBSCOhost makes available more than a dozen databases with a range of topic coverage from disciplinary specialization (e.g. ERIC, Health Source Plus) to broad subject coverage (e.g. MasterFILE Premier, Academic Search Elite). Databases may include citation and indexing information, with most providing full-text documents. ProQuest provides access to a number of newspapers, news journals and ABI/Inform Global for business information. Its contents are more specialized in that each of the newspaper databases contains records of the full-text of a single newspaper, or set of newspapers. In an earlier study of BadgerLink usage (Wolfram & Xie, 2002; Xie & Wolfram, 2002) revealed the value of the service to both libraries and residents of Wisconsin. Since becoming available in 1998, BadgerLink has been used by hundreds of libraries and countless users, generating millions of searches annually.

2. Research Questions

The purpose of the present study is to undertake a longitudinal analysis of full-text database usage in BadgerLink explore database usage in a general audience digital library. Specifically, this study intends to answer the following research questions: 1) Are shifts in resource usage and search interactions have occurred? 2) If yes, what characteristics of shifts in resource usage and search interactions have been observed?

3. Methodology

Among its many services, BadgerLink provides access to full-text databases through EBSCO and ProQuest database subscriptions. The focus of this investigation is on database usage, which represents a key service of BadgerLink. Vendor-based EBSCO usage data previously collected in 1999 and 2001 (Wolfram & Xie, 2000; Wolfram & Xie, 2002; Xie & Wolfram, 2002), representing six and eight months respectively, were analyzed with newer data for 2003 (eight months) and 2005 (seven months). The earlier data sets represented time frames for earlier studies. To keep the nature and seasonal characteristics of the data the same, similar time frames were selected for 2003 and 2005. These differences in the time frame of each dataset required that results be averaged by month for both the EBSCO and ProQuest data. Although some of the logging features have changed over the data collection period, most of the data fields collected were limited to those available for the full time frame.

Usage data for each year were processed in MS Excel. Data were summarized and tabulated based on organizational users (primarily libraries or users associated with different types of libraries), session and query counts, and title requests from databases available through EBSCO. Identifiable institutions were grouped into broad categories (four-year colleges/universities, junior/technical colleges, public libraries & systems, kindergarten through grade 12 (K-12) schools and systems, special libraries/corporate environments). The top 200 journal titles requested were categorized into ten broad categories (arts & humanities, business & finance, education, government & politics, health & medicine, leisure & entertainment, news & commentary, reference, science & technology, social science & society). Outcomes were compared across the four time periods for changes in relative activity, topicality of resources requested, and organization type. Data from BadgerLink's ProQuest Newsstand subscriptions were also analyzed for the period 2002-2005 and included title requests and feature usage.

4. Results

The findings reveal that shifts in resource usage and search interactions have occurred. Results are presented using the two broad areas of resource usage and search interactions. Resource usage includes most frequently accessed titles and requests by subject. Search interaction consists of searches per session, institutional usage, document format access, and search feature usage. Note that data summaries may be reported as absolute values or relative values (i.e., percentages) over time that reveal changes over time in the proportional allocation of values across categories of interest. Reporting changes in relative values can be more revealing of change, particularly if there is a general trend towards increasing absolute values over time across all categories, where these differences may not be as apparent.

4.1 Database Availability

The number of databases available through EBSCO and their usage has grown since 1999. Databases continue to be added, while some have been dropped or merged during the time period of the study. These databases represent a broad range of subject

areas. A summary of the currently available databases appears in Table 1. Note also that some databases have been enlarged over this time period. Those that have been dropped have not been included. ProQuest resources offered through BadgerLink represent full-text access to numerous newspapers, and are, therefore, not presented.

Table 1. EBSCO Database Availability Summary

Database	Availability
Academic Search Elite	1999-2005
Business Source Elite	1999-2005
Clinical Reference Systems	1999-2005
EBSCO Animals	1999-2005
ERIC	1999-2005
Funk & Wagnalls New World Encyclopedia	1999-2005
MasterFILE Premier	1999-2005
Middle Search Plus	1999-2005
Primary Search	1999-2005
Corporate ResourceNet	2001-2005
Health Source: Consumer Edition	2001-2005
Health Source: Nursing/Academic Edition	2001-2005
MEDLINE	2001-2005
American Heritage Children's Dictionary	2003-2005
Clinical Pharmacology	2003-2005
Knight Ridder Collection	2003-2005
MAS Ultra - School Edition	2003-2005
Military & Government Collection	2003-2005
Newspaper Source	2003-2005
Professional Development Collection	2003-2005
Regional Business News	2003-2005
Biomedical Reference Collection: Basic	2005
Columbia Encyclopedia	2005
Nursing and Allied Health Collection: Basic	2005

4.2 Most Frequently Accessed Titles

For both EBSCO and ProQuest, there have been shifts in the most frequently accessed titles over the study period. This reflects the broader array of titles made available as well as possible shifts in user requests. The ten most frequently accessed titles based on monthly abstract requests for EBSCO titles appear in Table 2. In general, news sources have become more popular while news magazines have lost some favor. Due to changes in the way ProQuest collected usage data, comparable equivalent data is only available for each year between 2002 and 2005 (Table 3). The list also reveals some changes, although regional news sources still top the list.

Table 2. Most Frequently Requested Abstracts Based on EBSCO Titles

	Rank			
Title	**1999**	**2001**	**2003**	**2005**
New York Times	*	1	1	1
ERIC Documents	*	7	2	2
USA Today	*	*	*	3
Economist	5	8	7	4
Time	3	5	4	5
Newsweek	2	3	3	6
Wall Street Journal - Eastern Edition	*	6	9	7
Dissertation Abstracts International	*	*	*	8
U.S. News & World Report	4	4	6	9
Christian Science Monitor	1	2	5	10
Lancet	6	9	*	*
Business Week	*	10	*	*
People	7	*	*	*
Billboard	8	*	*	*
Alberta Report / Western Report	9	*	*	*
Time South Pacific	10	*	*	*

Note: An "*" indicates it did not appear in the top 10 rank in a given year.

