

# ***Discoverability***

## **Phase 1**

### **Final Report**

**University of Minnesota Libraries**

**Submitted to the Web Services Steering Committee: February 13, 2009**

Cody Hanson (co-chair)  
Heather Hessel (co-chair)  
John Barneson  
Deborah Boudewyns  
Jan Fransen  
Lara Friedman-Shedlov  
Martha Hardy  
Chris Rose  
Barb Stelmasik  
Stacie Traill

## Table of Contents

Executive Summary .....	1
Introduction.....	5
Trends.....	7
External factors.....	7
Trend #1: Users are discovering relevant resources outside traditional library systems.....	8
Trend #2: Users expect discovery and delivery to coincide.....	11
Trend #3: Increasing usage of portable Internet-capable devices .....	14
Trend #4: Discovery increasingly happens through recommending.....	18
Trend #5: Our users increasingly rely on emerging nontraditional information objects .....	20
Principles .....	23
Analysis of Local Systems .....	25
A Month of Library Discovery: October 2008 .....	25
Statistics Recommendations .....	29
Affinity Strings Baseline Report .....	31
AgEcon Search Baseline Report.....	34
Aleph – OPAC Baseline Report .....	38
DLXS – Finding Aid Search Baseline Report .....	46
Interlibrary Loan ILLiad Baseline Report.....	47
Minnesota Digital Library Baseline Report .....	51
MetaLib – Metasearch Baseline Report .....	53
Primo Baseline Report .....	55
SFX – OpenURL Link Resolver Baseline Report .....	59
University Digital Conservancy Baseline Report.....	61
Vendor Statistics Baseline Report.....	66

Libraries' Website Statistics Baseline Report .....	68
Appendix A: Charter .....	A-1
Appendix B: Discoverability Principles from Peer Institutions .....	B-1
Appendix C: Subdomain Statistics .....	C-1
ames.lib.umn.edu Website Statistics .....	C-1
andersen.lib.umn.edu Website Statistics .....	C-4
arch.lib.umn.edu Website Statistics .....	C-6
bell.lib.umn.edu Website Statistics .....	C-9
biomed.lib.umn.edu Website Statistics .....	C-12
busref.lib.umn.edu Website Statistics .....	C-14
cbi.lib.umn.edu Website Statistics .....	C-17
courses.lib.umn.edu Website Statistics .....	C-20
digital.lib.umn.edu Website Statistics .....	C-23
dsrc.lib.umn.edu Website Statistics .....	C-25
eastasian.lib.umn.edu Website Statistics .....	C-28
efw.lib.umn.edu Website Statistics .....	C-30
ej.lib.umn.edu Website Statistics .....	C-33
etrc.lib.umn.edu Website Statistics .....	C-34
forestry.lib.umn.edu Website Statistics .....	C-36
friends.lib.umn.edu Website Statistics .....	C-38
govpubs.lib.umn.edu Website Statistics .....	C-39
infopoint.lib.umn.edu Website Statistics .....	C-41
Irc.lib.umn.edu Website Statistics .....	C-42
magrath.lib.umn.edu Website Statistics .....	C-44
map.lib.umn.edu Website Statistics .....	C-47
math.lib.umn.edu Website Statistics .....	C-49
music.lib.umn.edu Website Statistics .....	C-52

plant.lib.umn.edu Website Statistics ..... C-54

sciweb.lib.umn.edu Website Statistics ..... C-57

special.lib.umn.edu Website Statistics ..... C-59

vetmed.lib.umn.edu Website Statistics ..... C-62

walter.lib.umn.edu Website Statistics ..... C-64

wilson.lib.umn.edu Website Statistics ..... C-66



## EXECUTIVE SUMMARY

In October 2008, the Web Services Steering Committee at the University of Minnesota Libraries created the Discoverability exploratory subgroup, charged to recommend ways to make relevant resources more visible and easier to find, particularly within the user's workflow. This report shares the findings of Phase 1, in which the primary activity was data-gathering and analysis. Phase 2 of the group's work will take the discovery principles identified here and recommend specific strategies for the future.

The report consists of four main sections. The first section is a brief description of the process and methodology. The second is a discussion of five key trends related to discovery that were identified in the literature, including a description of how each trend is reflected in current use of local systems. The third section contains a set of suggested principles to guide future decisions related to discovery. Finally, we have collected and analyzed usage data from many of our local systems. These reports are collected in our fourth section and are summarized in "A Month of Library Discovery". We have also included specific recommendations regarding future data-gathering and analysis. Our appendices include a copy of the group's charge, a review of discovery principles at peer institutions, and a set of web statistics reports for the University Libraries' many websites.

### *Core Values*

At the outset of this project, we identified a set of core values to serve as a foundation for our process, to guide our work, and to keep us focused on our goals:

- **Objective analysis:** An open-minded and disinterested analysis will yield more accurate findings, leading to stronger and more relevant recommendations. Our work should be evidence-based.
- **User-centricity:** We are interested in user behavior as it is, not as we might wish it to be.
- **Beyond a "system":** The goal is not to replace Tool A with Tool B, but to identify a new set of strategies for exposing our resources to researchers as effectively as possible.
- **Efficiencies:** We are committed to exploring possibilities for collaboration at the network and/or consortial level where appropriate.

### *Phase 1 Methodology*

The group chose to break the work of Phase 1 down into two separate threads: Trends and Statistics, and divided into two separate teams accordingly. The Trends team identified and reviewed trends in user behavior as drawn from user studies; their work resulted in the identification of five broad trends. The Statistics team gathered and analyzed statistics on our existing local discovery systems. The outcome of their work was an overview of all systems and a set of baseline reports, which describe the use of our systems and a progress report on our statistics-gathering activity. The two teams then converged to synthesize all findings, and to create the set of Guiding Principles.

## *Trends*

### **1. Users are discovering relevant resources outside traditional library systems**

Search, once one of the key skills and specialties of librarians, is now a daily activity for the vast majority of our users. Our users approach their research with an established history of search success that gives them confidence in their search skills.

### **2. Users expect discovery and delivery to coincide**

Searchers do not distinguish between discovery and delivery in their web searches and increasingly find it discordant to experience this disconnect in the library environment.

### **3. Usage of portable Internet-capable devices is expanding**

Rather than just supplementing the desktop computer, mobile devices are poised to become the primary means of Internet access for a critical mass of our users.

### **4. Discovery increasingly happens through recommending**

Facilitating discovery requires us to develop and implement systems that push relevant content to users and that allows users to share content with others.

### **5. Our users increasingly rely on emerging nontraditional information objects**

The format of useful and discoverable information objects is much broader than those traditionally offered through libraries; users increasingly rely upon multimedia objects, data sets, blogs, and other “grey” objects to meet their information needs.

## *Guiding Principles*

The subgroup was charged with the task of developing “a set of principles related to discovery to help guide the Libraries’ strategic decisions about the selection, development, and support of relevant tools and services.” Based on the trends above and our analysis of local systems, we recommend the following principles:

- Users draw little distinction between discovery and delivery; systems, data, and information objects should be optimized for fulfillment. (*Trend 2*)
- In order to remain agile and responsive in a rapidly changing information environment, our systems and data structures should provide us with the greatest possible flexibility for frequent iteration and reuse by ourselves and others. (*External factors, Trend 5*)
- In order to best facilitate our users' discovery of relevant information, we should strive to be end-user device/platform agnostic. (*Trends 1, 3*)

- In order to remain responsive, relevant, and useful to our users, we must aggressively measure and analyze user behavior through local system statistics. These efforts will complement our ongoing assessments utilizing focus groups, usability studies, and reviews of applicable literature. (*External factors, Statistics recommendations*)
- Discovery should be organized around users rather than collections or systems. This organization should be based on realistic, evidence-based models of our users and their research tasks, e.g. task-based, persona-based, audience-targeted. (*Trends 1, 4, 5*)
- Users are successfully discovering relevant resources through non-library systems (e.g., general web searches, e-commerce sites, and social networking applications). We need to ensure that items in our collections and licensed resources are discoverable in non-library environments. (*Trends 1, 4*)
- Making collections discoverable requires optimizing for access by local and non-local user populations; being good stewards of our collections means participating in cooperative ventures that provide broad access to our collections. (*External factors, Trend 1*)
- Users rely on system- and peer-generated recommendations to discover relevant resources. We should capture the data necessary to provide targeted suggestions to users and defer to network-level systems where a critical mass already exists. (*Trends 1, 4*)

### *Analysis of Local Systems*

This report contains analysis of usage statistics for many of the Libraries' public-facing online systems. Included are baseline reports on the following environments:

- Affinity Strings data
- AgEcon Search
- Aleph OPAC (MNCAT Classic)
- Finding Aids (DLXS)
- Interlibrary Loan (ILLiad)
- MetaLib
- Minnesota Digital Library
- Primo (MNCAT Plus)
- SFX
- University Digital Conservancy
- Vendor Databases statistics
- Website statistics

In an attempt to gain insight into how users experience our discovery systems as a whole, we compare and contrast data between different systems in “A Month of Library Discovery”. Findings included:

- Our catalogs (MNCAT Classic and MNCAT Plus) and website continue to be heavily used. The SFX link resolver is a critical component of fulfillment, undergirding much of our users’ discovery activity.
- Using SFX and proxy data, we have found that 65% of information requests originate off-campus, with the other 35% originating on-campus primarily during class hours.
- External systems and search engines are driving an increasing volume of traffic to our systems: Google is responsible for more traffic to our websites than any other referrer. Google, PubMed, and the large databases and indexes account for ~75% of all SFX requests.

Our review of local systems also uncovered a number of areas where our current data gathering and analysis practices are incomplete, insufficient, or hampered by configuration choices. In order to improve the Libraries’ ability to make data-driven strategic decisions, we include a number of recommendations to augment the quality of the data we gather. These recommendations include:

- Enable the logging of referring sites on all web-facing systems.
- Enable logging of third-party discovery tools (e.g. LibX) and mobile browsers on all web-facing systems.
- Track web statistics using a single system with a common configuration across all web sites and systems.
- Track usage of the ExLibris X-Server on our Aleph and Metalib systems.
- Engage in search engine optimization to improve access to our sites and to better manage the load created by search engine spiders.

Our overarching recommendation to the Libraries is that we implement systems and practices, including dedicated staff time, to ongoing and rigorous analysis of our system usage statistics across the Libraries, including, but not limited to, our catalogs, ILS, ILL, link resolver, proxy server, websites, and repositories.

## INTRODUCTION

In October 2008, the Web Services Steering Committee at the University of Minnesota Libraries formed three exploratory subgroups, each charged with research and brainstorming in key areas: Current Awareness/Personal Information Management, Course Integration, and Discoverability. The focus of the Discoverability Subgroup charge was to recommend ways to make relevant resources more visible and easier to find, particularly within the user's workflow.

The Discoverability Subgroup charge – available in Appendix A – defines two phases for the overall project. Phase 1 was designed as a set of data-gathering activities to inform future decision-making. This was a unique opportunity to engage in a broad and disinterested review of how local systems are used, how users are approaching technology today, and how peer institutions are making decisions regarding discovery systems. The Phase 1 deliverables explicitly defined in the Discoverability Subgroup charter include:

- A set of guiding principles related to user discovery.
- A summary of guiding principles for user discovery that are applied at several other institutions.
- A summary of interviews with appropriate University partners and key stakeholders.
- A summary of relevant reports on campus and beyond.
- A summary of baseline data on library usage patterns based on data analysis and recommendations for action.
- Recommendations concerning how statistical analysis might be done on a regular basis.
- Trends in user behavior that are relevant to discovery (in “Objectives”).

One of the requests explicitly laid out in the charge was a consideration of grant resources. With the approval of our oversight committee, we have omitted direct discussion of grants.

### *“Discoverability”*

“Discoverability” is a neologism derived from the contemporary library science notion of discovery. The Discoverability Subgroup’s charge asks that we “recommend ways to make relevant resources more visible and easier to find, particularly within the user’s workflow.” Though the term “discoverability” refers to a quality or set of qualities of an information object, collection, or system, we have endeavored in Phase 1 of our work to focus primarily on the user. Our presumption is that in the second phase of our project we can work to describe a system that makes our owned and licensed information resources more discoverable using the work products from Phase 1.

Our charge also specifies that we are to take into account the discoverability not only of “library-owned collections, but [of] potentially relevant resources from any available source...” We have interpreted this to include both owned and licensed resources in our local collections as well as in collections at the network and consortial level. In addition, based on indications in the literature that scholars are relying on resources like blogs and other websites, we include discussion of resources that fall outside of traditional notions of “collectable” information objects.

Finally, we have interpreted discoverability to include the discovery of information objects not only in broad topical or subject-based research, but also in known-item searches.

## *Core Values*

The group understood this project to be an important component of future decision-making. As such, we identified a set of core values to serve as a foundation for our process, to guide our work, and to keep us focused on our goals. The guiding principles that we articulated are as follows:

- **Objective analysis:** An open-minded and disinterested analysis will yield more accurate findings, leading to stronger and more relevant recommendations. Our work should be evidence-based.
- **User-centricity:** We are interested in the research process writ large, both where it intersects with our current library environment and where it does not. We are interested in user behavior as it is, not as we might wish it to be, and as such those users who begin their research at Google or a similar network-level resource are as important to us as those who may be devotees of current library tools.
- **Beyond a “system”:** The goal is not to replace Tool A with Tool B, but to identify a new set of strategies for exposing our resources to researchers as effectively as possible.
- **Efficiencies:** We are committed to exploring possibilities for collaboration at the network and/or consortial level where appropriate.

## *Phase 1*

We chose to approach the challenges of this very broad topic by first concentrating on two separate threads: Trends and Statistics. We split our team of ten staff members into two groups based on skill sets, experience, and interests, with the co-chairs participating in both groups. For the first four weeks, each team concentrated on separate goals. The Trends Team identified and reviewed trends in user behavior – both broadly and for specific communities – as drawn from user studies. The objective for this group was to compile and analyze contemporary literature describing relevant discovery activity, ultimately culminating in a description of trajectories illustrating discovery behavior. The trends that we identified are not meant to represent the totality of the discovery experience, but simply reflect some of the broader shifts that we uncovered in the literature. The consensus on the core focus of each of the trends was unanimous among the Trends group members. The team feels very strongly that these trends should be considered in local discussions of discovery systems.

During the first four weeks of this project, the objective for the “Stats” group was to gather and analyze statistics on our current local discovery systems. The Stats Team was tasked with identifying current systems and pathways; gathering, parsing, and analyzing statistical data; and uncovering trends from that data. Each Stats group member was assigned to one or more local systems and was responsible for becoming a local expert on the statistics associated with that system. The outcome of this work was the set of baseline reports included in this report and its appendices that describe the use of our systems and a progress report on our statistics-gathering activity. In order to provide a cohesive picture of our discovery environment, rather than

just disconnected pieces, we have chosen to give a general overview of all the systems in the section entitled “A Month of Library Discovery”.

At the end of the four-week period, we had a set of trends from the literature and a draft set of baseline reports on our systems. At that point, we reorganized our working groups so that small groups of both Trends and Stats members worked together in an attempt to obtain evidence for each of the trends from statistical data on our own local systems. Thus each trend has a section describing evidence of this trend in the literature and a separate section showing evidence – or lack of evidence – of this trend locally. In our attempt to bridge the broader concepts with daily operations, we noted questions that we were unable to fully operationalize and lapses in data gathering. It was our sense that this kind of analysis would greatly benefit the Libraries if conducted more frequently. Concrete recommendations to assist with future data gathering and analysis have been included at the end of each of the baseline reports, and compiled into a general Recommendations section. Because Phase 1 is structured as a data-gathering period, we have refrained from making strategic suggestions outside of those affecting local statistics-gathering practices.

The co-chairs expressed a strong desire for all group members to be engaged with the process and to produce deliverables that were accurate expressions of what was learned by the whole group. As such, we have tried to create opportunities for maximum involvement; everyone has participated by completing explicit tasks and in shaping the ideas that are described in this report.

### *Looking ahead to Phase 2*

The deliverables for Phase 1 stand on their own as informational pieces, but are particularly useful as a base when considering the future challenges of Phase 2. In Phase 2 we will identify and prioritize recommendations for facilitating discovery according to the trends and data from Phase 1. We will also evaluate specific systems and opportunities for collaboration on discovery. We have endeavored to maintain a bright line between our Phase 1 and Phase 2 work. This means that we have avoided any discussion of strategic responses to the trends identified in this report.

## TRENDS

### *External factors*

As mentioned at the outset of this report, we have taken a very user-centric approach to our work on this project. This extends to the trends that we’ve identified, each of which describes changes in user behavior. With that said, we felt strongly the need to describe two of the external factors which don’t fit neatly into user behaviors, but which affect the information environment in which our users operate.

The first of these factors is the recent Google Book Search settlement with publishers and the Author’s Guild, which provides a subscription model for access to digitized in-copyright books. The settlement also establishes a revenue sharing model in which Google, publishers, and authors share in the proceeds of subscription fees.

In this new profit model we see echoes of iTunes, which revolutionized the music industry by creating an innovative means of connecting users to digital content in a system that adequately compensated labels and artists.

It is our expectation that the Google Book Search revenue model will continue to attract more publishers and authors, and that those authors whose work is available through the service will see that work cited more frequently. We also expect that the availability of Google Book Search, whether the University Libraries subscribes or not, will alter our users' expectations of the amount, quality, and authority of full-text works available freely on the open web.

The second external factor we wish to acknowledge is the increasing pace of change in technology and user behavior. Ray Kurzweil and others have noted that what may appear to us in the moment as a steady rate of innovation, when viewed over many years proves instead to be an exponential increase. We expect that keeping pace will require adoption of a deliberately iterative approach to discoverability. Past precedents for technology adoption will prove inadequate models for the future. We must approach each purchase, development, and organizational decision with the assumption that we will need to dismantle and repurpose sooner than we expect.

### *Trend #1: Users are discovering relevant resources outside traditional library systems*

Search, once one of the key skills and specialties of librarians, is now a daily activity for the vast majority of our users. This is due in large part to the increased use of search in non-academic research contexts. What once required specialized knowledge and mediated access is now accomplished millions of times a day via Google using simple keyword searching, with results that are perceived as "good enough." Our users approach their research with an established history of search success which gives them confidence in their search skills.

### **Review of the literature**

The ubiquity of search is reflected in several studies. The ECAR Study of Undergraduate Students and Information Technology, 2008, found that "A full 79.5% [of undergraduates] give themselves glowing reports about their ability to 'use the Internet effectively and efficiently to search for information,' with half rating themselves as 'very skilled' and another third rating themselves as 'experts'" (Caruso and Salaway 11). Furthermore, over 80 percent of respondents to the ECAR survey reported that Internet searching was a preferred method of learning (Caruso and Salaway 49). A study jointly funded by the British Library and Joint Information Systems Committee (JISC) found that, "Many young people do not find library-sponsored resources intuitive and therefore prefer to use Google or Yahoo! instead" (Rowlands et al. 296). Taking a pessimistic view of this trend, the study went on to note that, "intervention at university age is too late: these students have already developed an ingrained coping behaviour: they have learned to 'get by' with Google" (Rowlands et al. 303).



Usability tests by the University of Washington demonstrated undergraduate students' reliance on Google for academic research: "Many students reported that when preparing to do research, they use Google or Amazon as a starting point to help learn more about a subject they were unfamiliar with (several cited Wikipedia as a specific resource they would look for in Google search results) or to help them identify a research topic" (Shadle 85). Even those students who were familiar with library resources and who had been trained to begin their research within that scope, "expressed a preference for the ease of Google searching and [for] the fact that the information returned with the link would often provide enough context to support decision-making" (Shadle 86).

These behaviors are not restricted to undergraduates. Nor, would it seem, are undergraduates alone in their view that Google can provide quality results for academic research. When interviewed for our recent science assessment, a professor of physiology here at the University of Minnesota noted that Google is "surprisingly good, but also lamentably good...I mean, I know I shouldn't use Google, but it's too good not to" (Marcus et al. 13). Focus groups convened by OCLC researchers and comprised of faculty, graduate students, and undergraduate students spoke with one voice regarding Internet searching: "All the participants in the focus group interviews said that the first place they look for information is the Internet, closely followed by human sources." (Prabha et al. 15). This finding is supported by a study funded by publishers of academic journals, which found that compared to library websites and OPACs, "generalist search engines have gained in popularity since 2005 to the relative detriment of all the other potential starting points"(Inger, S and Gardner, T 7). This same study found that researchers were almost as likely to start at a general web search engine (this includes both Google and Google Scholar) as at a library website or OPAC when finding a known item from a reference or citation (Inger, S and Gardner, T 8). Further, when searching for unknown articles on a specific subject, researchers were nearly twice as likely to begin at a general web search engine than at a library website or OPAC (Inger, S and Gardner, T 10).

"No one goes to libraries to find things anymore. That is the rightful business of search engines; it is what they do well. Libraries must work to make their collections easily discovered by people -- not just by sophisticated librarians. They must be able to open up access to their collections through many different doors. Libraries must be available everywhere" (Brantley).

### **Evidence in local systems**

We would expect to find evidence of this trend in several of our local systems. SFX, our link-resolution system, acts as a conduit for users to connect to our licensed content from OpenURL-enabled abstracting, indexing, and search tools. Logs in SFX record the referring site, and we would expect to find that an increasing number of users are arriving at our licensed resources via Google. In fact, for the period of July - November 2008, the top five referrers to our SFX system were, in order, the SFX A-Z listing of e-journals, PubMed, MNCAT, Google Scholar, and the SFX citation linker. Google Scholar has risen to fourth place in this list from fifth place behind OVID Medline during the Libraries' fiscal year July 2007-June 2008, and seventh place during July 2006-June 2007.

SFX also logs the number of referrals that result in a click-through, indicating the number of users who were directed via the SFX menu to a licensed resource, InterLibrary Loan, or to the OPAC. Our statistics show that referrals to SFX from Google Scholar result in click-throughs at a rate comparable to the A-Z listing and MNCAT, indicating that Google Scholar users are likely as successful in finding our licensed resources as users of our local systems. (See SFX baseline statistics report).

Evidence of user preference for Google searching as a means of navigating the web is more pronounced in statistics gathered from our web server logs. Users arrive at the Libraries' websites from Google in greater numbers than from all other sources. In October, 2008 visitors from Google viewed 30,283 pages on the www.lib domain, surpassing the number of visitors from the University's domain (www.umn.edu) at 18,648 pages, and dwarfing the page views of visitors from our branch library subdomains. For example, visitors from the Bio-Medical library subdomain ([www.biomed.lib.umn.edu](http://www.biomed.lib.umn.edu)) to the Libraries' main web domain viewed 2,089 pages, and visitors from the Science and Engineering Library subdomain viewed 2,017 pages. During October 2008 fully 79% of page views on our Government Publications site were referred by Google searches.

Not surprisingly, the volume of traffic referred by search engines to a given subdomain does seem to vary according to the type of information present on that subdomain. Sites that contain rich information sources or guides benefit from greater search engine traffic than do others. For instance, the Learning Resource Center subdomain which otherwise receives little traffic, contains a series of lectures in mp3 format which have been indexed by several multimedia search engines. During October 2008, the LRC subdomain had 50 referrals from the Libraries' primary web domain, and 45,989 referrals from the Tagoo multimedia search engine.

Users of AgEcon Search, a document repository run by the Libraries on the DSpace platform, arrive from Google almost exclusively. In October of 2008, the top websites referring users to AgEcon Search each generated fewer than 300 page views. During that same period, visitors from Google viewed 110,491 pages. The bandwidth used by Google's indexing spiders on AgEcon Search is massive, 14.83 GB in October 2008. In contrast, the Google spider accounted for only 828 MB during the same period on the Libraries' main website (<http://www.lib.umn.edu>). We speculate that this bandwidth use is due to wholesale and repeated downloading of the PDF files stored in AgEcon Search. (Note: It is apparent from the logs that some searchers are using web search engines merely as a navigation device to access the AgEcon search page, and not as a discovery tool.)

User preference for alternative discovery mechanisms is visible in statistics from our Aleph OPAC as well. Though the OPAC has been deemphasized since late summer 2008 when we launched MNCAT Plus, it still sees significant traffic, of which 10% originates outside of our web catalog interfaces. Most of the external requests are generated by the Twin Cities instance of SFX and by various OCLC services.

The University Libraries provide a LibX web browser extension that will connect users to items in our catalog from external websites. For example, when a user with the LibX extension installed visits an item page for a book on Amazon.com, the extension will query our catalog to determine if the book is in our collection. If so, it will place a link to the item inline on the Amazon.com page. Unfortunately, at this time we are unable to isolate statistics on users who arrive at our resources via other discovery or search interfaces using the LibX

extension. These statistics would provide some indication of the research habits of users who are intentionally bypassing our search and discovery interfaces.

### *Trend #2: Users expect discovery and delivery to coincide*

Searchers do not distinguish between discovery and delivery in their web searches and increasingly find it discordant to experience this disconnect in the library environment.

#### **Review of the literature**

Today's searchers increasingly expect discovery of relevant resources and delivery of those resources to occur simultaneously. In the past, library users could anticipate the discovery of a surrogate record in a catalog, perform a preliminary assessment of relevance using that surrogate record, and then use an identification key, such as a call number, to locate the physical material. Many library systems are still designed using this mental model. With the availability of digital and digitized content online today, however, searchers increasingly expect – and prefer – to access the resource itself immediately online, or have it delivered to them conveniently as a second choice. Online searchers expect to find not only metadata describing an information resource, but also expect to quickly and easily access the information resource itself without having to do a lot of clicking or go to another system for delivery.

A survey conducted by the Research Information Network illustrates the growing expectation that discovery and delivery are melding as a concept: “Many, if not most, researchers do not draw a hard and fast distinction between ‘means’ and ‘ends’, between discovery services and what is being discovered. Initially, it is possible to see this as based on lack of familiarity with discovery services or misunderstanding, but it is actually a reflection of the real experience of research in the digital environment and of the design of services which seamlessly (at least in theory) move from reference to abstract to full-text in a way that was not possible in the print-on-paper world of indexes and card catalogues” (Research Information Network).

In situations where immediate delivery is not possible, seamless integration with physical delivery systems is expected. One participant in the University of Minnesota's “Understanding Scientists” report complains about “too much clicking” to get to services/resources. This problem is exacerbated for Academic Health Center scholars working at the University of Minnesota - Duluth, who have to go through extra “clicks” to access Interlibrary Loan, full-text electronic resources, and other library holdings” (Marcus et al. 11). The 2006 OCLC report, “College Students’ Perceptions of Libraries and Information Resources” discusses users’ expectations of “ubiquitous access to content” (DeRosa et al. 31). Many of the negative associations reported in this study pertain to information resources obtained via libraries being difficult to find, access and/or use.

In “The Library Catalogue in the New Discovery Environment: Some Thoughts” from 2006, Lorcan Dempsey emphasizes the importance of providing easy, convenient delivery tools integrated with discovery. In the pre-network era, when searchers’ information universe was largely confined to those materials owned by the library and information was “scarce”, searchers were more willing to work hard to acquire the actual information resources. Now, in an age of information “abundance” and networked systems, in which the

information universe is much larger than library collections, searchers are less willing to spend extensive effort to acquire the actual content. Also, users who are familiar with the aggregated, consolidated information resources and services provided by entities such as Amazon, Google and iTunes have less patience with the fragmented, disaggregated discovery and delivery systems provided by libraries. In other words, users no longer build their workflows around the library. According to Dempsey, libraries therefore need to figure out how to integrate our discovery and delivery systems into the workflows of contemporary users (Dempsey).

### **What convenience and timeliness now mean**

Even if the library doesn't own the item, users expect to be able to access or request the item easily and quickly. In his article, "The Local Catalog is Dead! Long Live the Local Catalog", Steve Shadle describes a typical response from study participants when they reached a dead-end using the available library services: "If I really wanted it, I would go to Amazon and buy it" (Shadle 86) By incorporating a seamless delivery flow into their implementation of WorldCat Local, the University of Washington Libraries saw their consortial borrowing increase more than 60% and interlibrary loan requests nearly triple. Libraries are rightly concerned about ownership of resources, but many users are concerned about ownership only as it impacts the availability of the material.

Anecdotally, members of our reference staff have observed students who do not want to come to the library to pull a print volume from the stacks and will give up on an information resource if the full text is not available electronically, perhaps even changing their topic because they "can't find anything". A participant in the 2007 University Libraries study "Understanding Research Behaviors, Information Resources, and Service Needs of Scientists and Graduate Students: A Study by the University of Minnesota Libraries" states: "Scholars still use the physical library to check out older books, journals, and conference proceedings that are not online, but the overwhelming preference is to access resources electronically because of it is efficient, easy, and convenient" (Marcus et al. 12). This convenience is especially important to students, staff and faculty who are distributed geographically across multiple campuses and/or who are distance learners.

We should point out that convenience and timeliness have always been primary factors for users as described in the information-seeking behavior literature. In "What is enough? Satisficing information needs", Prabha et al, quote Leckie et al. (1996, p.185) in a description of the critical nature of timely delivery of information for engineers and attorneys: "It is often important that the information be obtained immediately or within an acceptable period of time. Its usefulness and impact will decrease if it is obtained too early or too late" (Prabha et al. 74). Furthermore, in the same article Prabha et al, report results on their own study showing that "Faculty's criteria for stopping an information search are based on publication requirements and deadlines and the amount of time available for preparing and delivering lectures and presenting papers. Time constraints are an overwhelming factor for faculty in deciding how much effort they are willing to invest in satisficing their information needs." (Prabha et al. 15). The imperative for timeliness and convenience has not changed in the information-seeking process; what is changing is the shift in what convenience and timeliness mean. Discovering the surrogate for the item is not sufficient for today's users – they expect to be linked directly to the electronic full text of the item or to a convenient, simple way of placing a hold request or an interlibrary loan.

## **Evidence in local systems**

We had hoped to see evidence of this trend in our local systems, specifically:

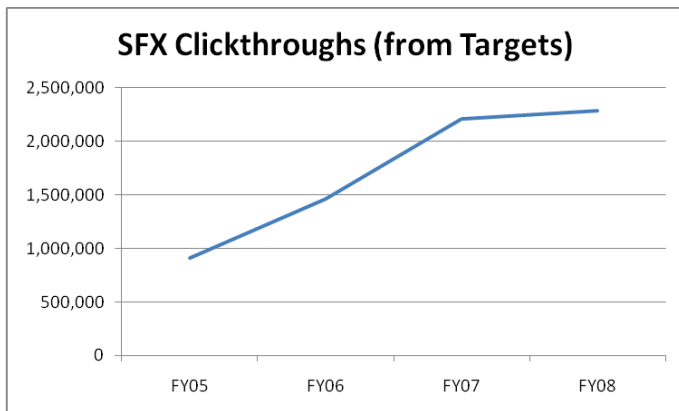
### *1) Statistics on the use of Recalls and the new Get It delivery service available via the catalog*

The University of Minnesota Libraries Twin Cities implemented a streamlined book delivery system in March 2008. Upon discovering desired material within the catalog, searchers could use the “Get It” link to request that material be delivered to the library of their choice or to their office (for faculty and staff). This feature complemented the “Recall” function already available within the catalog, which allows users to recall already checked out to another patron.

In general, overall circulation numbers have been decreasing over the past five years: Faculty by 20.7%, undergraduates by 20.3%, and graduate students by 39.3%. This decrease is particularly striking in that we have an increase in the overall number of active faculty and graduate students patrons (17.5% and 9.6%, respectively, from five years ago). Even given this decrease in general circulation, recalls have continued to increase in popularity. Undergraduate requests are up 296% over five years ago, while graduate student numbers are up 69%, and faculty recalls are up 87.5%. Since the inception of Get It in March 2008, 15.5% of loans made in UMTC libraries result from a Get It request.

### *2) Growth in use of SFX*

Tracking the “click-through rate” over the past few years based on the targets accessed shows a distinct increase in usage. While this change is to be expected, perhaps, given the increased volume of materials available, our group believes that the fact that more resources are available is also an indication of this trend.



### *3) A review of the users who choose not to pursue print access when they discover that full-text is not available via SFX*

We know that during October 2008, our users were presented with full-text (FT) options 73% of the times they viewed the SFX Find It menu.

FT option(s) presented:	288,085
No FT option presented:	104,375
Total presentations:	392,460

Of those 104,375 non-FT options, it would be valuable to know if those searchers decided to stop their search at that point, or if they continued to pursue other delivery options. While we know that the general click-through rate for the SFX menu is also at 73%, we do not have the data to make a correlation between full-text options being presented (or not) and when users click through the menu. If we can find a way to track what users do when presented with non-full text options in the SFX menu, our group believes that this would be valuable data to have, particularly as it may reveal situations in which a searcher perceives her search attempt to have failed.

#### *4) A review of the use of AgEcon Search and UDC as local examples of immediate fulfillment services*

The University of Minnesota Libraries sponsors the University Digital Conservancy and AgEcon Search as two open access repositories of full-text scholarly literature, the UDC containing work of the University of Minnesota and AgEcon housing papers specific to Agriculture and Applied Economics. As two of the few local examples of an immediate fulfillment service, we thought it might be valuable to include general usage statistics for these systems.

During October 2008, over 52% of the page views (113,674) for AgEcon came from web search engines. And as we know from other statistics such as the search terms and referring pages, users often get to directly to AgEcon resources through means other than the main search screen. According to local statistics, 274,975 .pdf files were downloaded during October 2008, discounting robot traffic or automated downloads. One item was downloaded 1,123 times; the next highest counts are in the 200's, suggesting a very "long tail".

As opposed to AgEcon, the UDC statistics have not been adjusted to account for multiple downloads being triggered automatically via scripting (in order to push up specific articles in rank). It is assumed that the predominance of Direct Address as a starting point is due, at least to some degree, to this lack of adjustment (as it was for AgEcon). Even so, it is interesting to note that over 12% of page views are generated by referrers, and at least a few page views are coming from Wikipedia. Looking at the full list, three different Wikipedia pages send traffic to papers or maps in the UDC.

### ***Trend #3: Increasing usage of portable Internet-capable devices***

Rather than just supplementing the desktop computer, mobile devices are poised to become the primary means of Internet access for a critical mass of our users.

## **Review of the literature**

While Americans have thus far lagged behind the rest of the world in using mobile phones for Internet access, the advent of the iPhone and other mobile devices with larger, touch-screen interfaces seems to have pushed us past the tipping point at last. The 2008 Horizon Report lists mobile broadband as one of the key emerging technologies likely to have an impact on teaching, learning, or creative expression within learning-focused organizations in the next two to three years. Citing the rapid innovation and falling prices that characterize this market today, the report finds mobiles are "quickly becoming the most affordable portable platform for staying networked on the go,"(p.4) and predicts that "over the time frame of this adoption horizon, it is expected that mobile broadband, full-featured Internet, touch-screen interfaces, remotely upgradeable software, and high-quality displays will become as common as cameras are today" (The New Media Consortium and the EDUCAUSE Learning Initiative 17). While claiming only a more conservative time frame of ten to twelve years, the Pew Internet & American Life Project's recently released report, "The Future of the Internet III," is just as unequivocal, finding that "the mobile device will be the primary connection tool to the Internet for most people in the world in 2020" (Anderson and Rainie).

The mobile technology trend also features prominently in the ECAR Study of Undergraduate Students and Information Technology: "One of the most significant trends we report this year is the continuing 'mobilization' of the student body. Laptops continue to gain as the computer platform of choice, and two-thirds of our respondents report owning an Internet-capable, though a variety of barriers -- probably high monthly data-access fees above all -- mean that only a fraction use their portable devices to access the Internet. That caveat notwithstanding, it appears the stage is set for a potential tidal wave of new student demands for mobile service support" (Caruso and Salaway 6)

The recently released preview of the 2009 Horizon Report moves the adoption time frame up to one year or less, citing innovations such as touch screen interfaces, accelerometers, and built-in GPS, as well as the ability to connect to the Internet via both wi-fi and increasingly higher speed 3G networks, which have greatly improved their functionality and usability as Internet access devices. Another key factor is the opening up of the devices to third-party developers. "New interfaces, the ability to connect to wi-fi and GPS in addition to a variety of cellular networks, and the availability of third-party applications have created an almost entirely new device with nearly infinite possibilities for education, networking, and personal productivity" (The New Media Consortium 2).

There is growing evidence that people are beginning to use their devices for Internet access. Two surveys give slightly different statistics, but draw a similar picture of the trend:

14% of 18-27 year olds (9% of all adults with a mobile phone) use cell phone or wireless device to search for information. (North American Technographics Benchmark Survey, 2007, quoted in (Kroski January 29, 2008))

11% of US mobile phone users age 18-64 access the Internet (44% text message). 20% of users in Japan (Office of Communications (OfCom), UK, "The International Communications Market, 2007," December 2007, quoted in (Fox)).

Almost 25% of undergraduates access the Internet from a mobile phone or PDA at least monthly, and 17.5% do so at least weekly (Caruso and Salaway 10).

The data suggests that as newer interfaces as epitomized by the iPhone are beginning to catch on, mobile Internet use is increasing markedly:

58.6% of iPhone, 37% of "Smartphone" users accessed web search (M:Metrics, 2008, quoted in (Fox)).

This trend moves beyond the traditional cell phone or PDA device into any technology where discovery occurs in a critical mass on a platform or using a form factor beyond the traditional web browser. One arena that has particular relevance for discovery is that of e-book readers. The group notes the ascendance of the Kindle, which has gained intense popularity over the last year. Although Amazon doesn't report sales figures itself, industry analyst Mark Mahaney of Citigroup estimated in August that Amazon would sell 380,000 Kindles in 2008. According to Mahaney, the Kindle adoption rate is similar to that of the Apple iPod during its first year (Mahaney). Software for reading ebooks on other devices such as cell phones is also becoming increasingly popular (Kharif). We can no longer expect that the traditional web browser will continue be the favored or most progressive interface for access to our systems.

Research done on the University of Minnesota campus corroborates this trend: "There is very strong support among students for the use of mobile technologies. Large majorities of survey respondents reported owning cell phones, laptops, and iPods, and nearly all of the rest reported wanting to own them. Personal digital assistants and smart phones formed a second, though noticeably less popular, grouping" (Walker and Jorn 6). The report goes on to recommend that the University "leverage mobile technologies. The immense popularity of cell phones, MP3 players, and laptop computers means that most students are well prepared for the integration of these devices into their education" (Walker and Jorn 8).

Other academic libraries working actively in this area include

- Ball State University (<http://www.bsu.edu/libraries/mobile>)
- Boston University Medical Center (<http://med-libwww.bu.edu/mobile>)
- University of Richmond (<http://oncampus.richmond.edu/academics/library/mobile/index.htm>)
- University of Virginia (<http://mobile.virginia.edu/library.php>)
- Wayne State University (<http://elibrary.wayne.edu:6060/airpac/>)
- North Carolina State University (<http://www.lib.ncsu.edu/m/catalog>)

All have developed mobile interfaces to their library catalog, and many also include links to library hours and other basic information about library services. Ball State University also allows searching journal articles or a web repository and includes links to "how-to" videos. The University of Richmond's mobile interface includes reference service via text messaging, email, or AIM. Simmons College is taking a different approach: Their library catalog (<http://library.simmons.edu/search>) provides a button for users to send the title and location of items found to their mobile devices as text messages.



## **Evidence in local systems**

We had hoped to see evidence of this trend in our local systems, specifically:

### *1) Accessing the University Libraries website and sub sites using a mobile device*

The University of Minnesota Libraries has almost no statistics on mobile use of its resources. Our current web statistics software flags major operating systems and browsers. Unknown operating systems accounted for 1.6% of usage at [www.lib.umn.edu](http://www.lib.umn.edu) in October 2008. There is a decent chance that mobile browsers could make up a significant portion of this statistic. Even so, mobile use currently represents a tiny fraction of library website usage. The Symbian mobile operating system registered 101 hits in October and is the only known mobile browser whose use we currently track. 6.2% of the traffic in October was via the Safari browser, so it is reasonable to postulate that a certain percentage of that is from the iPhone. There is little reason to expect much traffic from mobile devices when the design of our web pages is not optimized and mobile usability is so poor.

### *2) Use of the Aleph OPAC (MNCAT Classic) via a mobile device*

In order to get statistics on the number of times users are accessing the OPAC via a mobile device, we need to be able to track and parse the User-Agent lines from our Apache logs. Unfortunately, the OPAC's current configuration does not log User-Agent, so we do not have any of these numbers.

### *3) Use of Primo (MNCAT Plus) via a mobile device*

As with the Aleph OPAC, we need to track and parse the User-Agent lines from our Apache logs in order to gather statistics on mobile devices. Like the OPAC, Primo is not currently configured to track that information.

Although there is little data available to suggest any significant use of mobile devices to access U of M Libraries resources currently, there is a great deal of data from both the University of Minnesota and outside sources to support the growing popularity of mobile devices as a platform for Internet access and information-seeking. Nationally, 9 out of 10 (=5.2 million) students own cell phones (Student Monitor marketing research, [www.studentmonitor.com/telecom.php](http://www.studentmonitor.com/telecom.php), accessed 17 December 2008). According to the ECAR Study of Undergraduate Students and Information Technology (2008), just over 66% of these phones are Internet-capable. As of 2007, of students (both graduate and undergraduate) at the University of Minnesota:

- 23.8% own a PDA (another 22.1% would like to)
- 5.8% own a Smartphone (another 36.6% would like to)
- 92.3% own a cell phone (another 2.4% would like to)
- 68.6% own an iPod or MP3 player (another 18.7% would like to)

(Walker and Jorn 12)

This is a trend that seems self-evident, and yet we have done little to optimize our systems for mobile use. Our group highly recommends that we take steps to reconfigure our systems to ensure that we can begin to track usage of mobile devices. For example, there is a plug-in for the AWStats software package that adds support for over 300 phone types. We recommend installing and using this tool to begin providing us detailed

information on the mobile usage patterns of our users. Furthermore, we suggest reviewing the system configuration for our primary discovery systems in order to gather the data we need for effective decision-making.

#### *Trend #4: Discovery increasingly happens through recommending*

Facilitating discovery requires us to develop and implement systems that push relevant content to users and that allows users to share content with others.

Discovery of relevant information resources has traditionally been understood as the result of an active process of searching or browsing, either in a physical or a virtual environment. The vast array of information sources available to today's scholars requires that they rely increasingly on formal or informal recommendation systems for having relevant resources pushed to them, in addition to actively pulling information through searching and browsing. These recommendations can be loosely categorized as either system-generated or social/peer-generated.

E-commerce has introduced most of us to highly-targeted system-generated recommendations intended to connect us to products. These recommendations are typically based on the characteristics of a given product (e.g. "if you like this product you may like this similar product," accessories for the product you are currently viewing), or the characteristics of a user's behavior (e.g. "people who bought A also bought B"). These recommendations are made possible by the collective intelligence inherent in a system with a critical mass of users and critical mass of data.

The opportunities for social networking created by ubiquitous communication technology have transformed the peer-to-peer recommendations that have long been a crucial part of scholarship. Listservs of like-minded researchers have been joined by blogs, microblogging services like Tumblr and Twitter, and online social networking communities as outlets for scholars to actively or passively make colleagues aware of relevant information.

#### **Review of the Literature**

A task force report on resource discovery from the University of Wisconsin identifies that a notable 72% of the surveyed group reported they rely on some form of recommendations to share information, with a strong emphasis on peer-to-peer recommendations. "...nearly half of users on social networking sites share info on subjects of interest and books read, that they discover new info by talking to peers, and that they value personal recommendations" (Dentinger et al.).

As Peter Brantley notes in his article, Architectures for Collaboration, libraries need to support the idea that it is the connecting of people and communities that help to enrich collections, change knowledge and transform information. According to Brantley, building "paths for exploration" rather than just collections is a critical concept for how libraries can better prepare for and serve the future. He writes, "The world is moving to broad-based participation in information creation and information discovery."

A study of the "Google generation's" information seeking behavior, funded by the British Library and JISC, determined that "social networking is part of a wider trend" and that "...users [are] creating and posting content for themselves, blurring the age-old distinction between information producers and information consumers." The ECAR study of undergraduate use of technology substantiates that there is a current trend in the use of social networking sites. The ECAR study found that out of their 27,000 respondents, 49.7% claimed that they communicate with peers about coursework through social networking sites. The depth of these communications about coursework is unknown, but the potential is certainly significant (Caruso and Salaway).

In a survey of "knowledge workers," the Outsell research firm found (Information Management Best Practices: The Search For Search. Outsell Vol 9, 7/7/2006) that "users don't know if the information they want is available in a nontraditional format, like a colleague's email (listserv), blog post." The firm noted "a shift in preferences away from traditional citadels" of information to "wider adoption of peer-to-peer exchange and primary, real-time data."

The 2008 Horizon Report forecasts a two-to-five year horizon for the widespread adoption of collective intelligence, indicating that, "new forms of information stores are being created in real time by thousands of people in the course of their daily activities, some explicitly collaborating to create collective knowledge stores like the Wikipedia and Freebase, some contributing implicitly through the patterns of their choices and actions" (The New Media Consortium and the EDUCAUSE Learning Initiative 23). Peter Brantley encourages libraries to take advantage of collective intelligence, writing that through our services, "people change knowledge and transform information. People—not library curators—enrich our collections. Libraries must be designed to get better through use" (Brantley 32).