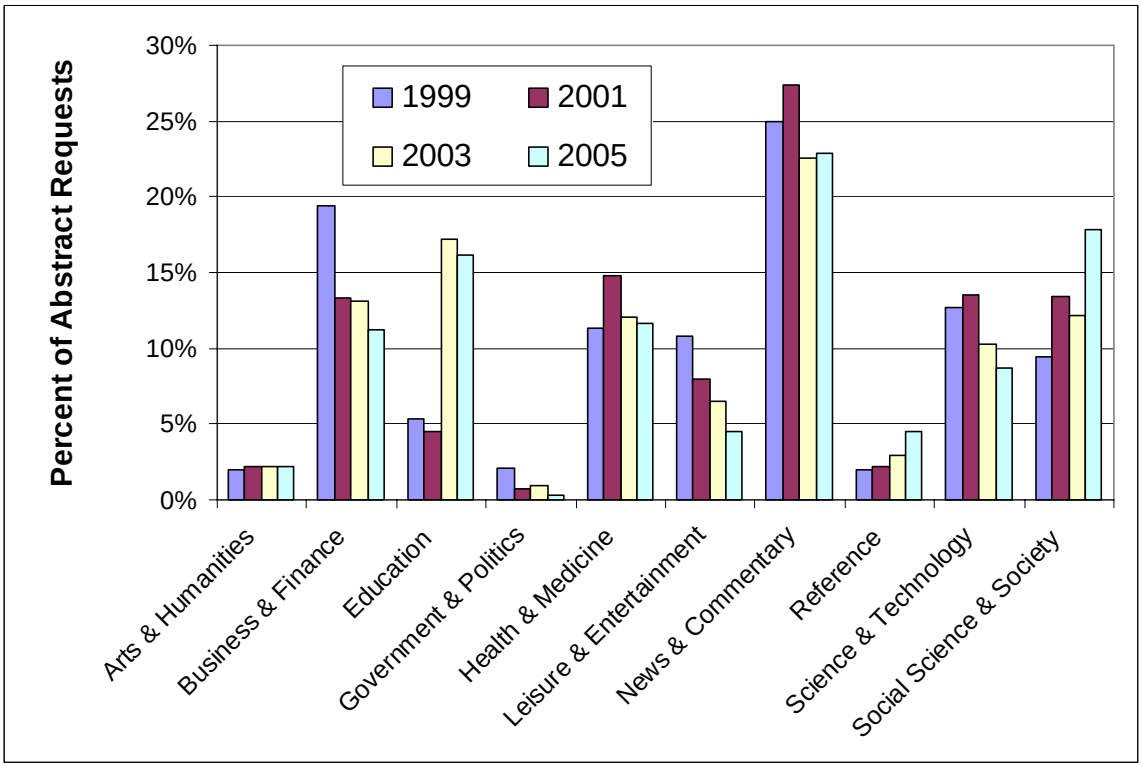
Table 3. Most Frequently Requested Abstracts Based on Titles for ProQuest

	Rank			
Title	**2002**	**2003**	**2004**	**2005**
Wisconsin State Journal	2	1	1	1
Milwaukee Journal Sentinel	1	2	2	2
Madison Capital Times	4	3	3	3
Wall Street Journal	5	5	4	4
BBC Monitoring Middle East	*	*	6	5
New York Times	3	4	5	6
Barron's	*	*	8	7
Chicago Tribune	*	10	9	8
The Washington Post	7	6	7	9
Los Angeles Times	10	7	10	10
Financial Times	6	9	*	*
PR Newswire	8	8	*	*
Boston Globe	9		*	*

4.3 Requests by Subject

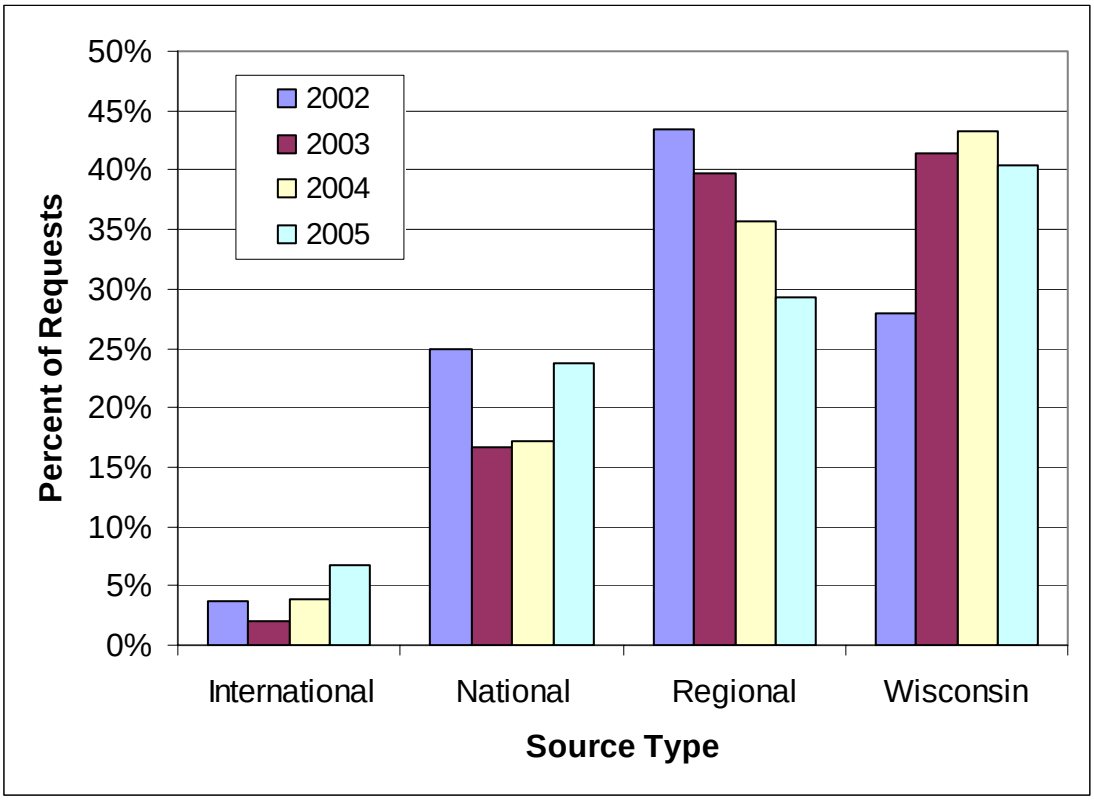
A broad subject categorization of titles offered through EBSCO databases was developed based on the titles of requested documents. Titles were grouped into one of ten subject categories. The number of abstract requests for titles in each group was tallied. The results appear in Figure 1. Several shifts are apparent over the six years. There are more requests (proportionately) for the areas of Education, Social Science & Society, and Reference. Requests for Business as well as Leisure & Entertainment titles have gone down.

Figure 1 – Topic Breakdown of Relative Abstract Requests for EBSCO Titles



Given the newspaper and general news journal content of ProQuest, topic analysis of titles could not be duplicated. Geographic analysis of titles could be provided (Figure 2). Full-text requests constitute the vast majority of requests (over 90%) for ProQuest sources. For EBSCO titles, requests for abstracts exceed those for full-text. Interest in Wisconsin-based and national news periodicals remains strong. Interest in international sources is growing, whereas requests for newspaper content from regional sources (city newspapers outside of Wisconsin) appears to be declining, although it still represents the second most popular category.

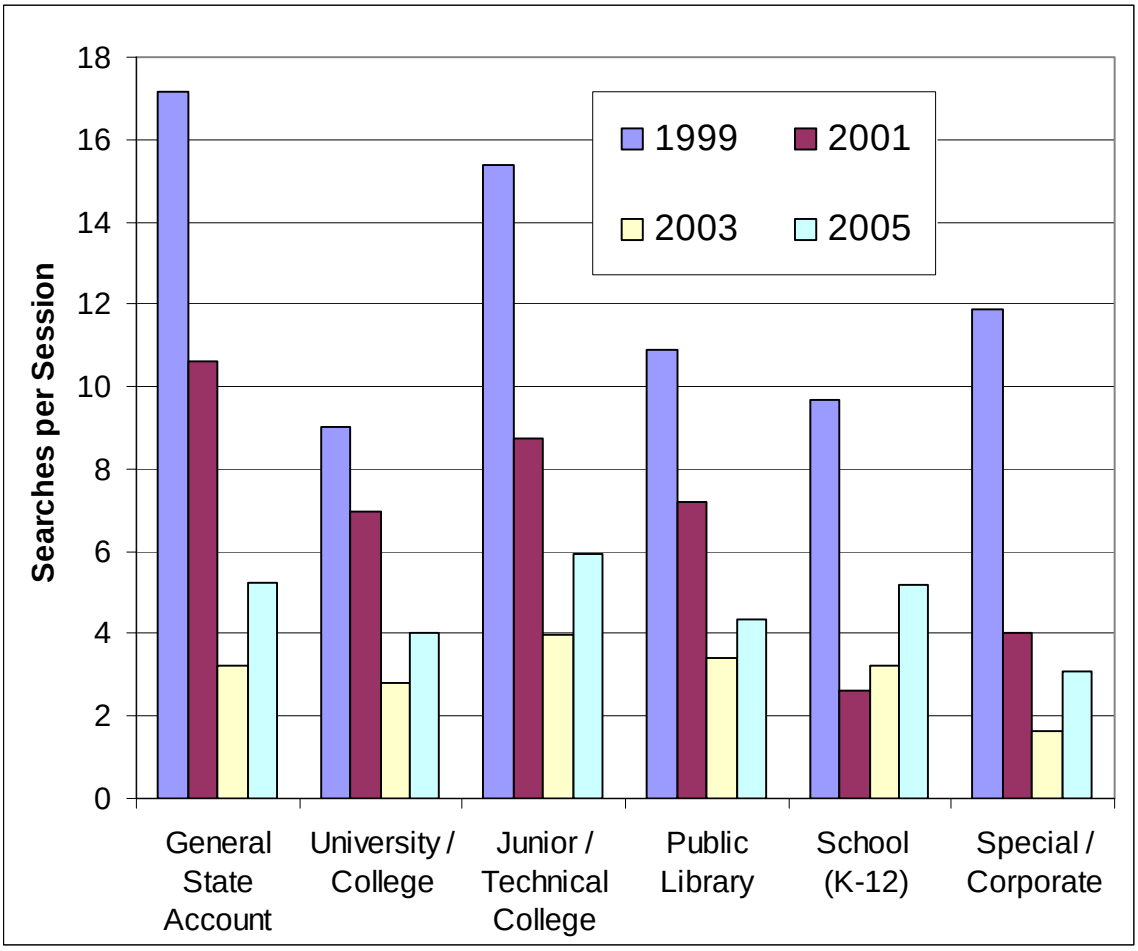
Figure 2 – Geographic Breakdown of Full-text Requests for ProQuest Titles



4.4 Searches per Session

Search interactions cover a variety of interaction types. One area of interest relates to changes in the number of searches conducted per session. If there are many searches per session, then users are either engaging in multiple search topics or modifying their searches frequently possibly because they are not finding what they are looking for. Without knowing the user motivation, this is difficult to determine. Any change, however, is revealing of a shift in user search tactics. The average number of searches per session has declined since 1999 across all institution types although this has rebounded somewhat in 2005 (Figure 3). One possible reason is that users have gained experience in searching for information across time, which requires fewer searches to complete their session activity.