### **Evidence in local systems**

With the adoption of Ex Libris's Primo discovery interface, branded locally as MNCAT Plus, the University Libraries now provides our users with simple, built-in tools for sharing and recommending resources from our collections. MNCAT Plus users can email records to colleagues, post links to resources directly to Del.icio.us or Connotea, and tag items in the catalog. MNCAT Plus also provides users with suggestions for new searches in the form of links to subject headings relevant to the user's search terms. We would hope to see evidence of the use of these recommendation tools in our MNCAT Plus statistics, showing not only how often the recommendation features in MNCAT Plus were used, but also how frequently users were referred back to MNCAT Plus from social bookmarking sites. However, our MNCAT Plus statistics do not provide answers to all the questions that we would like to ask. We do know that in October 2008, users clicked on the "Add Tags" link 731 times, and click on the "Send Email" link 1374 times.

We also have data from our Aleph OPAC, branded locally as MNCAT Classic. During October 2008, searchers used the email feature 12,194 times. Our statistics do not indicate how often users email records to themselves, making this statistic only a marginally useful indicator of our users' tendencies to recommend our resources to others.

Statistics from our web servers indicate the power of recommendations and collective intelligence to drive traffic to information sources. For example, our analysis of October 2008 statistics for the Archives and Special

Collections subdomain shows that 14.6% of the 5,050 people who entered our site from another web page got there from a Wikipedia entry. The Electronic Text Resource Center (etrc.lib.umn.edu) is an earlier project of the University Libraries, now officially “closed.” Although the Libraries no longer updates or actively supports the site it still draws significant traffic from other academic institutions that post links to ETRC collections, such as the Women’s Travel Writing Project, in syllabi and pathfinders.

The University Libraries have two ongoing projects that attempt to address this trend. Our tracking of affinity strings is laying the groundwork for providing users with system-generated recommendations based on the research behaviors of their peers (see Baseline Reports). The EthicShare project is aimed at creating a social network for researchers that allows users to share citations, recommending them to one another. These projects are both in very early stages and will provide useful statistics relating to this trend in the future, but do not yet have significant data to analyze.

### *Trend #5: Our users increasingly rely on emerging nontraditional information objects*

The format of useful and discoverable information objects is much broader than those traditionally offered through libraries; users increasingly rely upon multimedia objects, data sets, blogs, and other “grey” objects to meet their information needs.

#### **Evidence of the Trend in the Literature**

The digital information universe of the average person has expanded well beyond traditional, text-based information. Free or inexpensive tools have facilitated the ability for the average person to create and discover sound, images, video, geographical data, and other non-text-based objects. In "Beyond the Hybrid Library: Libraries in a Web 2.0 World", Derek Law argues that “the nature of content has also progressively changed while libraries have not. The nineteenth and much of the twentieth century can be defined in terms of words, whether spoken or written... Conversely the last fifty or so years can be defined almost entirely in images: film of the burning airship Hindenburg; the Dunkirk beaches; the mushroom cloud of an atomic bomb, the assassination of JFK; Neil Armstrong stepping on the moon; the beauty of fractal images; the obscenity of the aircraft crashing into the Twin Towers. Digital natives expect image content.”(Law 107-118). Locally, we also find demand for use of non-textual objects. In the University of Minnesota study, “A Multi-Dimensional Framework for Academic Support”, published materials were the most heavily used for research projects (>98% for all disciplines interviewed typically used published materials in projects). However, the study also found that “images and audio and video recordings also represent well-used materials (images 41.3%; film/video recordings 25% total; audio 12.9%).” (University of Minnesota Libraries)

In addition to images, audio, and video, searchers are increasingly seeking to discover data sets and information about data sets. The Mellon study found that the use of data sets was considerable in one discipline of the disciplines included in the research: “48.1% of social scientists rely on data sets and/or statistics...”. The University of Minnesota study, “Understanding Research Behaviors, Information Resources,

and Service Needs of Scientists and Graduate Students: A Study by the University of Minnesota Libraries,” includes a relevant quote from a bioinformatics faculty member asking about better ways of working with data sets: “If I am taking raw data – sequencing work of a genome --- can I Google the data to find out what’s known about it? ... Are there search engines just for data sets, even the ones that are constantly changing?” (Marcus et al. 18).

The proliferation of new publishing platforms such as blogs and wikis, and increasing availability of pre-print archives, has transformed even the realm of text-based information. For some researchers, the move toward using non-traditional objects for discovery may be driven by the view that books and articles are outdated by the time they are published. The University of Minnesota Libraries found this opinion regarding lack of timeliness of published materials in its study of Scientists and Graduate Students (Marcus et al. 13). The University of Minnesota Science Assessment also found that discovery is occurring in “websites of professional societies, researchers’ personal websites, pre-print archives ..., Google (for images and video clips for teaching, as well as hot topics), Google Scholar, personal libraries, Amazon.com (to read table of contents and receive notification of when a book will be published), and personal subscriptions to databases.”(Marcus et al. 11) Born-digital objects that meet a searcher’s information need are increasingly available quickly via search engine results.

Data in the literature provides ample evidence that users increasingly are not just seeking new forms of content, but are themselves creating them. In “Current Models of Digital Scholarly Communication”, it is posited that “the networked digital environment has enabled the creation of many new kinds of works that are accessible to end users directly. The decentralized distribution of these new digital resources can make it difficult to fully appreciate their range and number, even for university librarians tasked with being familiar with valuable resources across the disciplines.” (Maron and Smith 9) The ECAR Study of Undergraduate Students and Information Technology found that nearly half of students contribute content to photo or video websites, 32.9% use video creation software, 32.5% use audio creation software, 29.1% use podcasts, and 25% use webcasts. (Caruso and Salaway 4) A 2007 survey of students at the University of Minnesota showed that while students were less likely to have used media manipulation tools and newer communication tools to produce content than other educational technology, their average comfort-level with these technologies was quite high. (Walker and Jorn 17)

Two of the trends identified in recent editions of the "Horizon Report" relate closely to the growing importance of non-traditional data objects. The 2008 report highlights data mashups, which frequently involve data visualizations incorporating images and geographical data. "Data mashups are powerful tools for navigating and visualizing datasets; understanding connections between different dimensions such as time, distance, and location; juxtaposing data from different sources to reveal new relationships; and other purposes." (The New Media Consortium and the EDUCAUSE Learning Initiative) "Mapping geolocative data is not new, but the ability to easily create map mashups online using multimedia and geotagged data is...Many free or very low-cost tools to capture and display geolocative data are available online and are much easier to use than previously." (The New Media Consortium) If these applications meet their suggested potential as

educational tools, the ability to discover the images, datasets, and other "mashupable" information will become increasingly crucial.

### **Evidence in local systems**

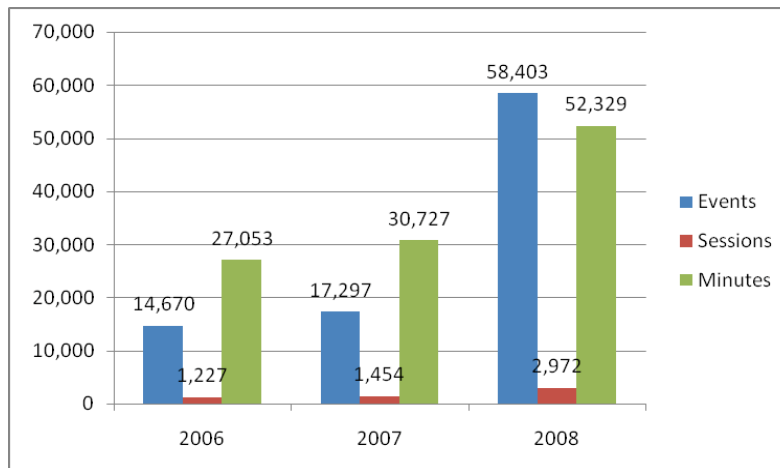
We would have expected to see evidence of this trend in the following ways:

*1. The popularity of the Libraries-sponsored UThink blog tool.*

University Libraries does not collect statistics on the focus of UThink blogs, but their wide-spread use suggests that students, staff, and faculty are increasingly comfortable creating their own online content. A survey conducted at the University of Minnesota in Spring 2006 indicated that over 42% of the active UThink blogs are related to courses. (Nackerud and Scaletta 81) Anecdotal evidence also supports the idea that blogs are used for all kinds of information-sharing and discovery activities. For example, the University of Minnesota Chair of the Department of African American and African Studies has stated, "I constantly encourage my students and colleagues to explore UThink for multiple purposes" (Jacobs, as cited in University of Minnesota Libraries 7).

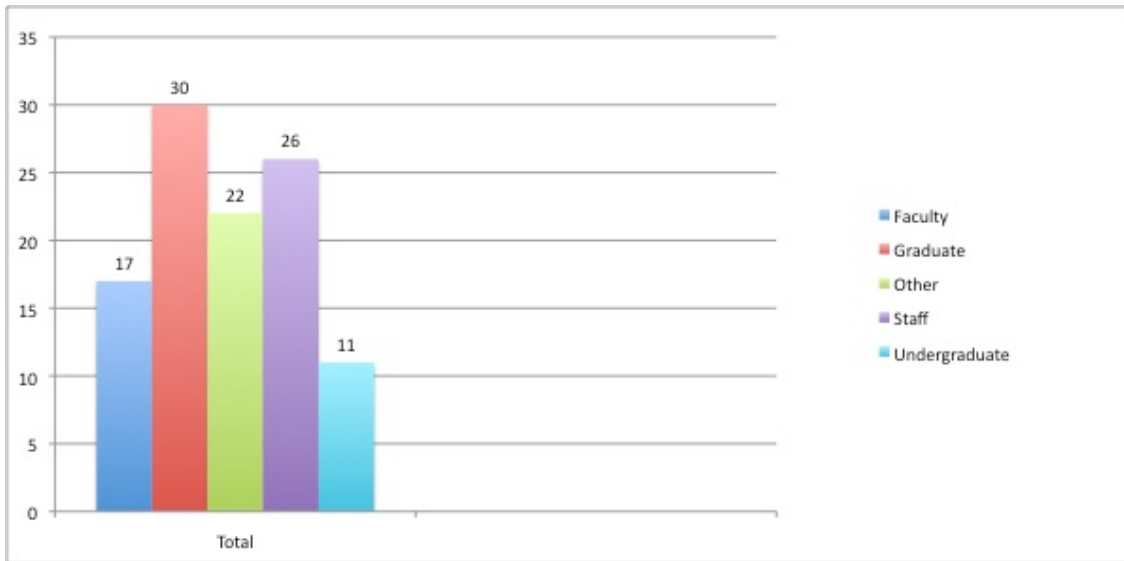
*2. Use of ARTstor*

The statistics for the ARTstor digital library clearly indicate increasing use of these objects by the University of Minnesota community over the last few years.



*3. Data sets*

The University Libraries' Data Services initiative is still in its infancy, but we are seeing measurable interest from our users in the availability of data sets for reuse. In the past 18 months, Amy West, our Data Services Librarian, has fielded over 100 inquiries from users about data. Of note is that inquiries from graduate students and University staff are more common than those from faculty.



Data Services reference statistics from our Desk Tracker system, 8/07-2/09

#### 4. Digitized content on the James Ford Bell Library site

Web server statistics from the James Ford Bell Library’s site illustrate the power of digitized images to draw users to even the more neglected corners of our websites. A collection of information on Trade Products in Early Modern History (<http://bell.lib.umn.edu/Products/Products.html>), a collection of images of Historical Maps (<http://bell.lib.umn.edu/hist/>), and a collection of images of Portolan Charts (<http://bell.lib.umn.edu/map/PORTO/porto.html>) draw thousands of visits to the Bell Library site, even though some of these pages don’t appear to have been updated in several years. Unsurprisingly, the vast majority of these visitors arrive through Google searches rather than through our own discovery tools. The popularity of these images has made the James Ford Bell Library site one of the most frequently visited of all of the Libraries’ websites.

## PRINCIPLES

The Discoverability charge requests that the group “develop a set of principles related to discovery to help guide the Libraries’ strategic decisions about the selection, development, and support of relevant tools and services.” We chose to complete this work as the final task before the writing of our report, thus adopting an approach that best utilized all the good work previously done with trend identification/scoping and statistical analysis. While a trend describes a trajectory for user behavior, a principle suggests how the trends, external factors, and statistical data should inform our choices and recommendations in Phase 2 and beyond.

Our interpretation of our charge was that we were being asked to look at discovery from the perspective of our users, no matter whether the object of their discovery was a part of our collections. Likewise, we wanted to consider discovery of resources in our collections from the perspective of both affiliated and unaffiliated users. Finally, knowing that our principles will be used for decisions about systems, we wanted to consider the

characteristics of an ideal system. In hopes of capturing each of these aspects of discoverability in our principles, we chose to open our discussion of potential principles with these three framing devices:

- “In order to best facilitate our users' discovery of relevant information, we must...”
- “Making collections discoverable requires...”
- “Any discovery system must (be, have, do)...”

The Discoverability team recommends the following eight Discovery Principles to be used when evaluating discovery technology at the University of Minnesota Libraries.

1. Users draw little distinction between discovery and delivery; systems, data, and information objects should be optimized for fulfillment. (*Trend 2*)
2. In order to remain agile and responsive in a rapidly changing information environment, our systems and data structures should provide us with the greatest possible flexibility for frequent iteration and reuse by ourselves and others. (*External factors, Trend 5*)
3. In order to best facilitate our users' discovery of relevant information, we should strive to be end-user device/platform agnostic. (*Trends 1, 3*)
4. In order to remain responsive, relevant, and useful to our users, we must aggressively measure and analyze user behavior through local system statistics. These efforts will complement our ongoing assessments utilizing focus groups, usability studies, and reviews of applicable literature. (*External factors, Statistics recommendations*)
5. Discovery should be organized around users rather than collections or systems. This organization should be based on realistic, evidence-based models of our users and their research tasks, e.g. task-based, persona-based, audience-targeted. (*Trends 1, 4, 5*)
6. Users are successfully discovering relevant resources through non-library systems (e.g., general web searches, e-commerce sites, and social networking applications). We need to ensure that items in our collections and licensed resources are discoverable in non-library environments. (*Trends 1, 4*)
7. Making collections discoverable requires optimizing for access by local and non-local user populations; being good stewards of our collections means participating in cooperative ventures that provide broad access to our collections. (*External factors, Trend 1*)
8. Users rely on system- and peer-generated recommendations to discover relevant resources. We should capture the data necessary to provide targeted suggestions to users and defer to network-level systems where a critical mass already exists. (*Trends 1, 4*)

The Discoverability group was also asked to provide a summary of guiding principles related to discovery used at our peer institutions. The group chose to glean this information from the ample literature available. Although efforts were made to identify broad strategic aims, some of the recommendations reflect specific system requirements. We have made this data available in Appendix B.



## ANALYSIS OF LOCAL SYSTEMS

The Discoverability charge is explicit in its request for the compilation and analysis of statistics on current usage, in order to make the most informed decisions regarding future discovery tools and systems. Consequently, the group reviewed data for all major discovery systems at the University of Minnesota Libraries, in addition to the Libraries' primary website and subsites. A baseline report has been prepared for each system, including a general introduction to the system, relevant data, any significant findings, and recommendations for future data gathering or analysis. To supplement these discrete sections, we provide a more cohesive picture by pulling together one month of data from each system, telling a story about current usage: *A Month of Library Discovery*.

This section of the report is arranged as follows:

### A Month of Library Discovery: October 2008

Baseline reports for primary discovery environments:

- Affinity Strings data
- AgEcon Search
- Aleph OPAC (MNCAT Classic)
- Finding Aids (DLXS)
- Interlibrary Loan (ILLiad)
- MetaLib
- Primo (MNCAT Plus)
- SFX
- University Digital Conservancy
- Vendor Databases statistics
- Website statistics
  - o A summary of all the website data
  - o Relevant data from the Libraries' home page
  - o Reports on 29 of the Libraries' web subdomains can be found in Appendix C

Summary and full set of statistics recommendations

### *A Month of Library Discovery: October 2008*

During a typical month, the University of Minnesota Libraries receives millions of requests for digital information. The Discoverability Team believes that mining statistics from our local systems can – and should – inform decision-making because it provides a unique view of user behavior as it currently exists. October 2008 is the most recent month with a rich set of data and thus serves as our model month for this exercise. **Using SFX and Proxy URLs as our data source, we have found that 65% of information requests originate off-campus, with the other 35% originating on-campus primarily during class hours.** We had a sample size of 1.5

million requests, and this 65/35 ratio normalized rather quickly. The ubiquity of SFX and proxying as an intermediary service for users seeking full-text articles, regardless of where they discover the item, allows us to use these statistics to generalize about user behavior. Not surprisingly, it appears the Libraries' user base is active 24/7 and has a marked preference for working off-campus.

**Three main services dominate use of library-controlled resources:**

- 1) the Libraries' **catalogs** (both MNCAT Plus and Classic),
- 2) the **SFX** link resolver, and
- 3) a few of the Libraries' myriad **web pages**

Primo (MNCAT Plus) produced the largest number of transactions: 367,981 searches, with 307,173 (83%) of them against the local catalog and 60,808 (17%) using the Primo "Articles" tab to find and filter results from remote targets via the Primo interface to MetaLib. The Aleph OPAC (MNCAT Classic, Twin Cities) had 151,850 search and browse transactions for the month and received 259,282 requests via the Z39.50 protocol. **Since our catalog data is not directly exposed to web search engines, we would expect to see significant activity occurring within the native interfaces. Even so, there is evidence of demand by external systems, e.g.,** within the Aleph OPAC. 10% of the referrers are from systems outside of Primo or the Aleph server alias. Most of the external requests are generated by the Twin Cities instance of SFX and by various OCLC services. (Our data is exported to OCLC on a regular basis, and made available to services such as Google Book Search.) Furthermore, none of this data includes the impact of Aleph X services in the discovery process, which can be used to query the catalog outside the native interface. For example, the Duluth campus is using X services to generate lists of bookable resources and New Books feeds. *More data is available in the Aleph and Primo Baseline Reports.*

AgEcon Search also provides some striking statistics regarding external demand. **Slightly over half of all AgEcon traffic entered via a search engine:** 113,674 out of 213,359 page views. Additionally, since Google downloads and indexes PDF files, it is possible that users choose the *view as html* option in their Google search results and never visit the site; we do not have access to data to determine the percentage of searchers who choose this option. AgEcon users downloaded 274,975 PDF files during the month. *More data is available in the AgEcon Baseline Report.*

The majority of SFX requests originate from external sources. **Google, PubMed, and the large databases and indexes account for ~75% of all SFX requests.** However, what jumps out here is that 3 of the top 5 sources are local in origin: the A to Z Journal List, MNCat, and the SFX Citation Linker. The libraries still exert a fair amount of control over where full-text is initially discovered. Our vendor-supplied statistics show that users spend a significant amount of time searching inside vended portals and then linking back to us for fulfillment. Academic Search Premier handled 131,219 searches in October from U of MN users. EBSCO Academic Search Premier and Business Source Premier, Science Direct, Journals @ Ovid, and Wiley Interscience together accounted for 226,531 searches within their portals.

SFX handled 392,460 user requests overall, with 285,452 of these requests resulting in a 'click-through' to a service offering. Of the 392,460 SFX requests, 288,085 provided a full-text option. This almost matches the

actual click-through rate. Unfortunately, we do not have the data to make a correlation between full-text options being presented (or not) and when users click through the menu. Users showed a strong preference for using the SFX E-Journal A to Z List and the Citation Linker for locating full-text. The libraries also benefit from having proxy links for resources on our website. *More data is available in the SFX Baseline Report.*

In order to develop a collective view of use of external resources, we have aggregated statistics from many sources, some at the origin of discovery and others at fulfillment. Many of these statistics overlap significantly. Comparing this crossover yields some interesting results. As an example, we have proxy, SFX, and vendor statistics for two large EBSCO resources in Academic Search Premier and Business Source Premier. Below is a chart comparing the statistics at various points in the information retrieval process during October 2008.

Resource	Vendor Sessions	Vendor Searches	SFX Click-Throughs	Proxy Clicks	MetaLib Searches
EBSCO Academic Search Premier	85430	131219	10448	14664	43578
EBSCO Business Source Premier	51266	60603	1965	4675	38033
<b>Totals</b>	<b>136696</b>	<b>191822</b>	<b>12413</b>	<b>19339</b>	<b>81611</b>

**These numbers present EBSCO as a significant discovery service in its own right.** There were 19,339 users who went through the front door of these resources via our proxy links, yet the vendor shows 136,696 sessions. This discrepancy can be mostly accounted for by MetaLib searching and SFX article resolving. In both cases the proxy clicks, SFX clicks, and MetaLib searches do represent a large percentage of the Vendor reported sessions. With Academic Search Premier, our stats account for 68,690 of the reported 85,430 sessions (80%). While with Business Source Premier, our stats account for 44,673 of the reported 51,266 sessions (87%). The missing 10 to 13% probably represents on-campus users who go directly to EBSCO and are authenticated via IP address. Data mining our statistics from different use angles presents a more detailed picture of our users' discovery and fulfillment preferences.

Nowhere was users' reliance on external systems in accessing the Libraries more evident than in our web statistics. With a few anomalous exceptions, **our sites see more visitors from Google than from any other measurable source**, including other of the Libraries' sites and the University's sites. Some of this traffic may be accounted for by visitors using Google for navigation rather than true search, entering "UMN library" as a search term, for instance. However, those of our sites that contain information objects rather than pathfinders or information about our libraries or collections benefit greatly from search engine traffic, giving sites such as the Bell Library online traffic that is clearly disproportionate to their in-person visits.

**The majority of our web pages are visited infrequently.** In comparing our various subdomains, the majority get less than a tenth the traffic of the main web site. Within each subdomain, with very few exceptions, **users**

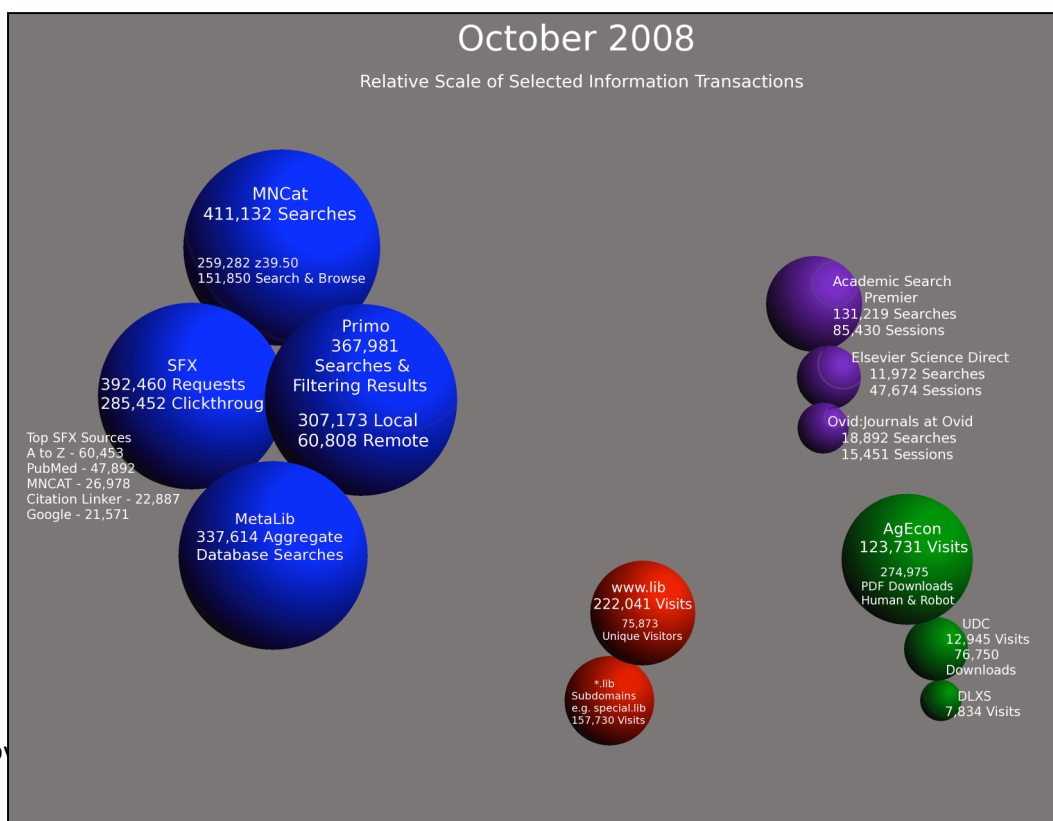
rarely visit any page but the home page. More information is available in the Website statistics baseline reports and subdomain reports.

One of the more interesting discoveries during data analysis was the prevalence of non-human traffic across our systems; we calculate that non-human traffic accounted for ~36% of all traffic on the AgEcon portal. Much of this usage can be attributed to the fact that Google now ingests and indexes PDF files. Robots and spiders activity shows strongly in many of the statistical reports and yet none of our local systems are optimized for this kind of traffic; **Search Engine Optimization (SEO) is an area that the group identifies as requiring attention from Libraries staff.** More data is available in the AgEcon and Website baseline reports and subdomain reports.

The Libraries home page at www.lib.umn.edu had 75,873 unique visitors during the month. The Bio-Med site had 33,318 unique visitors, special.lib.umn.edu had 28,194, bell.lib.umn.edu had 12,084, and no other site had more than 10,000. The majority of our sites had fewer than 4,000 unique visitors during the month and nearly a dozen had fewer than 1,000.

**Another interesting discovery was the emergence of several types of “long tails” and “short tails” on many of our information systems.** Nearly all of the library subdomains (e.g., special.lib.umn.edu) demonstrated a short tail or hockey stick pattern. Users visit the top level of these sites but do not navigate deeper with any consistency, if at all. AgEcon shows a long tail with regards to usage; robots dominate, and then use dwindles off into a very long tail.

Trying to provide a uniform analysis of these disparate systems is difficult as we’re rarely ever comparing apples to apples. Catalog searches are quite different from SFX menu requests or web page views or visits. Thus, when comparing resources it’s important to remember the user intent and the interaction type that the resource offers. The figure below offers a high-level comparison of many of our systems and resources.



*Note: MetaLib conducted 337,614 database searches during the month, most of them as part of a Quickset search. (i.e., a single user search often results in multiple database searches. We do not have the number of user searches.)*

This data visualization is not intended to be comprehensive, and it only covers systems and portals for which we have acceptable data. For purposes of comparison we treat the primary function of each system as an “information transaction” and present the total usage for the month.

### ***Statistics Recommendations***

The Discoverability group was charged with providing recommendations for ongoing analysis of system statistics. Our review of the available statistics yielded some specific recommendations, most of which are contained in the various baseline reports. We have chosen to highlight some recommendations below that arose repeatedly or which represent quick wins for improving our statistics gathering. We have, in our usage statistics, an incredibly rich set of data, but our ability to use them for decision-making is severely limited by several factors:

- Inconsistent data gathering practices and system configurations that prohibit us from making valid comparisons across systems
- Data gathering that is bound to internal organizational structures and specific systems rather than to user behaviors
- Lack of synthesis and analysis of the data we gather

Our overarching recommendation to the Libraries is that we implement systems and practices, including dedicated staff time, to ongoing and rigorous analysis of our system usage statistics across the Libraries, including, but not limited to, our catalogs, ILS, ILL, link resolver, proxy server, websites, and repositories.

### **Web Server Log Recommendations**

All of our local and vended systems run the Apache web server ([www.apache.org](http://www.apache.org)). Such homogeneity allows us to recommend configurations that would be standard across library systems.

- Enable the logging of referrers on all web-facing systems.
- Whenever possible, enable logging of third party discovery tools that we use, e.g. LibX. LibX checks library availability of resources when users are on sites like Amazon.com.

- Develop standard methods to accurately measure robot and script (non-human) use of our web facing resources. This includes people “gaming the system” on portals like AgEcon to artificially inflate their own download statistics. Jeff Silvis has conducted preliminary research on UDC and AgEcon.
- Standardize on a single web statistics package for all library web servers. Currently, AWStats is used on most non-vended systems. Web statistics for all systems should be continually updated and available to authorized staff at all times.
- Install the AWStats plug-in for tracking mobile user agents. This software currently recognizes over 300 phone types.

### **Resource Use Statistics Recommendations**

The libraries have gathered local digital resource usage statistics for several years. We check all users to determine if they’re on or off-campus and then proxy accordingly. Such tracking gives us a fairly complete picture of resource traffic. However, we do miss on-campus users who navigate directly to a resource site and are then IP authenticated. There is some room to improve our data mining and we have to track the emergence of new trends and services.

- Continue to locally track access of vendor databases. Currently a local script stores proxy URL access data that is searchable via the CDM wiki page.
- CDM should continue to harvest monthly SFX statistics.
- Continue to gather other data to track user preferences i.e. the affinity string project associated with the MyU Portal. Other examples include search statistics, SFX click-throughs, portal downloads, etc.
- Continue to acquire statistics for systems that we don’t own or control but use to facilitate access. These often are provided to CDM by the vendor.
- Investigate ways to data mine crossover statistics to reveal new patterns or previously unknown user preferences. An example of this would be to compare our local proxy statistics, the vendor’s statistics, and SFX link resolver statistics.
- Track usage of the new map discovery interface available via the Science and Engineering website.
- Track development of new OpenURL features like getCitedGenome and getCitedDataSet.

### **Discoverability and Non-Human Use Recommendations**

In addition to statistics gathering, it is vital that the libraries’ systems are adequately configured for search engine indexing. Since most discovery is happening remotely via search engines, we need to make sure our systems are optimized for non-human (spider, robot) use.

- Track development and usage of iTunes U as a new discovery environment that offers searchers a broad array of formats.
- Track usage of the ExLibris X-Server on our Aleph and Metalib systems. The X-Server provides access to these systems via alternative portals like the Undergraduate Virtual Library and the mobile library website currently in development.

- Evaluate and optimize library usage of robots.txt files across all of our systems.
- Use standard tools to further optimize our site for search engines. This would include such options as submitting site maps, using analytics, indexing optimization, etc. Examples of such tools are those provided by the Google Webmaster toolbox.
- Enable the logging of user-agents to track mobile use

### *Affinity Strings Baseline Report*

The Libraries have been capturing electronic resource usage data since Spring 2008. These data are aggregated according to user affinity string(s). Every time an X.500-authenticated user accesses a database or E-Journal, a use of that resource is recorded for each of the user’s affinity strings.

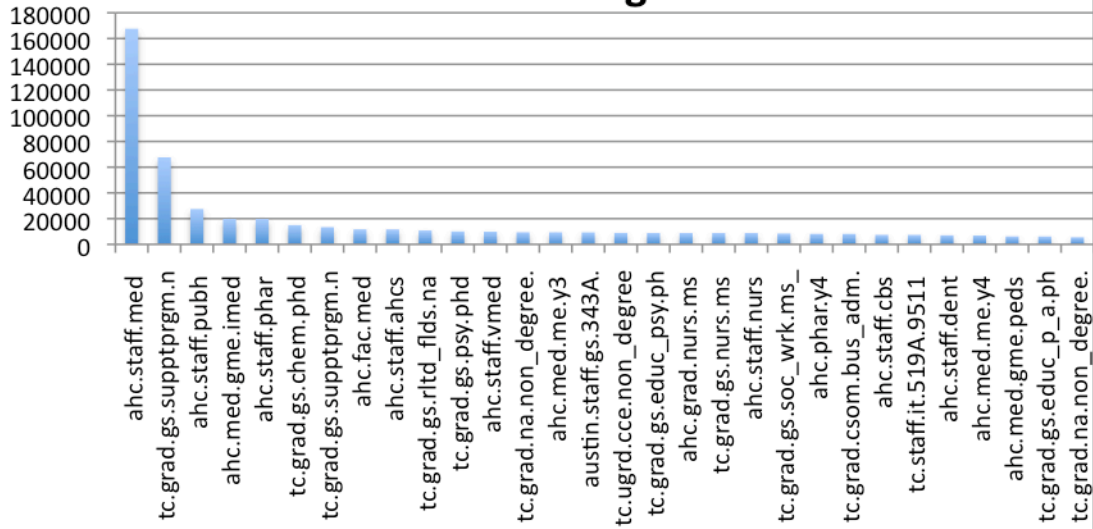
These statistics have great potential for providing insight into use (or non-use) of our resources by their intended audience. However, the data in its current form must be heavily qualified, and should not form the basis for any serious decision-making.

The current affinity string stats database (<https://www.lib.umn.edu/libstats/Statistics001.php>) does not report the population size of the affinity strings it records. Hence, the affinity strings with large populations skew the data. We suspect, for instance, that two of the most prominent affinity strings, “ahc.staff.med” and “tc.grad.gs.supptprgm.na” are assigned to many people, most of whom are also assigned at least one other string that would reflect their relationship with the University with greater fidelity. In effect then, the data from these strings represent an aggregation rather than a unique user population. If our suspicions are borne out, then the data on usage of individual E-Journals or Databases will also be skewed, because each use would be recorded more than once.

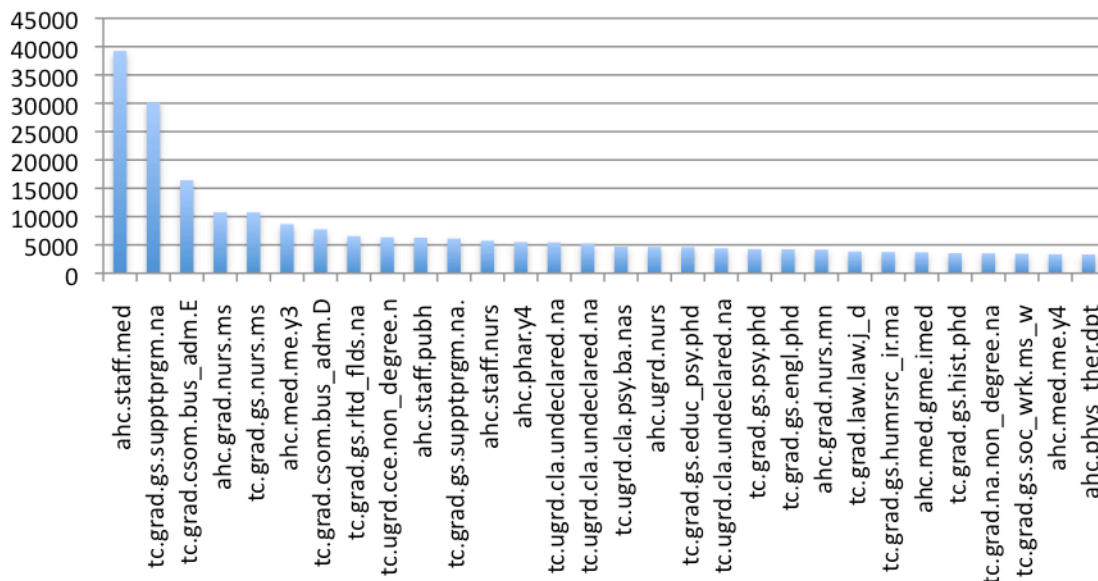
Presently e-resource usage is recorded in our database only when the user has an active X.500 cookie in their browser. This cookie will most often be obtained when the user authenticates to access the resource through our proxy server. As a result, some highly used resources, e.g., PubMed, will not be included in this data because they do not require authentication. Many of our resources provide IP address-based authentication for users on campus, which allows users to bypass our proxy servers. This means that the data we’ve captured reflects only those users who are not IP-authenticated (likely off-campus), or who have an active X.500 cookie from another application, such as the University’s web mail system. We cannot assume that the proportions of on-campus vs. off-campus usage of resources is uniform across all user populations, and as a result, we cannot take the data we collect this way as a representative sample of usage. (Hanson, 2008)

Below are graphs representing the top 30 affinity strings using our databases and e-journals, as well as the top 30 databases and e-journals by usage.

## E-Journal usage

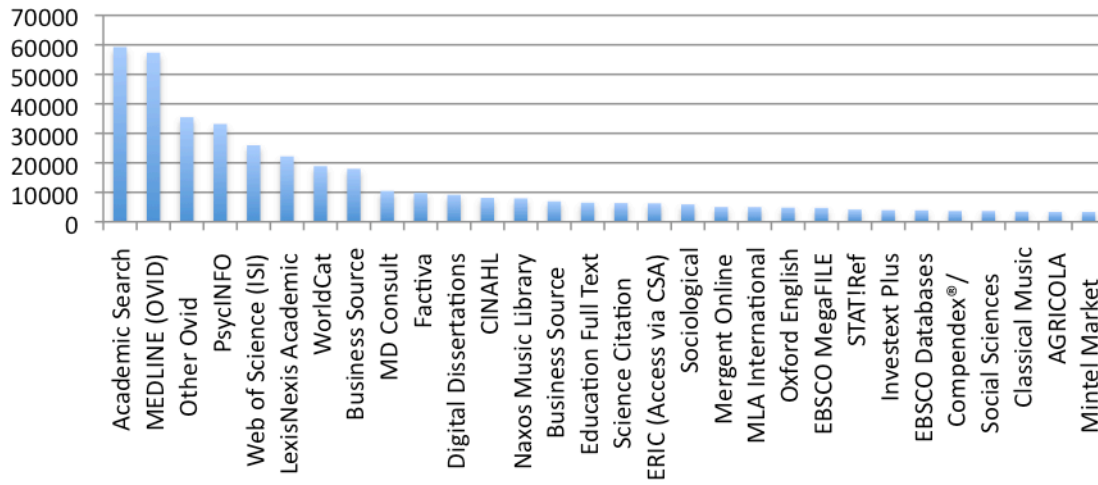


## Database usage

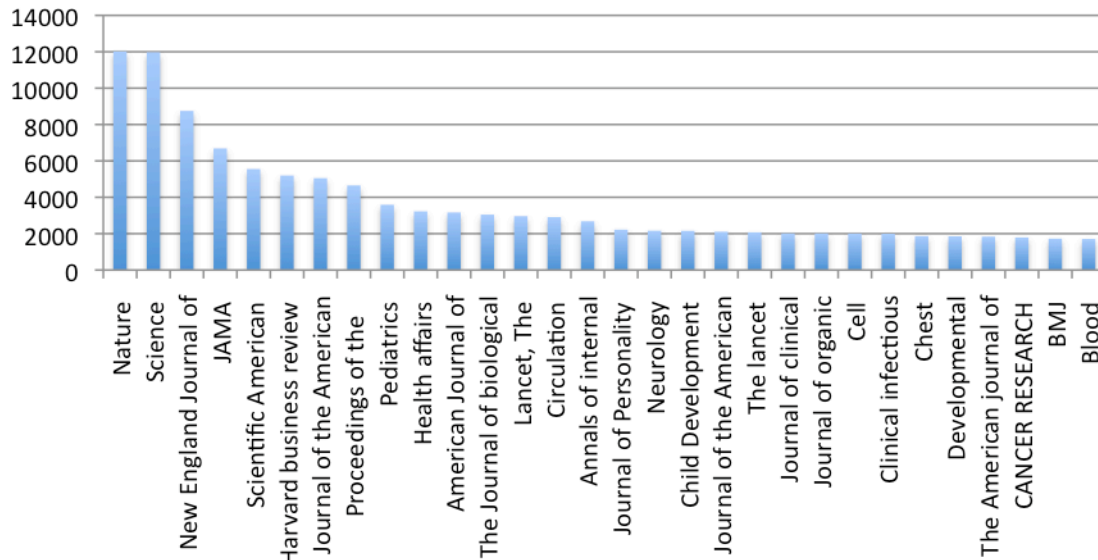




## Top Databases



## Top E-Journals



## *AgEcon Search Baseline Report*

AgEcon Search is an open access repository of full-text scholarly literature in agriculture and applied economics, including:

- Working papers
- Conference papers
- Journal articles

AgEcon Search is sponsored by the University's Department of Applied Economics and the University Libraries, as well as the Agricultural and Applied Economics Association. The repository is not limited to University of Minnesota authors.

In addition to discovering items in the AgEcon collection through the AgEcon Search interface, users find AgEcon items through Web searches (Google, etc.) or direct links from other Web sites, email, etc. The AgEcon collection is not exposed through the MNCAT Classic or MNCAT Plus (Primo) interfaces.

AgEcon Search is built on the DSpace platform, and statistics are gathered and presented using AWStats. AgEcon currently contains about 31295 items.

### **Findings Summary**

Apache stats are available in AWStats from 5/12/2008-present: [http://ageconsearch.umn.edu/stats\\_main.jsp](http://ageconsearch.umn.edu/stats_main.jsp)

The following statistics are for October 2008. Note that traffic from IP addresses suspected of "gaming" the statistics (repeatedly downloading files to increase download counts) are filtered out.

#### ***Bandwidth***

Because Google (and perhaps other search engines) download PDFs in order to provide full text indexing, a large percentage of the bandwidth consumed serves non-human users. It's worth exploring whether any changes can be made to the site's configuration that will make robot access more efficient.

<b>Human bandwidth</b>	72.87 GB
<b>Non-human bandwidth</b>	42.05 GB

#### ***Connect to Site From***

For any Web page, a user might get to the page through:

- **Direct address.** The user might type the address directly into the browser Address box, choose a saved bookmark, or click a link in an e-mail message or in a file that is not a domain-hosted Web page. AWStats places site entries made through one of these methods in one category, called Direct Address in this report, and cannot differentiate by the actual method used.
- **Search engine.** The user can search using keywords or phrases and click a link to an AgEcon page in the search results. AWStats provides a list of all phrases and individual words used during the month, along with the number of searches that used that keyword or phrase.
- **Referring page.** The user may arrive on the AgEcon Search site or download a specific item after clicking a link on another Web site. AWStats provides a list of sites that refer traffic to AgEcon Search, but doesn't provide details on which page(s) of AgEcon Search the user entered from the referring site.

In the October 2008 timeframe, more traffic came to AgEcon from search engines than from direct addresses, but not much more. It's likely that many users of AgEcon Search come to the main site frequently and either type in the address from memory or select a bookmark in their browser to open the page.

Starting Point	Page Views	Percent of Total
Referring Page	9511	4.4%
Search Engine	113674	52.3%
Direct Address	90174	42.2%

Although a significant number of page views came from another Web page (referring page), such views were not a significant percentage of the total.

#### ***Referring Page***

Although users came to AgEcon through many different Web sites (percents even at the top of the list are tiny), the top references are either directly to the AgEcon Search main page or are from pages that list many items in the AgEcon repository rather than a single item.

Referring Page	Description	Page Views
<a href="http://www.aaea.org/flash/aaea_menu.swf">http://www.aaea.org/flash/aaea_menu.swf</a>	Sidebar on the Agriculture and Applied Economics Association site. Lists AgEcon Search under "Search & Research" choice.	258
<a href="http://agecon.lib.umn.edu">http://agecon.lib.umn.edu</a>	Redirects to <a href="http://ageconsearch.lib.umn.edu">ageconsearch.lib.umn.edu</a>	229
<a href="http://www.apec.umn.edu/Staff_Papers_1990-1999.html">http://www.apec.umn.edu/Staff_Papers_1990-1999.html</a>	U of M Dept. of Applied Economics lists individual items in the AgEcon repository.	197
<a href="http://www.apec.umn.edu/Working_Papers_from_The_Center_for_Inter...">http://www.apec.umn.edu/Working_Papers_from_The_Center_for_Inter...</a>	Same as above	160
<a href="http://www.apec.umn.edu/SESWlist.html">http://www.apec.umn.edu/SESWlist.html</a>	Same as above	147

<a href="http://www.apec.umn.edu/Economic_Reports_1973_-_1983.html">http://www.apec.umn.edu/Economic_Reports_1973_-_1983.html</a>	Same as above	132
<a href="http://www.apec.umn.edu/Staff_Papers_2000_-_2008.html">http://www.apec.umn.edu/Staff_Papers_2000_-_2008.html</a>	Same as above	131
<a href="http://www.ers.usda.gov/Data/CostsAndReturns/">http://www.ers.usda.gov/Data/CostsAndReturns/</a>	USDA Commodity Costs and Returns Datasets page links to specific items in the AgEcon repository.	131
<a href="http://www.narea.org/ARER.HTML">http://www.narea.org/ARER.HTML</a>	Agriculture and Resource Economic Review publication. Links to the journal in AgEcon Search. (ARER is a community; each issue is a collection.)	125
<a href="http://www.saea.org/jaae/issues.htm">http://www.saea.org/jaae/issues.htm</a>	Agricultural and Applied Economics. Refers user to AgEcon Search to find full text of articles in PDF format.	102
Others		7899

### ***Search Engines***

Google is far and away the most popular general search engine for those finding items in the AgEcon repository.

<b>Search Engine</b>	<b>Searches ending in an AgEcon Search page</b>
Google	110491
Yahoo!	592
Unknown search engines	545
AOL	497
Windows Live	387
MSN Search	313
Google (cache)	274
Ask	197
Dogpile	64
AT&T search (powered by Google)	63
Others	251

### ***Phrases Resulting in Click-throughs by Searchers using Search Engines***

As with referring pages, percents for any particular search term or phrase are tiny. But it's worth noting that four versions of "AgEcon" are in the top six. Such searches are most likely the users' alternative to

bookmarking the site; they know where the item(s) they want are stored, or they want to see what's new on the site. They need to get to the AgEcon Search site and choose not to use a bookmark or type the address into the browser's address bar. It's apparent that the AgEcon name is something of a brand within its user community. That said, the majority of the searches are for terms that occur in a repository item's abstract or full text.