Figure 3 – Average EBSCO Searches per Session by Institution Type

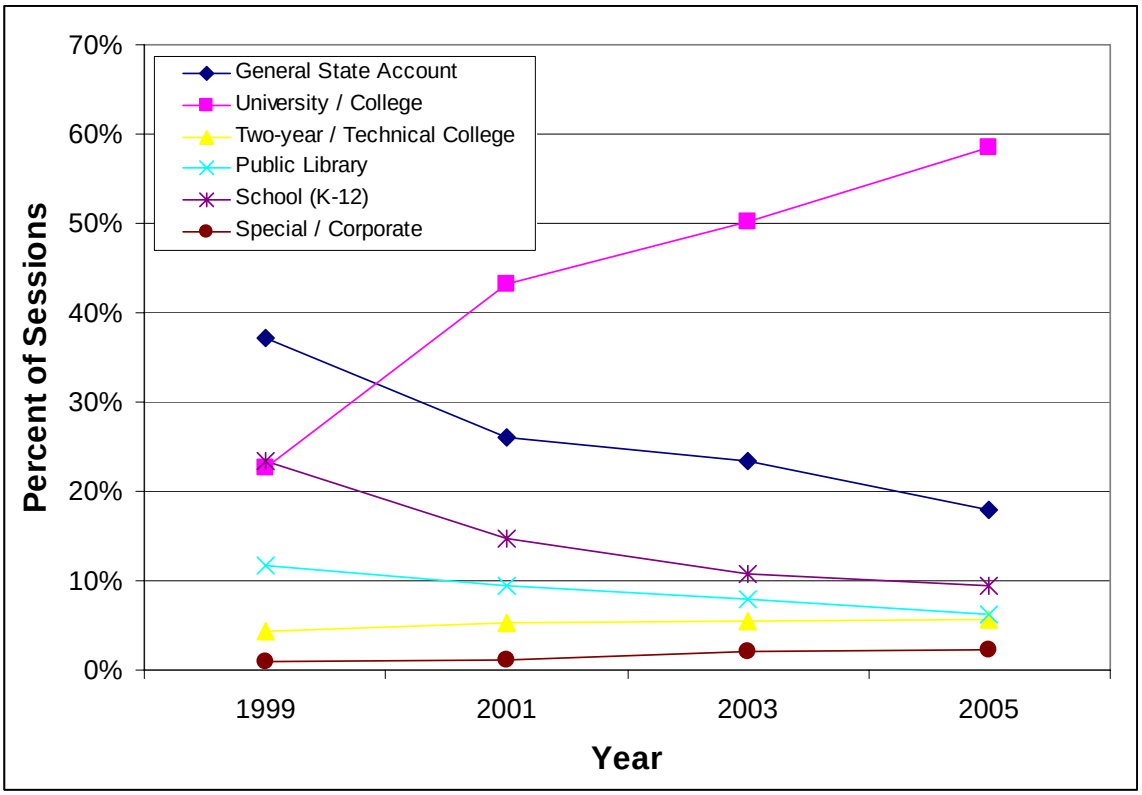


4.5 Institutional Usage

In addition to the five institution types identified above, there is also a general state library account that covers users, libraries and organizations that cannot be identified because they do not have dedicated or static Internet Protocol (IP) addresses. The data can be summarized by the number of searches conducted or the number of sessions undertaken. Session data at the institutional level is considered to be more reliable because it represents a complete set of search actions and not just a single search by an individual.