Phrase	Search	Percent
Agecon	778	0.7 %
agecon search	653	0.5 %
marine resource economics	232	0.2 %
the economics of thinking pdf	196	0.1 %
ag econ search	188	0.1 %
ageconsearch	153	0.1 %
sample case study on recreation design filetype pdf	113	0.1 %
apple production in tropics	112	0.1 %
ppt public economics	99	0 %
meaning agricultural technology	88	0 %
Other phrases	108262	97.6 %

### Entry pages

As we know from other statistics such as the search terms and referring pages, users often get directly to AgEcon resources through means other than the main search screen. Still, the main page is the first thing about 7.8% of users see.

Page	Description	Entries	Percent
/	Main page	8387	7.83
/feed/rss_2.0/site	RSS feed. Probably triggers an entry each time an RSS reader pings the site for updates.	837	0.78
/simple-search	Search result page	700	0.65
/items-by-subject	Result of a Subject browse. The URL for a particular Subject browse is easy to construct, so users may be bookmarking the page for their subject and entering that page directly rather than the main page.	667	0.62
/bitstream/31045/1/2602387.pdf	JARE article on genetically modified food	201	0.19
/bitstream/43678/2/revista_v6_n1_jan-jun_2004_9.pdf	Spanish language article on a personnel department case study	185	0.17

**Note:** Views of *styles.css.jsp* have been removed from this summary.

### ***“Recommended Item” Page Views***

Although AgEcon Search does not offer customized recommendations for users, it does include PDF files that summarize which papers have been most popular. Two of these, the Top Thirty downloads from 1/1/01 to 4/15/08 and the Top 35 for 2007, are among the 20 most viewed pages on the site. Although the PDF files list persistent URLs for the papers, they are not hyperlinked. Therefore, there is no way to know whether the lists actually drive traffic to the articles listed.

### **Further Explorations**

1. Is there any way to know what the effect of adding the AgEcon “silo” to our current primary discovery tool (Primo) would have?
2. Is it worth exploring the idea of adding links to relevant AgEcon work to appropriate Wikipedia pages?
3. Can we capture click-throughs from RSS feeds or the Recent Submissions list on the main page? Doing so would provide evidence regarding the value of “recommending” new items.

### ***Aleph – OPAC Baseline Report***

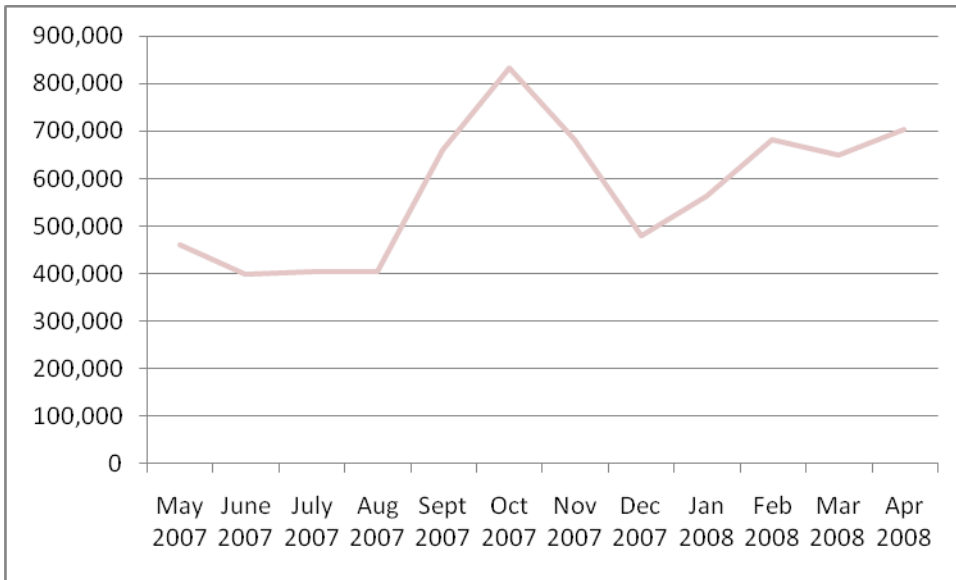
The Aleph OPAC was the primary discovery interface for the University of Minnesota Libraries’ catalog from the ILS migration in 2002 through July 2008. The features available within the Aleph OPAC grew extensively over this period of time to include both browse and keyword searching, a large number of pre-limiting options, as well as other basic functionality found in traditional library discovery environments (e.g., My Basket, email results, etc.). In August 2008, the University of Minnesota Libraries switched the default interface to the catalog from Aleph to Primo, although accessibility to the OPAC has been maintained via a separate link on the Libraries’ home page.

**Note:** Interactions between Primo and Aleph are tightly interwoven: 1) through the use of PDS as the means of coordinating authentication, and 2) in that Primo relies on Aleph for availability data. Using a “search” as a basic unit of measurement enables us to make better comparisons between these two systems, and allows for a more accurate interpretation of the data.

### **Findings**

#### ***Annual OPAC Activity***

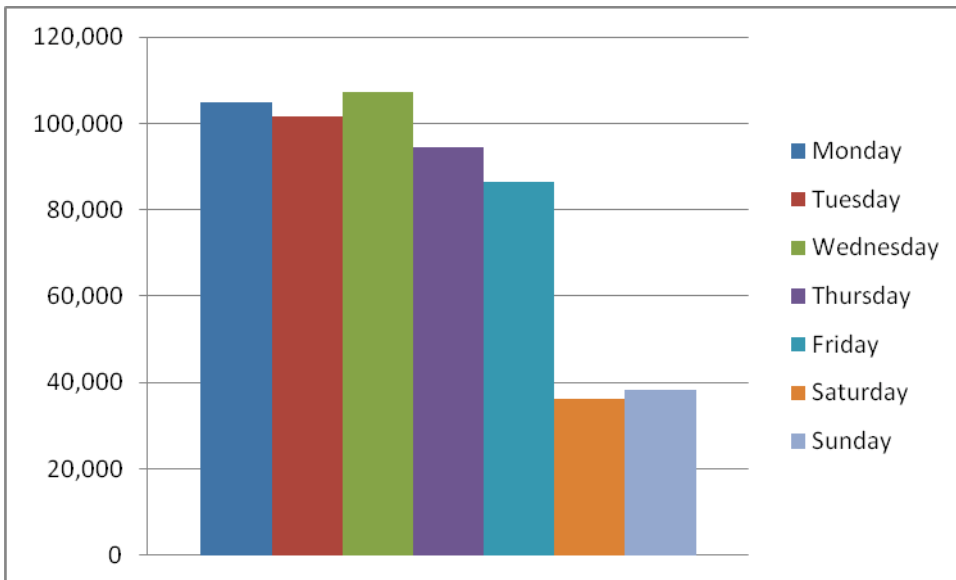
Annual usage of the Library OPAC follows a predictable trend based on a standard academic year, peaking in October. The data represented in the chart below counts Aleph OPAC “events”, which are categorized activities such as searching, browsing, or accessing help, etc. Note that this data represents a period when the Aleph OPAC was the default interface to the UMN catalog. It includes nearly all web-generated activities against the catalog, including coordinate campuses, course reserves, etc. Z39.50 activity, however, could not be tracked during this time period and is not included in these numbers. X Services transactions are also not included.



**Note:** The raw numbers above cannot be compared directly with numbers in any of the following charts due to a system problem that overcounted scan activities for the Firefox browser. A bug fix was implemented in May 2008.

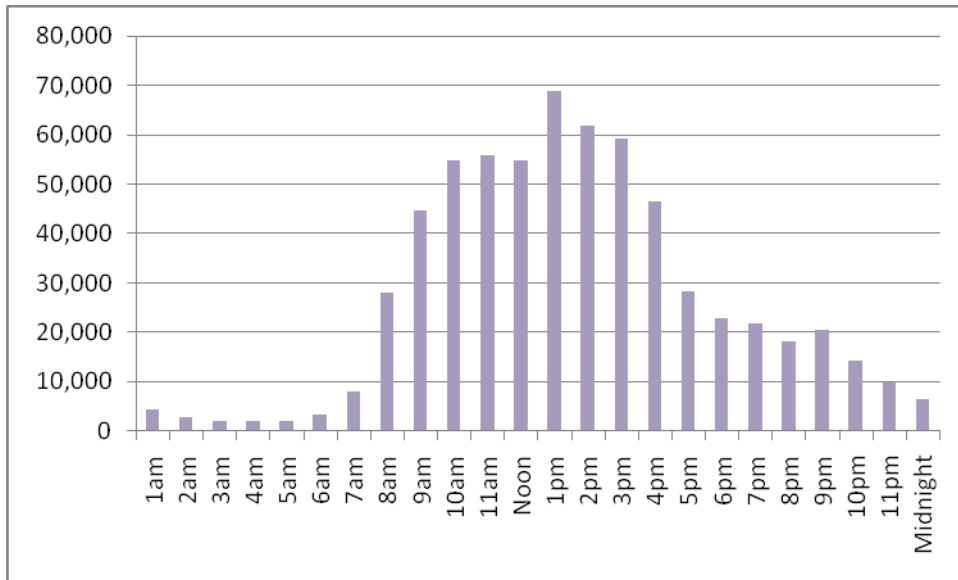
**Weekly OPAC Activity: October 2008**

Usage of the web OPAC follows a Monday-Friday trend, with significantly reduced use over the weekend. The graph below charts data from four weeks in October 2008. This chart includes all event types, including incoming Z39.50 searches. It does not include X Services activity.



**Daily OPAC Activity: October 2008**

The majority of OPAC usage occurs during regular business hours, beginning around 8am and gradually trickling off throughout the evening. The chart below includes data for all OPAC events occurring during October 2008, including incoming Z39.50 searches. It does not include X Services activity.



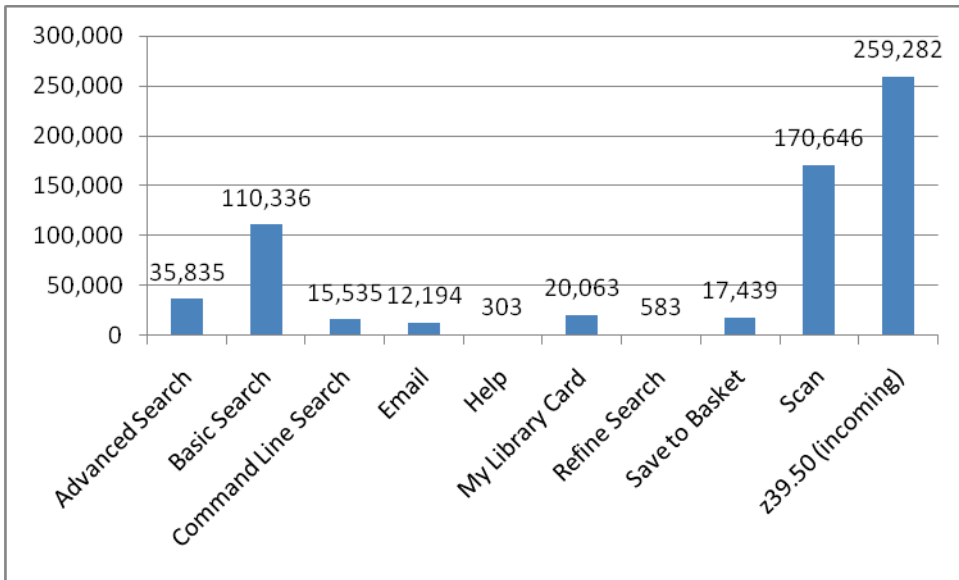
**Breakdown of Event Types: October 2008**

OPAC usage is categorized by different “event” types, shown in the chart below. The numbers here reflect a breakdown of activity during October 2008 (excluding X Services). There are at least two aspects of this chart that are noteworthy: 1) Volume of activity, and 2) Large number of Z39.50 searches.

As a non-default interface to the catalog, the OPAC still gets a tremendous amount of use. In fact, comparing raw numbers of OPAC searches, browses, and incoming Z39.50 transactions with the number of Primo interface searches during the same period, shows nearly equal use. There are a few reasonable explanations for this continued reliance on the OPAC. It is the interface most familiar to longstanding patrons. Staff members have built mechanisms for OPAC data retrieval, such as preconstructed URLs triggering a search or display of a specific title. Furthermore, some external tools and interfaces still direct people to the OPAC , e.g., WorldCat.

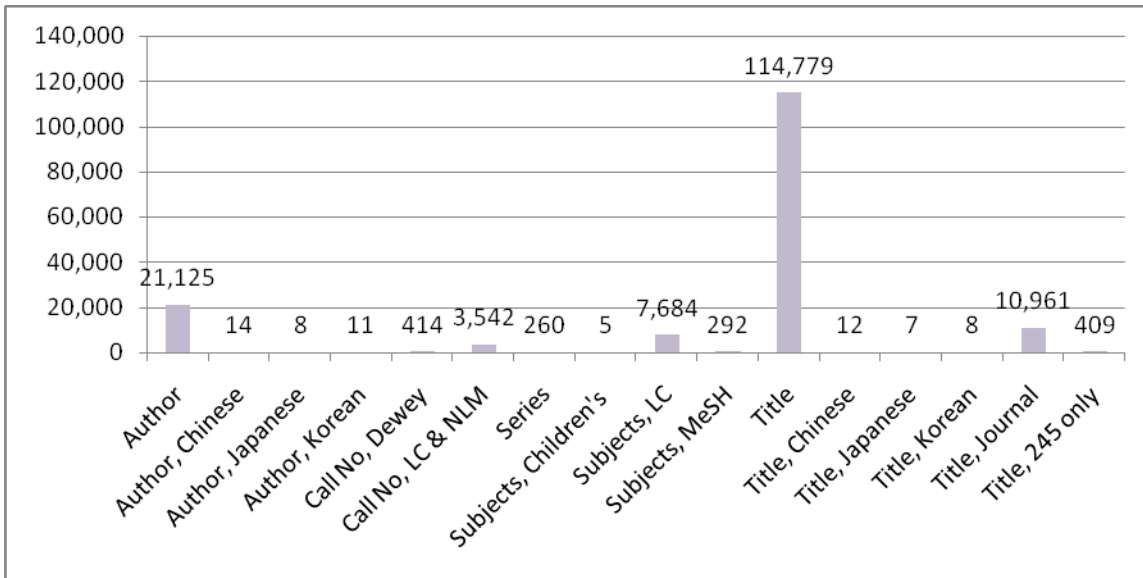
The large number of incoming Z39.50 searches are due, at least in part, to the fact that MNCAT is set as a Z39.50 target in the MnLINK gateway. As a federated search interface to many of the libraries within Minnesota, this gateway has wide-ranging use and satisfies diverse needs. So the number of Z39.50 transactions includes many users that would not be considered part of the University Libraries’ primary audience. Of the 259,282 searches, 106,042 of them were produced as a result of our participation in the MnLINK gateway. (Note: The gateway polls MNCAT multiple times during one user search, so this data point does not reflect the number of times a user has clicked on the “Search” button.)





#### **Preferred Browse Indexes: October 2008**

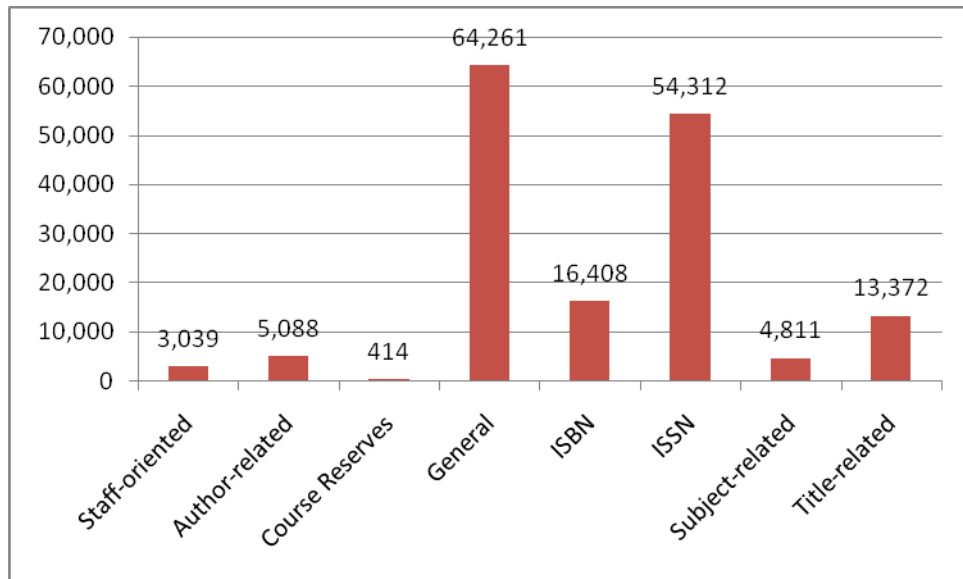
Of the browsing options available for use in the Aleph OPAC, the Title index is set as the default option and is the most popular choice by a significant margin. The graph and numbers below include data from the Course Reserves system, as well as coordinate campuses. They do not include any Z39.50 or X Services activity.



#### **Preferred Search Indexes: October 2008**

The General and ISxN search options account for well over 80% of the searches conducted against the catalog. It is suspected that a number of the ISxN searches originate elsewhere and are directed to the catalog via SFX, Worldcat, and similar services. Referring sites are discussed in the section "Sites driving traffic to the Aleph OPAC."

The term “Staff-oriented” has been used here to signify a group of indexes that would be of particular use to staff members. It includes many different keyword indexes accessed via the Command Search interface, e.g., System No, Sub Library/Collection. Use of any of the other mentioned indexes – including General – could have been via the Command Search interface as well. The graph and numbers below include data from the Course Reserves system, as well as coordinate campuses. They do not include any Z39.50 or X Services activity.



#### ***Searches Resulting in Zero Hits: October 2008***

Not including Z39.50 transactions, 44,327 searches resulted in 0 hits in October 2008. The breakdown of these transactions by search tool is as follows:

<b>Search Tool</b>	<b>Count of Zero Results</b>
Advanced Search	9,397
Command Search	4,457
Basic Search	30,473

The 30,473 Basic Search queries break down further by index. The following categories of indexes represent the majority of the Basic Search failures.

<b>Type of Index</b>	<b>Count</b>
Author-related index	1,092
ISxN index	10,679
Subject-related index	1,467
Title-related index	4,951
General Keywords	12,187
	Of this number, 740 were clearly ISxN searches

**Note:** There is some evidence that people thought they could access general websites, e.g., Google, Yahoo, Hotmail, through the catalog search box.

***Preference for Pre-limiting Options: October 2008***

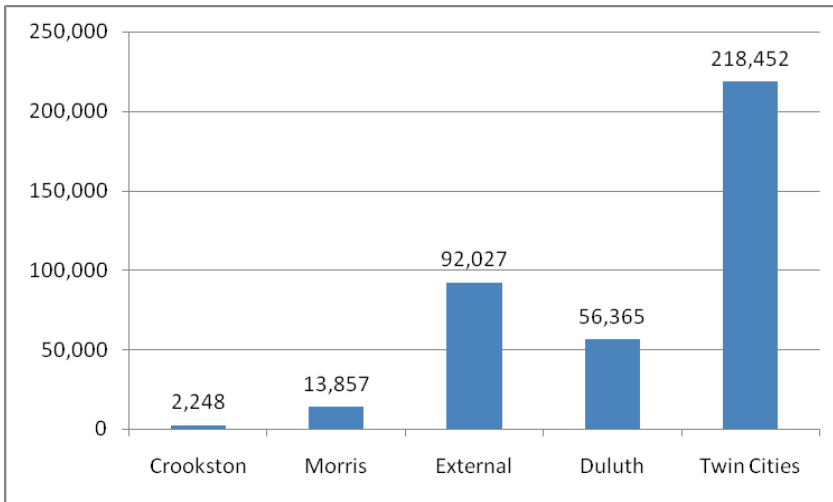
A breakdown of the preferred options for pre-limiting searches appears below. These numbers apply to all kinds of native-interface searches, including command search (i.e., excluding Z39.50 and X Services).

<b>Limit</b>	<b>Times used in October 2008</b>
by Format	5,234
by Location (sub library / collection)	2,751
by Language	1,223
by Publication Year(s)	535

***OPAC Activity Based on IP Addresses: October 2008***

When applicable, the Aleph system tracks the IP address of the client conducting an OPAC transaction. A rough breakdown of the IP addresses using the Aleph OPAC is shown below. These numbers do not include Z39.50 transactions, primarily because all Z39.50 transactions have a UMN IP address associated with them, and thus do not accurately reflect the searcher.

<b>Source</b>	<b>Number of OPAC Transactions</b>	<b>Percentage</b>
Crookston	2,248	0.59%
Morris	13,857	3.62%
Duluth	56,365	14.72%
External	92,027	24.03%
Twin Cities	218,452	57.04%



Within the External group, the seven IP addresses most frequently conducting native-interface transactions during October 2008 were as follows:

1. Comcast
2. Qwest
3. Hopkins School District
4. Anoka School District
5. Macalester College
6. Fairview
7. USDA

We would expect to see Internet Service Providers, such as Comcast and Qwest, appearing at the top of this list. It is unclear why some of the other organizations are generating notable traffic to MNCAT.

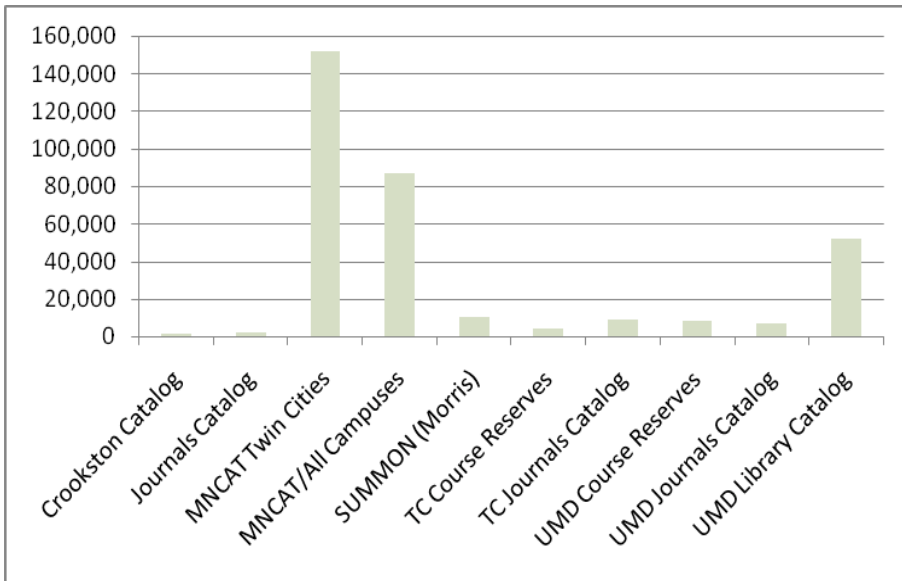
**Activity by Aleph Base: October 2008**

99.5% of the search and scan tasks for the month of October 2008 were conducted against the following Aleph “bases.” Aleph bases are used to create subsets of data for particular audiences or tasks. For example, the catalog for each coordinate campus is configured as a separate base. Z39.50 and X Services transactions are not included in these numbers.

**Search and Scan Transactions: October 2008**

Crookston Catalog	1,591	0.5%
Journals Catalog	1,913	0.6%
MNCAT Twin Cities	151,850	45.6%
MNCAT/All Campuses	86,935	26.1%
SUMMON (Morris)	10,097	3.0%
TC Course Reserves	3,944	1.2%
TC Journals Catalog	8,645	2.6%
UMD Course Reserves	8,018	2.4%

UMD Journals Catalog	6,539	2.0%
UMD Library Catalog	51,769	15.6%
<b>Total</b>		<b>99.5%</b>



**Sites driving traffic to the Aleph OPAC**

The Aleph OPAC tracks usage in a verbose log format that includes the referring URL, when applicable. While these logs are not always clean and uniform in format, they are the best method for determining where contact is made with the OPAC from external sources, given that referrers are not currently configured to appear in the standard apache logs. A brief analysis of the referrers for the month of October 2008 yields some information on links into the system.

This analysis differs from “OPAC Activity Based on IP Addresses” in that it includes usage outside Aleph’s particular categories of events. Examples of activities that would be uniquely logged here include accessing bibliographic information via direct linking to a particular resource (rather than using the native Aleph search interface), and simply bringing up the Aleph OPAC main page prior to conducting a search. Conversely, referrers do not represent a complete picture of usage as they only apply when a user clicks on a link to access the Aleph OPAC.

As expected, the two top referrers are an alias for the Aleph OPAC and UMN’s instance of Primo, particularly given the tight interplay between these two systems. Of the remaining referrers, the breakdown for 95% of activity in October 2008 is as follows:

Referrer	Percent of activity where referrers are logged, October 2008
Duluth Library website	21.0%
HSL Library website	1.9%

Law Library website	1.3%
Morris Library website	3.7%
OCLC, in its various manifestations	26.3%
SFX	26.8% (of which TC SFX represents 87%)
Twin Cities Library website	15.8%

### ***Statistics on use of X Services***

We are not currently tracking the use of X Services to request data from the Aleph system. However, we do have some information about the use of X Services in regards to MetaLib, the Ex Libris federated search system. Two local MetaLib interfaces, OneSearch and the Undergraduate Virtual Library (UGVL), include the local Aleph catalog as a target in its General quickset, which is set as the default quickset in those environments. Although MNCAT is not included in the default General quickset used in the Primo “Articles” tab, it is available further down in the dropdown list, so there is some potential use coming from Primo as well. The overall number of MetaLib quickset and category searches for October 2008 is 182,114; 13,602 of them included MNCAT as a target.

### **Future Explorations**

1. Continue to monitor activity to compare with adoption of Primo as default interface.
2. Attempt to determine the breakdown of staff versus patron in regards to continued use of OPAC. For what types of activities is the OPAC still preferred?
3. Look more closely at Z39.50 traffic. Where is this traffic originating?
4. X Services transactions are not tracked. This may be an area worth exploring further.

### ***DLXS – Finding Aid Search Baseline Report***

DLXS is the system used by the University Libraries to make our Archives and Special Collections holdings searchable online. DLXS, developed by the University of Michigan Digital Library eXtension Service, indexes Encoded Archival Description (EAD)-formatted finding aids and presents them as HTML web pages with a custom search box. Our statistics for DLXS consist of AWStats data on these web pages. However, the server is not configured to track referring sites or common search phrases, metrics which we have found useful in analyzing use of other systems.

### **Key Findings**

#### ***Usage Statistics***

During October 2008, DLXS received 2,002 unique visits. During the first 11 months of 2008, DLXS averaged 1,515 unique visits per month.

### ***Sticky Usage***

Usage of DLXS appeared to be “sticky”. During October 2008 users averaged 3.91 visits and viewed 11.9 pages per visit. By comparison, the main library website averaged 1.37 visits and 3.9 pages viewed per user. www.lib visits averaged 560 seconds, whereas DLXS averaged 775 seconds. 25.6% of DLXS visits lasted longer than 15 minutes.

### ***Low Non-Human Traffic***

Compared to what was discovered with UDC and AgEcon, the robot traffic on discover.lib.umn.edu was quite low. DLXS viewed traffic in October, 2008 was 3.39 GB, not viewed traffic (robots, spiders, worms, etc.) was 2.16 MB (~1,500 to 1). By comparison, www.lib had viewed traffic of 68.73 GB and unviewed of 11.56 GB (5.95 to 1).

### ***Browsing is Popular***

The most popular page was *cgi/f/findaid/findaid-idx*, which is the A-Z browse page for DLXS. It was accessed 64,309 out of 93,234 total page views (69%).

Unfortunately, we don't have referrer statistics for DLXS.

### **Future Explorations**

1. Why is the robot traffic so low? Apache logs do not accurately cover OAI harvesting, so further analysis of OAI stats may shed some light on usage patterns.
2. Configuring DLXS to capture referrer statistics would be helpful.

## ***Interlibrary Loan ILLiad Baseline Report***

ILLiad has been the system used for Interlibrary Loan borrowing since November 2004. ILLiad maintains a record of all requests and this data can be exported for report creation. Almost all information included in a request is available for reporting. The most notable exception is material format, which has been requested to OCLC but at this time is not yet reported. ILLiad statistics contain information regarding the item requested, the actions taken by ILL staff, as well as basic demographic information about the requestor.

For purposes of discoverability, this analysis is primarily focused on cancelled interlibrary loan requests. In most cases, the request is cancelled due to the library's owning the requesting material in either print or electronic formats. This implies failed discoverability by the user; had the user known the material was available on campus, it is unlikely an ILL request would have been created. By looking at who is failing to discover and where these users are failing, it may be possible to identify areas to focus on improving user's awareness of library resources.

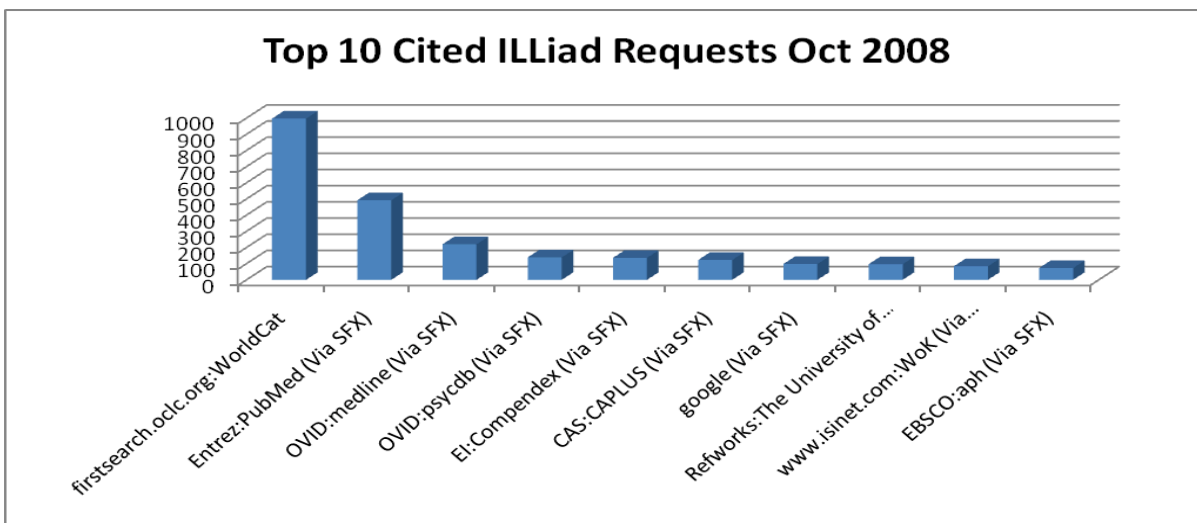
Cancellation statistics are based on the Interlibrary Loan staff manually cancelling a request in most instances – although users can also choose to cancel. Interlibrary Loan staff cancel for a variety of reasons: the university

owns the material, the citation cannot be recognized, no libraries are willing to lend the material, the item falls out of academic scope, etc. This information is stored with the request and assist in identifying if a request was cancelled due to failed discoverability or if it was for other reasons. Cancelling is a manual process by ILL staff and does introduce the element of human error; some requests are processed that should have been cancelled and vice versa. However, given the expertise of our ILL staff, the cancelation statistics should be considered a very through but not 100% complete snapshot of discoverability.

**Key Findings**

***Cited In Statistics***

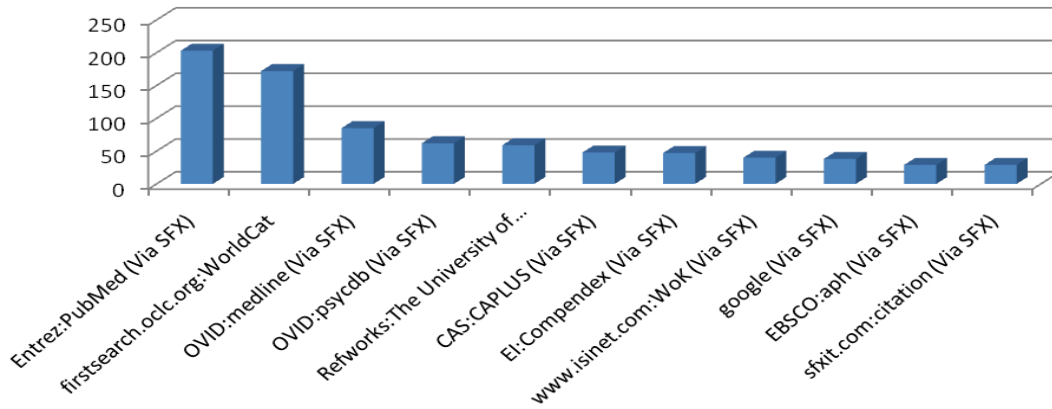
If an interlibrary loan request is created from a database and/or SFX, Illiad captures this information as ‘Cited In’. Below is a list of the top 10 referrals for all interlibrary loan requests in October 2008.



Cancelled requests show that requests coming from WorldCat are much more likely to be successful requests. As shown below, only 13.4% of the cancellations come from WorldCat despite being 25% of all requests. This is likely due to the fact that WorldCat shows local holdings from Minnesota. Additionally, the link to create an ILL request is in very close proximity to the holdings information on the web page. Other than the WorldCat exception, the other databases tend to show a close proportional relationship between total number of requests and number of cancellations.



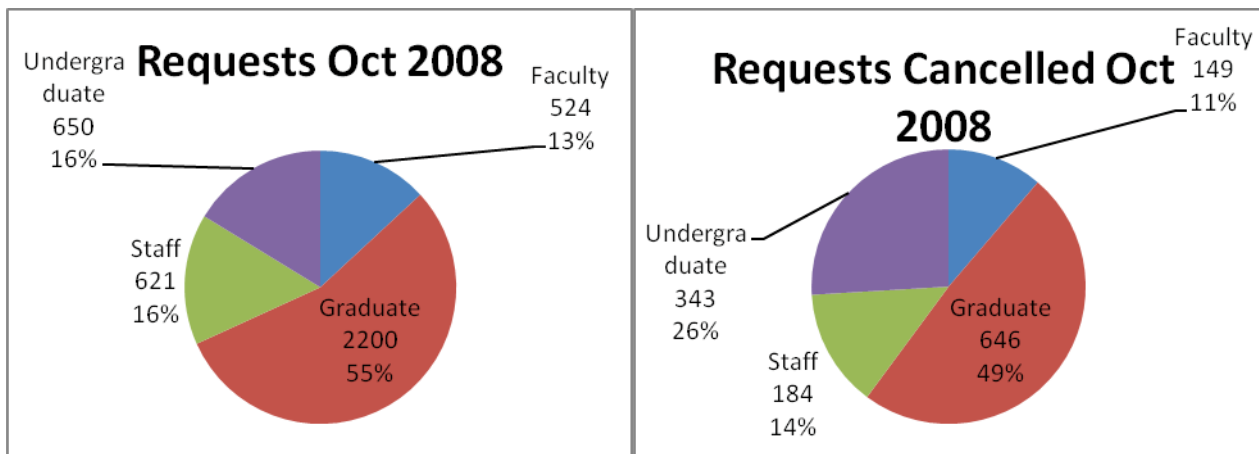
### Top 10 Cited ILLiad Cancellations Oct 2008



The October trends hold over the full scope of ILLiad data. Requests from WorldCat tend to be much more successful than others. The other databases have cancellations statistics that match their overall totals. Overall, WorldCat is the source of 30% of all requests that come from a database but just 20.6% of the cancellations.

### User Statistics

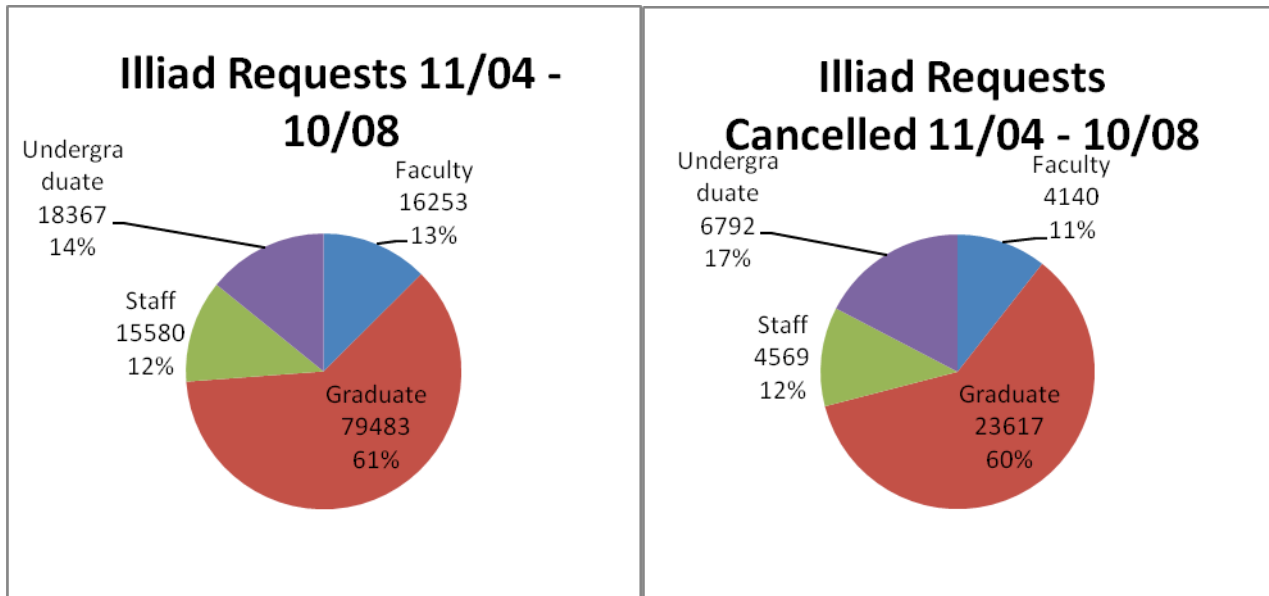
When users create an ILLiad account, they identify themselves as undergraduate, graduate, faculty, or staff. They also select their department in the university from a pre-populated list of options. This status is then recorded with every request the user makes. There is not a verification mechanism to confirm that the user selected their “official” status at the university. Additionally, since status at the university can change for individuals, it is quite likely that the undergraduate and graduate samples are slightly inflated as some of these individuals have moved into different roles at the university. Below is a breakdown by status of all requests created in October 2008:



Looking at the requests that were cancelled shows that a disproportionate number of the cancellations come from undergraduates. Only 16% of October 2008 requests come from undergraduates but 26% of

cancellations come from them. The ratios for graduate students, faculty, and staff are relatively consistent between number of total requests and cancellations.

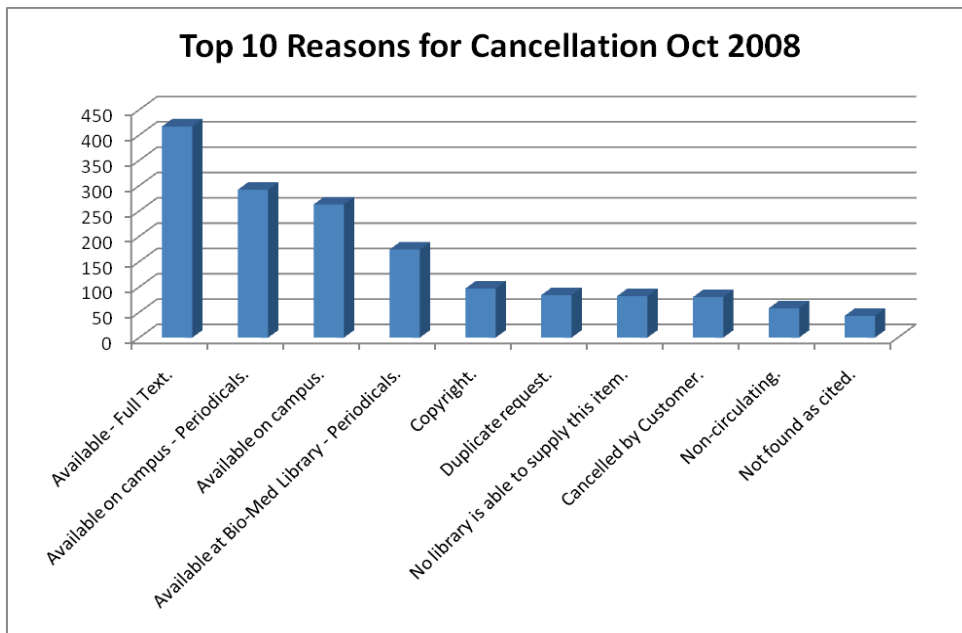
While the October 2008 statistics indicate that undergraduates are much more likely to create an unsuccessful request, the entirety of ILLiad statistics show that while undergraduates are more likely to create an unsuccessful request, it is not at the October levels. Overall undergraduates create 14% of the ILLiad requests and comprise 17% of the cancellations.



### ***Reasons for Cancellation***

When Interlibrary Loan staff cancel a request, they indicate the reason for why the request was cancelled. Below are the most popular reasons for cancellations in October 2008. This includes all user types. When a request is cancelled, 65% of the time the reason is because the resource is already available at the Twin Cities libraries either in print or electronic form. Of the total requests for the month, over 31% were cancelled.

In the chart below, items available in the Twin Cities are broken down between four different types (listed in order of frequency): Available Electronically in Full Text, Available on Campus – Periodicals, Available on Campus, and Available on Campus – Bio-Med periodicals.

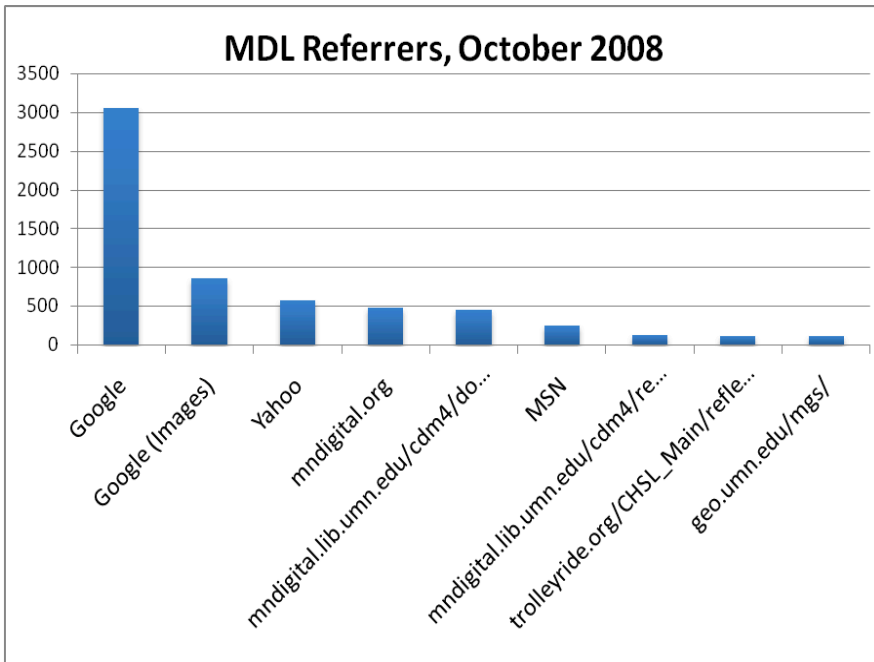


### *Minnesota Digital Library Baseline Report*

The Minnesota Digital Library is a consortial partnership presenting digitized artifacts related to Minnesota history online. The University Libraries host the Minnesota Digital Library website, and we track usage using the same AWStats system that we use for our other local web sites. We have statistics available from August 2005 to the present, with the exception of the months of September and October 2006, for which data is missing from AWStats.

#### **Referrers**

Statistics for visitor referrals to the Minnesota Digital Library site indicate that when there is a referring site recorded, that site is virtually guaranteed to be a search engine. The chart below shows the top ten referrers to the site in October 2008. During that month search engines accounted for four of the top ten referrers, and the vast majority of referrals. Several of the referrers recorded during this month are also on the mndigital.org domain, and thus do not represent true referrals, but traffic moving within the Minnesota Digital Library site.



### **Entry Pages**

The dominance of traffic driven by search engines is reinforced by the list of Entry Pages in AWStats, indicating the page where users begin a session on the site. The data below show that during October 2008, the MDL home page ranked only third among entry pages, meaning that most visitors arrived in the middle of the site, likely from a search engine result.

#### ***Entry pages***

/cdm4/results.php	4720
/cdm4/browse.php	3562
/cdm4/item_viewer.php	2893
/	<b>2087</b>

### **Future exploration**

Despite the dominance of search engine traffic on the Minnesota Digital Library site, our statistics hold one piece of damning evidence that indicates that we are not optimizing the site properly for search engines. AWStats records those URLs for which visitors receive “404” errors, indicating a page or file a visitor attempted to access, but which does not exist. During October 2008, the most common 404 error was for the site’s robots.txt file, which is a file that contains indexing instructions for search engine spiders. These spiders looked for the non-existent MDL robots.txt file over 17,000 times during the month. Were we to create that file and look at other search engine optimization measures for this and other sites, we could expect more and better traffic from search engine users.

## ***MetaLib – Metasearch Baseline Report***

MetaLib is a licensed software product developed by the ExLibris Group - <http://www.exlibrisgroup.com/>. The University of Minnesota Libraries supports MetaLib for all four University of Minnesota campuses, as well as for the bridge consortium of Carleton College and St. Olaf. MetaLib allows for the simultaneous search of multiple sources with ranked results returned to a single interface.