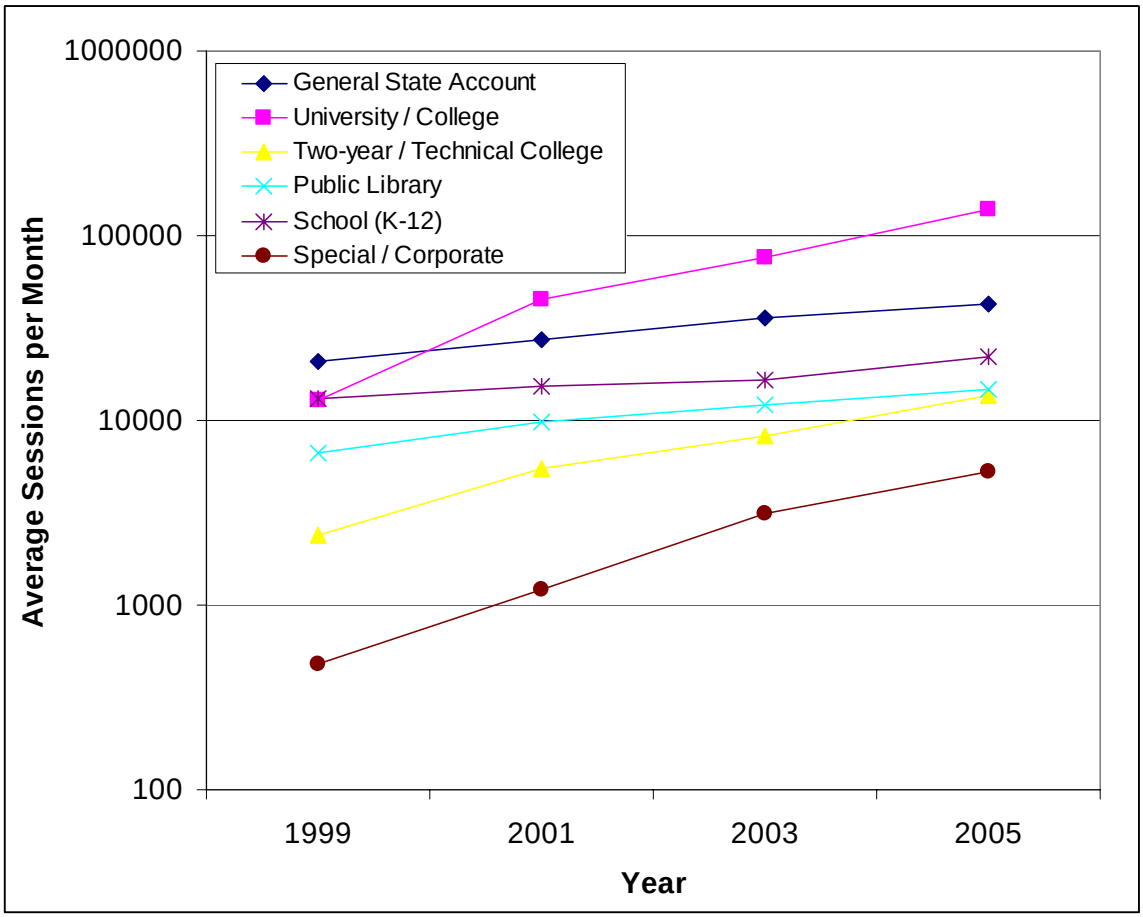
Data for institution type may be broken down a number of ways. To provide sense of the change in relative usage by institution type, the percentage breakdowns of overall usage for 1999 to 2005 are presented in Figure 4.

Figure 4 – Change in EBSCO Relative Usage by Institution Type



Academic users continued to grow in terms of the percentage of sessions conducted between 1999 and 2005, while public library and K-12 school usage has declined as an overall percentage. Relative usage through the state library account has also declined since 1999, which is undoubtedly attributable to an increase in the number of libraries and organizations with identifiable IP addresses in that time. Overall usage by all institutional types in absolute terms is up over the past six years when the raw figures are compared (Figure 5). Note that the vertical axis on this figure uses logarithmic scaling based on powers of 10 to make the changes within each institution type more visible. Smaller values would not be discernible without this scaling.

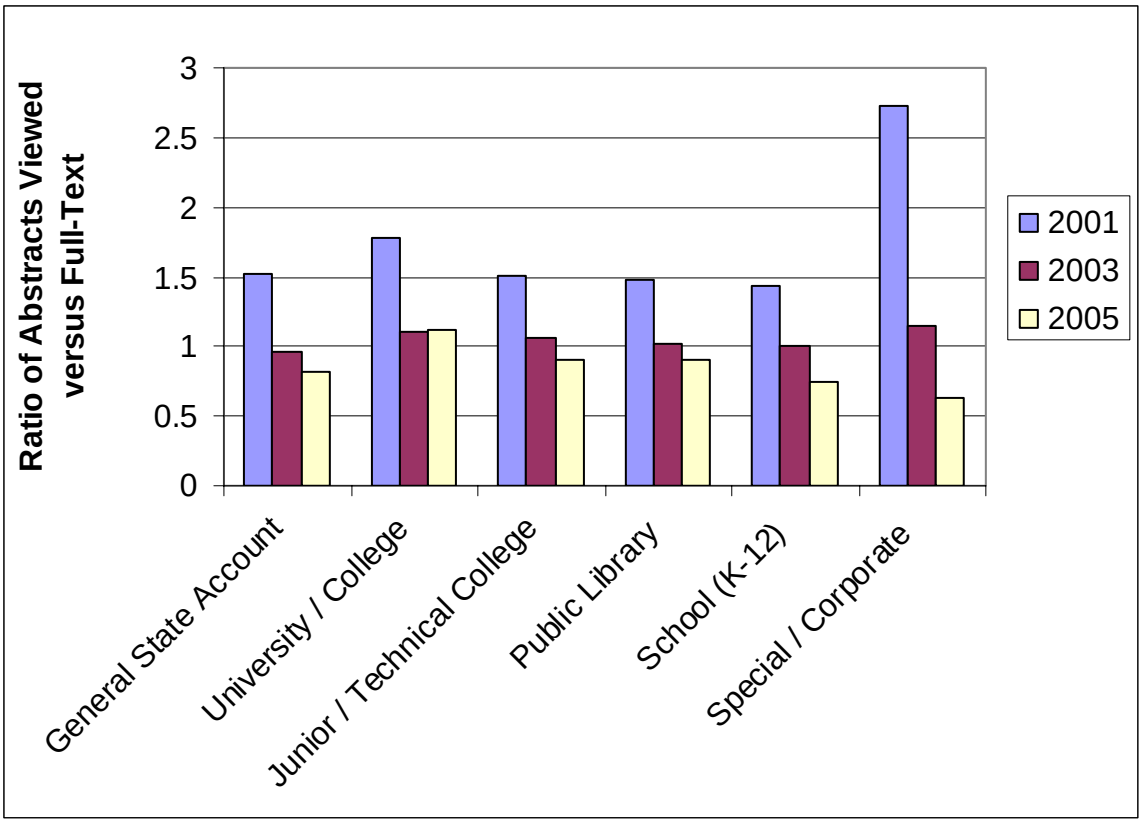
Figure 5 – Average Sessions Conducted per Month



4.6 Document Format Access

With increasing full-text access to EBSCO resources, users across the different institution types are shifting towards more frequent full-text requests as opposed to abstracts only. As a result of changes in how EBSCO keeps track of its data, data were only available for 2001, 2003 and 2005. Data comparing abstract to full-text access appears in Figure 6. A value greater than one along the vertical access indicates more abstracts are being requested than full-text items per session. Values less than one indicate more full-text articles are being requested. The trend across all institution types is towards more full-text viewing. This is particularly true for the special/corporate environment, which has seen the greatest drop in viewing abstracts. For four-year university/college environments there is still a slight preference towards viewing more abstracts over full-text results.