### **Usage Overview**

Between January 1, 2008 and October 31, 2008, MetaLib had 125,930 unique sessions. This averages about 414 sessions/day, and during peak usage MetaLib can experience over 1,000 sessions per day. MetaLib is embedded in the Primo Article tab and in the main search box of the UGVL, but we have yet to determine what percentage of MetaLib use is generated via these tools. The only other highly visible link to MetaLib is via the Libraries' home page, where MetaLib is branded as 'Libraries OneSearch'. The Twin Cities campus has 93 resources configured for search via MetaLib. While we have made MetaLib available in multiple locations, the University of Minnesota has not strongly promoted the product.

### **Findings**

#### ***Top Search Resources (>5,000 searches) – January 1, 2008 – October 31, 2008***

<u>Resource</u>	<u>Searches</u>
Academic Search Premier (EBSCO)	219106
PsycINFO (Ovid)	205088
Education Full Text (Wilson)	185310
Business Source Premier(EBSCO)	177906
AGRICOLA (Ovid)	173479
MLA International Bibliography	171263
Applied Science Abstr (OCLC)	146902
MNCAT	131332
Web of Science (ISI)	24245
PubMed	22492
BIOSIS Previews (Ovid)	18849
Linguistics + Language (CSA)	18096
ERIC (CSA)	14511
Physical Education Index (CSA)	14167
CAB Abstracts (Ovid)	13394
Inspec (Ei Village 2)	12231
PAIS International (CSA)	9837
Worldwide Political Science Ab	9800
Zoological Record Plus (CSA)	9421
ASFA Aquatic Sciences (CSA)	9408
GEOBASE (OCLC)	9219
MEDLINE (Ovid)	9013
Sociological Abstracts (CSA)	8909

ABSEES (EBSCO)	7286
CINAHL (Ovid)	6942
SPORTDiscus (EBSCO)	6862
Periodicals Contents Index (Pr	6676
Proquest Newsstand	6121

The large database vendors dominate MetaLib searches. Many of the large indexes are a part of multiple Quick Sets.

***Top Quick Sets: January 1, 2008 – October 31, 2008***

<u>QuickSet Name</u>	<u>Number of Searches</u>
General Resources	116383
Education/Psyc/Sport	12938
Health + Medicine	8312
Soc Sci/Anthro/Area	6196
Biological Sciences	5706
Business/Economics	4143
Humanities/History	3497
Language/Lit/Comm	3311
Engineering + CSci	2857
General Indexes	2697
Ag/Food/Nutrition	2582
Arts/Architec/Design	2397
Law/Govt/Poli Sci	2325
Environ/Natural Res	2271
Physical Sci/Math	1686
Ethnic/Gen Studies	1577
Newspapers	1537
MNCAT	958

The General Resources QuickSet is used in Primo and the UGVL and thus dominates searches. A variety of Quick Sets are used quite frequently.

***Links to Native Interfaces***

<u>Database Name</u>	<u>Number of Click-throughs</u>
Total	19114
Academic Search Premier (EBSCO)	4671
Education Full Text (Wilson)	3428
PsycINFO (Ovid)	1906
Business Source Premier(EBSCO)	1734
MNCAT	1684
MLA International Bibliography (PQ)	704
PubMed	445
AGRICOLA (Ovid)	391

New York Times (1851 - 2003) (ProQuest)	298
Web of Science (ISI)	289
Art Full Text (Wilson)	252
ERIC (CSA)	244
Periodicals Contents Index (ProQuest)	221
EconLit (CSA)	213
CAB Abstracts (Ovid)	174
Applied Science Abstr (OCLC)	158
Proquest Newsstand	150
MEDLINE (Ovid)	143
Sociological Abstracts (CSA)	135
PAIS International (CSA)	112

MetaLib provides links to the native search interfaces of the databases that it queries. Users often launch from MetaLib into these resources.

### **Future Explorations**

1. Users are not using the built-in customization features of MetaLib. Are there any widgets we could build that would make this easier?
2. If vendors start to allow meta-vendors to index content, is there any point to providing metasearching locally?

### ***Primo Baseline Report***

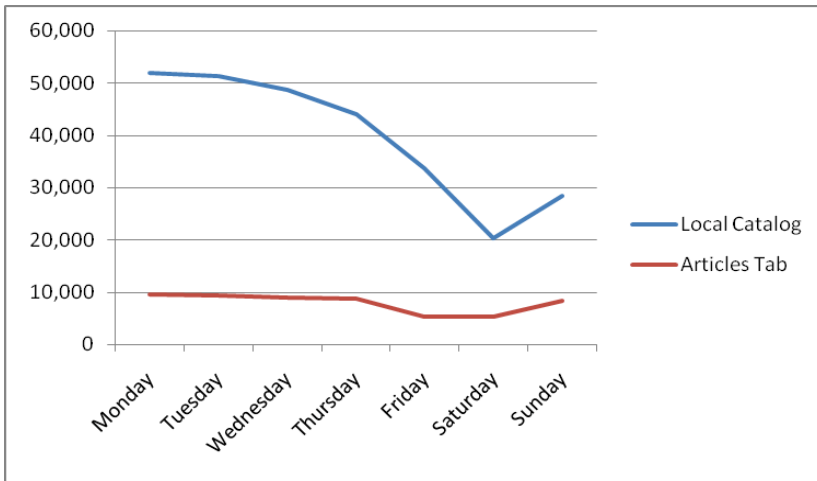
Primo, branded as MNCAT Plus, replaced the Aleph OPAC as the primary front-end discovery interface for the University of Minnesota Libraries' catalog in August 2008. Primo features many options touted as "NextGen OPAC" functionality, including a simple search interface, faceted browsing, and Web 2.0 tools like tagging and reviews. This new functionality was important in the selection of Primo as a discovery interface; Primo was designed to better meet the needs and expectations of contemporary searchers.

### **Findings**

There are two sources for Primo usage statistics: "Canned" reports using report specifications created by Ex Libris, and the JBoss web server logs. Both sources were reviewed for the purpose of this project. The standard Primo reports use discrete data elements captured and stored in Oracle as search summaries or click-through statistics.

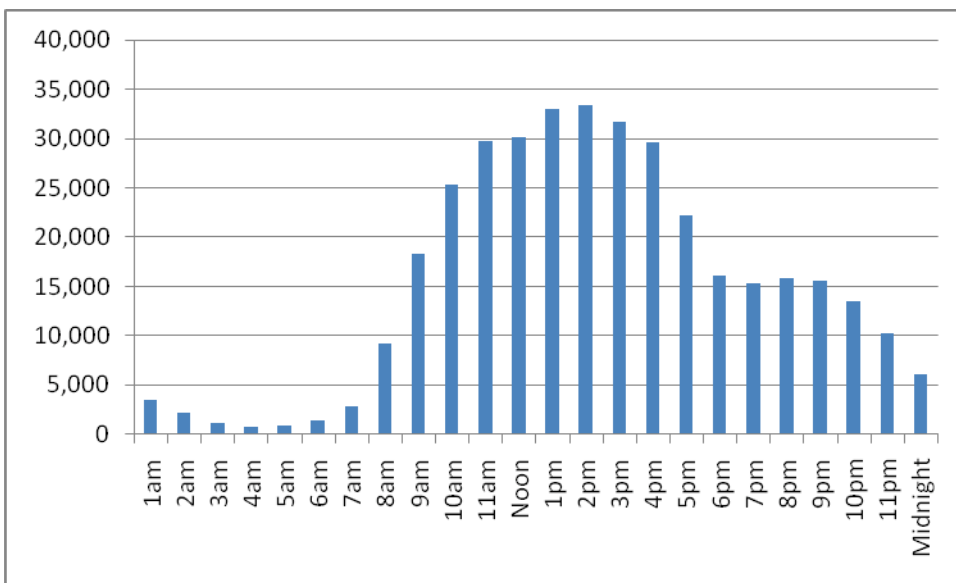
#### ***Weekly Primo Activity: October 2008***

Most Primo activity closely matches that of the Aleph OPAC, with more activity occurring Monday through Friday. However, the graph below shows a slightly different usage pattern for the "Articles" tab within Primo, which is another gateway to the MetaLib federated search engine. The data reflected below is taken from a four-week period in October 2008.



**Daily Primo Activity**

The data in the chart below shows the total number of searches per hour of the day. This breakdown mirrors similar activity in the Aleph OPAC.



**Breakdown of Primo Activity: October 2008**

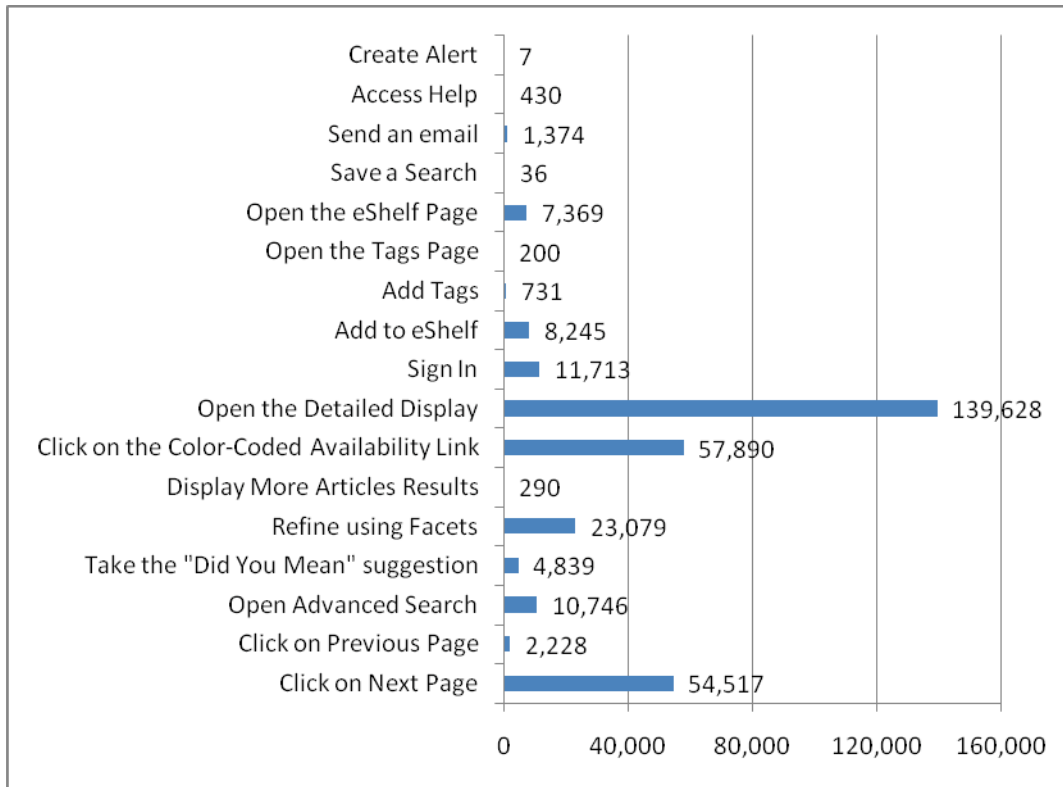
A breakdown of activities happening within Primo can be viewed in two ways: as a discrete set of searches, and as “click activity” that occurs within the application. The overall number of searches can be seen in the table below. Please note that use of a facet/filter is counted as a separate search in Primo, even though one can argue that it represents qualitatively different user intent.

<u>Catalog searches/filters</u> <u>(“Local”)</u>	<u>“Articles” tab searches/filters</u> <u>(“Remote”)</u>	<u>Total number of</u> <u>searches/filters</u>
307, 173 (83%)	60,808 (17%)	367,981



**Primo searches: October 2008**

A more complete picture of events can be seen in the chart below, showing number of times a particular link was clicked on during October 2008. For example, comparing the number of Detailed Displays (clicking on the title link) to the number of clicks on the color-coded availability link (“Get It”), possibly implying that people did not have enough information from the brief results list to make a determination of relevance. Also of note is the number of times people clicked on the “Next Page” link, indicating that they did not think that the first page of results fully addressed their information need.



**Where Help is accessed**

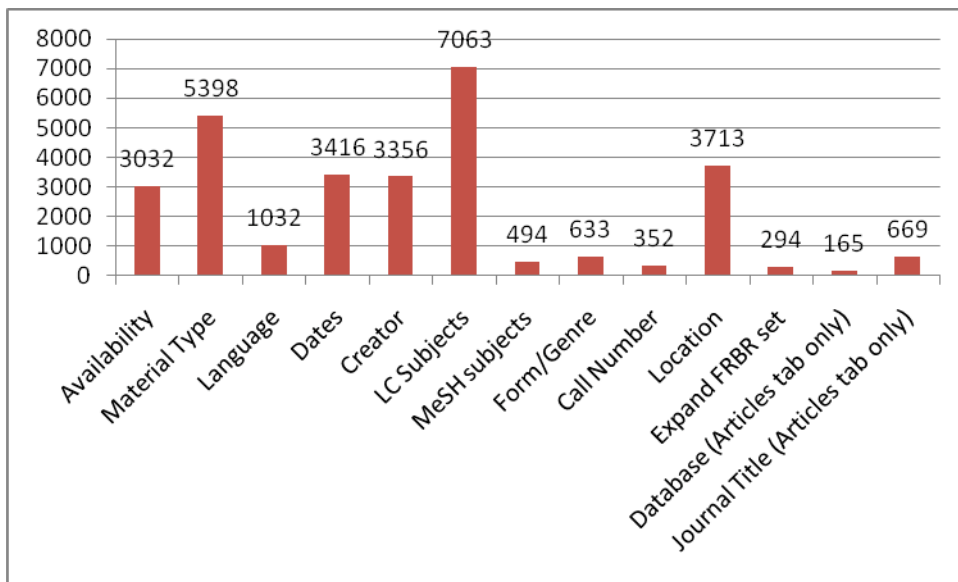
Primo logs the number of times people clicked on the Help text and from where that link is accessed. Below is a breakdown of where people felt they needed help during their use of the application. The number of times people felt the need of assistance in the eShelf is striking because the second most accessed help page – for “How to search” – is accessed from multiple locations, whereas the eShelf help is only accessed from within the eShelf itself.

Number of Times Topic Accessed in Oct 2008	
Why sign in?	71
What is eShelf?	163

What is tagging?	81
What are preferences?	8
How to search	103
How to save a query	4

**Preferred facets: October 2008**

The chart below shows a tokenized view of the preferred facets during the month of October 2008, showing use of several key facets.



**Future Explorations**

1. We have logs on search summaries and click throughs, but can we get more information about direct linking activities?
2. Investigate the current configuration of the robots.txt file.
3. Look at referrers in order to see where traffic is originating.
4. Look for ip addresses to determine off-campus/on-campus usage.
5. Look for evidence of pre-limit preferences. For example, are people choosing to limit by material type in advance?
6. Ex Libris provides data on search performance. Begin to look at, and compare, search performance between all discovery systems.
7. See if we can get more information on the use of "Push to"/"Save to". Are people saving to external sources, like delicious or RefWorks?

## SFX – OpenURL Link Resolver Baseline Report

SFX is a licensed software product developed by the ExLibris Group - <http://www.exlibrisgroup.com/>. The University of Minnesota Libraries deploys a complex SFX configuration that supports 15 unique instances. These include the four University of Minnesota campuses: Twin-Cities, Duluth, Crookston, and Morris; as well as Carleton College, Macalester College, Gustavus Adolphus, and several other Minnesota colleges and Universities. SFX is open source and lends itself to customization. Over the last five years SFX use at the University of Minnesota has shown a marked increase. In response, the libraries have tried to continually improve and enhance the SFX experience.

OpenURL is a standard that allows resource metadata to be embedded in the path portion of a URL.

Example:

`<http host>/<sfx instance>?<open url metadata>`

`http://resolver.example.edu/cgi?url_ver=Z39.88-`

`2004&rft_val_fmt=info:ofi/fmt:kev:mtx:book&rft.isbn=0836218310&rft.btitle=The+Far+Side+Gallery+3`

See <http://en.wikipedia.org/wiki/OpenURL> for an introduction.

### **Key Usage Statistics**

During October 2008, SFX received 392,460 user requests. 285,452 of these requests led to a 'click-through' to a service offering. This represents a 73% click-through rate. Historically, our click-through rate for SFX has been higher. Our suspicion is that the request statistics are being artificially inflated by the SFX menu integration in MNCat Plus, where the menu is regularly generated in the background without intentional action by the user. This would account for the lower click-through rate. Approximately 65% of user requests originate off-campus. By comparison, the libraries main web page was visited 222,041 times (2.9 visits/visitor) during the same time period. CDM has been harvesting statistics from SFX for over four years:

<https://wiki.lib.umn.edu/CDM/SFXStatistics>

### **Findings**

#### ***Top Sources***

SFX requests all originate from a known source. Sources include vendor databases, A to Z lists, Google Scholar, etc. It is impossible for OpenURLs to work properly without a preconfigured source. Thus, source usage gives a pretty clear indication of user patterns.

#### ***Sources with more than 10,000 Click-throughs: July 1, 2008 – October 31, 2008***

Source	Requests	Click-thrus
<a href="#">info:sid/sfxit.com:azlist – The SFX A to Z List</a>	208928	191690
<a href="#">info:sid/Entrez:PubMed - Pubmed</a>	155762	132795
<a href="#">info:sid/sfxit.com:opac_856 - MNCat</a>	86742	69599
<a href="#">info:sid/google – Google Scholar</a>	62057	54059
<a href="#">info:sid/sfxit.com:citation – SFX Citation Linker</a>	62186	53116
<a href="#">info:sid/OVID:medline</a>	65951	52821

info:sid/www.isinet.com:WoK:WOS	42778	35669
info:sid/CAS:CAPLUS	39914	32714
info:sid/OVID:psycdb	40843	29704
info:sid/EBSCO:aph	24313	21435
info:sid/Refworks:The University of Minneso	21197	16289
info:sid/primo.exlibrisgroup.com:primoJournal-umn_aleph	103372	10805
info:sid/metalib.com:EBSCO_APH	10989	10484
info:sid/EI:Compendex	17993	10407

The majority of SFX requests originate from external sources. Google, PubMed, and the large databases and indexes account for ~75% of all SFX requests. However, what jumps out here is that 3 of the top 5 sources are local in origin: the A to Z Journal List, MNCat, and the SFX Citation Linker. The libraries still exert a fair amount of control over where full-text is initially discovered.

**Top Targets (>10,000 requests): July 1, 2008 – October 31, 2008**

Target	Total
EBSCOHOST_ACADEMIC_SEARCH_PREMIER	110716
LOCAL_CATALOGUE_EX_LIBRIS_ALEPH	68144
ELSEVIER_SD_ELSEVIER	40320
OVID_JOURNALS_AT_OVID	27511
AMERICAN_CHEMICAL_SOCIETY_JOURNALS	25027
WILEY_INTERSCIENCE_JOURNALS	23394
EBSCOHOST_BUSINESS_SOURCE_PREMIER	21577
SPRINGER_LINK_JOURNALS_STANDARD	18670
HIGHWIRE_PRESS_FREE	18181
MISCELLANEOUS_FREE_EJOURNALS	17035
SAGE_PREMIER_2008	15919
ELSEVIER_SD_PERGAMON	15320
EBSCOHOST_MEGAFILE	15011
APA_PSYCARTICLES	14907
SYNERGY_BLACKWELL_PREMIUM	14438
NATURE	13371
DOAJ_DIRECTORY_OPEN_ACCESS_JOURNALS_FREE	12792
MISCELLANEOUS_EJOURNALS	11801
ELSEVIER_SD_ACADEMIC_PRESS	11318
INFORMAWORLD_JOURNALS	10539

Not surprisingly, the big vendors dominate. Users clicked into the catalog 68,000 times, but this was a small percentage of overall usage.

**Full Text vs. Other**

In October 2008, 288,085 SFX requests resulted in full-text, 104,375 did not. Interestingly, this is also a 73% ratio (the same as the click-through rate).

## **Future Explorations**

1. What does the 73% click-through rate represent? The lower recent click-through rate could partially be explained by Primo. Note the following statistic:

Source	Requests	Click-thrus
info:sid/primo.exlibrisgroup.com:primoJournal-umn_aleph	103372	10805

From July through October, 2008 Primo users only clicked through 10% of the time. This substantially lowered the overall click-through rate. Primo presents users with a three-tabbed pop-up window. The third tab is the SFX menu. Primo likely renders the SFX menu behind the scenes and it counts as a hit. This probably explains the low click-through rate. Interpreting the 10% figure is a matter for further investigation.

2. Is there a correlation between the 73% click-through rate and the 73% full-text rate? Perhaps. Since we can assume that the true click-through rate minus Primo is probably higher, users are still showing an extreme fondness for full-text.
3. How can you explain the Journal A to Z lists enduring popularity? There has been a lot of discussion about this on the SFX mailing lists. Most institutions report that their journal browse lists are also far and away the most popular way to find articles. Is this library site inertia or true user preference?

## ***University Digital Conservancy Baseline Report***

The University Digital Conservancy (UDC) is an open access repository of full-text scholarly literature and other work of the University of Minnesota. Items in the repository include:

- Working papers
- Conference papers
- Journal articles
- Slide presentations (typically in PDF format)
- Maps
- Institutional resources that would have traditionally gone to the University Archives.

UDC is sponsored by the University Libraries. Although not strictly limited to UMN authors, most papers and presentations are by people affiliated with the University.

The UDC co-directors and working group have always operated under the assumption that most discovery will be through a search engine such as Google, but users can search within the UDC interface as well as browse by collection, author, title, or subject. Users also come to the UDC through direct links from other Web sites, email, etc. The UDC collection is not exposed through the MNCAT or Primo interfaces.

UDC is built on the DSpace platform, and statistics are gathered and presented using AWStats. UDC currently contains about 6812 items. Growth tends to be in fits and starts, with whole collections added at once. With such a small set and a significantly shorter history, comparing UDC statistics with AgEcon statistics may be apples to oranges. But since the overall purpose of the two repositories is similar and they share a platform, I have included a few comparisons.

## **Findings Summary**

Apache stats available in AWStats from 1/1/2008-present: [https://conservancy.umn.edu/stats\\_main.jsp](https://conservancy.umn.edu/stats_main.jsp)  
The following statistics are for October 2008. Note that traffic from IP addresses suspected of “gaming” statistics (repeatedly downloading files to increase download counts) have **NOT** yet been filtered out.

### ***Bandwidth***

Because Google (and perhaps other search engines) download PDFs in order to provide full text indexing, a large percentage of the bandwidth consumed serves non-human users. It’s worth exploring whether any changes can be made to the site’s configuration that will make robot access more efficient.

<b>Human bandwidth</b>	23.27 GB
<b>Non-human bandwidth</b>	73.26 GB

### ***Connect to site from***

For any Web page, a user might get to the page through:

- **Direct address.** The user might type the address directly into the browser Address box, choose a saved bookmark, or click a link in an e-mail message or in a file that is not a domain-hosted Web page. AWStats places site entries made through one of these methods in one category, called Direct Address in this report, and cannot differentiate by the actual method used.
- **Search engine.** The user can search using keywords or phrases and click a link to an UDC page in the search results. AWStats provides a list of all phrases and individual words used during the month, along with the number of searches that used that keyword or phrase.
- **Referring page.** The user may arrive on the UDC site or download a specific item after clicking a link on another Web site. AWStats provides a list of sites that refer traffic to UDC, but doesn’t provide details on which page(s) of UDC the user entered from the referring site.

Only 8.4% of users connecting to the UDC site came through search engines, versus 52.3% for AgEcon Search. It’s very likely that the Direct Address numbers would drop considerably if “gaming” IP addresses were filtered out of the statistics set.

<b>Starting Point</b>	<b>Page Views</b>	<b>Percent of Total</b>
Referring Page	16354	12.6
Search Engine	10929	8.4
Direct Address	101598	78.7

### ***Referring Page***

Although links to the UDC’s main page predominate at the top of the list, it’s interesting to see that a few page views are coming from Wikipedia. Looking at the full list, three different Wikipedia pages send traffic to papers or maps in the UDC.

Referring Page	Description	Page Views	Percent of Total
<a href="http://special.lib.umn.edu/uarch/">http://special.lib.umn.edu/uarch/</a>	University Archives	132	14%
<a href="http://www.lib.umn.edu">http://www.lib.umn.edu</a>	University Libraries	92	10%
<a href="http://blog.lib.umn.edu/moore144/ahcarc-hives/">http://blog.lib.umn.edu/moore144/ahcarc-hives/</a>	Academic Health Center History Project blog	67	7%
<a href="http://www1.umn.edu/usenate/committees/scfp.html">http://www1.umn.edu/usenate/committees/scfp.html</a>	University Senate Financial Planning Committee	45	5%
<a href="http://en.wikipedia.org/wiki/Duluth_Complex">http://en.wikipedia.org/wiki/Duluth_Complex</a>	Wikipedia entry on the Duluth Complex, a rock formation.	27	3%
<a href="http://www1.umn.edu/usenate/committees/scit.html">http://www1.umn.edu/usenate/committees/scit.html</a>	University Senate Information Technologies Committee	21	2%
<a href="http://purl.umn.edu/1237">http://purl.umn.edu/1237</a>	Bibliographic information for an item in the UDC	20	2%
<a href="http://blog.lib.umn.edu/moore144/ahcarc-hives/2008/10/masonic_can...">http://blog.lib.umn.edu/moore144/ahcarc-hives/2008/10/masonic_can...</a>	Academic Health Center History Project blog: Entry on the Masonic Cancer Hospital's 50 <sup>th</sup> anniversary	19	2%
<a href="http://www1.umn.edu/usenate/committees/fcc.html">http://www1.umn.edu/usenate/committees/fcc.html</a>	University Senate Faculty Consultative Committee	18	2%
<a href="http://searchservice.myspace.com/index.cfm">http://searchservice.myspace.com/index.cfm</a>	MySpace Search Service	18	2%
Others		452	50%

### ***Search Engines***

Google tops the list of search engines.

Search Engine	Searches ending in an AgEcon Search Page
Google	10386
AOL	107
Unknown search engines	103
Yahoo!	98

MSN Search	57
Windows Live	47
Ask	35
Google (cache)	23
Earth Link	12
AT&T search (powered by Google)	9
Others	53

### ***Phrases Resulting in Click-throughs by Searchers using Search Engines***

Although most of the top-of-the-list searches appear to be for specific items, there were a few searches farther down for variations on “university digital conservancy.” Still, it’s obvious that UDC is not the brand that AgEcon is. It’s impossible to tell the searcher’s intent, but finding known items may be somewhat serendipitous: The searcher knows the item exists, but may or may not know it’s in the UDC.

Phrase	Search	Percent
scalar wave equation finite difference approximations	80	0.7 %
nomenclador de los censos	41	0.3 %
single-intersection evaluation of real-time adaptive traffic signal control algorithms	41	0.3 %
fault diagnosis in dynamic systems using analytical and knowledge based redundancy—a survey and some new results	41	0.3 %
advantage of pc based truck weighing system	40	0.3 %
cindy swanlaw	34	0.3 %
rec center	30	0.2 %
functional analysis filetype pdf normed space operator	27	0.2 %
compact linear operators filetype pdf	27	0.2 %
nomenclador de los censos de poblacion	24	0.2 %
Other phrases	10323	96.4 %

### ***Entry pages (the first page the user sees)***

Because a significant number of the page views may be due to repeated downloads to bump up counts, it’s difficult to draw conclusions from the entry page counts.

Page	Description	Entries	Percent
/	Main page	589	4.69
/items-by-subject	Subject Browse result	190	1.51
/items-by-author	Author Browse result	171	1.36
/feed/rss_1.0/3	RSS feed. Each ping by an RSS reader probably	115	0.92



	counts as an entry.		
/dissertations/	Information on submitting your dissertation to the UDC.	114	0.91
/feed/rss_2.0/site	RSS feed.	112	0.89
/feed/rss_2.0/37067	RSS feed	108	0.86
/browse-title	Browse by Title page	98	0.78
/handle/3/	University Archives community	92	0.73

### **Browsers**

The browser list isn't particularly relevant to our task, but in this case it's worth noting that 60% of the page views came from a browser called "Curl." This is actually a command line tool that downloads a specified URL, and it's what people trying to "game" the statistics would probably use. At this point, it's difficult to know what percent of the Curl hits are legitimate.

<b>Browser</b>	<b>Page Views</b>	<b>Percent</b>
Curl	93870	60.4 %
MS Internet Explorer	32619	20.9 %
Firefox	22505	14.4 %
Safari	3964	2.5 %
Mozilla	1373	0.8 %
Netscape	358	0.2 %
Unknown	273	0.1 %
Opera	259	0.1 %
Konqueror	91	0 %
Microsoft Data Access Component Internet Publishing Provider Protocol Discovery	32	0 %
Others	58	0 %

### **Further Explorations**

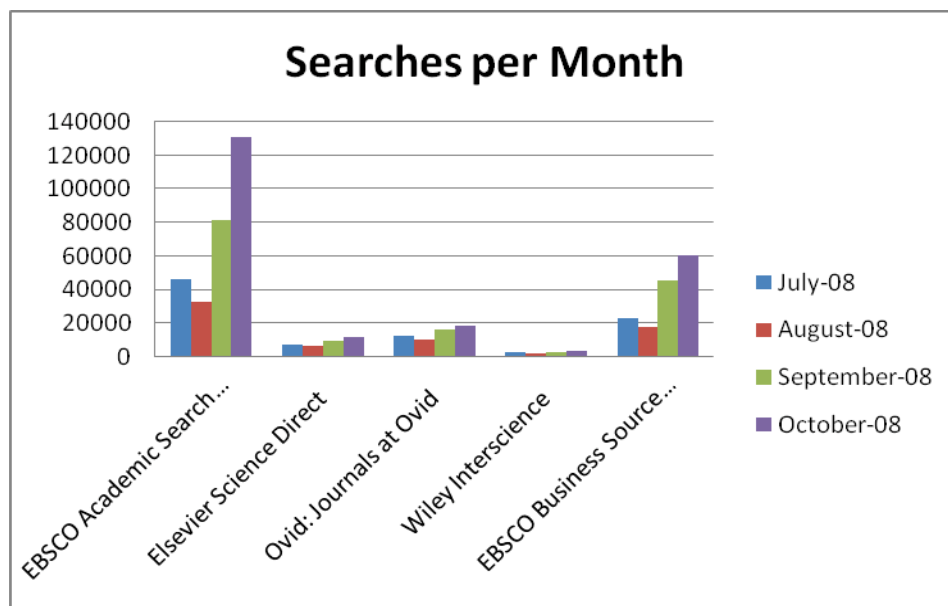
1. Is there any way to know what the effect of adding the UDC "silo" to our current primary discovery tool (Primo) would have?
2. What happens to these statistics when "gaming" hits are filtered out?
3. Is it worth discussing the merits of the UDC's own search as a discovery tool when it's not considered important by those who "own" the platform?
4. Is it worth exploring the idea of adding links to relevant UDC work to appropriate Wikipedia pages?

## Vendor Statistics Baseline Report

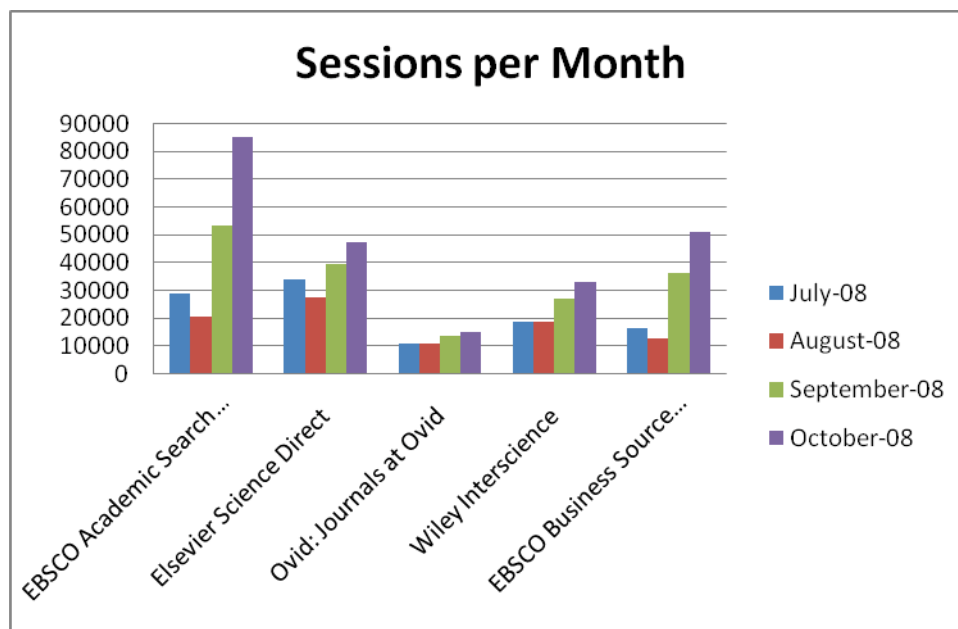
In examining vendor-supplied statistics, we chose to use only statistics from those vendors who are COUNTER compliant. COUNTER (Counting Online Usage of NeTworked Electronic Resources) is an international, non-profit organization whose members include publishers, libraries, consortia, and other interested parties (see [www.projectcounter.org](http://www.projectcounter.org)). COUNTER has developed Codes of Practice which outline standards and protocols for the recording and exchange of online usage data. We have only examined COUNTER-compliant statistics for this project since we can have a high degree of confidence that COUNTER statistics are comparable across vendors.

Here are presented search and session statistics for five of the top 10 SFX targets (by number of requests) for July-October 2008. COUNTER defines “search” as “a specific intellectual query, typically equated to submitting the search form of the online service to the server.” A “session” is “a successful request of an online service. It is one cycle of user activities that typically starts when a user connects to the service or database and ends by terminating activity that is either explicit (by leaving the service through exit or logout) or implicit (timeout due to user inactivity).”

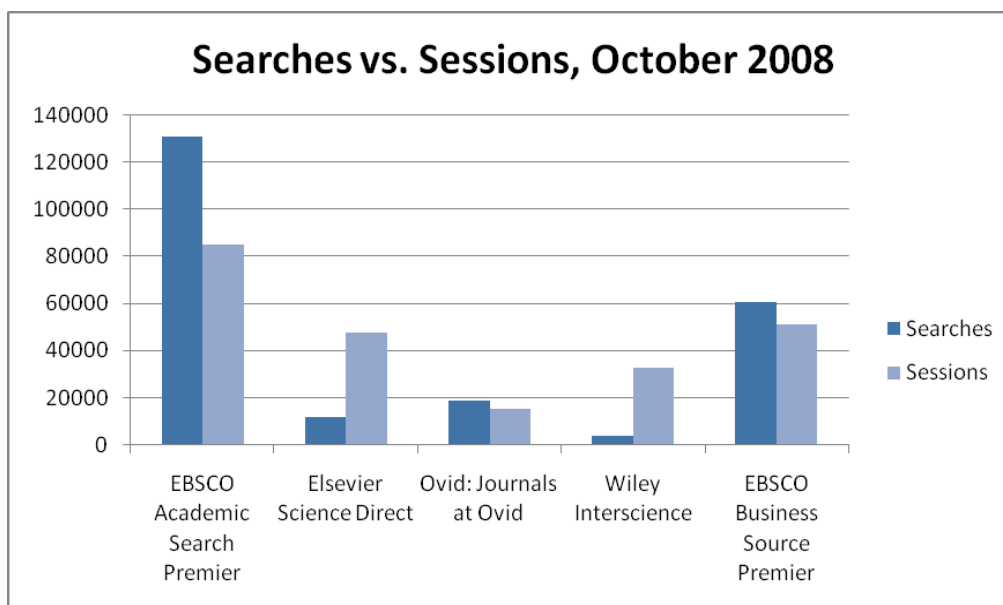
Searches	July-08	Aug-08	Sep-08	Oct-08	Resource Totals
EBSCO Academic Search Premier	46696	32705	81885	131219	<b>292505</b>
Elsevier Science Direct	7343	6607	10099	11972	<b>36021</b>
Ovid: Journals at Ovid	12513	10324	16815	18892	<b>58544</b>
Wiley Interscience	2699	2068	3025	3845	<b>11637</b>
EBSCO Business Source Premier	23542	17796	45981	60603	<b>147922</b>
<b>Monthly Totals</b>	<b>92793</b>	<b>69500</b>	<b>157805</b>	<b>226531</b>	<b>546629</b>



Sessions	July-08	Aug-08	Sep-08	Oct-08	Resource Totals
EBSCO Academic Search Premier	29048	20809	53309	85430	<b>188596</b>
Elsevier Science Direct	34190	27737	39765	47674	<b>149366</b>
Ovid: Journals at Ovid	11039	11038	14036	15451	<b>51564</b>
Wiley Interscience	19070	18851	27433	33076	<b>98430</b>
EBSCO Business Source Premier	16579	12702	36662	51266	<b>117209</b>
<b>Monthly Totals</b>	<b>109926</b>	<b>91137</b>	<b>171205</b>	<b>232897</b>	<b>605165</b>



Among these five resources, four of them fall into two natural divisions: resources that appear to be frequently searched via the native interface, and resources that probably primarily provide direct full-text links and which are not often searched. For the EBSCO resources, the number of searches typically exceeds the number of sessions in a given month. Since they are databases that provide one of several logical starting points for many users, this matches expectations. For both Elsevier Science Direct and Wiley Interscience, the number of sessions greatly exceeds the number of searches in a typical month, which probably indicates that when users are connecting to these resources, they are linking directly to full text rather than using the resource's native search engine for discovery. The statistics for Ovid: Journals at Ovid align with neither of the other two categories seen in this sample, with the number of searches typically narrowly exceeding the number of sessions each month.



Unsurprisingly, the total number of searches and sessions appears in this four-month sample to rise and fall with the academic year. August is a low point (few classes in session), while the numbers are much larger in September and continue to grow in October, the midpoint of fall semester.

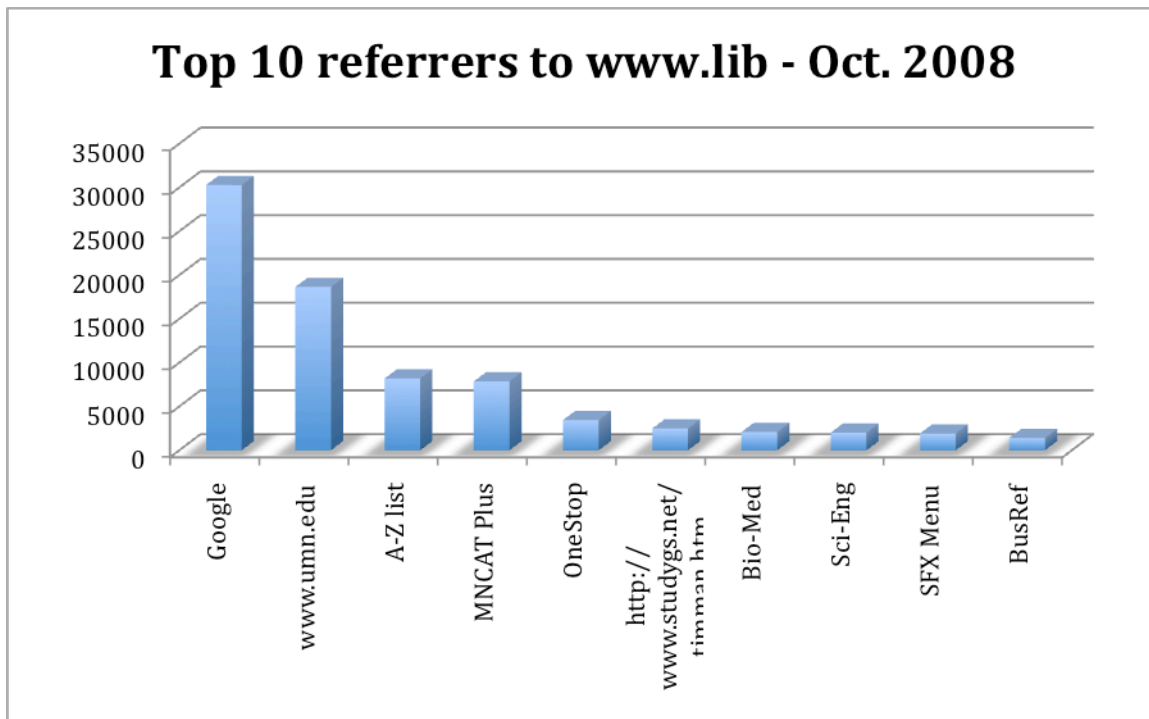
### *Libraries' Website Statistics Baseline Report*

In addition to our catalogs and other web-based discovery tools, the University Libraries maintain nearly 30 different websites. In addition to the Libraries' main site, <http://www.lib.umn.edu>, there are separate sites for most of the Libraries' collections, buildings, and many service points. These sites each have their own subdomain, which replaces "www" in the URL of the main site, meaning that, for instance, our Science and Engineering Library's website can be found at <http://sciweb.lib.umn.edu>.

The Libraries' IT staff have set up a program called AWStats, which is a web-based web statistics analysis program that presents site usage information gleaned from the log files of the Apache web server. The statistics for each subdomain are tracked separately. The Health Sciences Libraries' sites use a different system for tracking web statistics. Rather analyzing log files using AWStats, they track usage using Google Analytics, which relies on a small piece of JavaScript code embedded in each page to track users. We have prepared reports on each of the subdomains using data from AWStats or Google Analytics. In the sections below, we describe trends observed across our subdomains and note those sites that deviate from those trends. The complete subdomain reports are collected in Appendix C.

### **Referrers**

The Apache web server records the URL of any page that leads someone directly to one of our sites. So, if a user visits the University's home page and clicks on "Libraries", Apache would log a referral from <http://www.umn.edu>. Google Analytics records the same metric using JavaScript on each page. Google dominates referral statistics on almost every one of our websites. The chart below shows the top 10 pages that referred users to the Libraries' main website during October 2008.



This chart shows our non-Google referrals as primarily coming from elsewhere at the University (www.umn.edu, OneStop) and from some of our other websites and applications (A-Z list, Bio-Med, etc.). The amount of traffic between our various sites and applications is indicative of the complex environment our users must traverse in order to use our resources successfully. Unfortunately, we cannot glean much information about these interactions because there is no tracking continuity between subdomains.

Several of our sites do buck the trend of Google as top referrer. Our Friends of the Library site (friends.lib.umn.edu) sees more referrals from the Libraries' main website than from Google. This may simply be an indication of the mental model of users interested in our Friends site: that they associate the Friends first with our Libraries, arriving at our main site via Google, and then clicking through to the Friends subdomain. Regardless, Google is the second most frequent referrer of visitors to this site.

Users rarely arrive at our InfoPoint digital reference site (infopoint.lib.umn.edu) through Google. This may have to do with a lack of descriptive content on the site, or the fact that the InfoPoint brand isn't recognized or sought out by our users. It may also be that the services on the site are easily accessed at the point of need through links elsewhere on our site.

The (now deprecated) Learning Resource Center (LRC) site (lrc.lib.umn.edu) presents another interesting case. The top referrer to this site during October 2008 was not Google, but Tagoo, which appears to be a multimedia search site originating in Russia. Tagoo has indexed a series of recordings of the Pushto (or Pashto) language hosted as mp3s on the LRC site. These files were accessed over 45,000 times through Tagoo in October 2008. While this obviously represents a unique scenario among our sites, it is a clear indication of the power of information delivered on the web when it finds its proper audience.

Each of the abovementioned sites is different from the majority of our websites in that they have a primary referrer other than Google, but also in that the content of the sites don't relate directly to one of our collections or subject areas, as do most of our sites. The lone exception to this correlation is our Math Library site ([math.lib.umn.edu](http://math.lib.umn.edu)), to which more traffic is referred from the University's Math Department site ([www.math.umn.edu](http://www.math.umn.edu)) than from any other site. Google ranks second, but at least two other pages on the Math Department site make an appearance among the top ten referrers to the Math Library site.

We believe it is worth studying the relationship between the Math Library and Math Department as a model, if we wish to encourage more traffic from academic departments. Our understanding is that the Math Library creates CourseLib guides for each class section offered by the Math Department each semester. The presence of <http://courses.lib.umn.edu/page.phtml>, a page on our CourseLib site, among the top referrers to the Math Library site may be an indicator that these guides are driving traffic to the site.

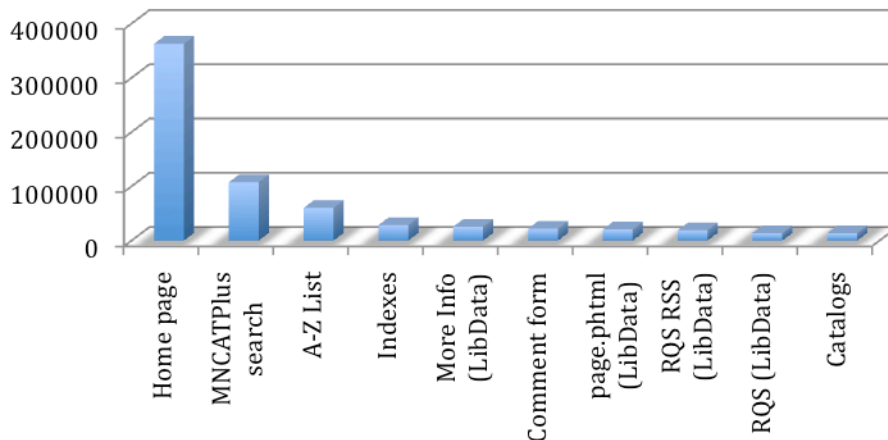
It is important to note that a great number of visits to our site do not log a referrer, indicating that the user accessed the site from a bookmark, had the page set as their browser home page, or that their referring URL was somehow obfuscated. In fact, "no referrer" is our top referrer.

Because we run web applications on several domains and subdomains, we see a notable amount of circular referrals from our own services, i.e. MNCAT Plus, SFX, and branch libraries. With that said, the sum of all our internal referrals does not equal the number of referrals from Google. We have been unable to ascertain whether referrals from the University's Google Search Appliance are included among the Google or [www.umn.edu](http://www.umn.edu) referrals, are logged somewhere else, or are not logged at all. We recommend further investigation.

### **Top Pages**

AWStats and Google Analytics provide a listing of the pages (or files) most frequently accessed on each subdomain. The chart below shows the top pages viewed on the [www.lib.umn.edu](http://www.lib.umn.edu) domain during October 2008. The drop-off in traffic between the home page and any internal page is dramatic, with the next most popular page seeing less than 1/3 the visits of our home page. This dramatic curve is a pattern that repeats on nearly all of our websites. Internal pages rarely see a fraction of the traffic of the home page.

## Top 10 pages on www.lib - Oct. 2008



In some cases, the dominance of site home pages may be explained by the public terminals (and staff computers) that have site home pages set as their default launch page. For example, AWStats logs a visit to the scieng subdomain every time someone launches a web browser on a public machine in Walter Library. We do have some statistics on the IP addresses of computers that visit our sites, but to filter out our public terminals and staff computers would be to eliminate legitimate traffic along with those transient users. Likewise, AWStats and Google Analytics provide information on how frequently users leave a site from a given page. From this information we could try to determine how many users are delving deeper into our site. However, because we log stats for each subdomain separately, what appears to be a user leaving the site may be that user searching our OPAC or visiting another of our websites.

Worth noting here is that several of the pages in the chart above, indicated by “(LibData)” actually represent the cumulative visits to many pages. LibData URLs contain arguments that are not logged by Apache, and as a result, every visit to every RQS page is logged as a visit to a single page, here represented as “RQS (LibData)”. Likewise on the Archives and Special Collections site, the most frequently accessed page represents every visit to any Encoded Archival Description file on the site. A similar problem is evident on our Ames Library website, where AWStats is logging the site’s layout template files as separate pages, meaning that a visit to any page on the site is counted as many as four times. It is fairly easy to filter out this noise from the Ames site statistics, but in future it would be helpful to use layout elements more consistently across our sites.

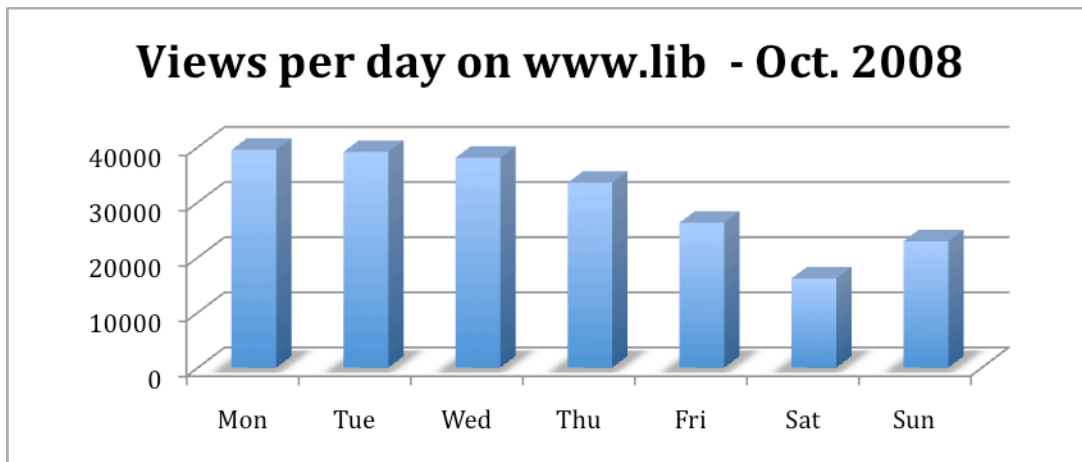
We do have several interesting exceptions to the usage pattern illustrated in the chart above. Magrath Library has created an RSS feed for their home page, which is hit more than four times as frequently as the site’s home page. The nature of RSS feed readers is that the feed will be fetched at a set frequency, which artificially inflates these statistics, but this does illustrate a clear demand for RSS content.

The James Ford Bell Library (bell.lib.umn.edu) site is another interesting counter-example. The library’s home page is the most frequently visited, but the drop-off in traffic on internal pages is much more gradual. As many

as ten pages on the site get one-third the traffic of the home page or more, dramatically more than many of our sites. These popular pages contain digitized collections of historical materials, some of which were accessed over 1,000 times in October 2008. These collections are exceedingly difficult to navigate to using the Bell Library site, so it comes as no surprise that referrals to this site come from external sources, primarily search engines. In fact, there is no page on the umn.edu domain among the top ten referrers to the Bell Library site.

### **Day of week**

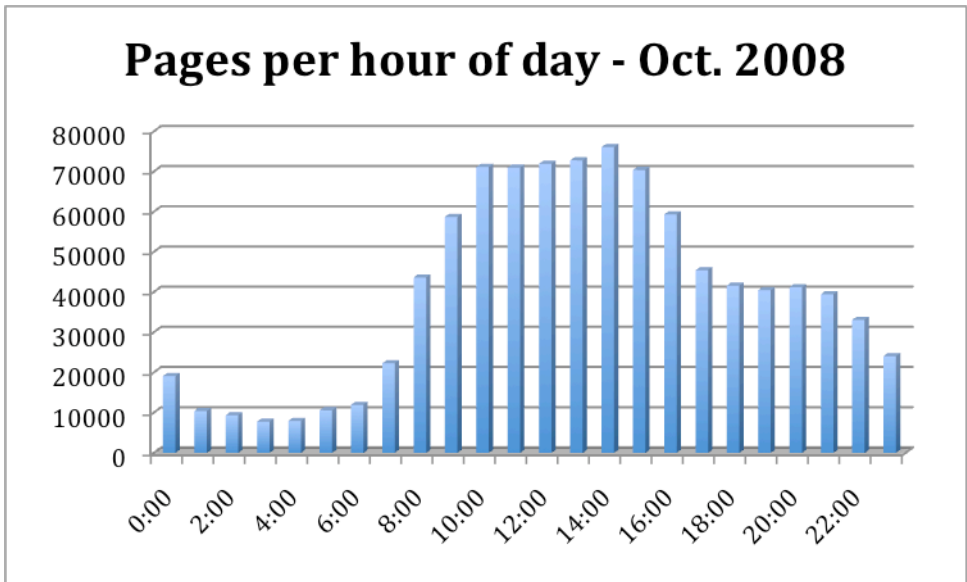
AWStats provides a listing of page views by day of the week. The chart below shows visits to the www.lib.umn.edu domain by the day of the week during October 2008. There were no significant aberrations from this pattern on any of our other sites.



### **Hour of day**

This chart depicts the number of pages viewed on the www.lib.umn.edu domain by the hour of the day for October 2008. Usage of the site peaked at 2:00 p.m. There are few visits to the site before 8:00 a.m., and while traffic declines after 4:00 p.m., it remains significant through 10:00 p.m. Again, there was little deviation from this pattern on any of our other websites.



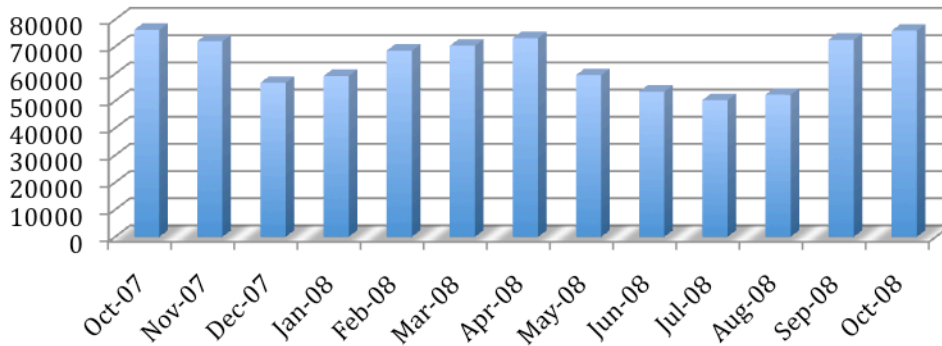


**Visitors by Month**

This chart shows unique visits to pages on the [www.lib.umn.edu](http://www.lib.umn.edu) domain between October 2007 and October 2008. During that period October and April saw the heaviest traffic, and December, June, July, and August saw the fewest visits. This school-year cycle of usage was evident on all of our sites, though it was least dramatic on the VetMed site, which sees fairly steady traffic year-round.

AWStats’ visitors per month tracking alerted us to the fact that from late January to mid-March 2008, web statistics are missing for many of our sites. Likewise, there are a number of our sites, such as AgEcon Search ([agecon.lib.umn.edu](http://agecon.lib.umn.edu)) that have a break in statistic tracking continuity. In May of 2008 AgEcon Search statistics were moved from the Libraries’ main AWStats instance into a separate version of AWStats. This lack of complete and continuous statistics frustrates efforts to make comprehensive and effective comparisons from year-to-year and site-to-site.

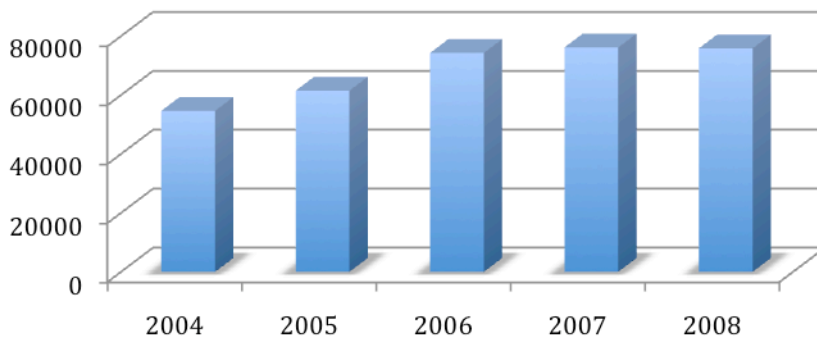
## Unique visitors by month - Oct. 2007- Oct. 2008

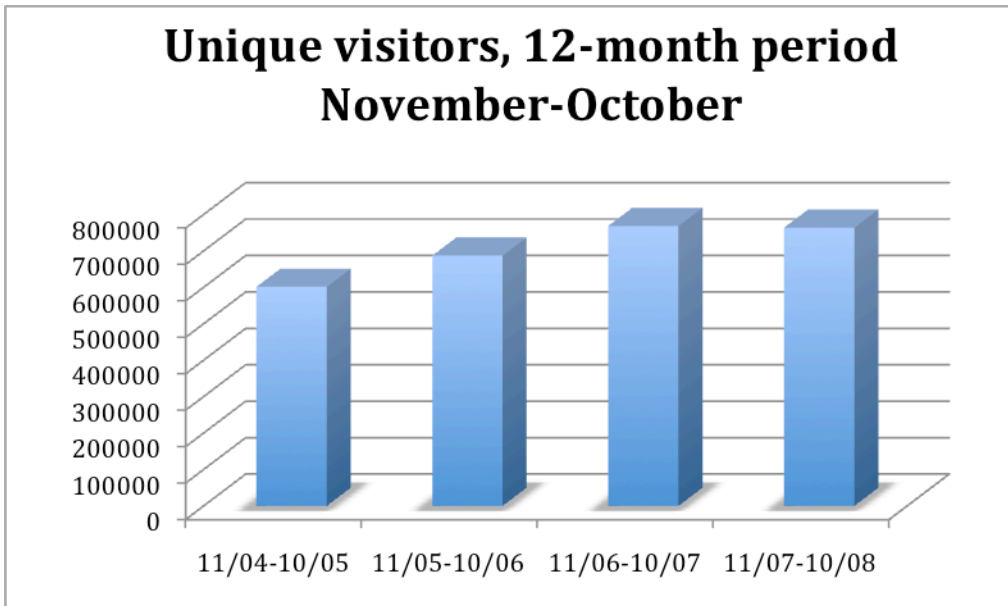


### Historical Unique Visits

The first chart below shows unique visits to the [www.lib.umn.edu](http://www.lib.umn.edu) site in October over the past several years. The Libraries' main site is the only one of our sites for which we have consistent data extending back this far. The graph is relatively flat since 2006, and 2008 showed a small decline from 2007. This trend is reinforced by the second chart below, which shows unique visits in a November-October 12-month period over the past several years. After rising steadily, visits to the site peaked during the November 2006-October 2007 period and declined slightly in the November 2007-October 2008 period.

### October unique visits

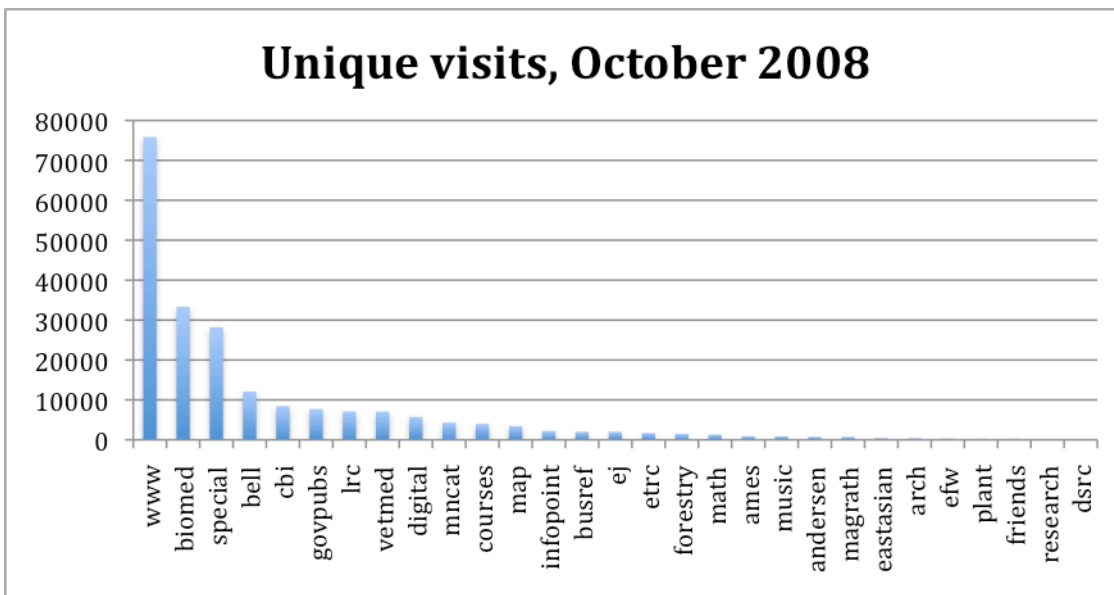




This flattening in usage of our websites would seem to support the idea of discovery happening elsewhere.

#### **Website Visits**

We have used statistics from the Libraries' main website as examples in this section in part because the trends in these data tend to exemplify those we see in our other sites. The other reason we have focused on the main website is that much as Google dominates our referrals and our home pages dominate our page views, the traffic on most of our subdomains is dwarfed by traffic to the main site. The chart below shows unique visitors our sites in October 2008.



While most of these sites have a smaller target audience than our main site, the distinct lack of traffic on some sites is striking. As we consider evolution of the Libraries' web presence, we should take care to consider the costs and effort required to maintain separate sites for our smaller collections and branches when a page or two on the main site might suffice.

## Works Cited

Anderson, Janna Quitney, and Lee Rainie. "The Future of the Internet III." 2008.

<[http://pewinternet.org/PPF/r/270/report\\_display.asp](http://pewinternet.org/PPF/r/270/report_display.asp)>.

Brantley, Peter. "Architectures for Collaboration: Roles and Expectations for Digital Libraries." EDUCAUSE Review 43.2 (2008).

Caruso, Judith Borreson, and Gail Salaway. The ECAR Study of Undergraduate Students and Information Technology, 2008. Educause Center for Applied Research, 2008.

<<http://net.educause.edu/ir/library/pdf/EKF/ekf0808.pdf>>.

Dempsey, Lorcan. "The Library Catalogue in the New Discovery Environment: Some Thoughts." Ariadne.48 (2006) <<http://www.ariadne.ac.uk/issue48/dempsey/>>.

Dentinger, Sue, et al. Resource Discovery Exploratory Group Task Force Final Report. Madison, WI: University of Wisconsin - Madison Libraries, 2008. <<http://staff.library.wisc.edu/rdetf/RDETF-final-report.pdf>>.

DeRosa, Cathy, et al. College Students' Perceptions of Libraries and Information Resources: A Report to the OCLC Membership. Dublin, OH: OCLC Online Computer Library Center, Inc., 2006.

<<http://www.oclc.org/reports/perceptionscollege.htm>>.

Fox, Megan. Trends in Mobile Tools and Applications for Libraries., 2008.

Hanson, Cody, et al. Affinity Strings: Enterprise Data for Resource Recommendations. The Code4Lib Journal. 5 (2008) <<http://journal.code4lib.org/articles/501>>.

Inger, S and Gardner, T. "How Readers Navigate to Scholarly Content." White Paper research funded by MetaPress, National Academy of Sciences, Annual Reviews, Nature Publishing Group (2008).

Kharif, Olga. "Move Over Kindle - E-Books Hit Cell Phones." 2009.

<<http://www.ecommercetimes.com/story/Move-Over-Kindle---E-Books-Hit-Cell-Phones-65685.html>>.

Kroski, Ellyssa. "On the Move with the Mobile Web: Libraries and Mobile Technologies." Library Technology Reports 44.5 (2008): January 29, 2008. <[www.techsource.ala.org](http://www.techsource.ala.org)>.

Law, Derek. "Beyond the Hybrid Library: Libraries in a Web 2.0 World." Digital Convergence - Libraries of the Future. Ed. Rae Earnshaw and John Vince. London: Springer-Verlag London Limited, 2007. 107-118.

<<https://www.lib.umn.edu/slog.phtml?url=http://dx.doi.org/10.1007/978-1-84628-903-3>;

<https://login.libpdb.d.umn.edu:2443/login?url=http://dx.doi.org/10.1007/978-1-84628-903-3>>.

Mahaney, Mark. "Citigroup sees better Amazon Kindle Sales." 2008.

<<http://www.reuters.com/article/internetNews/idUSWNA802020080811>>.

Marcus, Cecily, et al. Understanding Research Behaviors, Information Resources, and Service Needs of Scientists and Graduate Students: A Study by the University of Minnesota Libraries. Minneapolis: University of Minnesota Libraries, 2007. <<http://www.lib.umn.edu/about/scieval/Sci%20Report%20Final.pdf>>.

Maron, Nancy L., and K. Kirby Smith. Current Models of Digital Scholarly Communication. Washington, DC:

Association of Research Libraries, 2008. <<http://www.arl.org/bm~doc/current-models-report.pdf>>.

Nackerud, Shane and Scaletta, K. "Blogging in the academy." New Directions for Student Services, Winter (2008).

Prabha, Chandra, et al. "What is enough? Satisficing Information Needs." Journal of Documentation 63.1 (2007): 74. <<http://www.oclc.org/research/publications/archive/2007/prabha-satisficing.pdf>>.

Research Information Network. Discovering Physical Objects: Meeting Researchers' Needs. Research Information Network, 2008.

Rowlands, Ian, et al. "The Google Generation: The Information Behaviour of the Researcher of the Future." Aslib Proceedings 60.4 (2008): 290-310. Emerald.

Shadle, Steve. "The Local Catalog is Dead! Long Live the Local Catalog!" Serials Review 34.2 (2008): 85-7.

The New Media Consortium and the EDUCAUSE Learning Initiative. 2008 Horizon Report. The New Media Consortium, 2008. <<http://www.nmc.org/pdf/2008-Horizon-Report.pdf>>.

The New Media Consortium. 2009 Horizon Report Preview. The New Media Consortium, 2008. <[http://net.educause.edu/elements/attachments/eli/2009\\_hr\\_preview.pdf](http://net.educause.edu/elements/attachments/eli/2009_hr_preview.pdf)>.

University of California Libraries Bibliographic Services Task Force. Rethinking how we Provide Bibliographic Services for the University of California., 2005. <<http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf>>.

University of Minnesota Libraries. A Multi-Dimensional Framework for Academic Support. Minneapolis: University of Minnesota, 2006a. <[http://www.lib.umn.edu/about/mellon/UMN\\_Multi-dimensional\\_Framework\\_Final\\_Report.pdf](http://www.lib.umn.edu/about/mellon/UMN_Multi-dimensional_Framework_Final_Report.pdf)>.

---. The University of Minnesota Libraries: Changing the Paradigm. Minneapolis: University of Minnesota, 2008b. <<https://wiki.lib.umn.edu/wupl/Staff.HomePage/2009%20ACRL%20Award%20Application.pdf>>.

Walker, J. D., and Linda Jorn. Net Generation Students at the University of Minnesota, Twin Cities.

Minneapolis: University of Minnesota Office of Information Technology, 2007.

<<http://dmc.umn.edu/surveys/students/ssreport07.pdf>>.

Ward, Jennifer L., Steve Shadel, and Pam Mofjeld. "WorldCat Local at the University of Washington Libraries."

Library Technology Reports 44.6 (2008): 4-41.



## APPENDIX A: CHARTER

### DISCOVERABILITY EXPLORATORY GROUP CHARTER (28 October 2008)

#### Scope

This subgroup is focused on the library user and will recommend ways to make relevant resources more visible and easier to find, particularly within the user's workflow. In developing its recommendations, the subgroup should consider use cases for a range of users, research needs, and resource or collection types. It should take into account not only library-owned collections, but potentially relevant resources from any available source (e.g., electronic books, special collections, grant resources).

#### Sponsor

Web Services Steering Committee (Janice Jaguszewski and Shane Nackerud, primary contacts)

#### Membership

Heather Hessel (co-chair)

Cody Hanson (co-chair)

John Barneson

Deborah Boudewyns

Jan Fransen

Lara Friedman-Shedlov

Martha Hardy

Chris Rose

Barb Stelmasik

Stacie Traill

Consult as appropriate: Jim Stemper

#### Phase 1 (2 months)

##### Objectives

- Principles and guidelines. Develop a set of principles related to discovery to help guide the Libraries' strategic decisions about the selection, development, and support of relevant tools and services.
- Data Analysis. Analyze evidence from sources such as SFX logs, referrer logs, proxy server logs, affinity strings, Google search appliance statistics, Interlibrary Loan statistics, and MetaLib search logs to draw conclusions about patterns of user behavior. Usage patterns to look at may include but are not limited to:
  - Find out where people are initiating their interaction with desired content.
  - Identify patterns of resource usage by population
  - Identify resources which are not being found or are found but not used
  - Identify which library websites are getting the most use
  - Compare on- vs off-campus usage of various resources
- Trends. Identify trends in user behavior.

## **Deliverables**

The team should provide regular reports to WSSC addressing each of the objectives outlined above and should maintain a wiki to keep Libraries staff informed of the group's progress.

A written report with recommendations should follow two months after the group begins its work. The final report should include:

- A set of guiding principles related to user discovery.
- A summary of guiding principles for user discovery that are applied at several other institutions.
- A summary of interviews with appropriate University partners and key stakeholders.
- A summary of relevant reports on campus and beyond.
- A summary of baseline data on library usage patterns based on data analysis and recommendations for action.
- Recommendations concerning how statistical analysis might be done on a regular basis.

## **Phase 2 (3 months)**

### **Objectives**

- Based on evidence about user behavior and preferences, identify and prioritize the types of discovery activity that should be supported. Consider the context of these activities, e.g., how far is discovery happening at the global (e.g. Google, WorldCat) rather than the local level?
- Evaluate the potential of specific tools and services for supporting key discovery needs.
  - Assess how well next-generation discovery systems such as WorldCat Local, BlackLight, VUfind, and Primo fit the discovery needs of UMN users, taking into account these systems' differing design objectives.
  - Analyze content (subject guides, gateways, the catalog, etc.) for opportunities to elevate high priority resources in search engine results via Search Engine Optimization (SEO) and provide recommendations for improvement.
  - Explore services that are customized to particular needs, e.g., recommendations and pre-emptive filtering.
- Identify developing trends and key opportunities for collaboration (e.g., recommender system among CIC institutions)

### **Deliverables**

The team should provide regular reports to WSSC addressing each of the objectives outlined above. A written report with recommendations should follow three months after phase 2 begins. Other specific deliverables:

- Summarize interviews with appropriate University partners and key stakeholders.
- Summarize relevant reports on campus and beyond.
- Provide a pro/con summary of available systems, and develop a set of criteria for selecting a system or array of systems that will meet evolving user needs for resource discovery. Consider whether specific needs would best be addressed through a system developed locally. Summarize next-gen system choices at 3-4 institutions.
- Recommend a timeline for determining the Libraries' future direction with regard to such systems.
- Make recommendations regarding what we should do to facilitate discovery that would be most impactful and should be prioritized. Recommendations that have been determined to be a higher priority should

rate highly using the WSSC Project Criteria -<https://wiki.lib.umn.edu/WebServices/ProjectCriteria>. Provide reasonable timelines for the prioritized items.

### **Possible partners**

Sandra Ecklein (OIT – portal usage)

Curt Squires (U of M Google Search Appliance)

Scott Barnard (Emerging Technologies, Digital Media Center)

### **Time Commitment**

The committee assignment is for five months; on average committee work will require 6 hours per week. Time commitments will vary according to project demands.

### **Stakeholders**

All Libraries staff and users.

### **Budget**

Budget requests developed as needed.

### **Responsibilities**

#### Committee Sponsors' Accountabilities (Web Services Steering Committee)

The sponsors are the executive champions for the program and have a responsibility to:

- Lay out expected outcomes, objectives, and high level deliverables
- Allocate resources – time, money, and assistance as needed
- Assist in project planning and control
- Monitor committee progress
- Identify and provide data and information sources and assure full participation and support from relevant units, both external and internal
- Provide reinforcement and feedback
- Work to resolve cross-functional issues and to remove any organizational barriers
- Provide reinforcement and feedback
- Approve recommendations
- Champion approved team recommendations

#### Co-chair Accountabilities

The co-chairs play an important role in coordinating committee meetings and activities and have a responsibility to:

- Clarify committee purpose and objectives and program scope and deliverables
- Manage committee meetings (schedule and communicate meeting logistics; assure appropriate meeting planning and follow-up; chair meetings)
- Assure preparation of action plans and the completion of action items and project deliverables
- Track and report committee progress
- Participate in quarterly review meetings with primary contact

### Member Accountabilities

Each group member has a responsibility to:

- Attend meetings, actively participate in meetings, and promptly complete assignments
- Collect, prepare, and analyze relevant data
- Participate in framing up recommendations and prepare project deliverables
- Represent stakeholders and library units as appropriate; solicit input and communicate information and feedback
- Work to ensure group effectiveness and adherence to established ground rules
- Support group decisions and actions

### **Suggested Ground Rules:**

- Start on time and work to stay on schedule
- Monitor digressions and refocus the discussion
- Make use of straw polling on important decisions
- Handle disagreements; poll other co-workers on controversial decisions and, in extreme cases, report a majority and minority opinion in the final report
- Treat other members with courtesy and respect regardless of opinions
- Take issues that involve a few people or further investigation off-line

## APPENDIX B: DISCOVERABILITY PRINCIPLES FROM PEER INSTITUTIONS

### *University of Wisconsin, Madison*

In June, 2008, a subgroup at the University of Wisconsin, Madison, Libraries issued the “Resource Discovery Exploratory Task Force Final Report” in response to a charge for “developing a vision for information resource discovery in the Libraries to support teaching, learning and research at UW-Madison.” p.6 The following criteria regarding discovery has been pulled from that report.

"The conclusions in this report are tuned toward the library catalog because the library has other projects underway for improving and implementing access to non-catalog data." However, the report goes on to state, "It is expected that these recommendations will apply to additional types of data beyond the library catalog."

Discovery should be aligned with user behaviors and expectations. This is achieved by supporting a culture of assessment in order to understand their users and to be where the users are. Discovery must be “fast, smart, engaging”(Dentinger et al.) . Implicit in this report was the suggestion that whatever choices UW Madison makes must be scalable.

Also in regards to specific system requirements, the new discovery environment must:

1. Decouple the interface from the ILS so that it is sleek, lean, and enabled for rapid change.
2. Maintain complete control over the discovery interface, data, and index. Nothing should be unchangeable.
3. Emphasize simplicity in the interface. As Lorcan Dempsey noted: "'simple search' but supported by smart results and rich browse" (single search box, single sign on, clean layout) {{35 Dempsey,Lorcan}}.
4. Include sophisticated search and result functionalities (faceted browsing and/or topical clustering, natural language, obvious relevancy ranking, searching within results, clarity via FRBRization).
5. Seamlessly integrate and deliver UW collections and resources at the campus and at the system level (library catalogs, library websites, digital collections, museums, archives).
6. Adapt to user behaviors and expectations (personalization, recommendations, "did you mean?" functionality, internationalization).
7. Encourage personalization and customization of the discovery environment in MyUW and course management systems, including Learn@UW and Moodle.
8. Deliver library search functionality, links and services where our users work and play, including off-campus resources (Amazon, iGoogle, Facebook, WorldCat).

9. Compare well in design and user experience to popular Internet destinations. Resource discovery in the libraries must become Fast, Smart, and Engaging to compete in the current and future information marketplace.

10. Be staffed for excellence and continuous change (developers, graphic and interaction designers, and public services staff). This includes collaboration and leadership within the Open Source community.

(Dentinger et al.), p. 3-4

### *University of Washington*

The University of Washington Libraries “Vision 2010 Strategic Plan” includes a goal to “meet user needs by providing access to resources and services at the point of need in the users' environments.” In regards to specific requirements, they identified the need for a tool that:

- provides a user-centric discovery interface
- provides access to resources beyond the physical walls of the library by representing institution, group, and WorldCat holdings
- incorporates library work flows without confusing the user
- syndicates content to other user environments

(Ward, Shadel and Mofjeld 4-41)

Steve Shadle expands and elaborates in his article, “The local catalog is dead! Long live the local catalog!” According to Shadle, the UW Libraries was interested in WorldCat Local as a primary discovery tool because it was:

- “intended to integrate discovery across a number of content silos...”
- “intended to integrate separate delivery services...”
- “user-centered”, in that research shows that “most users prefer simple/direct search mechanisms”
- informed by usability testing
- a product from OCLC, who they felt was “in a better position than any single library to work with national and international partners (Google, Amazon, and Microsoft) to support syndication and to move library content and services to the network level.”

(Shadle 85-87)

### *Minnesota Oberlin Group of Libraries (MnObe)*

In 2007, staff from the Minnesota Oberlin Group of Libraries (MnObe) met to create a set of guiding principles for future discovery tools. The five member libraries of the MnObe Group (<http://www.macalester.edu/mnobe/>) are: Gould Library, Carleton College; Folke Bernadotte Memorial Library, Gustavus Adolphus College; DeWitt Wallace Library, Macalester College; Alcuin/Clemens Libraries, College of St. Benedict /St. John’s University; and, St. Olaf College Libraries, St. Olaf College. The result of their

work can be found in a report to the MnObe directors, called "MnObe and the Future of the Catalog". The three Key Principles they defined are as follows:

#### 1. Flexible Data Feeding Flexible Tools

- MnObe librarians envision that users should be able to access data from the libraries' ILS in the same discovery environment that they can access data from other sources.
- "a next generation discovery tool should be based on flexible, open, standards-based data."
- "We recommend providing [a] discovery tool which is clearly identified as such and which provides an Application Programming interface (API) so that users can assimilate library data to their own "best" discovery tools."

#### 2. Intellectual Connectivity between Resources

- "any discovery tool that the library does provide should guide users through the intellectual connections that exist between resources."
- Such features include: recommender services, faceted searching, citation linking, and FRBRization.

#### 3. Interactivity

- "library-provided tools should allow flexible and robust interactivity between users and the system and between the system and other systems."
- Interactivity features include: tagging, reviewing, ratings, saving, sending, extracting data (records), and creating sets.
- Should also be able to feed information into a wide variety of non-library tools such as social networking applications (e.g., Facebook), communications applications (Twitter) or desktop applications

### *University of California*

The Bibliographic Services Task Force for the University of California Libraries released a final report in December 2005 entitled "In Rethinking How We Provide Bibliographic Services for the University of California" (University of California Libraries Bibliographic Services Task Force). Although this report is now several years old, the guidelines and criteria included in it are still relevant and useful today.

Specific recommendations include:

- Provide users with direct access to item
- Provide recommender features
- Support customization/personalization

- Offer alternative actions for failed or suspect searches
- Offer better navigation of large sets of search results
- Deliver bibliographic services where the users are
- Provide relevance ranking and leverage full-text
- Provide better searching for non-Roman materials

The group also identified a number of Design Principles for technology; the following list is a subset, related to the discovery process: (University of California Libraries Bibliographic Services Task Force 40)

- Enrich Bibliographic data with helpful, related information, linked to the broader information universe (e.g., TOC, citations, etc.).
- Support user supplied metadata. Design for user participation (p.43)
- Mine existing data and metadata to generate additional metadata
- Reduce the clickstream whenever possible, facilitate self-service, etc.
- "Discovery alone is not enough, we must provide the full cycle of Discover-Locate-Request-Deliver. An interaction is not successful until the user has access to the resources itself."
- "For electronic resources, provide immediate links to the content... For print resources, provide immediate information about where to get an item..."
- "Never send a user to a dead end..."
- Adopt perpetual test model
- Trust users as co-developers, i.e. release new features on a monthly, weekly, or even daily basis, and remove features that are not being adopted by users
- Support continuous assessment & improvement

### *University of Michigan*

In December 2007, the WorldCat Local Task Force at the University of Michigan Library issued a report to Executive Council on the advisability and implications of implementing WorldCat Local. Based on information found within this report, we have gleaned the following objectives. Looking at what the task force viewed as inadequacies of WorldCat Local, one could deduce that they need:

- a comprehensive representation of the collection (including vendor records, on order records,
- the ability to build custom tools (APIs)
- ability to customize and brand the interface
- support of Z39.50 and NCIP interactions
- support of complex and known item searching as essential for research



Looking at what the task force viewed as strengths, one could deduce that they would like:

- a networked, shared interface needing little to no development, maintenance, or support for the interface,
- ability for users to easily find nearby holdings as well as U Mich holdings
- desirable Web 2.0 features, such as:
  - single search box in simple search;
  - faceted browsing;
  - easy integration with social bookmarking sites; and
  - content enrichment features.

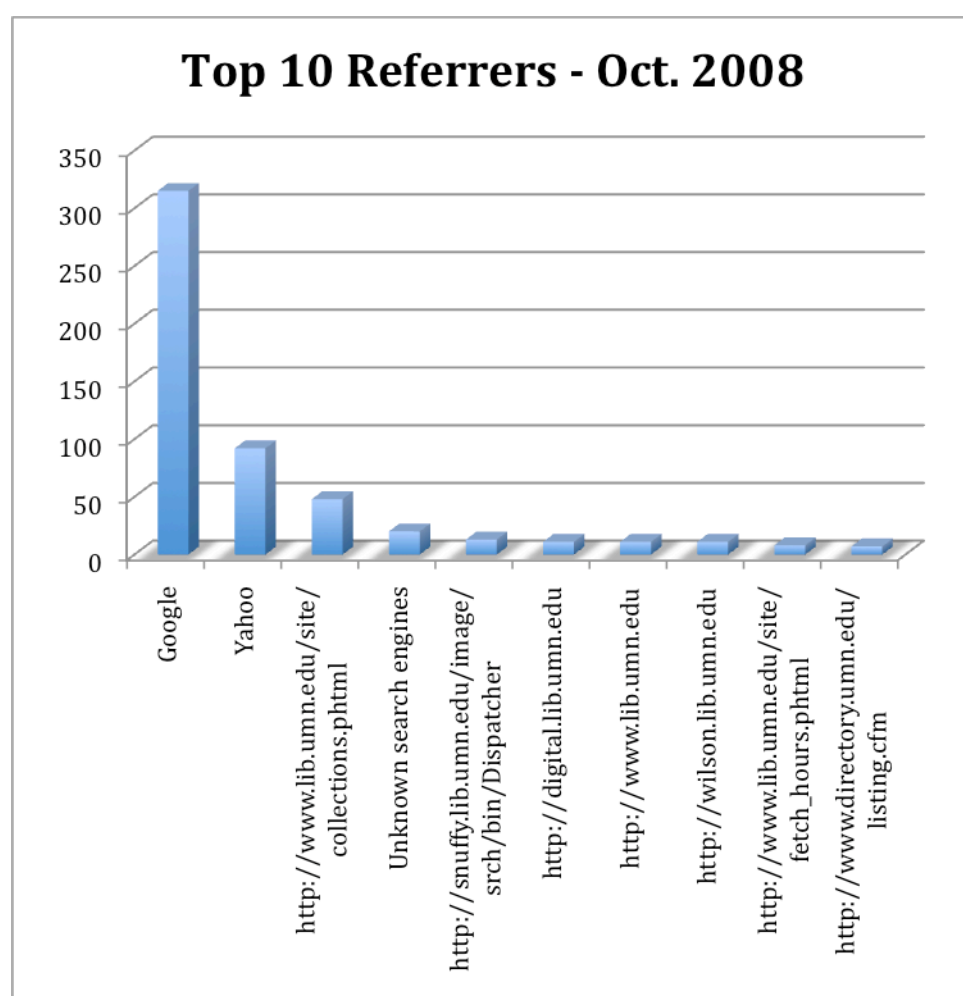
## APPENDIX C: SUBDOMAIN STATISTICS

### *ames.lib.umn.edu Website Statistics*

This set of statistics covers pages served on the ames.lib.umn.edu subdomain. These data were gathered from the Libraries' AWStats program, which ingests and parses Apache web server logs.

#### **Referrers**

This chart shows referring domains for visits to all pages on the ames.lib.umn.edu subdomain in October, 2008. In this chart I have combined search engine and web page referrals for those visits that log a referrer. Notable here is that Google drives more than twice the traffic to the ames.lib subdomain than the top ten non-search engine referrers combined.

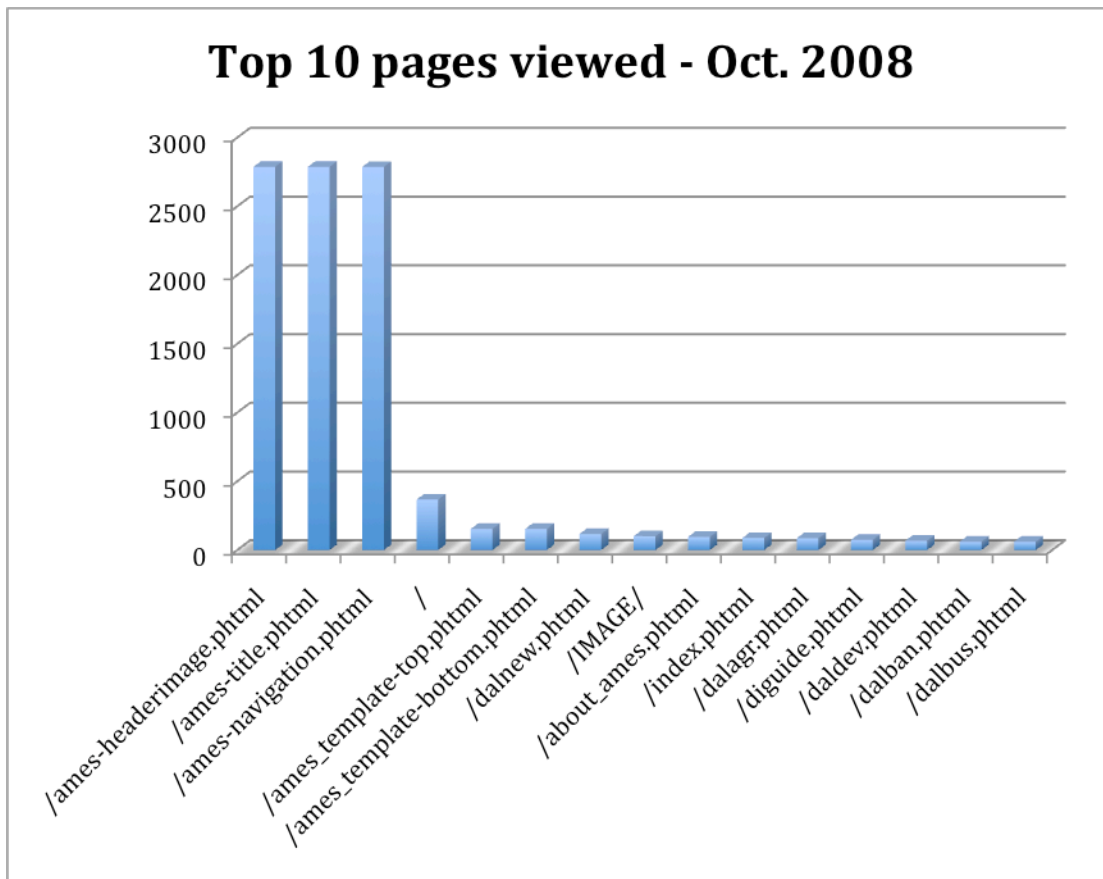


## Page views

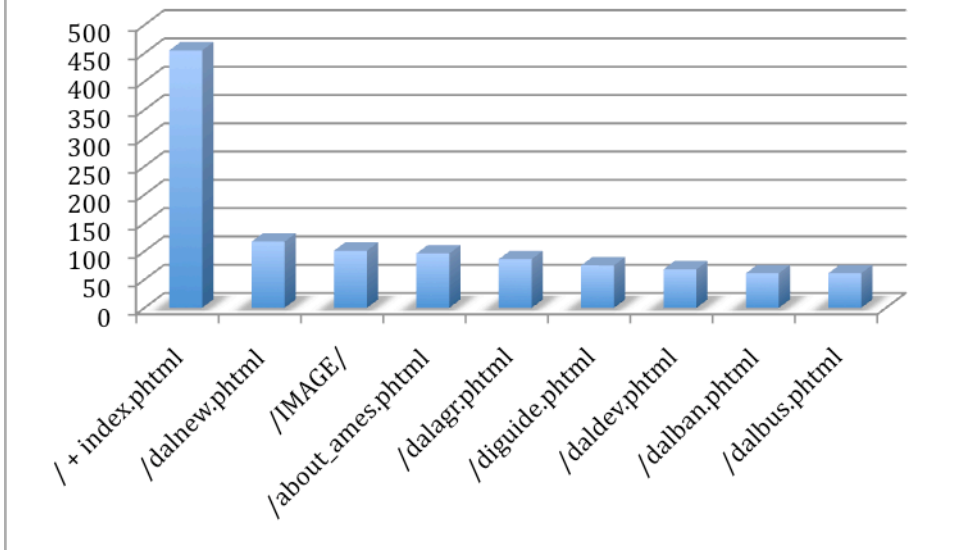
The following charts show the ten most viewed pages on the ames.lib.umn.edu subdomain. The first chart is heavily skewed by the presence of layout template files that are likely served on many, if not all, of the pages on the subdomain. As a result, a visit is logged for each of these files when any page on the subdomain is loaded. The second chart shows an adjusted view of these statistics, eliminating template files and combining the statistics for <http://ames.lib.umn.edu/> and <http://ames.lib.umn.edu/index.phtml>, which are the same page.

Notable in these statistics are the abovementioned idiosyncrasies of our template that can skew statistics in unforeseen ways. For instance, the default statistics for hour of day and day of week in AWStats rely on page views rather than unique visits. For the ames.lib.umn.edu subdomain and others with similar template statistics we ought to attempt to adjust these statistics if we compare to another subdomain.

Also notable is the lack of penetration by site visitors beyond the site's home page. Visitors may move from the ames.lib.umn.edu subdomain to other of our sites or applications, but most internal pages are viewed three times per day or fewer.