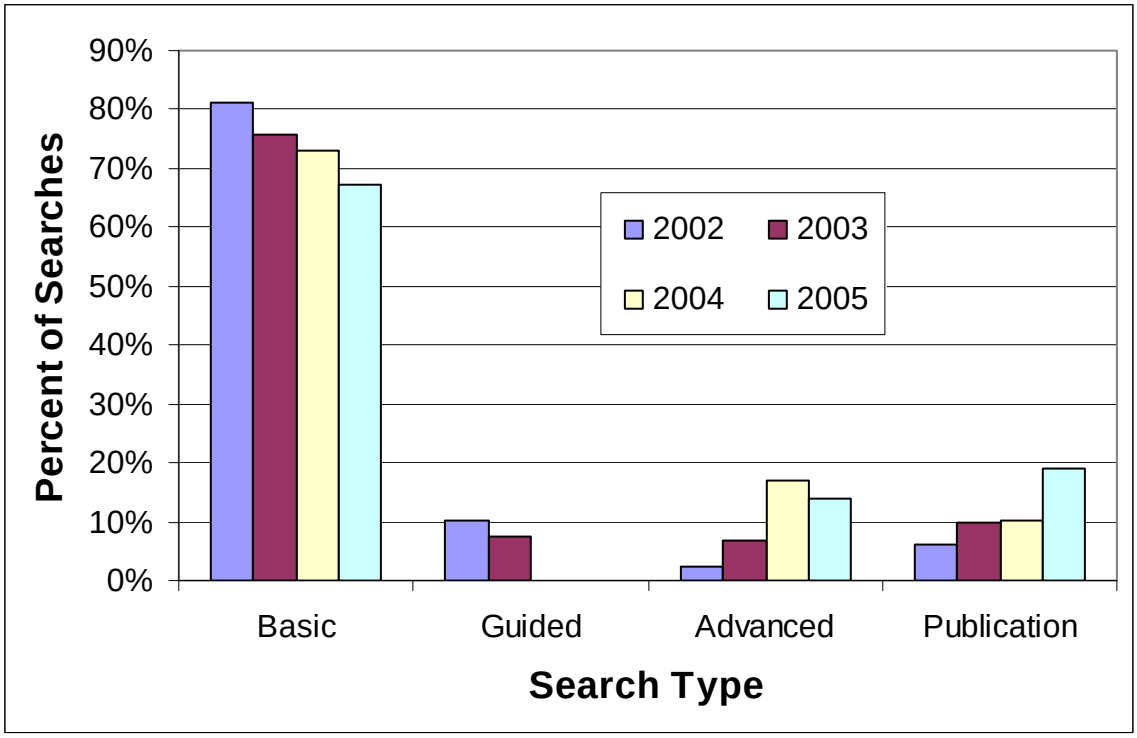
Figure 6 – Abstract to Full-text Requests by Institution Type



4.7 Search Feature Usage

ProQuest usage logs did record other types of information not available on EBSCO. Included in this additional data is the type of search users conducted. In addition to basic search mode, guided and natural search mode (discontinued in 2004), advanced search, and publication search (which facilitates searching and browsing within specific publications) provide users with different search options. The breakdown of search feature usage appears in Figure 7. Basic searching continues to be the most frequently used method, but advanced and publication search options have become more popular over time. The data aggregation reported by ProQuest, unfortunately, do not allow a further breakdown of the data by institution type, which would provide further insights into potential differences in searcher activities across institution types.

Figure 7 – Search Feature Usage on ProQuest



5. Discussion

When compared with longitudinal studies and log analysis of different digital environments, the findings of this study have yielded several notable results pertaining to search topics, institutional usage, search behaviors, search function usage, and document format access. First, Web search engine studies observed a shift in search topics from entertainment and recreation to e-commerce (Jansen, Spink, & Saracevic, 2000; Spink & Jansen, 2004; Spink, Wolfram, Jansen, & Saracevic, 2001) and people places or things (Jansen and Spink, 2006). In this study, some similar shifts have also been observed, but with some differences. For example, like the Excite studies, the relative proportion of requests for leisure and entertainment searching have declined (from 10.3% to 6.1% between 1999 and 2005), although the absolute values have increased as system usage has grown. However, business and finance resource requests were also down relatively (from 17.4% of requested titles to 13.5%). Similarly, requests for titles dealing with social sciences and society increased relatively (10.6% to 18.9%) as was the case for education (6.7% to 10.2%). Two reasons might account for the differences: the collections and main users of these systems. The diverse coverage, in particular more commercial product information and general public users, of Web search engines as well as the availability of more databases on education and social sciences and the increasing university users of these databases in BadgerLink determine the differences in search topics in these two search environments. The results of this study also have some overlap with Wang, Berry, and Yang's (2003) study of an academic Web environment in which the identified search topics were more related to academics, career, finances and sports.

Second, this study echoes Zhang, Lee and You's (2001) results for organizational usage of an electronic theses and dissertation digital library, and Nicholas, Huntington, and Watkinson's (2005) results of a scholarly journal digital library. In those two studies educational institutions were the dominant users of the systems (82.7% in Zhang, Lee and You's study; 68% in Nicholas, Huntington, and Watkinson's work). For the current study, users affiliated with higher education institutions constituted 64.2% of all sessions. The nature of collections in these three studies, to some extent, affects institutional usage. In addition, Jones, Cunningham, McNab and Boddie (2000) found that users from educational institutions (.edu) were ranked first compared with other domains in using digital libraries. Post-secondary institutions traditionally have been active users of bibliographic database resources through internal subscriptions to database vendor products. In 1999, for the current study, K-12 schools and public libraries represented approximately 35.2% of sessions conducted, whereas universities and technical colleges represented 26.9% of sessions. By 2005, K-12 schools and public libraries represented only 15.6% of all sessions, with universities and technical colleges growing to 64.2% of all sessions. Although sessions are up across all groups, the growth by users of institutional types that could benefit from these resources has been slower. This is not just a reflection of sources of academic interest included as part of the EBSCO databases in BadgerLink. Indeed, the breakdown of title requests reflects general public interest, and not just academic subject areas.

Third, regarding search behaviors, in Cooper's (2001) study of Web-based OPAC environment, the number of searches did not change even though users became more familiar with the system. The Excite data also showed that there was little change in numbers of queries per sessions. However, in this study, the results indicated that the average of searches per sessions declined in all institution types (from 12.3 searches per session in 1999 to 4.5 searches per session in 2005) probably because their experiences help them more effectively find relevant information. Also, the greater prevalence of dedicated Internet access in more recent years has made access to the database resources more convenient for more spontaneous searching. The more recent increase in searches per session across all environments for the 2005 data cannot be easily explained.

Fourth, whereas Zhang, Lee and You (2001) found that the search function was the most frequently used system function, this study further analyzed the breakdown of search feature usage and showed that basic search was the most frequently used the search feature over time in ProQuest. Usage of the basic search feature, which constituted 81.1% of searches in 2002, was down to 67.2% in 2005. Similarly, the advanced search feature usage saw more than a five-fold increase in use from 2.4% in 2002 to 13.8% in 2005. The greater use of the publication search feature would indicate that users are performing more specific searches on sources once they know they exist within the ProQuest collection.

Fifth, regarding document format access, the results of this study correspond to the findings of scholarly journal usage analysis from Nicholas, Huntington, and Watkinson's (2005) work. In the present study, more full-text than abstracts were requested and viewed in 2005 across all institution types except academic environments, where browsing of abstracts was still slightly more prevalent than full-text. In the scholarly journal usage study, viewing of full-text articles accounted for 31% of all views while abstract viewing only represented 20% of the views. Moreover, the current study's

findings are also similar to those of Ke et al. (2002), who analyzed log data of Elsevier's ScienceDirect. In their study, Ke et al. found that the usage of full-text viewing was 3.6 times more prevalent than abstract viewing. In addition, full-text viewing was identified as the most frequent employed behavior, even more than search submission. In the current study, the average ratio of abstract to full-text requests was 0.85 in 2005, down from 1.74 abstracts per full-text request in 2001. However, academic users still made slightly more abstract than full-text requests. The nature of the task that leads academic users to search for information may determine they have to view the abstracts to make relevance judgments before they access to the full-text documents.

As with any transaction log research, this research has its limitations. These limitations stem from the nature of log analysis itself and the available data. While log analysis provides detailed and unobtrusive data about longitudinal patterns of user interactions with IR systems, it cannot explain the reasons behind the usage patterns. It cannot present a direct and clear account of why users engage in the interactions they do. In addition, the available data only offers limited demographic profiles, along with aggregated data across different categories. Finally, the lack of availability of query data for this study limited the analysis and comparison of more detailed search behaviors. Still, the available longitudinal data has provided valuable insights into system usage over time from which comparisons may be made.

6. Conclusion

This longitudinal analysis of database usage associated with a general audience digital library reveals a picture of dynamic change of resources usage and search interactions over a six year period. The findings of this study identified changes in the distribution of search topics based on resource requests, with a growth in requests for resources from the social sciences and education. The proportional decrease in EBSCO database use by users affiliated with public library and K-12 schools with a corresponding increase by post-secondary academic users may be cause for concern because the former institutions, which traditionally have not had the same level of access to these resources, could benefit more fully from available databases. The increase in usage of advanced search features in ProQuest indicates that users are becoming more exploratory in their use of available features. The change in search behaviors might also be the result of more training offered by different institutions. Variation of search feature usage by institution type cannot be analyzed based on the current data availability. This area requires further investigation. Based on the comparison of this study with previous research in different digital environments, we can conclude that the findings of this study are more similar to the results reported in other online databases and digital library environments than Web search engine and Web page environment studies. Wolfram & Xie (2002) defined the context of digital libraries as representing a hybrid of both "traditional" IR, using primarily bibliographic resources provided by database vendors, and "popular" IR, exemplified by public search systems available on the World Wide Web. Usage of online databases in digital libraries reveals that users' search topics were close to online databases and digital libraries but their search behaviors were more similar to searching Web search engines.

Universal access is the objective of digital libraries and information services. Unfortunately, not all user groups benefit from general audience digital library and information services. The results of this longitudinal study indicate a possible gap in the services. There are multiple reasons for the gap. First, the resources in general audience digital libraries or information services have more academic related materials even though they do cover resources that attract general interest. This may be similar to the analogy of the "rich becoming richer", where users associated with post-secondary institutions who have traditionally had access to database resources, now have even more access and are taking greater advantage of the increased resources. More diverse collections are needed for general audience digital libraries and information services. Second, more promotions and training sessions are offered in academic settings than in other settings. Based on previous research (Buttenfield & Reitsma, 2002; Efthimiadis & Bruce, 2000; Xie & Wolfram, 2002), the imbalanced distribution of promotion and training among the users in digital libraries or information services was identified as a major problem. The main issue for promotion is how to promote the service to different user groups, especially those who rarely use physical libraries. Compounding this challenge is the ability for institutions with limited resources to be able to engage in promotional and training activities. More channels should be used to publicize the services, such as major media (newspaper, TV, etc.), Web, community newsletters, electronic mail, regular mailings, etc

The current study represented a high-level, longitudinal foray into general audience digital library usage. Because of the limitations of this study, future research should apply multiple methods, such as query transaction logs, think aloud, interviews, questionnaires, etc. to investigate not only how users interact with IR systems but also what lead to their interactions. Data collection should also extend to all the search behaviors during the search process instead of limiting it to vendor data. Future studies should also expand to other general audience digital libraries and information services to identify usage patterns as well as to other digital environments to compare the similarities and differences in search behaviors to help inform system design and improvement.