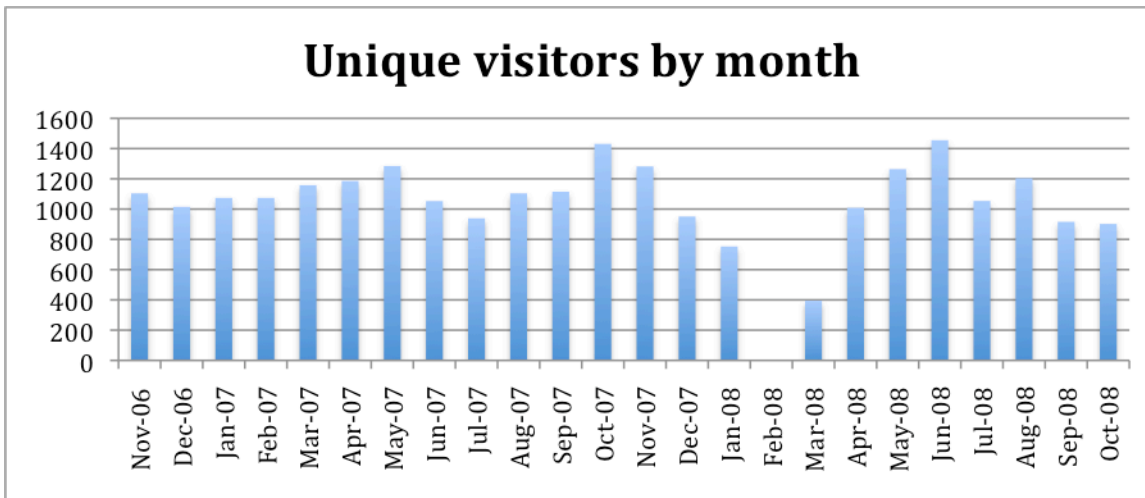


## Top 10 pages viewed - Oct. 2008 Adjusted



### Monthly visitors

This chart shows unique visitors per month from November 2006 to October 2008. The figures for January, February, and March, 2008 are incomplete. We have no statistics for the period of January 23 to March 20. Traffic to the ames.lib.umn.edu subdomain appears to be steady with little seasonal fluctuation.



### Conclusions

Generally speaking, traffic to the ames.lib.umn.edu is almost entirely search-engine driven, and directs users to the Ames Library's home page. Visits to pages within the ames.lib.umn.edu subdomain are so relatively

infrequent that staff usage could account for many or all of them. At best over the past twelve months this subdomain has seen less than three percent the traffic that the www.lib domain has seen.

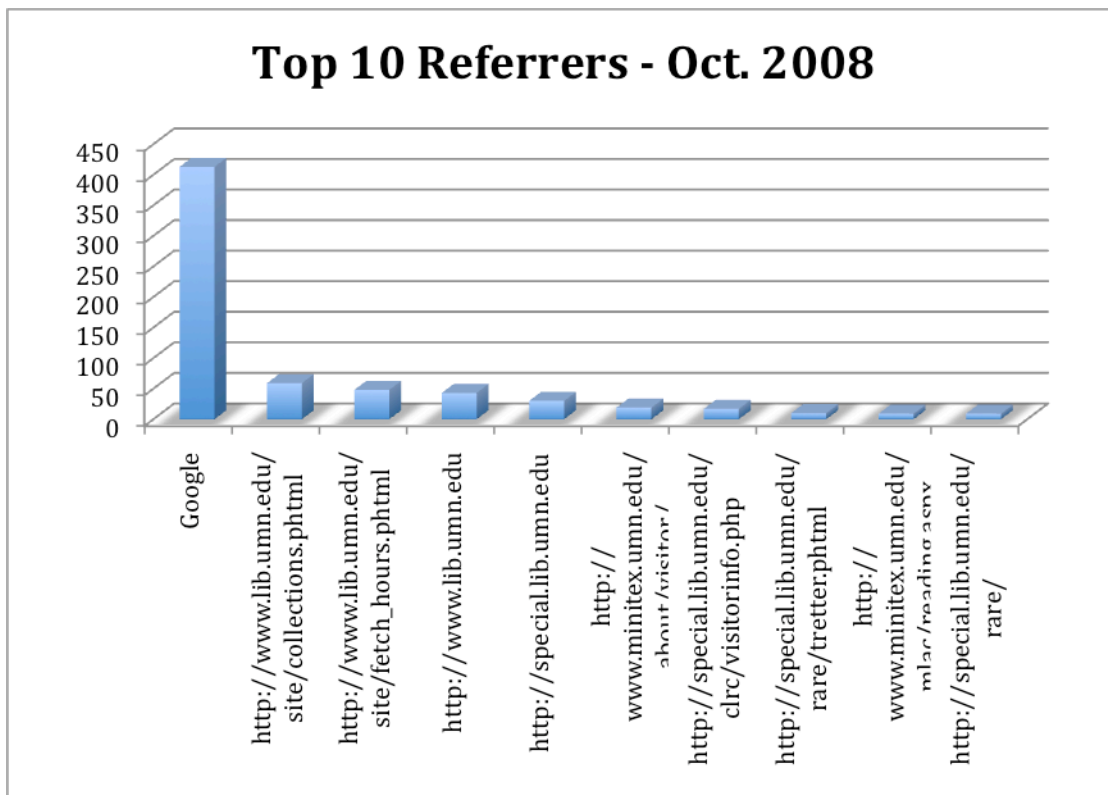
Future analysis of statistics from this subdomain should take care to address the skewing effect of the layout template files, either by excluding them from the logging process or removing them after the fact.

### *andersen.lib.umn.edu Website Statistics*

This set of statistics covers pages served on the andersen.lib.umn.edu subdomain, which primarily contains pages related to the Elmer L. Andersen library building, which houses many of the Libraries' archives and special collections. Most archives and special collections have their own websites which are not contained in this subdomain. These data were gathered from the Libraries' AWStats program, which ingests and parses Apache web server logs.

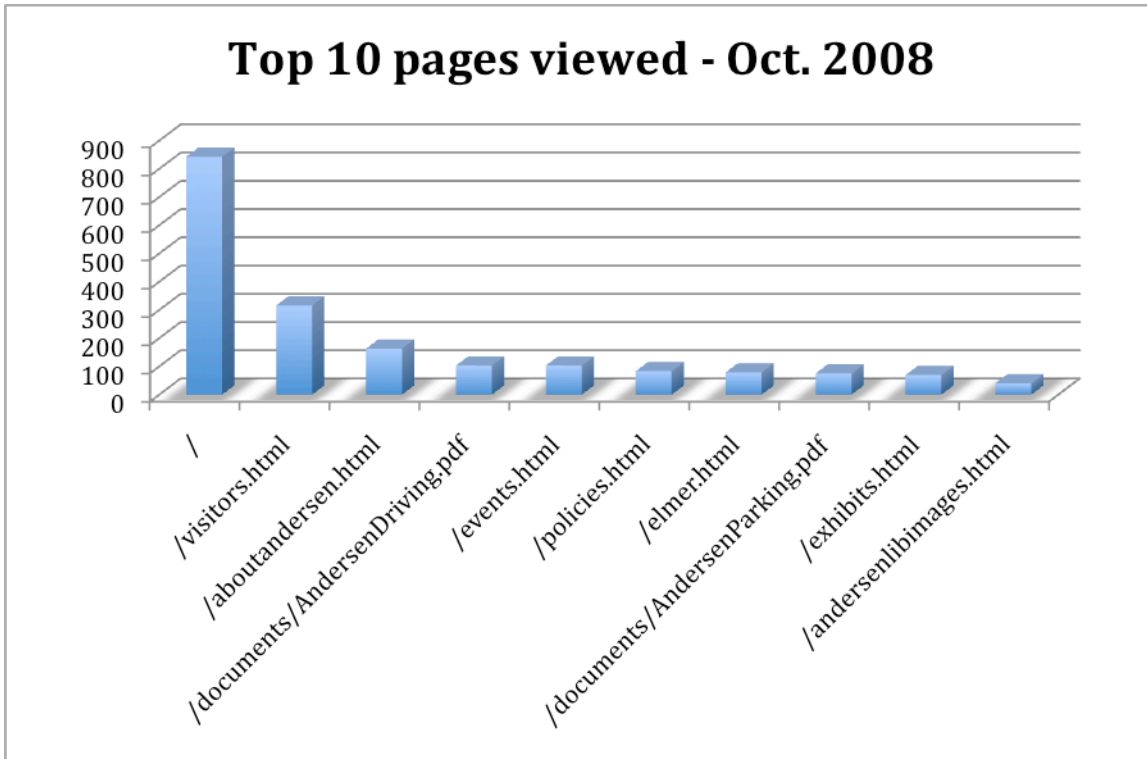
#### **Referrers**

This chart shows referring domains for visits to all pages on the ames.lib.umn.edu subdomain in October, 2008. In this chart I have combined search engine and web page referrals for those visits that log a referrer. As with many of our other subdomains, Google is the single most prominent referrer of traffic to this site.



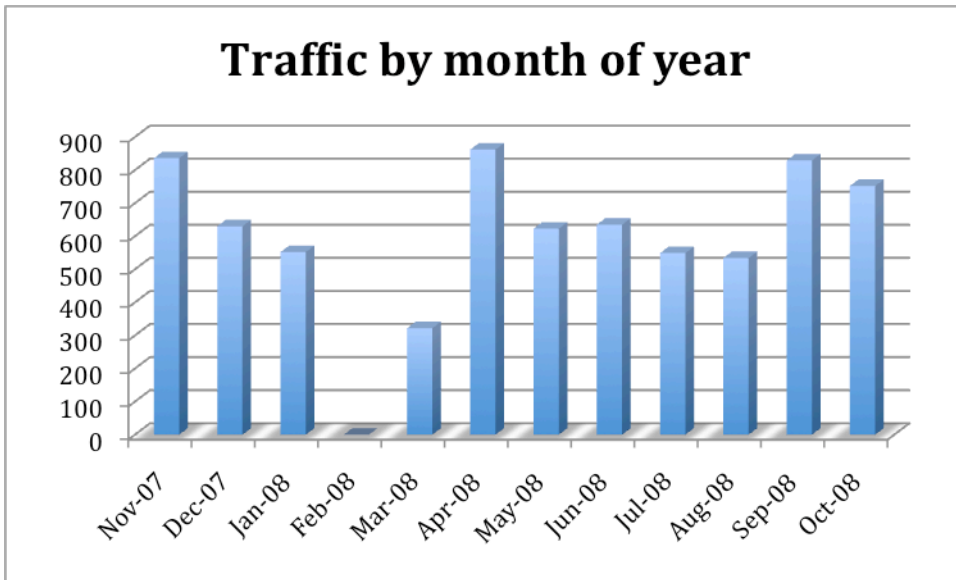
### Page views

The following chart shows the ten most viewed pages on the andersen.lib.umn.edu subdomain. As with many of our subdomains, the site's home page receives many more views than any other page on the site. Notable here is that the overall volume of traffic on this site is very low.



### Monthly visitors

This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. This site follows a similar pattern to other subdomains in that traffic mirrors the rhythm of the school year. April, 2008 was the busiest month of those for which we have data. It would be interesting to investigate if there was an event or promotion which drove traffic during that month.



**Conclusions**

Traffic to the andersen.lib.umn.edu subdomain is largely search-engine driven, and generally restricted to the site’s home page. Visits to pages within this site are infrequent, and may be accounted for by staff usage. It may be worth investigating the potential for folding the content from this site into the Archives and Special Collections site (special.lib.umn.edu).

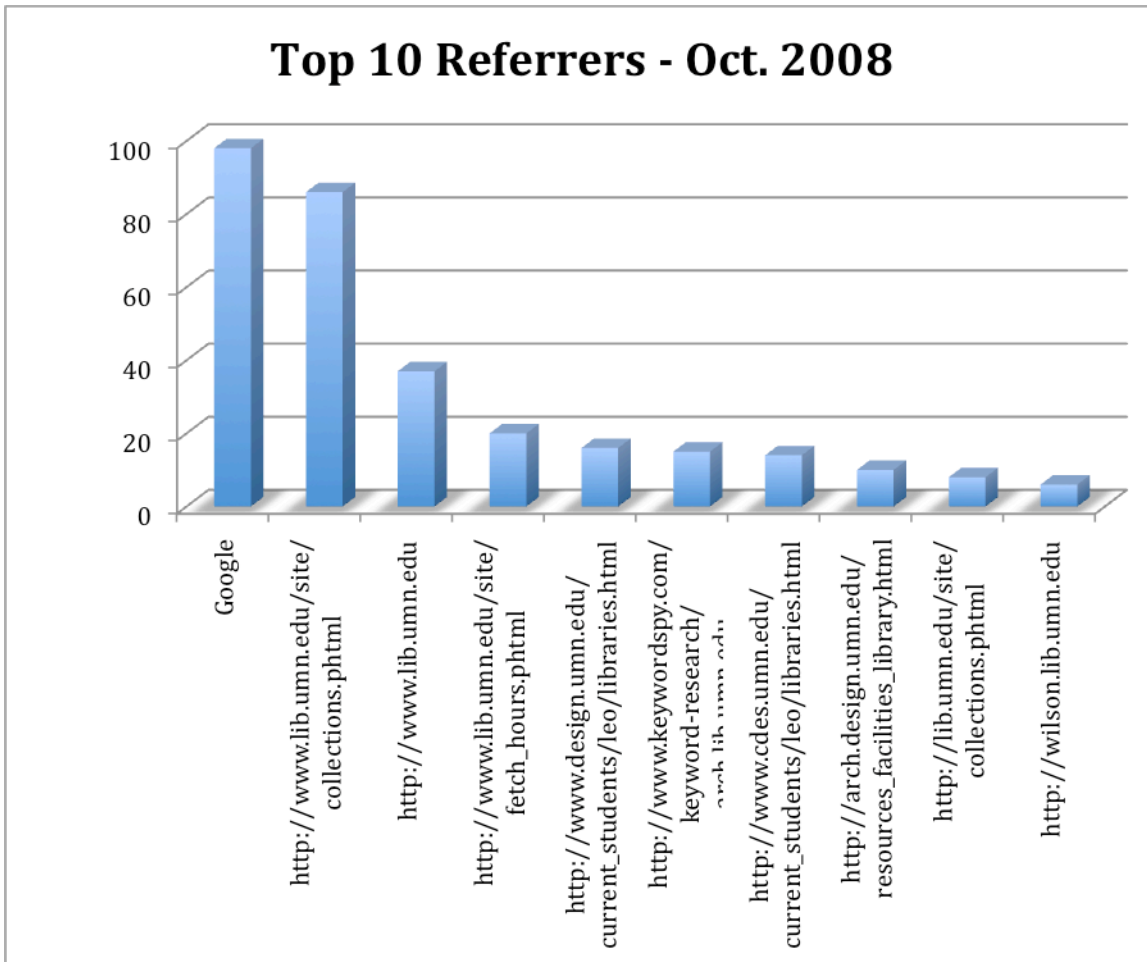
*arch.lib.umn.edu Website Statistics*

This report covers statistics from the arch.lib.umn.edu subdomain, containing pages related to the Architecture and Landscape Architecture Library.

**Referrers**

Like many of our subdomains, the top referrer to pages on the arch.lib.umn.edu subdomain is Google. However, the collections.phtml page on the Libraries’ main website, which contains a listing of all branches and collections, ranks a close second, followed by the Libraries’ home page and hours page. It is gratifying to find that a page on the College of Design’s site ranks in the top 10 pages referring users to this subdomain, as this library is intended primarily to serve the students, staff, and faculty of the college. Unfortunately, referral of fewer than twenty users in the month of October, 2008 was sufficient to put this page fifth among referring pages.

## Top 10 Referrers - Oct. 2008



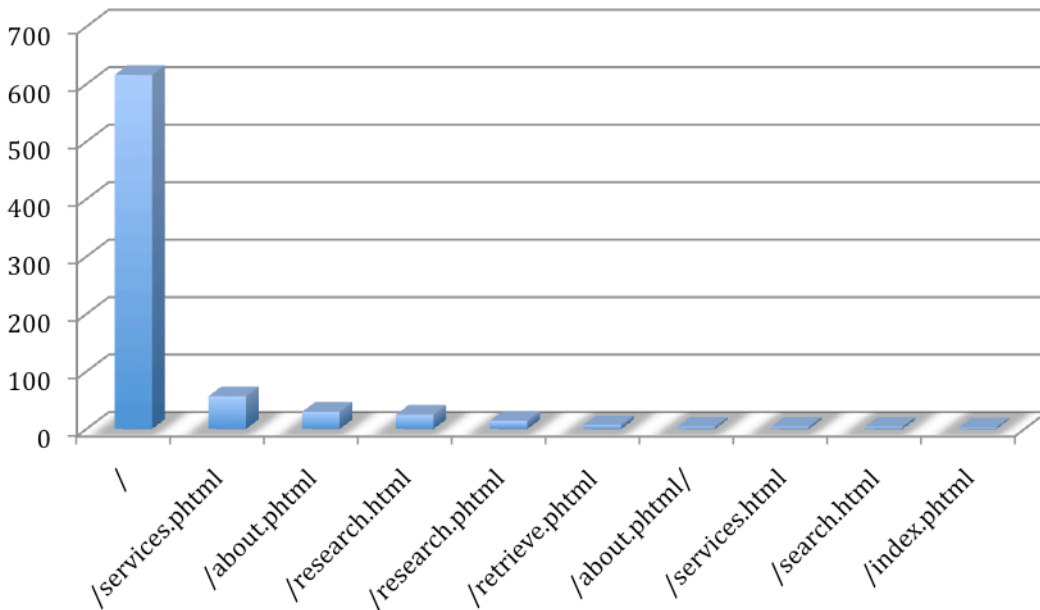
### Page views

The most frequently viewed pages on the Architecture and Landscape Architecture Library website show a pattern common among many of our subdomains. The home page was viewed 616 times in October, 2008, and the next most commonly viewed page registered 67 views during the same period, slightly more than one-tenth the traffic.

A visit to <http://arch.lib.umn.edu> shows that there is little need for navigation deep into this site. There are only two pages on this subdomain linked from the home page, <http://arch.lib.umn.edu/services.phtml>, which lists services and policies, and <http://arch.lib.umn.edu/about.phtml>, which describes the library and its staff, replicating much of the same information as the home page. The chart below shows at least five additional pages that are live on this domain but not linked from any page on the domain and which were viewed during October, 2008. These pages appear to be remaining from earlier iterations of this site.



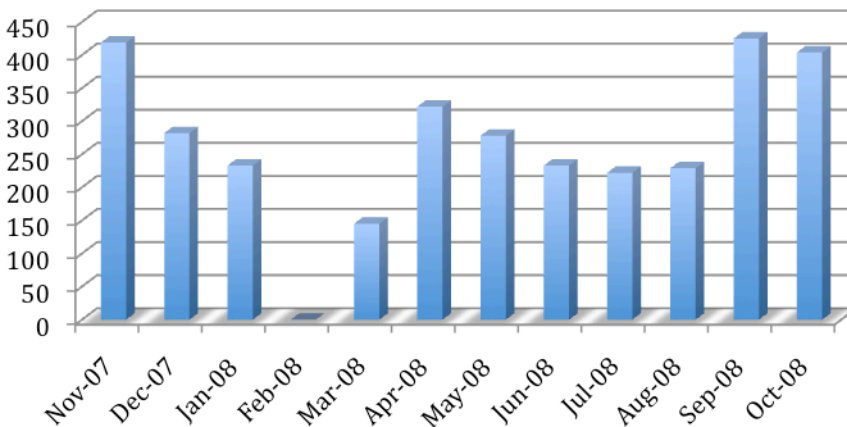
## Top 10 pages viewed - Oct. 2008



### Monthly visitors

Traffic from the past year seems to follow the same pattern as many of our other subdomains, peaking during the fall. Like some other of our subdomains, data from February and March, 2008 appears to be missing or corrupted.

## Traffic by month of year



## **Conclusion**

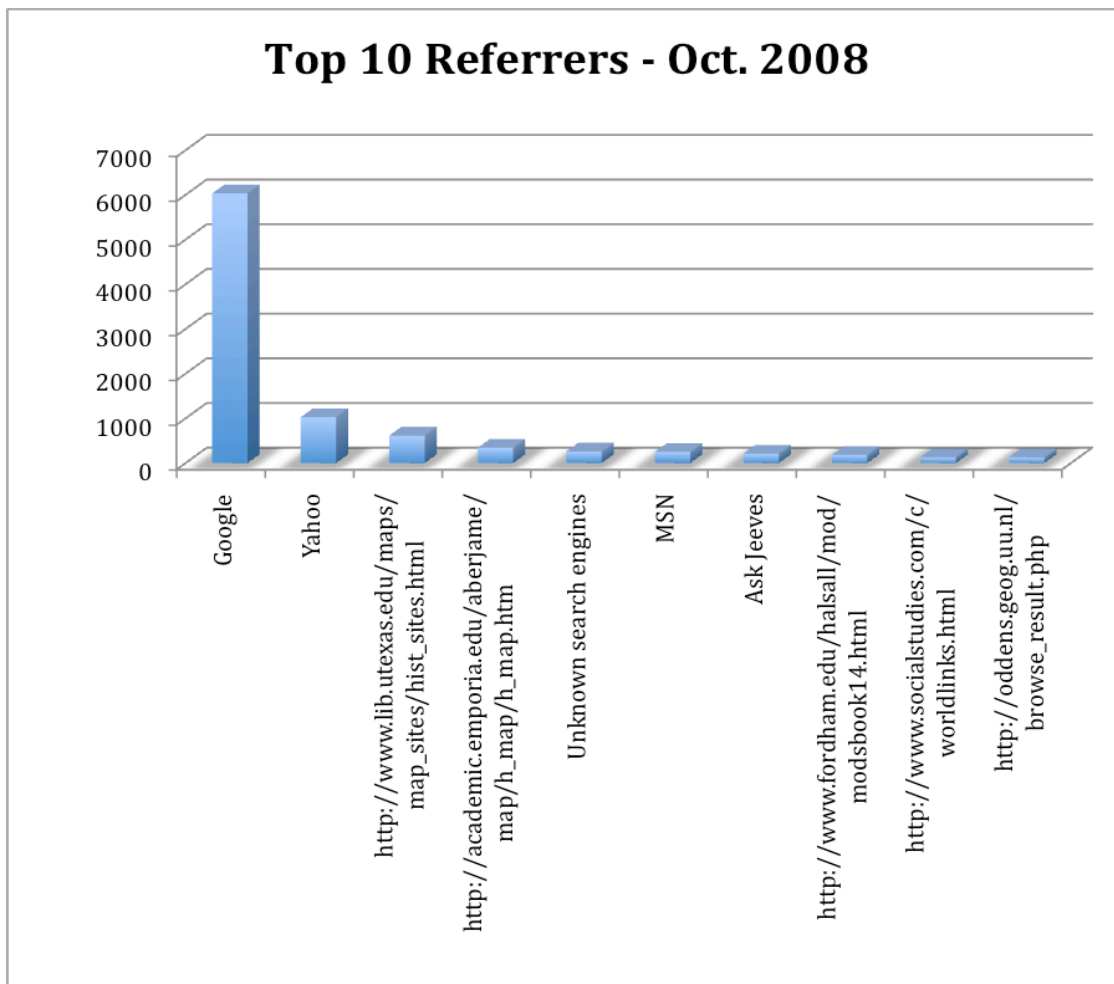
Traffic on this subdomain is very light, and were we able to filter the visits to the library's home page from public terminals in the library with <http://arch.lib.umn.edu> set as the home page, we may find that actual traffic is even less than it appears in these data. I would recommend this step. I would also recommend that those pages from earlier versions of the website be redirected to their current equivalents in order to ease future transitions and to ensure that users are presented with the most current information.

## ***bell.lib.umn.edu Website Statistics***

This report covers statistics from the [bell.lib.umn.edu](http://bell.lib.umn.edu) subdomain, containing pages related to the James Ford Bell Library. These pages have information related to the physical collection in the library, but many pages on this domain contain digitized images and documents from the collection.

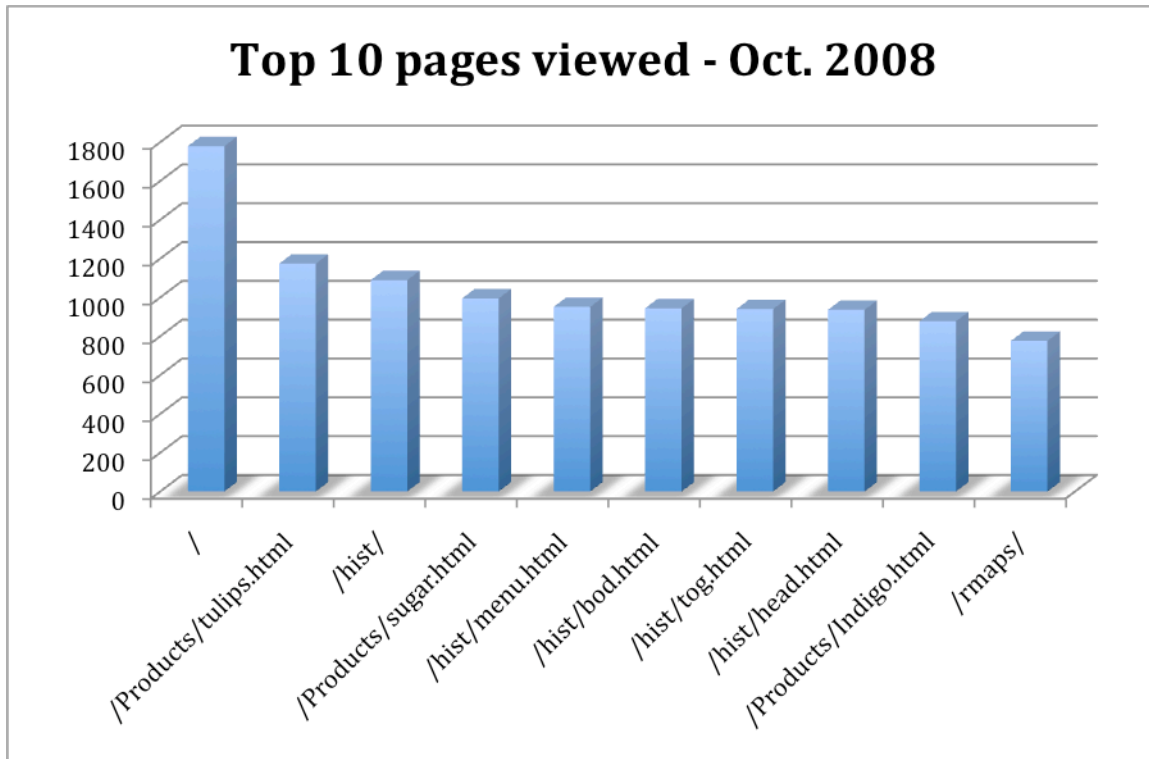
## **Referrers**

The chart below combines page views referred from search engines and from pages on other domains and subdomains during the month of October, 2008. As with many of the University Libraries' subdomains, a great deal of the traffic to pages on the [bell.lib.umn.edu](http://bell.lib.umn.edu) subdomain are referred from search engines. Unlike other of the Libraries' sites, however, none of the top ten referrers are pages on the [umn.edu](http://umn.edu) domain.



## Page views

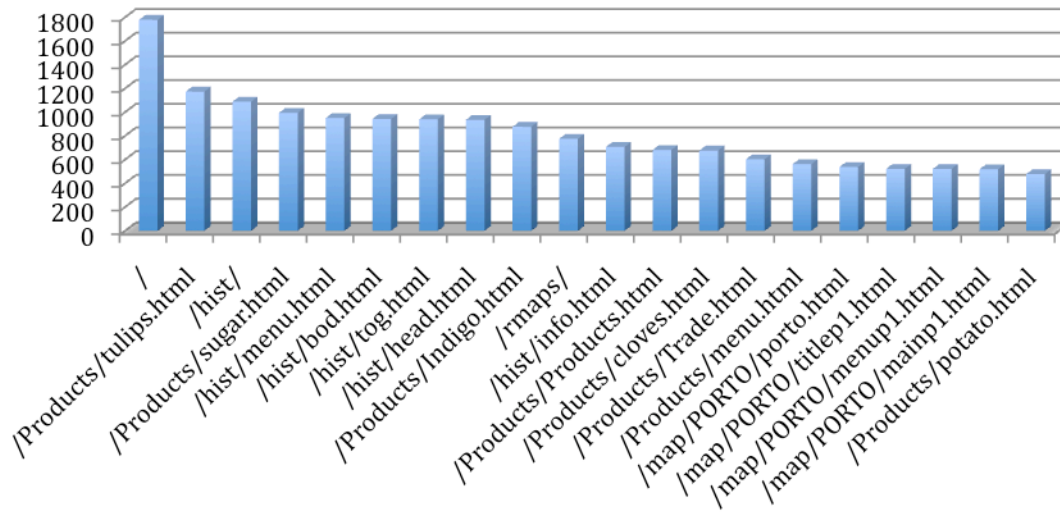
The discrepancy in referrers between the bell.lib.umn.edu subdomain and others is explained in part by studying the most frequently viewed pages on the site. Like our other subdomains, the bell.lib.umn.edu home page sees the most traffic on the site. However, unlike other subdomains, there are a significant number of pages within the site that have a great number of page views. The chart below shows the ten pages with the most views on the site.



The same list of the top pages on our other subdomains typically shows a precipitous drop-off in views after the top two or three pages. The Bell site, however, has even the tenth page in the list seeing hundreds of page views per month. In fact, this trend continues, as shown in the graph below of the top 20 pages on the site in October, 2008.

The pages that show up in these graphs are largely found in three directories on the domain, “/Products/”, “/hist/”, and “/map/PORTO/”. These directories correspond to three collections of historical materials available for viewing on the Bell site: a collection of information on Trade Products in Early Modern History (<http://bell.lib.umn.edu/Products/Products.html>), a collection of images of Historical Maps (<http://bell.lib.umn.edu/hist/>), and a collection of images of Portolan Charts (<http://bell.lib.umn.edu/map/PORTO/porto.html>).

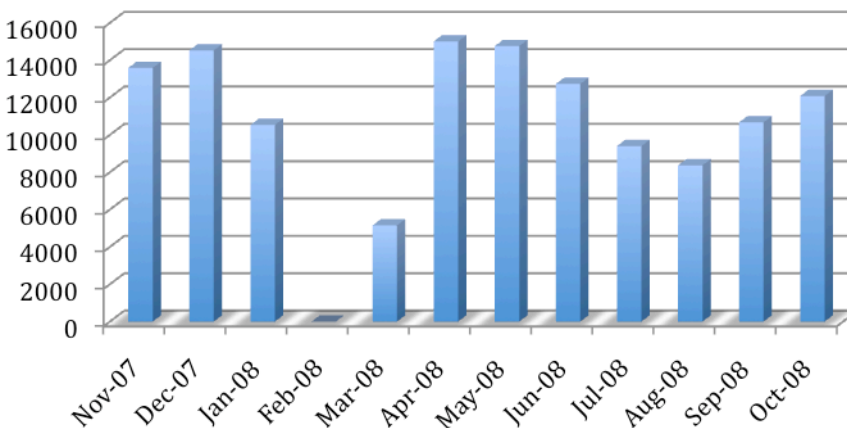
## Top 20 pages viewed - Oct. 2008



### Monthly visitors

Perhaps surprisingly, the traffic on the Bell site seems to follow the same school year pattern as many of our other subdomains. Perhaps this indicates that though many users of the site are not coming through the University Libraries that they are academic users nonetheless. As with some of our other subdomains, data for February and part of March, 2008 are missing.

## Traffic by month of year



## **Conclusion**

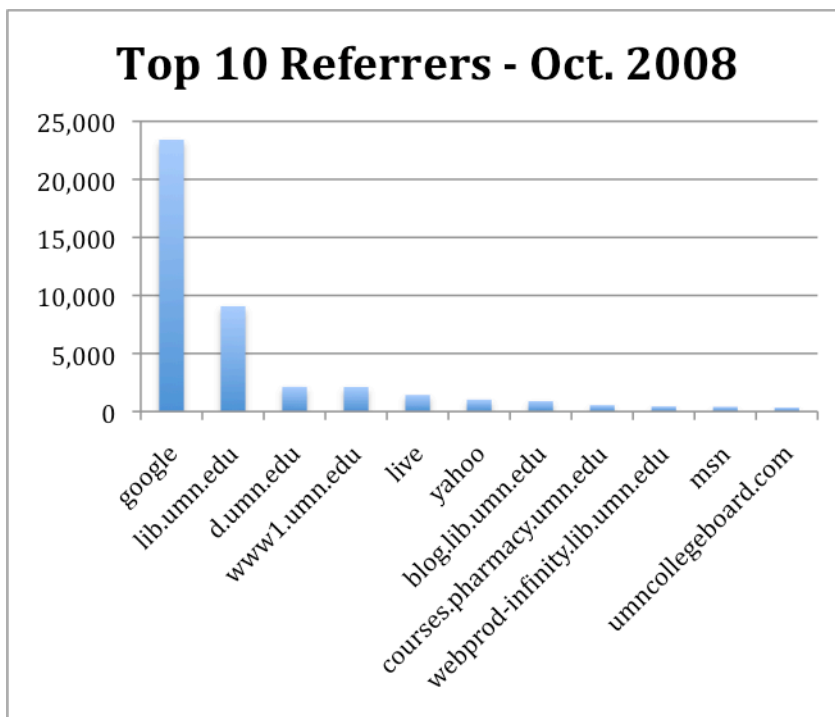
The James Ford Bell Library subdomain is a compelling example of the power of quality primary and secondary research materials to drive traffic to our websites. Several hundred words on early modern trade in tulips drive more traffic to this site in a month than some of our branch library subdomains see across all of their pages. The attention these pages see from external viewers stands in contrast to the apparent lack of attention that has been given to them internally. The traffic on this subdomain may warrant an update and revisions to the content and design of these collections.

## ***biomed.lib.umn.edu Website Statistics***

This set of statistics covers pages served on the [biomed.lib.umn.edu](http://biomed.lib.umn.edu) subdomain, the main site for the University's Health Science Libraries. The data here is from October 2008 and is drawn from the Health Science Libraries' web statistics as gathered using Google Analytics. This site and [vetmed.lib.umn.edu](http://vetmed.lib.umn.edu) are the only of the Libraries' sites that currently use Google Analytics.

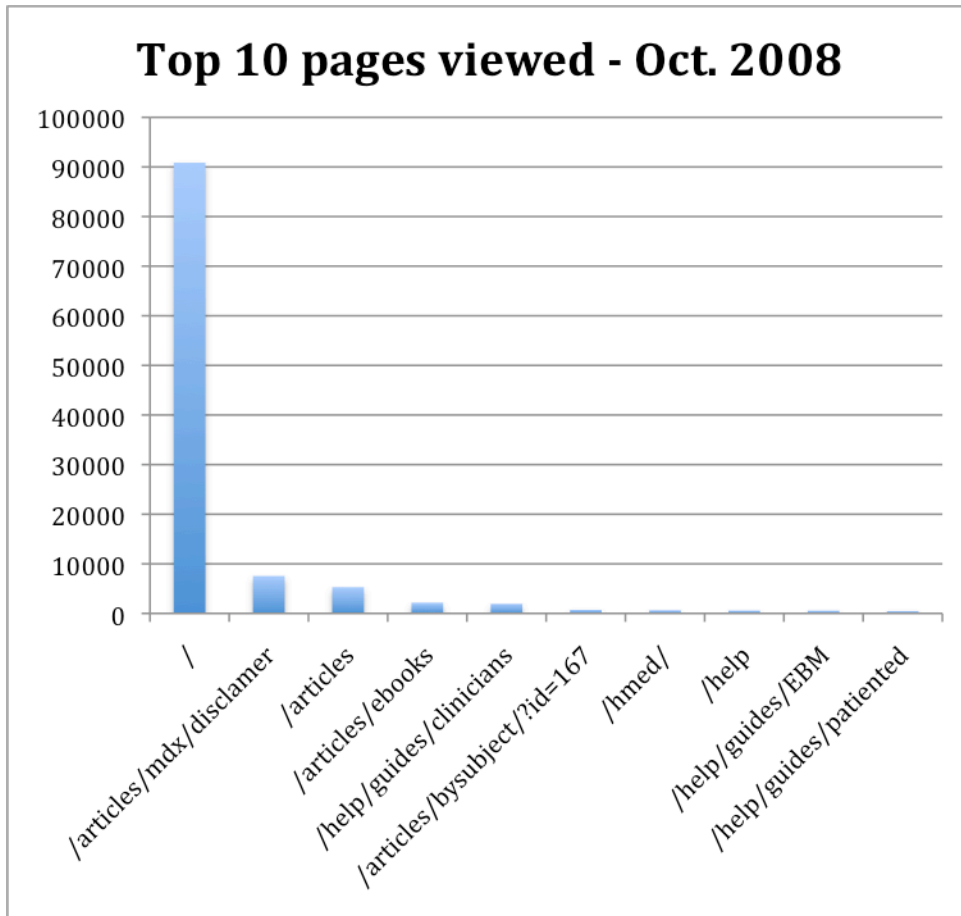
## **Referrers**

This chart shows the domains of sites referring visitors to pages on the [biomed.lib.umn.edu](http://biomed.lib.umn.edu) subdomain during October 2008. Google is the top referrer of traffic to this site, responsible for more than twice the traffic of the Libraries' main web site, which comes in second among referrers. The third and fourth sites on the chart are the University of Minnesota's Duluth campus site and main site, respectively. The domain listed here as "live" is Microsoft's Live Search. It is interesting to note here the presence of the Libraries' UThink blog system among the top referrers to the Bio-Med site. It would be interesting to learn if this is the result of blogs maintained by library staff or related to courses that link to the Bio-Med site.



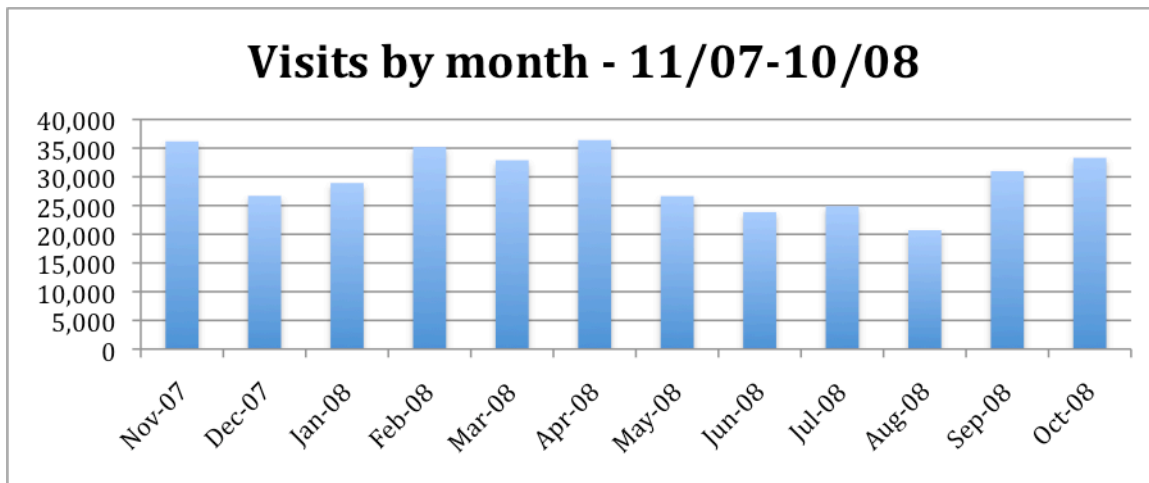
### **Page views**

The chart below shows the top ten most frequently viewed pages on the Bio-Med site in October 2008. The library's home page is most frequently viewed, receiving more than ten times the traffic of the second-place page (a page which at the time of this writing was not viewable). This pattern is common among our sites; however, the Bio-Med site is a particularly dramatic example of how little users navigate into deeper pages on our sites.



### **Monthly visitors**

The chart below tracks the number of visitors to the Bio-Med site each month for the year ending October 2008. The pattern reflected here is common among our sites, and seems to follow the rhythm of the school year. Notable here is the consistent volume of traffic month-to-month on the Bio-Med site, which is second only to the Libraries' main web site.



### **Future exploration**

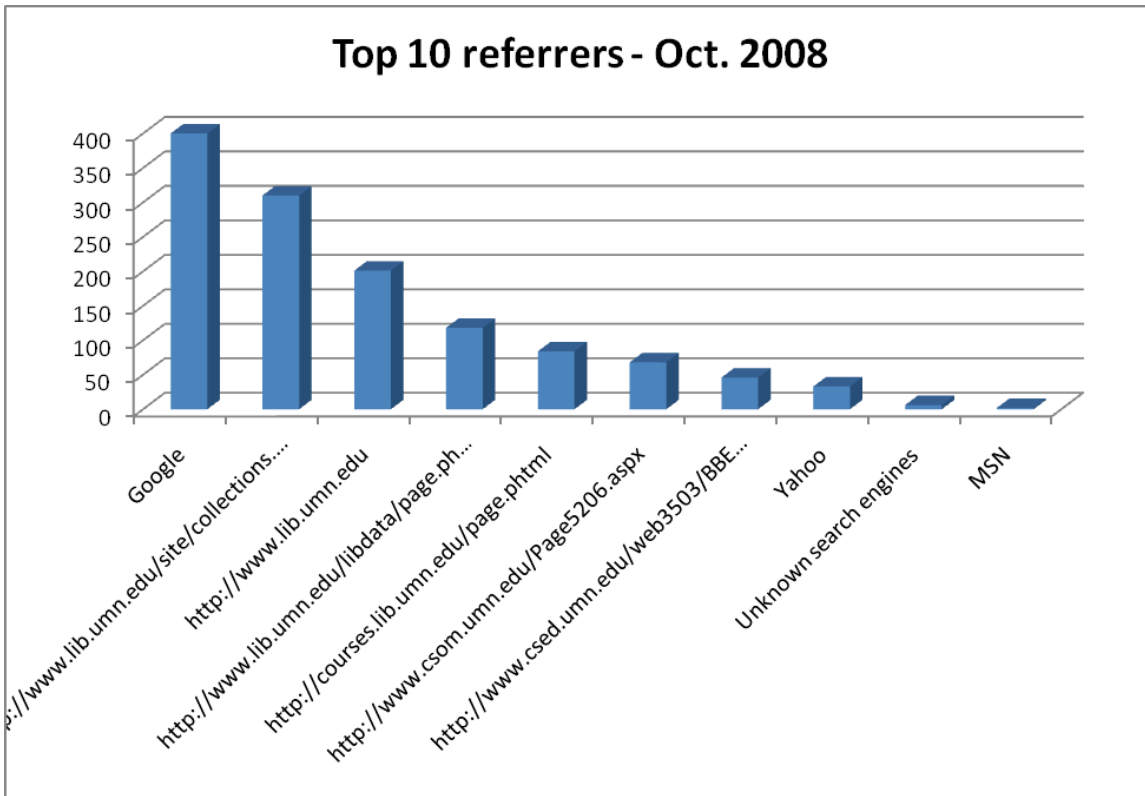
As mentioned above, the Bio-Med site is one of two of the Libraries' websites that use Google Analytics to track web traffic. While there have been some concerns raised at the University and at other Libraries around the privacy implications of using a third-party service to track web use, Google Analytics presents a compelling alternative to our AWStats tool, providing richer analysis and reporting. Regardless of which tool it is, use of a consistent system across all of our sites would make it easier to generate apples-to-apples comparisons of web traffic.

### ***busref.lib.umn.edu Website Statistics***

This set of statistics covers pages served on the busref.lib.umn.edu subdomain, which contains pages for the University Libraries' Business Reference Library. Data is from October of 2008 and was obtained from the Libraries' AWStats program, which ingests and parses Apache web server logs.

### **Referrers**

This chart shows referring domains for visits to all pages on the busref.lib.umn.edu subdomain in October, 2008. Google is the leading referrer of site traffic, however the cumulative total of referrals coming from University Libraries' web pages exceeds Google referrals. In addition to search engines and the Libraries' web pages, two university departmental pages also appear in the top 10 list. One of the departmental pages is for a specific course and another is a site providing students with links to research resources.

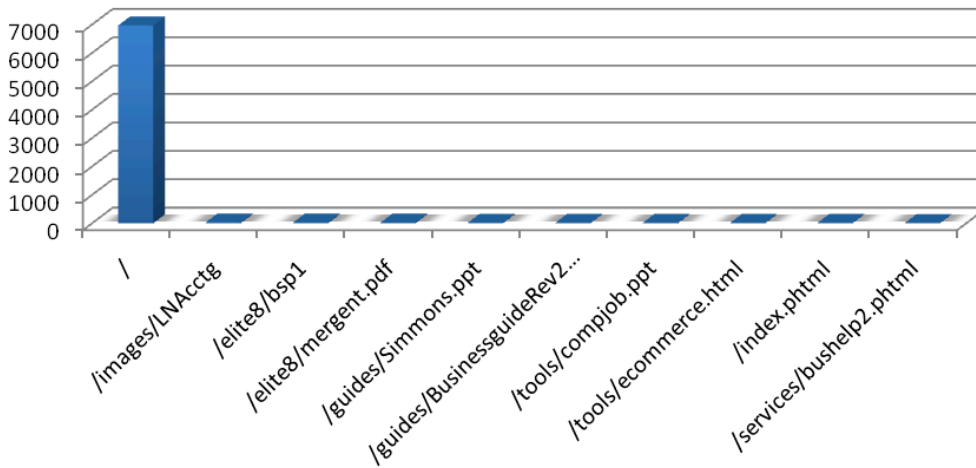


**Page views**

The following chart shows the ten most viewed pages on the busref.lib.umn.edu subdomain. The site’s homepage has an overwhelming majority of page views with just less than 7,000 for the month. No other page received more than 30 page views. These pages comprise a variety of sources: basic web pages, PDF documents, and PowerPoint presentations.