7. Acknowledgements

The authors thank the Wisconsin Department of Public Instruction Division for Libraries, Technology, and Community Learning for making the data available.

8. References

Blecic, D.D., Dorsch, J. L., Koenig, M. H., & Bangalore, N.S. (1999) "A longitudinal study for the effects of OPAC screen changes". *College and Research Libraries, 60*(6), 515-530.

Bollen, J., Luce, R., Vemulapalli, S. S., & Xu, W. (2003) "Usage analysis for the identification of research trends in digital libraries". *D-Lib Magazine, 9(5).* Retrieved July 20, 2008, from http://www.dlib.org/dlib/may03/bollen/05bollen.html

Buttenfield, B. P., & Reitsma, R. F. (2002) “Loglinear and multidimensional scaling models of digital library navigation”. *International Journal of Human-Computer Studies, 57*(2), 101-119.

Cockburn, A. & McKenzie, B. J. (2001) “What do web users do? An empirical analysis of web use”. *International Journal of Human Computer Studies, 54* (6), 903-922.

Cooper, M. D. (2001) “Usage patterns of a Web-based library catalog”. *Journal of the American Society for Information Science and Technology*, 52, 137–148.

Cothey, V. (2002) “A longitudinal study of Worl Wide Web users’ information searching behavior”. *Journal of the American Society of Information Science and Technology, 53*(2), 67-68.

Davis, S. E. (2006) “Identifying opinion leaders and elites: A longitudinal design”. *Library Trends, 55*(1), 140-157.

Efthimiadis, E. & Bruce, H. (2000) “A study of the impact of statewide database licensing on information provision in Washington state”. In D. H. Kraft (Ed.), *Proceedings of the ASIS Annual Meeting*, vol. 63, (pp. 47-62). Medford, NJ: Information Today.

Jansen, B. J., & Spink, A. (2006) “How are we searching the World Wide Web? A comparison of nine search engine transaction logs”. Information Processing & Management, 42(1), 248-263.

Jansen, B. J., Spink, A., & Saracevic, T. (2000) “Real life, real users, and real needs: A study and analysis of user queries on the Web”. *Information Processing and Management, 36*(2), 207-227.

Jones, S., Cunningham, S. J., McNab, R., & Boddie, S. (2000) “Human-computer interaction for digital libraries: A transaction log analysis of a digital library”. *International Journal on Digital Libraries, 3*(2), 152-169.

Ke, H. R., Kwakkelaar, R. , Tai, Y. M. & Chen, L.C. Chen (2002) “Exploring behavior of e-journal users in science and technology: Transaction log analysis of Elsevier’s ScienceDirect Onsite in Taiwan”. *Library and Information Science Research, 24*(3), 265-291.

Nicholas, D., Huntington, P., Monopoli, M., & Watkinson, A. (2006) “Engaging with scholarly digital libraries (publisher platforms): The extent to which 'added-value' functions are used”. *Information Processing and Management, 42*(3), 826-842.

Nicholas, D., Huntington, P., & Watkinson, A. (2005) “Scholarly journal usage: The results of deep log analysis”. *Journal of Documentation, 61*(2), 248-280.

Spink, A., & Jansen, B. J. (2004) "*Web search: Public searching of the Web*. Boston: Kluwer Academic Publishers".

Spink, A., Wolfram, D., Jansen, B. J., & Saracevic, T. (2001) "Searching the Web: The public and their queries". *Journal of the American Society for Information Science, 52*(3), 226-234.

Wang, P., Berry, M. W., & Yang, Y. (2003) "Mining longitudinal Web queries: Trends and patterns". *Journal of the American Society for Information Science and Technology*, 54(8), 743–758.

Wolfram, D. (2008) "Search characteristics in different types of Web-based IR environments: Are they the same?" *Information Processing & Management*, 44, 1279-1292.

Wolfram, D., & Xie, H. (2000) "End user database searching on the Internet: An analysis of the state of Wisconsin's BadgerLink service". In M.E. Williams (Ed.), *Proceedings of the 21st National Online Meeting.* (pp. 503-512). Medford, NJ: Information Today.

Wolfram, D., & Xie, H. (2002) "Traditional IR for Web users: A context for general audience digital libraries". *Information Processing & Management, 38*(5), 627-648.

Xie, H., & Wolfram, D. (2002) "State digital library usability: Contributing organizational factors". *Journal of the American Society for Information Science and Technology*. 53(13), 1085-1095.

Yi, K., Beheshti, J., Cole, C., Leide, J. E., & Large, A. (2006) "User search behavior of domain-specific information retrieval systems: An analysis of the query logs from PsycINFO and ABC-Clio's Historical Abstracts/America: History and Life: Research Articles". *Journal of the American Society for Information Science and Technology*, 57(9), 1208-1220.

Zhang, Y., Lee, K., & You, B. J. (2001) "Usage patterns of an electronic theses and dissertations system". *Online Information Review, 25*(6), 370-377.