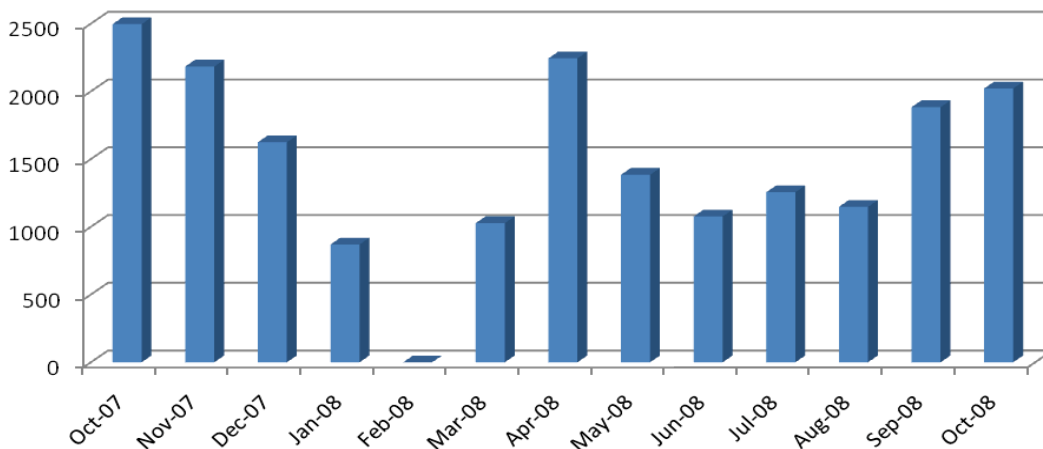
### Top 10 pages viewed - Oct. 2008



### Monthly visitors

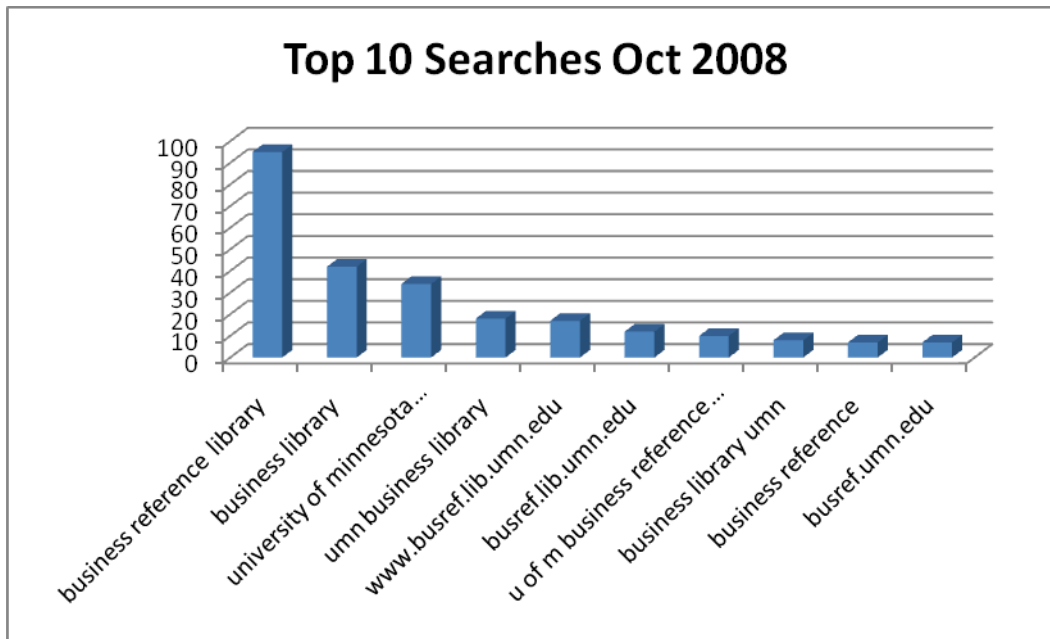
This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. Site traffic is highly correlated with the peaks and valleys of the academic year. Traffic is busiest in April, October, and November which are the months before semesters come to an end. Traffic peaks at approximately 2,500 users during these months. During non-peak months, there are approximately 1,000 unique visitors to the busref.lib.umn.edu pages.

### Traffic by month of year



## **Search Terms**

Searching in the busref.lib.umn.edu is almost exclusively for that of the Business Reference Library itself. Various permutations of Business Reference comprise the entire top 10 list.



## **Conclusions**

The majority of traffic to the busref.lib.umn.edu subdomain is largely driven either from referrals from university websites (primarily University Libraries) or from search engines (primarily Google). Monthly traffic varies between 1,000 to 2,500 unique users and the traffic is highly influenced by the academic calendar. Users accessing the busref.lib.umn.edu domain are almost exclusively using the resources available on the site's home page.

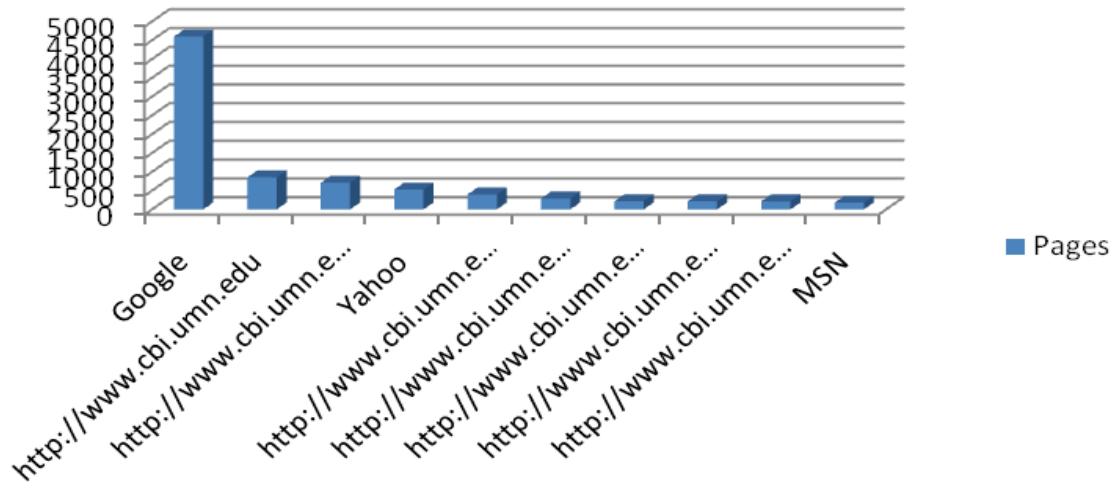
## ***cbi.lib.umn.edu Website Statistics***

This set of statistics covers pages served on the cbi.lib.umn.edu subdomain, which contains pages for the Charles Babbage Institute. Data is from October of 2008 and was obtained from the Libraries' AWStats program, which ingests and parses Apache web server logs.

## **Referrers**

This chart shows referring domains for visits to all pages on the cbi.lib.umn.edu subdomain in October, 2008. Google is the dominant referrer of site traffic with most of the other referrers from [www.cbi.umn.edu](http://www.cbi.umn.edu) pages.

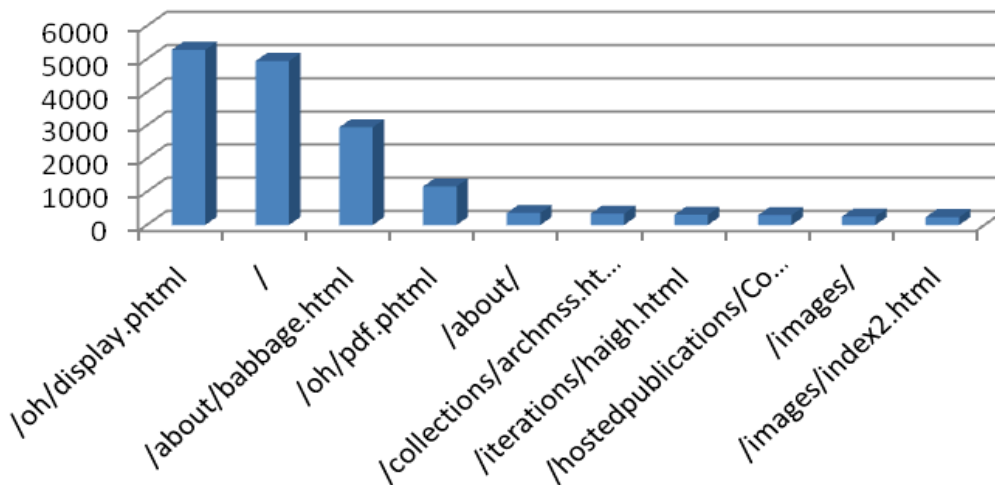
## Top 10 Referrers - Oct 2008



### Page views

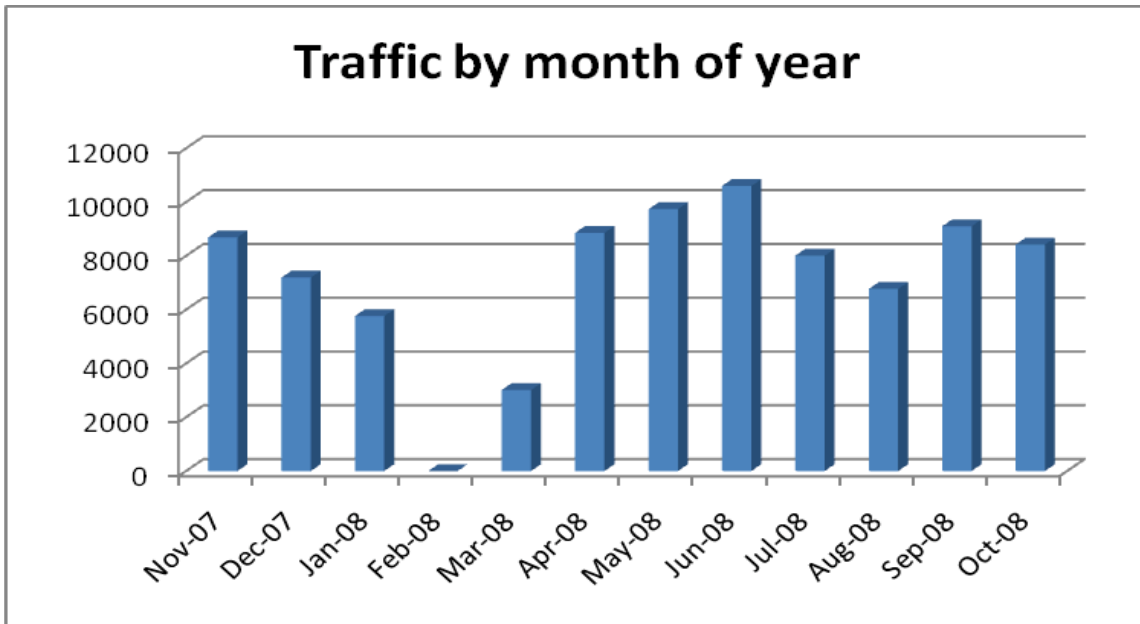
The following chart shows the ten most viewed pages on the cbi.lib.umn.edu subdomain. The search results from cbi.lib.umn.edu and the site's home page are the most viewed pages in the subdomain.

## Top 10 Page Views - Oct 2008



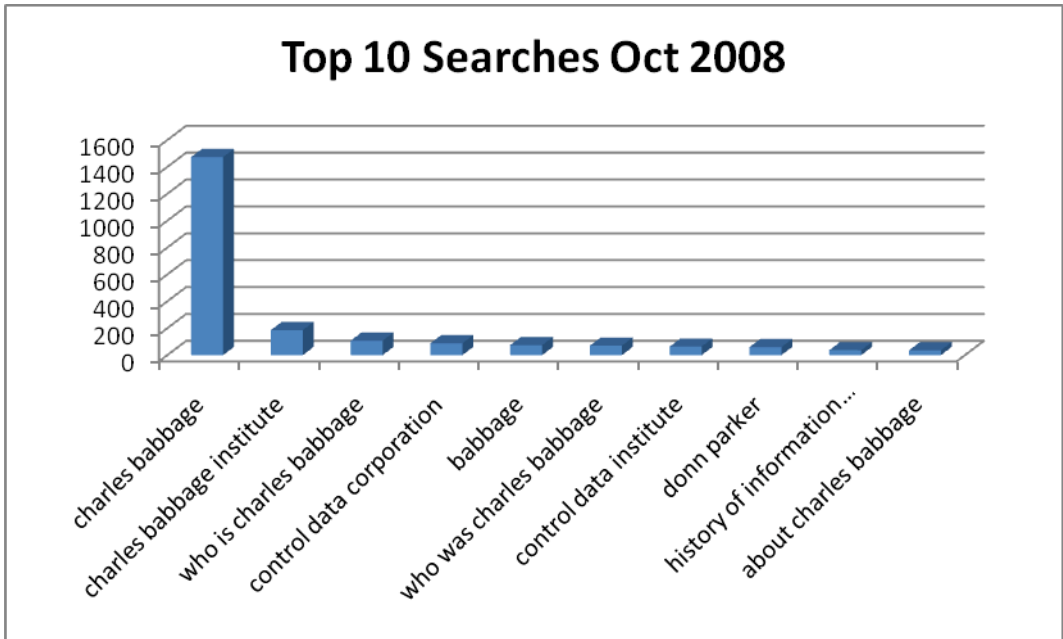
### **Monthly visitors**

This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. Unlike many of the other subdomains, site traffic peaks in June when school is not in session. “Peak” academic months (April, October, November) only showed slightly above average traffic. Users to the cbi.lib.umn.edu do not seem to be motivated by the academic calendar.



### **Search Terms**

Searching in the cbi.lib.umn.edu is almost exclusively searches on Charles Babbage. “Charles Babbage” is by far the most common phrase and several other permutations of Charles Babbage also make the top 10 list. No other search phrase on the cib.lib.umn.edu subdomain was used more than 100 times.



**Conclusions**

Traffic to the [cbi.lib.umn.edu](http://cbi.lib.umn.edu) subdomain is largely driven either from referrals from search engines (primarily Google) or from referrals from the Charles Babbage Institute’s own web pages. While number of monthly page hits varies between 6,000 and 10,000 a month, the subdomain is receiving a relatively constant stream of traffic throughout the year. Those searching the site are overwhelmingly searching for information on Charles Babbage. Site visits are not dependent on the academic calendar.

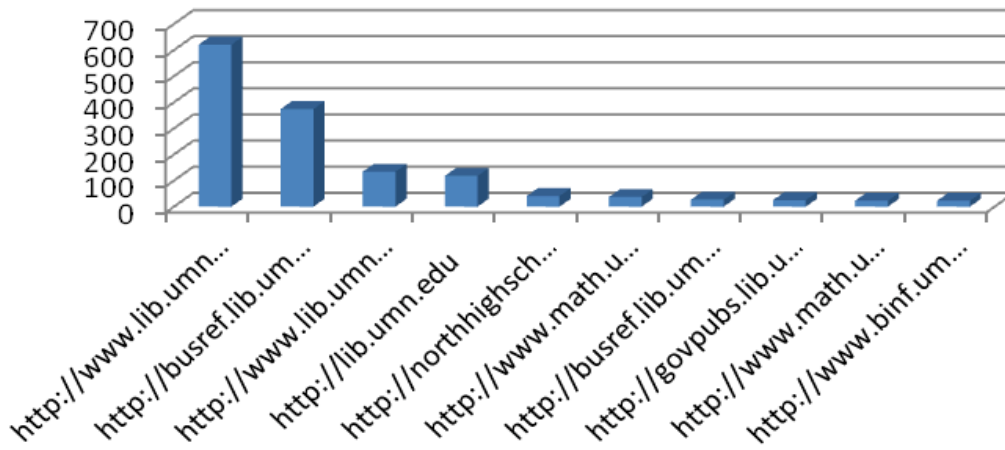
*[courses.lib.umn.edu Website Statistics](#)*

This set of statistics covers pages served on the [courses.lib.umn.edu](http://courses.lib.umn.edu) subdomain, which contains pages for the Libraries’ CourseLib program. Data is from October of 2008 and was obtained from the Libraries’ AWStats program, which ingests and parses Apache web server logs.

**Referrers**

This chart shows referring domains for visits to all pages on the [courses.lib.umn.edu](http://courses.lib.umn.edu) subdomain in October, 2008. The overwhelming majority of traffic to the [courses.lib.umn.edu](http://courses.lib.umn.edu) subdomain comes from the Libraries’ own web pages. There are referrals from search engines. However, none make the top 10 list.

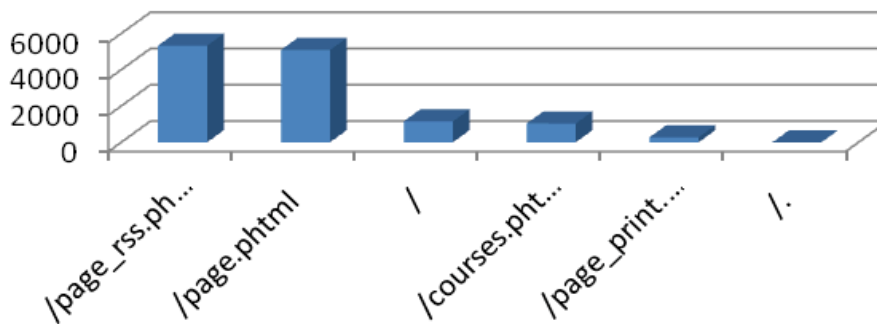
## Top 10 Referrers - Oct 2008



### Page views

The following chart shows all viewed pages on the course.lib.umn.edu subdomain. The contents on a course's Course Lib page are the most common views in the subdomain. Pages that search for a particular course are most of the remaining page views.

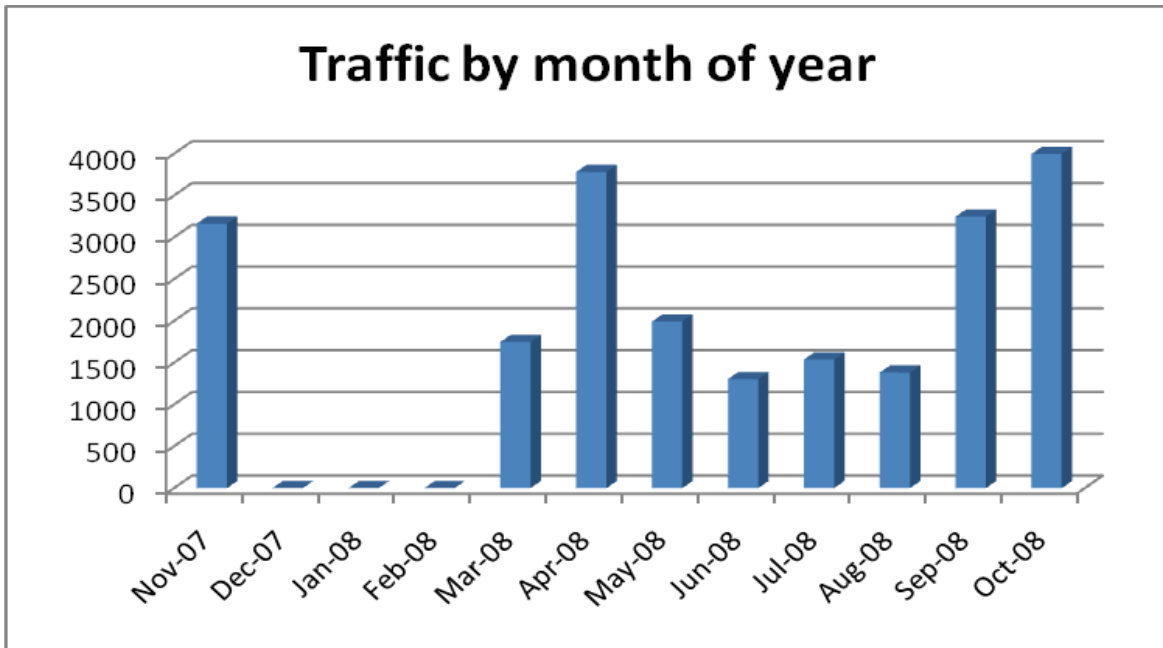
## Page Views - Oct 2008



### Monthly visitors

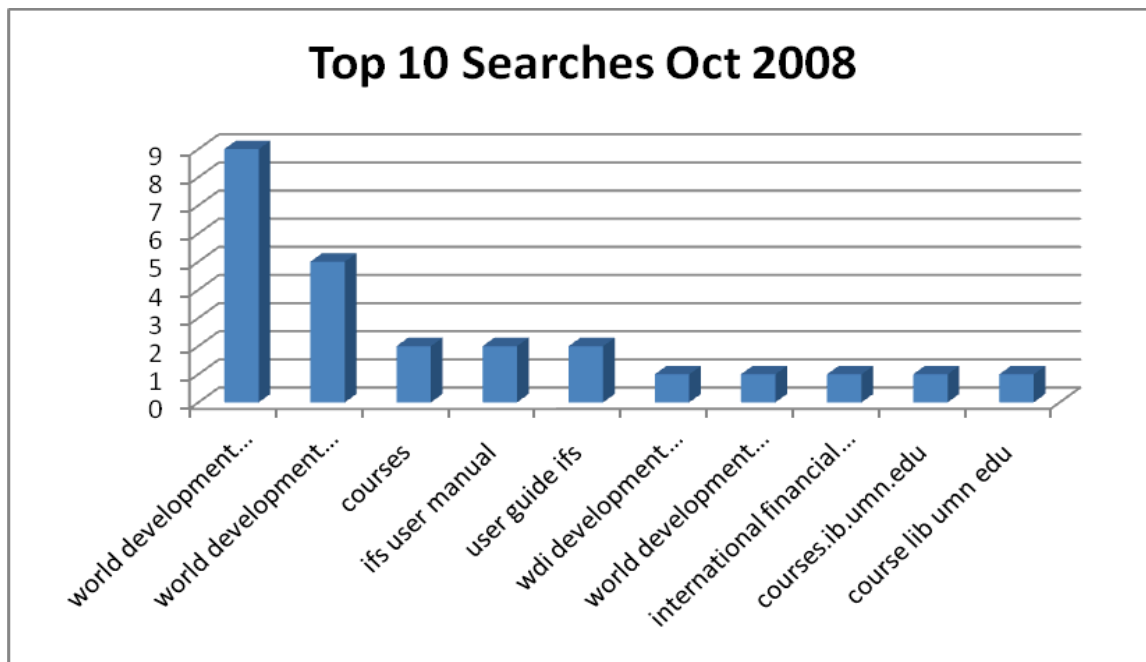
This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. Additionally CourseLib was down in December and January. For the months available, traffic is highly dependent on the academic calendar. Since these web pages are directly

supporting academic courses, this should be expected. Given that these pages are directly tied to the academic year, it is somewhat surprising to see that the summer month's maintain a relatively high volume of traffic.



**Search Terms**

Course.lib.umn.edu is not designed for searching. While possible, the site encourages users to select a class by clicking on dropdown menus or an A-Z list of courses. The result is that while searching is possible, the highest search phrase was only used 4 times in October 2008



## **Conclusions**

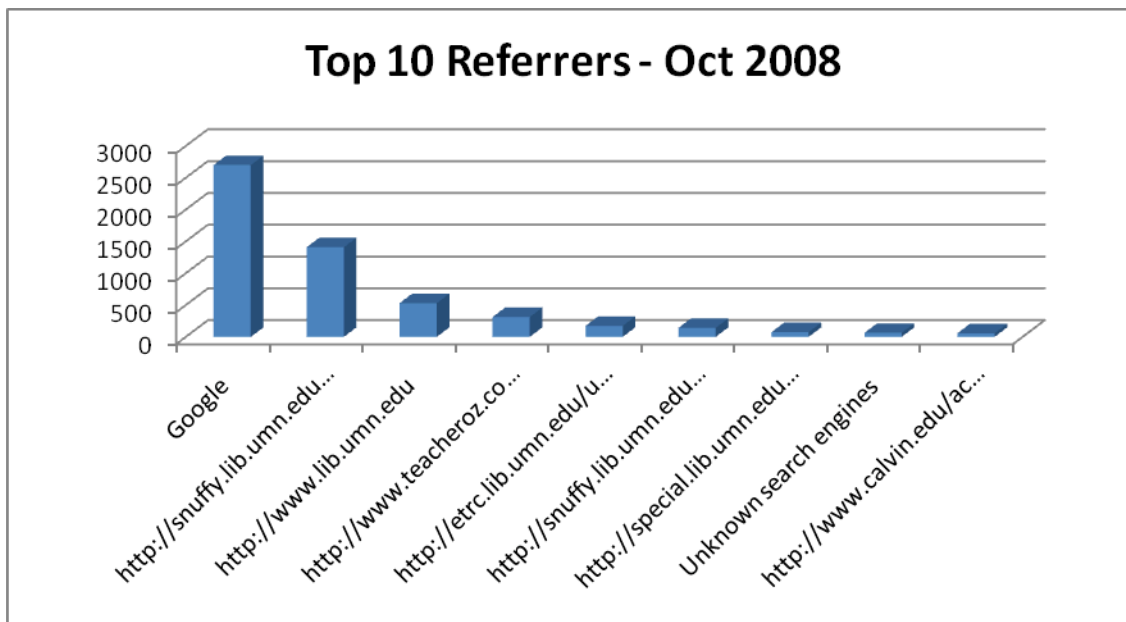
Traffic to the courses.lib.umn.edu subdomain is largely tied to the academic year and mostly comes from the Libraries' own web pages. Once at the courses.lib.umn.edu subdomain, users simply find their course and utilize the resources available on the page for the course.

## ***digital.lib.umn.edu Website Statistics***

This set of statistics covers pages served on the digital.lib.umn.edu subdomain, which contains pages for the Libraries' digital images. Data is from October of 2008 and was obtained from the Libraries' AWStats program, which ingests and parses Apache web server logs.

## **Referrers**

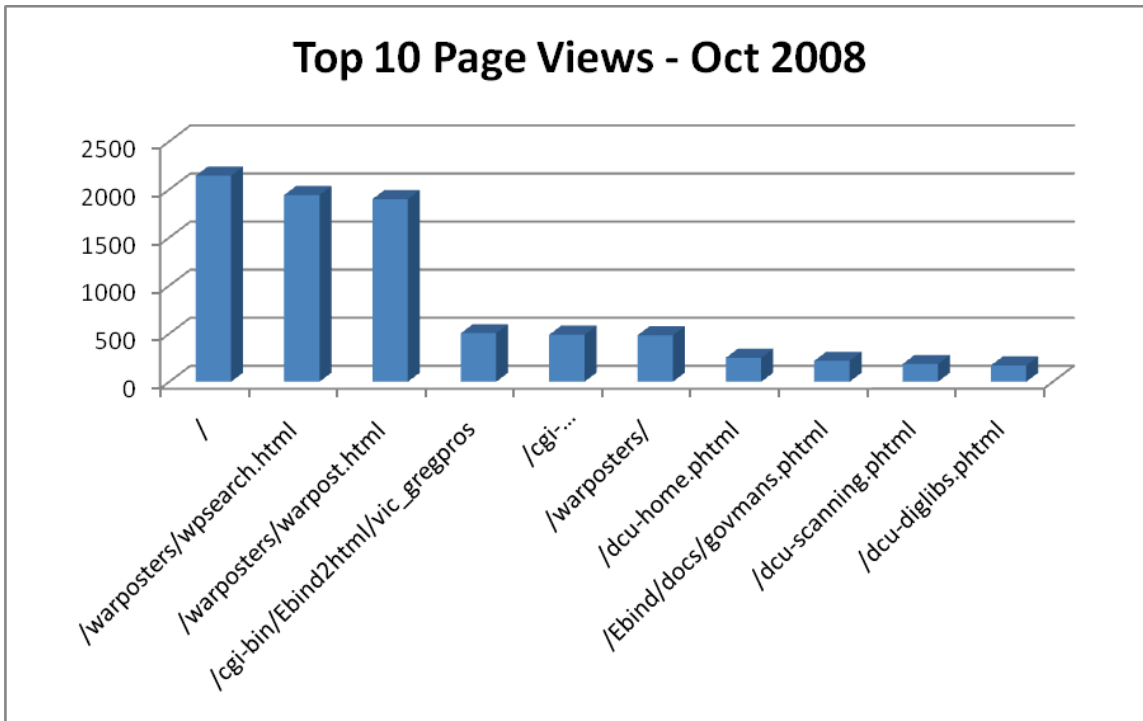
This chart shows referring domains for visits to all pages on the digital.lib.umn.edu subdomain in October, 2008. Google is the largest referrer. Most of the remaining referrals in the top 10 come from Libraries' websites, although two external sites also are found in the top 10.



## **Page views**

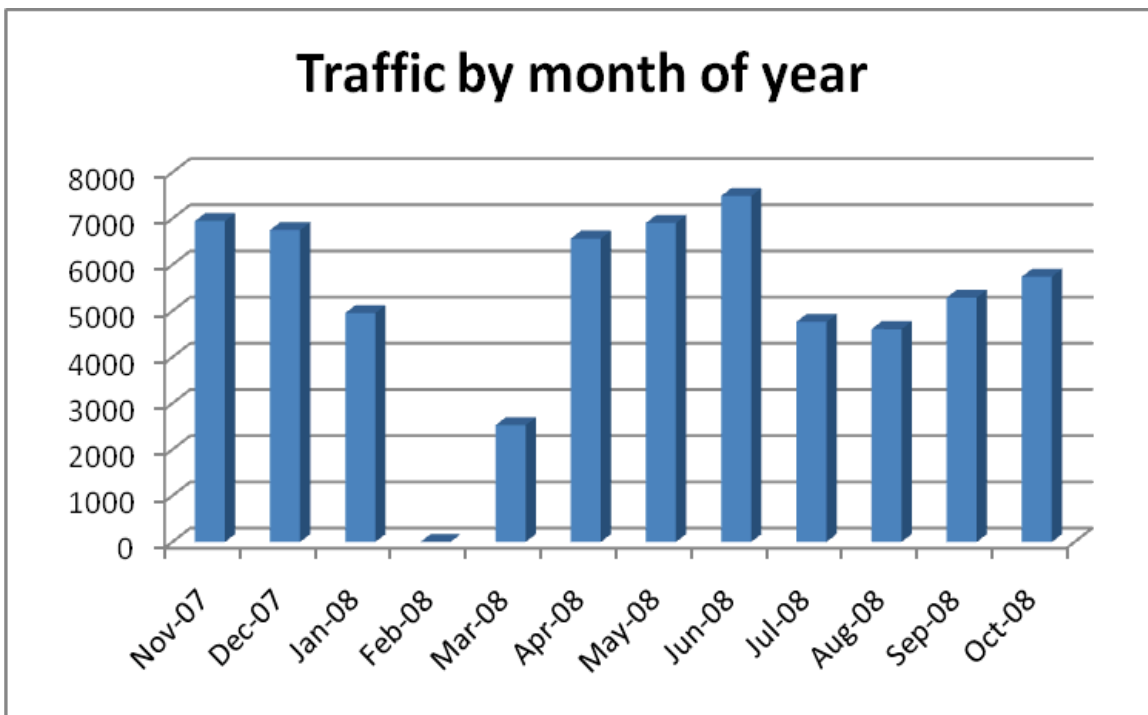
The following chart shows all viewed pages on the digital.lib.umn.edu subdomain. The subdomain's home page is the most highly viewed page but has only slightly more page views than does the home page for the war posters and the search page for the war posters. These three pages have four times as many views as any other, but there are several other pages with more than 100 page views in the month of October.





#### Monthly visitors

This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. Traffic dips during the summer months. During the months of Spring and Fall Semester, the digital.lib.umn.edu subdomain averages between 6-7,000 unique visitors. This dips to between 4-5,000 during the summer months.



## **Search Terms**

Many of the most popular search phrases in the digital.lib.um.edu subdomain are for war posters. This is not surprising given that the war posters home page and search page have page views that almost exceeds that of the home page for the subdomain. Search counts are relatively low given the number of page views for the subdomain.



## **Conclusions**

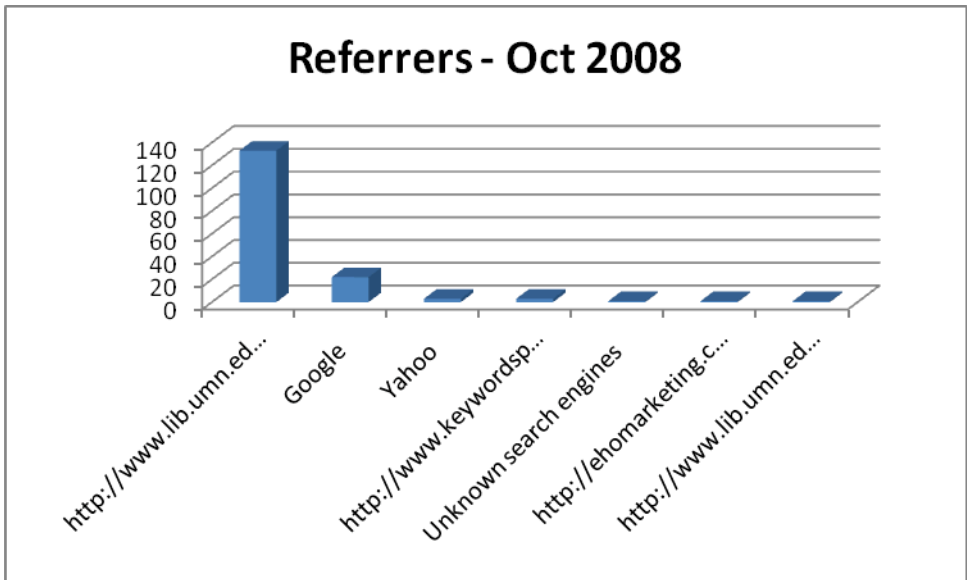
Traffic to the digital.lib.umn.edu subdomain is largely tied to searching for the Libraries' war posters collection. This is seen in the number of page views as well as in the search phrases used in the subdomain. Traffic to the site remains relatively constant throughout the year, although Spring and Fall Semester months show higher usage of the subdomain.

## ***dsrc.lib.umn.edu Website Statistics***

This set of statistics covers pages served on the dsrc.lib.umn.edu subdomain, which contains pages for the Libraries' Social Science Data site. Data is from October of 2008 and was obtained from the Libraries' AWStats program, which ingests and parses Apache web server logs.

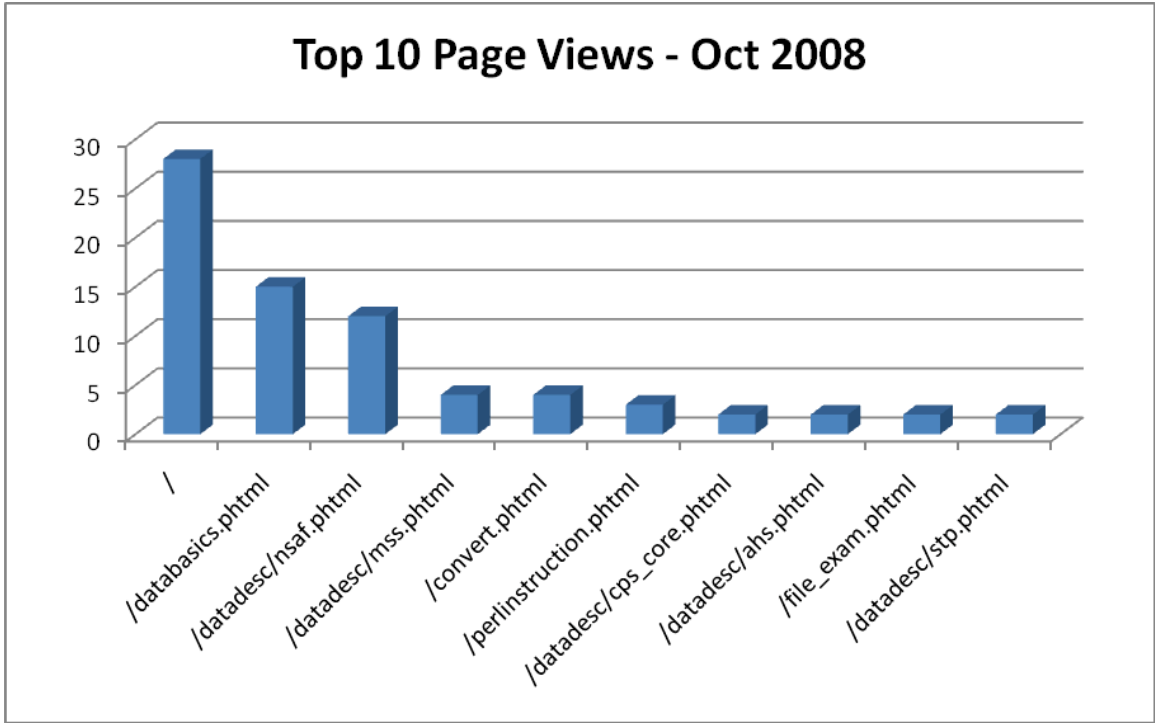
## **Referrers**

This chart shows referring domains for visits to all pages on the dsrc.lib.umn.edu subdomain in October, 2008. Almost all referrals to the subdomain come from the Libraries' LibData page. Google refers a small set of users to the dsrc.lib.umn.edu subdomain.



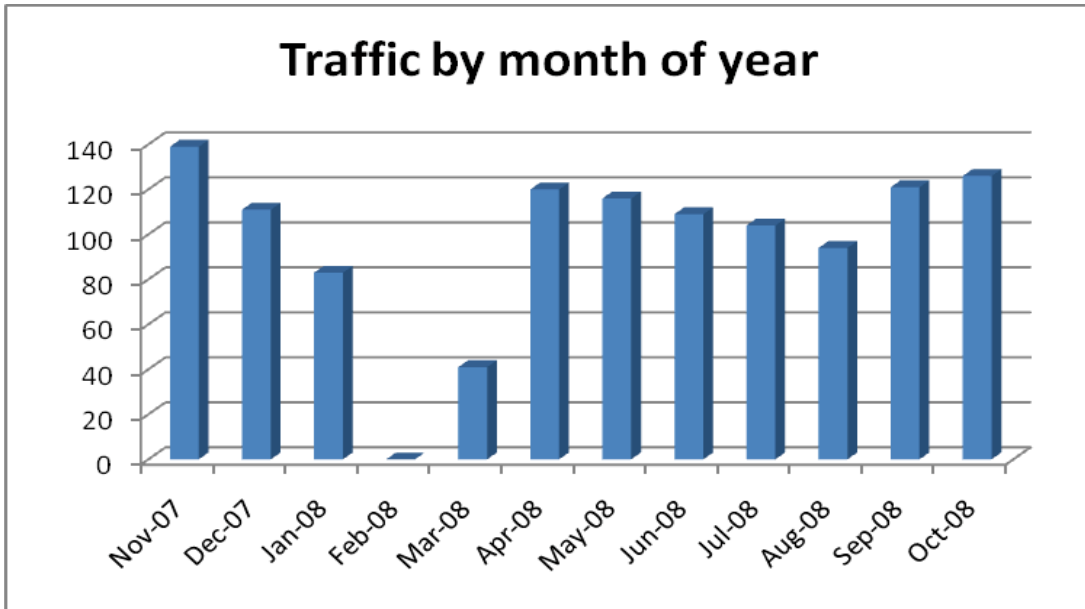
**Page views**

The following chart shows all viewed pages on the dsrc.lib.umn.edu subdomain. The subdomain is not highly used and even the most heavily used page (the home page) received less than 25 page views.



### Monthly visitors

This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. For the months with available data, the number of visitors remains quite consistent. In almost every month, slightly over 100 unique visitors came to the dsrc.lib.umn.edu subdomain.



### Search Terms

As to be expected for a subdomain with so few page views, no search term in the dsrc.lib.umn.edu is used with any frequency. "National survey of American families" was the most common search phrase and it was searched just four times in the month of October.



## **Conclusions**

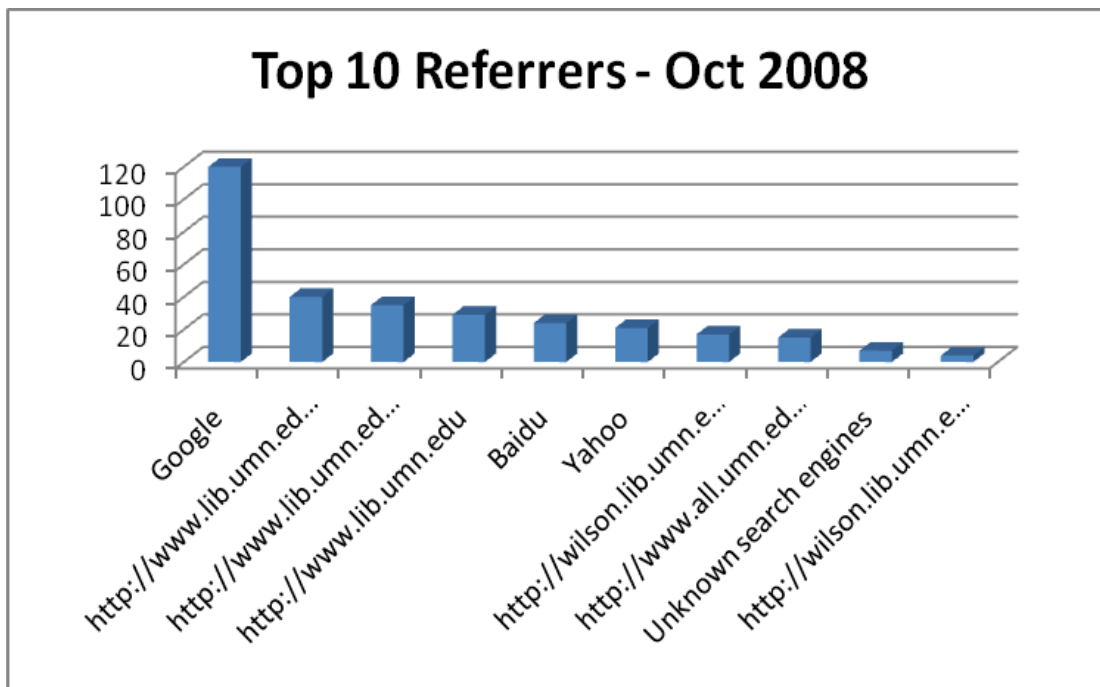
Traffic to the dsrc.lib.umn.edu subdomain is minimal. In no month in the year survey did 150 unique visitors use pages on the subdomain. No real trends are identified for pages or searches that are most prevalent in the subdomain but this in large part may be explained by the lack of use within it.

## ***eastasian.lib.umn.edu Website Statistics***

This set of statistics covers pages served on the eastasian.lib.umn.edu subdomain, which contains pages for the Libraries' East Asian Library website. Data is from October of 2008 and was obtained from the Libraries' AWStats program, which ingests and parses Apache web server logs.

## **Referrers**

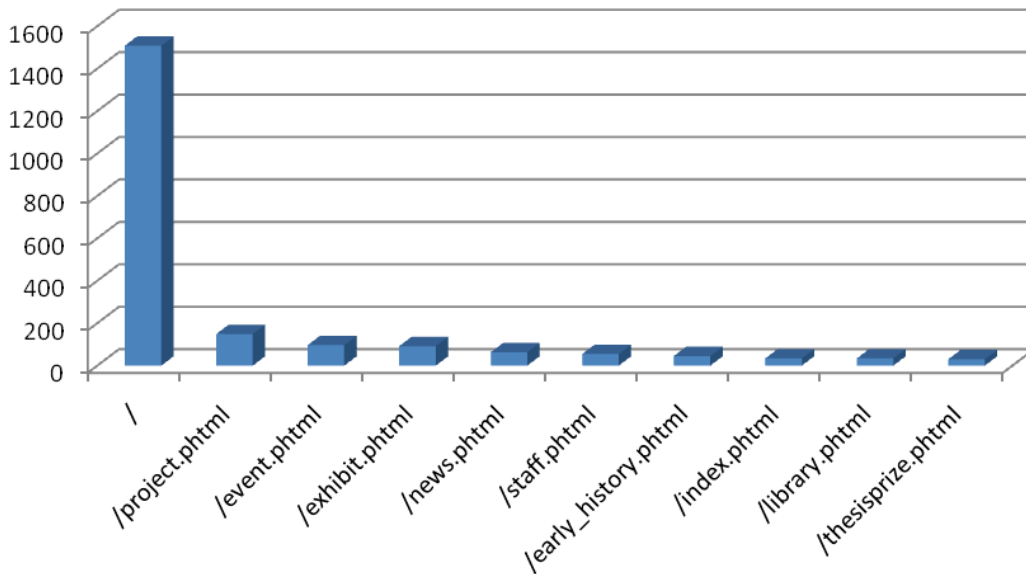
This chart shows referring domains for visits to all pages on the eastasian.lib.umn.edu subdomain in October, 2008. Almost all referrals to the subdomain come from either search engine traffic or from the Libraries' own web pages. Google is the most frequently used search engine but Baidu and Yahoo are also used.



## **Page views**

The following chart shows all viewed pages on the eastasian.lib.umn.edu subdomain. The overwhelming majority of page views are the subdomain's home page. The only other page to have received more than 100 views in the month of October was the project page, which highlights current East Asian Library projects. The other pages within the top 10 list received between 30 – 100 page views.

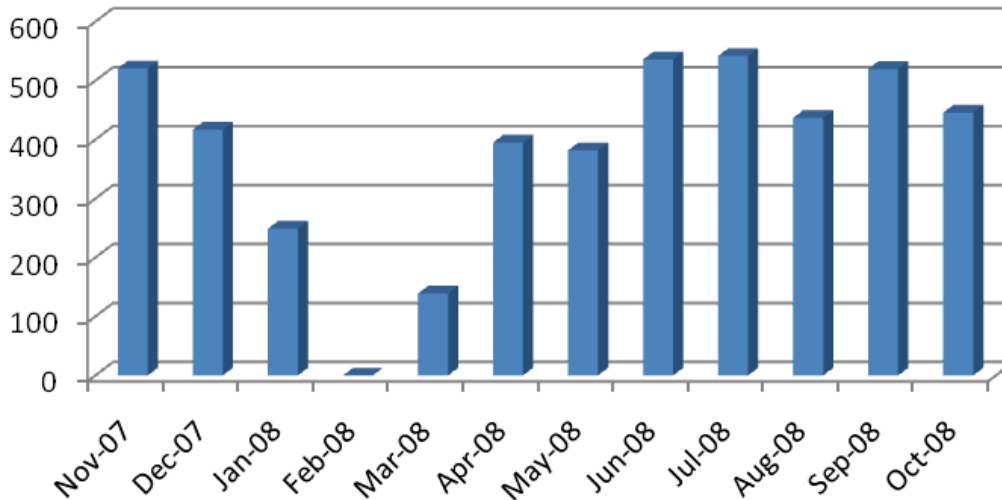
## Top 10 Page Views - Oct 2008



## Monthly visitors

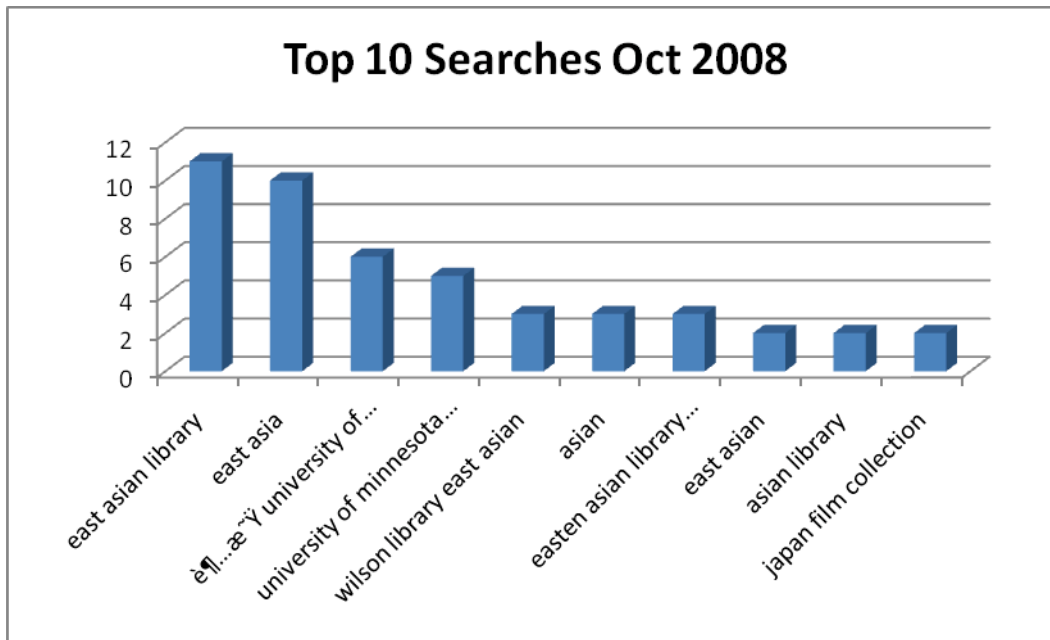
This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. Subdomain traffic is less than many other library sites. Traffic actually peaks during the summer months, which is a trend not seen by many of the other subdomains.

## Traffic by month of year



## **Search Terms**

The eastasian subdomain does not have many visitors each month, so it should not be surprising that there are few searches performed on the site. Most of the searches within the subdomain are searching for the East Asian Library itself.



## **Conclusions**

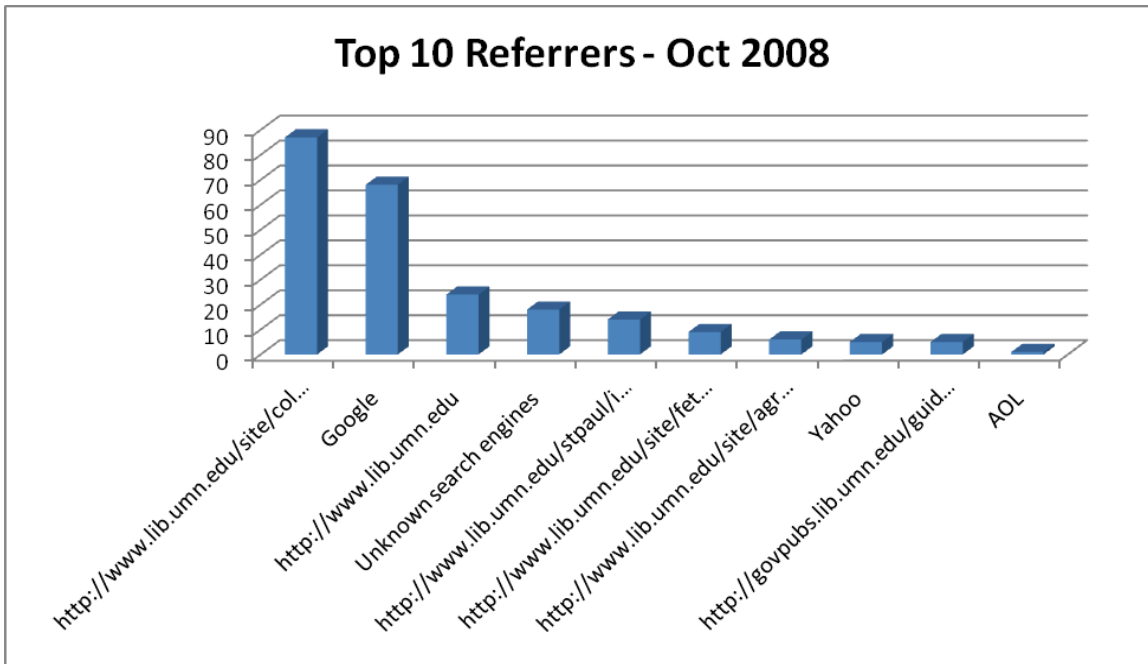
Traffic to the eastasian.lib.umn.edu subdomain is low but remains steady throughout the year. The vast majority of the subdomain traffic is on the site's home page; however, there are several other pages that are viewed dozens of times within a month. Users are overwhelmingly likely to be referred to the subdomain either by search engine or from a Libraries' web page.

## ***efw.lib.umn.edu Website Statistics***

This set of statistics covers pages served on the efw.lib.umn.edu subdomain, which contains pages for the Entomology, Fisheries, and Wildlife Library's website. Data is from October of 2008 and was obtained from the Libraries' AWStats program, which ingests and parses Apache web server logs.

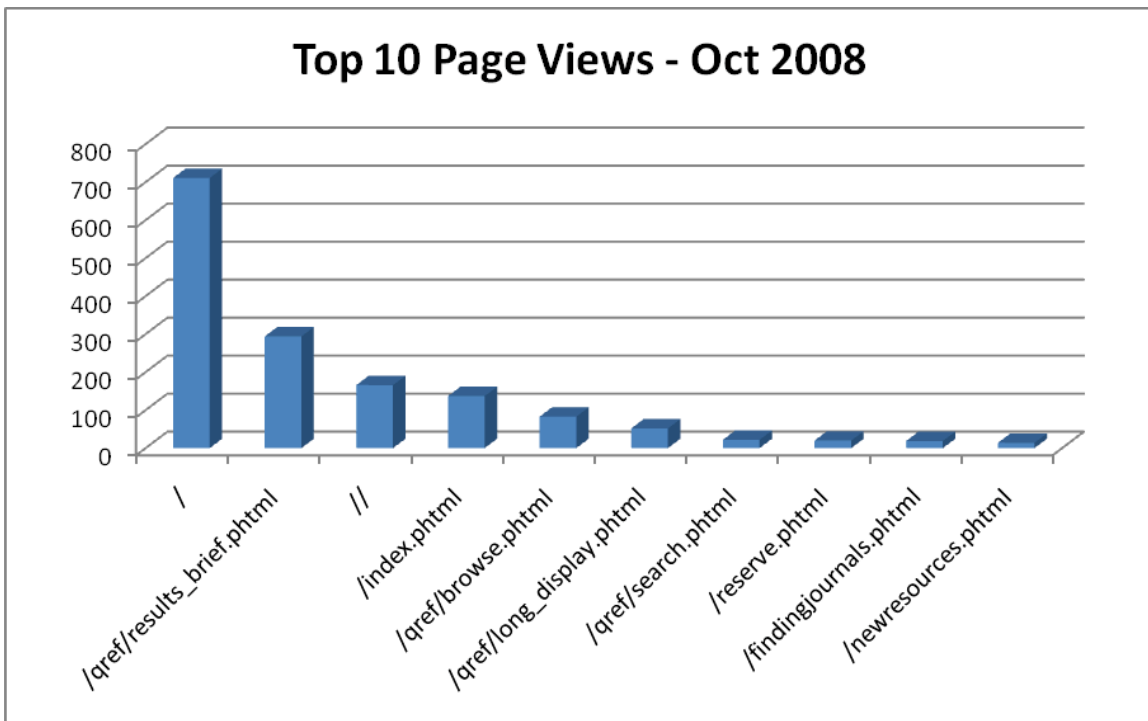
## **Referrers**

This chart shows referring domains for visits to all pages on the efw.lib.umn.edu subdomain in October, 2008. Users were directed to the efw subdomain either via search engine or from the Libraries' own web pages. The most frequent referral page is the libraries' web page listing the different libraries and collections. Google is the primary search engine that refers users to the efw.lib.umn.edu pages.



#### Page views

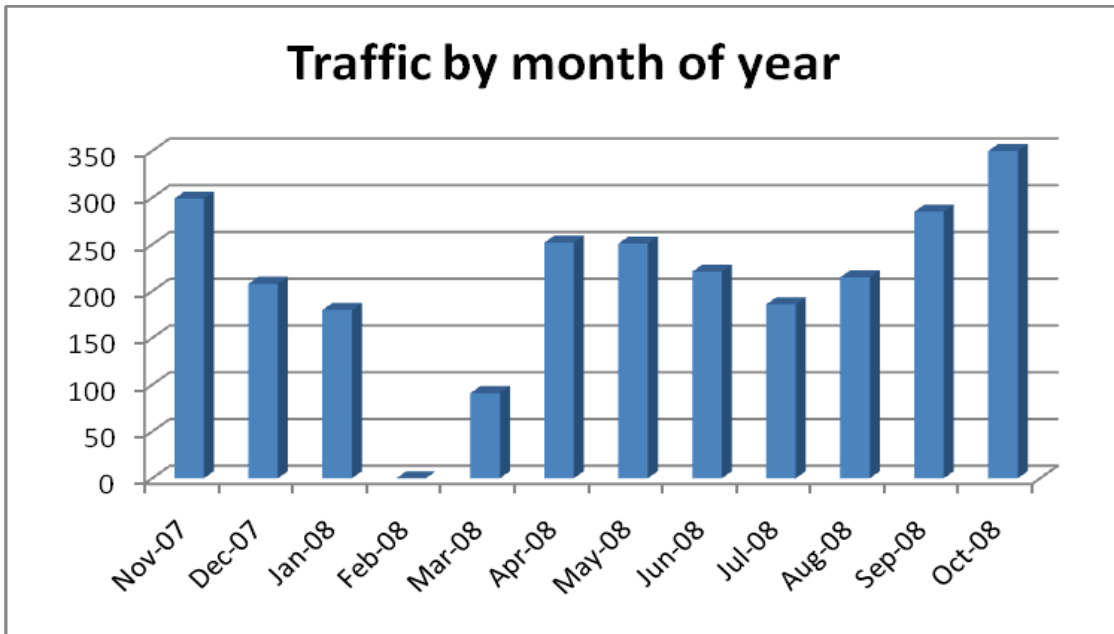
The following chart shows all viewed pages on the [efw.lib.umn.edu](http://efw.lib.umn.edu) subdomain. The majority of page views are for the EFW Library's home page. A substantial percentage of the page views are for pages that support the EFW Library's Quick Reference service. For the number of unique visitors to the [efw](http://efw.lib.umn.edu) subdomain, there are a large number of page views for the Quick Reference service. Informational resource pages comprise the remainder of the top 10 page view list.





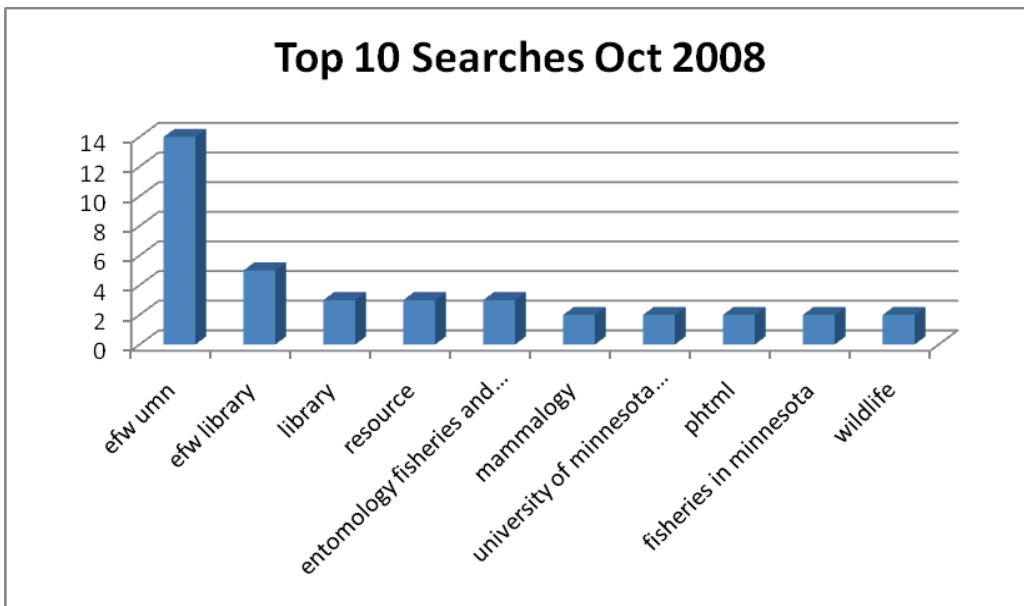
### Monthly visitors

This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. Subdomain traffic is less than many other lib.umn.edu sites. Throughout the year, monthly traffic is between 150 and 130 unique users. Site traffic is busiest during the busiest parts of the academic calendar.



### Search Terms

The efw subdomain does not have many visitors each month, so it should not be surprising that there are few searches performed on the site. Most of the searches within the subdomain are searching for the EFW Library itself.



## **Conclusions**

Traffic to the [efw.lib.umn.edu](http://efw.lib.umn.edu) subdomain is low. However, despite the low numbers, the subdomain contains a relatively high number of page views. The Quick Reference service receives the majority of the page views, but the subdomain's informational pages are also used frequently by those who visit the site. Users are overwhelming likely to be referred to the subdomain either by search engine or from a Libraries' web page.

## ***ej.lib.umn.edu Website Statistics***

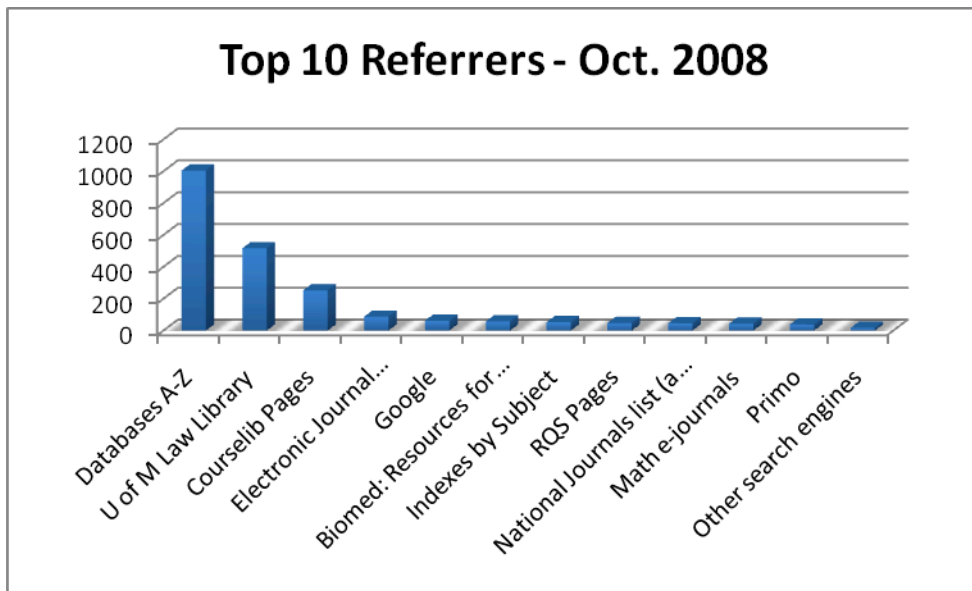
This set of statistics covers the Electronic Journals domain ([ej.lib.umn.edu](http://ej.lib.umn.edu)). Statistics reviewed here are for the month of October 2008 and come from the Libraries' AWStats program, which ingests and parses Apache web server logs.

Unlike most library subdomains, the Electronic Journals subdomain appears to be used primarily for directing the user to a login for a particular resource. There is a main Electronic Journals page, but it redirects to [http://tc.liblink.umn.edu/sfx\\_local/azlist/default](http://tc.liblink.umn.edu/sfx_local/azlist/default), shown below.



## **Referrers**

Due to its nature as a redirecting tool, [ej.lib](http://ej.lib) does not get significant traffic from search engines. Instead, its traffic comes primarily from links on other pages that are meant to direct the user to a particular electronic journal. For example, there are five links on the "Indexes and Databases A-Z" page that are routed through the [ej.lib](http://ej.lib) domain for login. These links account for just over 1000 page loads for [ej.lib](http://ej.lib) for the month of October.



ej.lib also receives traffic from direct address links such as bookmarks or links in e-mail. In October, 38% of the pages served were from direct links, vs. 60% from other Web pages and 1.9% for search engines.

#### **Traffic by Day, Day of Week, and Month**

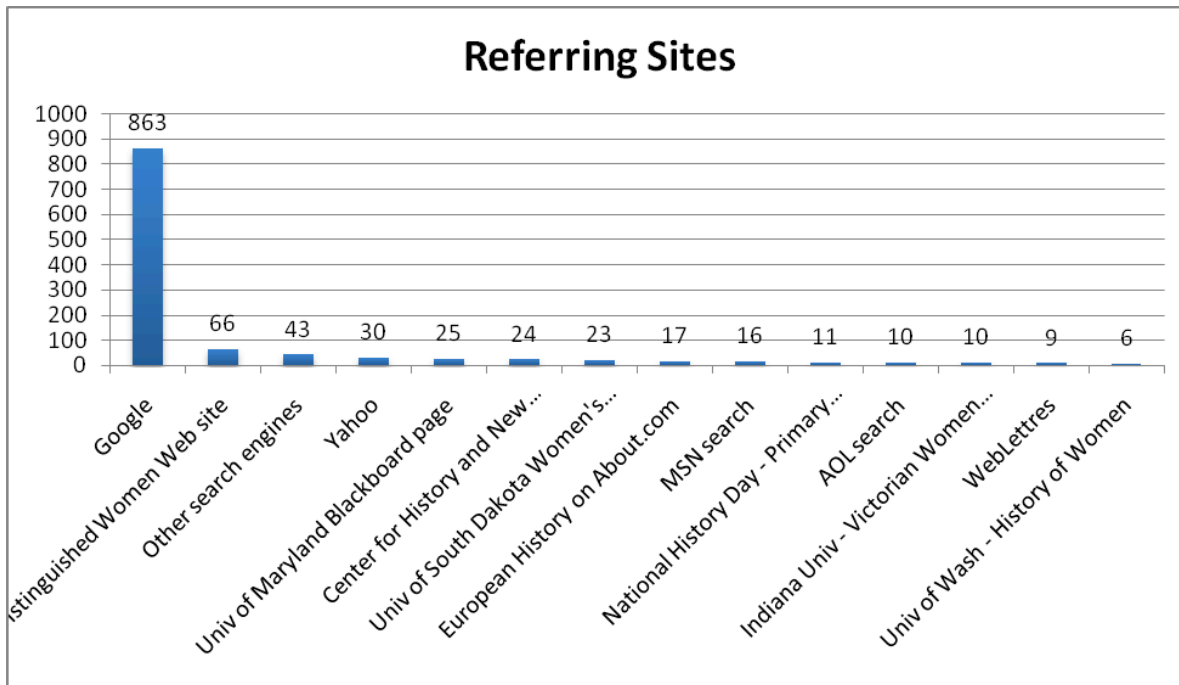
Although the amount of traffic to ej.lib is dwarfed by overall traffic to the main library site, the curves formed by the data points are a close match. For example, with both domains traffic is higher early in the week, peaking on Monday. Traffic then gradually dropping through the week, most significantly on Saturday. Traffic increases slightly on Sunday.

#### ***etrc.lib.umn.edu Website Statistics***

This set of statistics covers the Electronic Text Research Center domain (etrc.lib.umn.edu). Statistics reviewed here are for the month of October 2008 and come from the Libraries' AWStats program, which ingests and parses Apache web server logs. The ETRC is no longer in business, but the files and projects remain there for archiving. The statistics show that the site still receives traffic from other academic institutions, if not our own.

#### **Referrers**

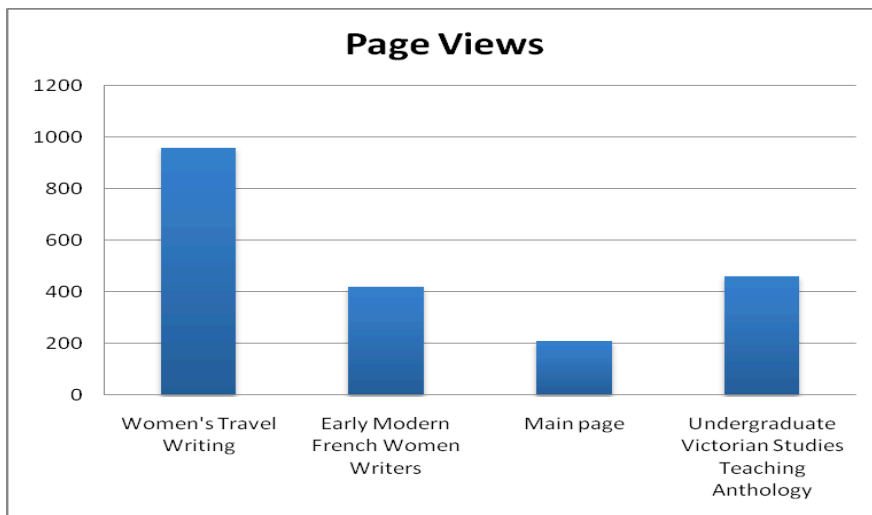
75% of the traffic the ETRC site receives comes from Google. In fact, if all search engine traffic is combined, the total is more than double the traffic referred from all Web sites combined. The chart also shows, though, that the resources found in the subdomain remain popular with other academic institutions.



At first glance, it seems odd that so little traffic (12 pages for October 2008) come from University of Minnesota-owned pages. However, 61% of the overall traffic is coming from direct address links, so it's likely that our own scholars have simply bookmarked the site.

### Pages Viewed

The majority of users view the pages that make up the Women's Travel Writing Project. Two other projects bring in readers, and the main page still gets a few hits.



### Keywords/Key Phrases

With 75% of the traffic to the site coming from Google, it's worth looking at what words people are using to get there. As one would expect from any list of keywords used to get to a site, even the top-ranked words and

phrases account for a small percentage of traffic on their own. The most popular single word was *women*, present in 5.1% of the searches. The words *French* and *Victorian* are also high on the single word list.

Each key phrase accounts for even less of the overall search traffic. Phrases tend to be very specific.

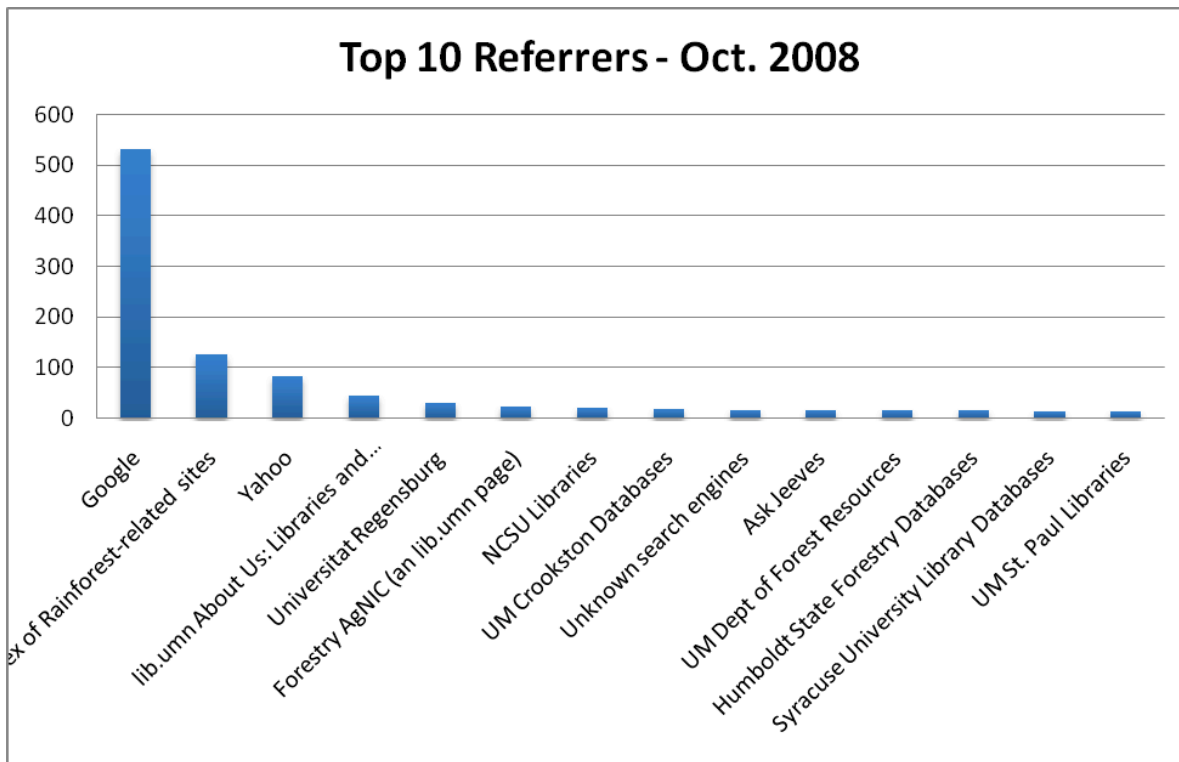
715 different keyphrases	Search	Percent
table of ranks	31	3.2 %
full text girl of the period linton	11	1.1 %
frances calderon de la barca	11	1.1 %
the table of ranks	11	1.1 %
christine de pizan biography	10	1 %
electronic text	9	0.9 %
women travel writers	8	0.8 %
mary french sheldon	7	0.7 %
travel writing	7	0.7 %
travbio	6	0.6 %
Other phrases	839	88.3 %

### *forestry.lib.umn.edu Website Statistics*

This set of statistics covers the Forestry subdomain (forestry.lib.umn.edu). Statistics reviewed here are for the month of October 2008 and come from the Libraries' AWStats program, which ingests and parses Apache web server logs.

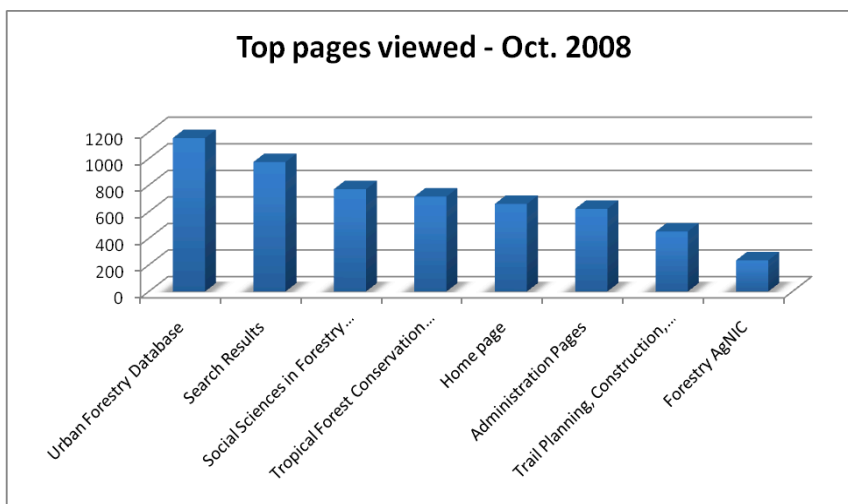
#### **Referrers**

71% of those coming to the Forestry subdomain do so through a direct link. Of the rest, Google accounts for 55% of the traffic to the site.



### Pages Viewed

Most pages viewed on the Forestry subdomain are the databases maintained by the Forestry Library.



### Keywords/Key Phrases

Looking at the top ten key phrases, it appears that a large number of the people finding the Forestry Library are looking for its databases. The phrase *Urban Forestry* leads the list at 25.1%. That's consistent with the number one page viewed within the site, the Urban Forestry Database page.

305 different keyphrases	Search	Percent
urban forestry	162	25.1 %
forestry	28	4.3 %
forestry conservation	22	3.4 %
urban forest	17	2.6 %
forestry library	16	2.4 %
forest conservation	15	2.3 %
indian journal of forestry	9	1.3 %
tropical forest	9	1.3 %
trail maintenance	7	1 %
history of urban forestry	5	0.7 %

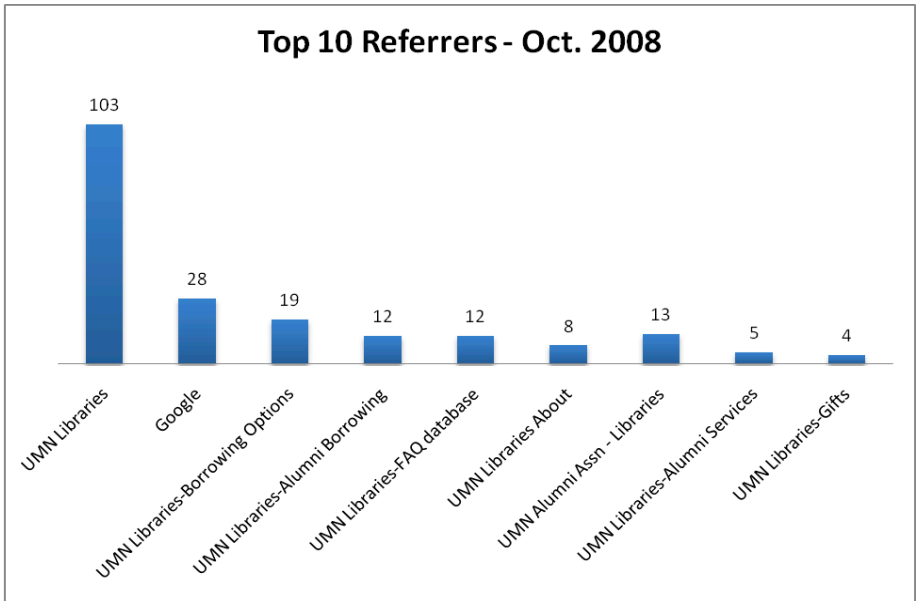
### *friends.lib.umn.edu Website Statistics*

This set of statistics covers the Friends subdomain (friends.lib.umn.edu). Statistics reviewed here are for the month of October 2008 and come from the Libraries' AWStats program, which ingests and parses Apache web server logs.

#### **Referrers**

Unlike most lib subdomains, most referrals to the Friends subdomain come from within the University. Note that total hits on the entire site are low: only 345 visitors (245 unique) for the month of October. October 2008 was among the highest month counts for the subdomain for all of 2008.

Consistent with the purpose of this subdomain, Google was the only external site sending any visitors; all other referrers are from the Libraries or the Alumni Association.



**Key Phrases**

Only 14% of referrals (just 28 searches) to the site come from Google. Most, if not all, appear to be searches for the site itself rather than searches for any particular content.

16 different keyphrases	Search	Percent
friends	4	14.2 %
libraries	4	14.2 %
the friends of university of minnesota libraries	3	10.7 %
friends of the university of minnesota libraries	3	10.7 %
univ of minn friends of the library	2	7.1 %
friends of the library	2	7.1 %
friends of the library university of minnesota	1	3.5 %
university of minnesota library friends	1	3.5 %
friends of the libraries	1	3.5 %
university of minnesota friends of the libraries	1	3.5 %
Other phrases	6	21.4 %
	28	

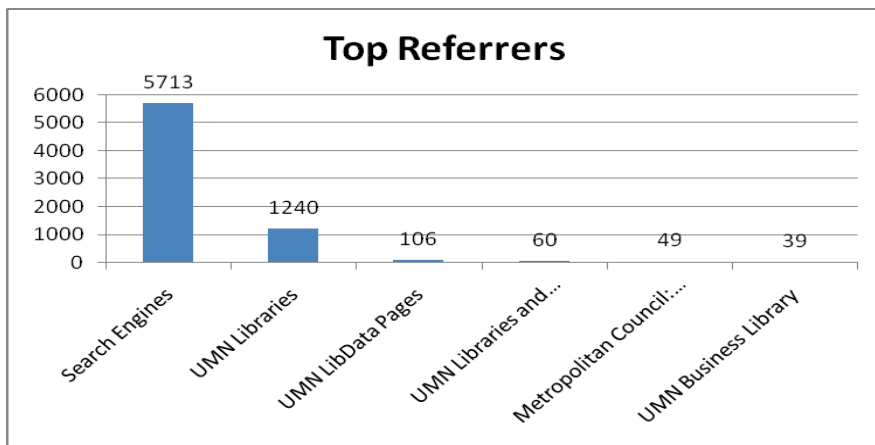
*govpubs.lib.umn.edu Website Statistics*

This set of statistics covers the Government Publications subdomain (govpubs.lib.umn.edu). Statistics reviewed here are for the month of October 2008 and come from the Libraries' AWStats program, which ingests and parses Apache web server logs.



## Referrers

Nearly 80% of the visitors referred to the Government Publications subdomain from another Web site come through a search engine.



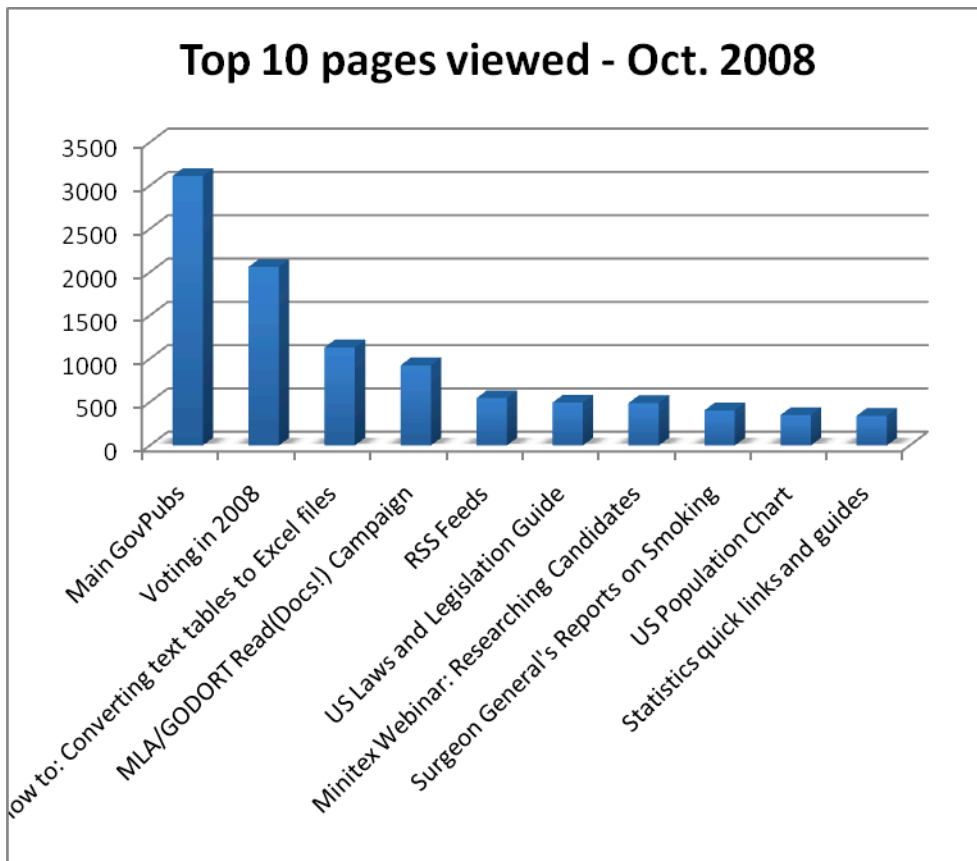
## Keywords/Key Phrases

Within the large number key phrases used to get to the Gov Pubs subdomain, very few appear to be searches for the Gov Pubs Library itself. The table shows a sample of the phrases used.

3033 different keyphrases	Search	Percent
money statistics	99	1.7 %
housing statistics	94	1.6 %
txt to excel	86	1.5 %
national trade data bank	71	1.2 %
inflation statistics	67	1.2 %
federal register citation	53	0.9 %
voting 2008	53	0.9 %
us population chart	49	0.8 %
convert notepad to excel	47	0.8 %
us population graph	44	0.7 %
Other phrases	4879	88 %

## Pages Viewed

Although the main GovPubs page draws the most views overall, many of the other pages shown on the chart are either discoverable material themselves (Voting in 2008) or sources for the user to find what they're looking for (a conduit to discovery?).

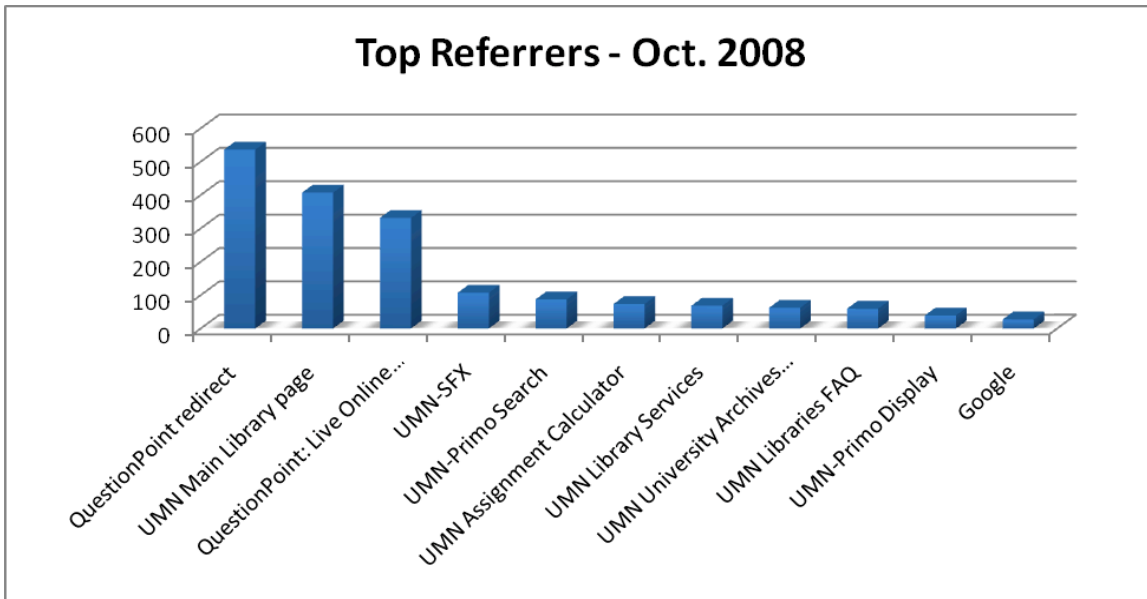


### *infopoint.lib.umn.edu Website Statistics*

This set of statistics covers the InfoPoint subdomain (infopoint.lib.umn.edu). Statistics reviewed here are for the month of October 2008 and come from the Libraries' AWStats program, which ingests and parses Apache web server logs.

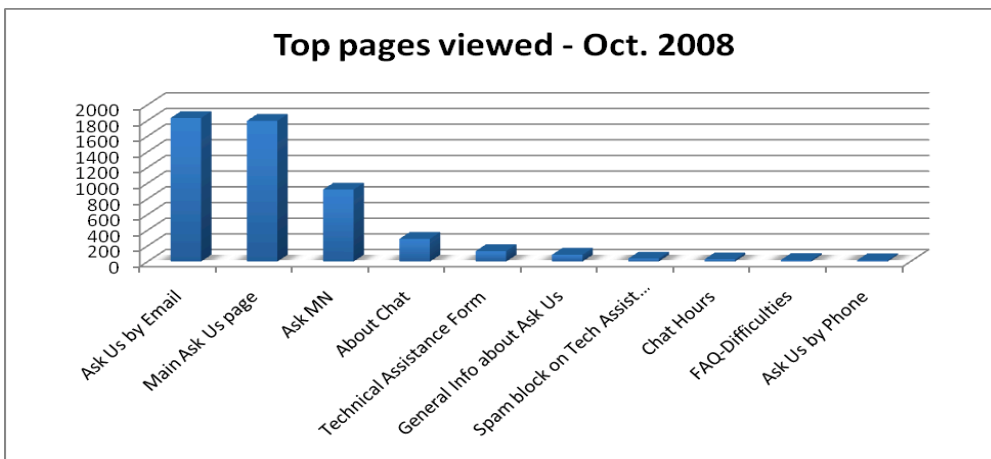
#### **Referrers**

The InfoPoint subdomain gets more traffic from direct address links (email, bookmarks) than from referrers, and almost no traffic from search engines. Most referrals come from University Libraries pages.



### Pages Viewed

Although not particularly relevant for discovery, the top page views are shown below.



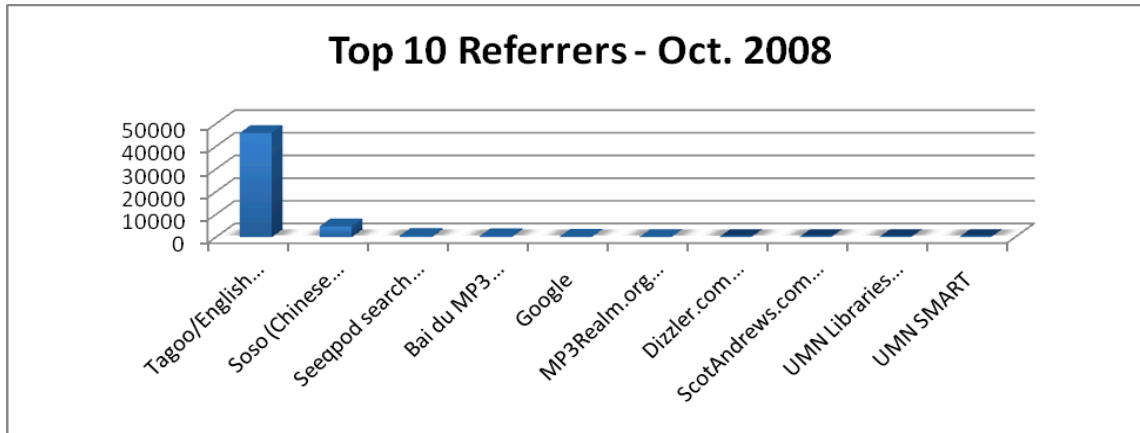
### *Irc.lib.umn.edu Website Statistics*

This set of statistics covers the Learning Resource Center subdomain ([irc.lib.umn.edu](http://irc.lib.umn.edu)). Statistics reviewed here are for the month of October 2008 and come from the Libraries' AWStats program, which ingests and parses Apache web server logs.

Although the main LRC page now redirects to Library Media Services ([lib.umn.edu/media](http://lib.umn.edu/media)), the LRC subdomain still hosts content such as audio recordings of lectures (English 3020 is most prominent) and audio language lessons (Pushto recordings and lessons from a 1965 Peace Corps project).

## Referrers

Sometime before October of 2008, the Russian media search engine called Tagoo indexed the English 3020 Science Fiction and Fantasy lectures in its database. The resulting hits—downloads of the lecture MP3s—seriously skew the results of this analysis, but do make a case that people come to our sites through referrals.



The data in table form, with a bit more information on referring sites:

Referrer Name	Pages
Tagoo/English (Russian media search site)	45989
Soso (Chinese media search site)	4811
Seeqpod search and recommendation engine	591
Bai du MP3 (Chinese?)	401
Google	235
MP3Realm.org (audio search engine)	99
Dizzler.com (media search and player)	73
ScotAndrews.com (blog post recommending lectures in June 2004)	58
UMN Libraries and Collections	50
UMN SMART	49

Looking back at the months before Tagoo, visits from media recommendation sites still far outnumbered visits from search engines or any internal site.

## Pages Viewed

Nearly all pages viewed were actually downloads of MP3 files. For October, the ENGL 3020 MP3s dominate.

## Key Phrases Used

Although visits via a traditional search engine (e.g., Google) accounted for little overall traffic, it's worth noting that a large percentage of those who found the site were looking for tools for learning Pashto (Pushto, Pashtun), the national language of Afghanistan.

146 different keyphrases	Search	Percent
--------------------------	--------	---------

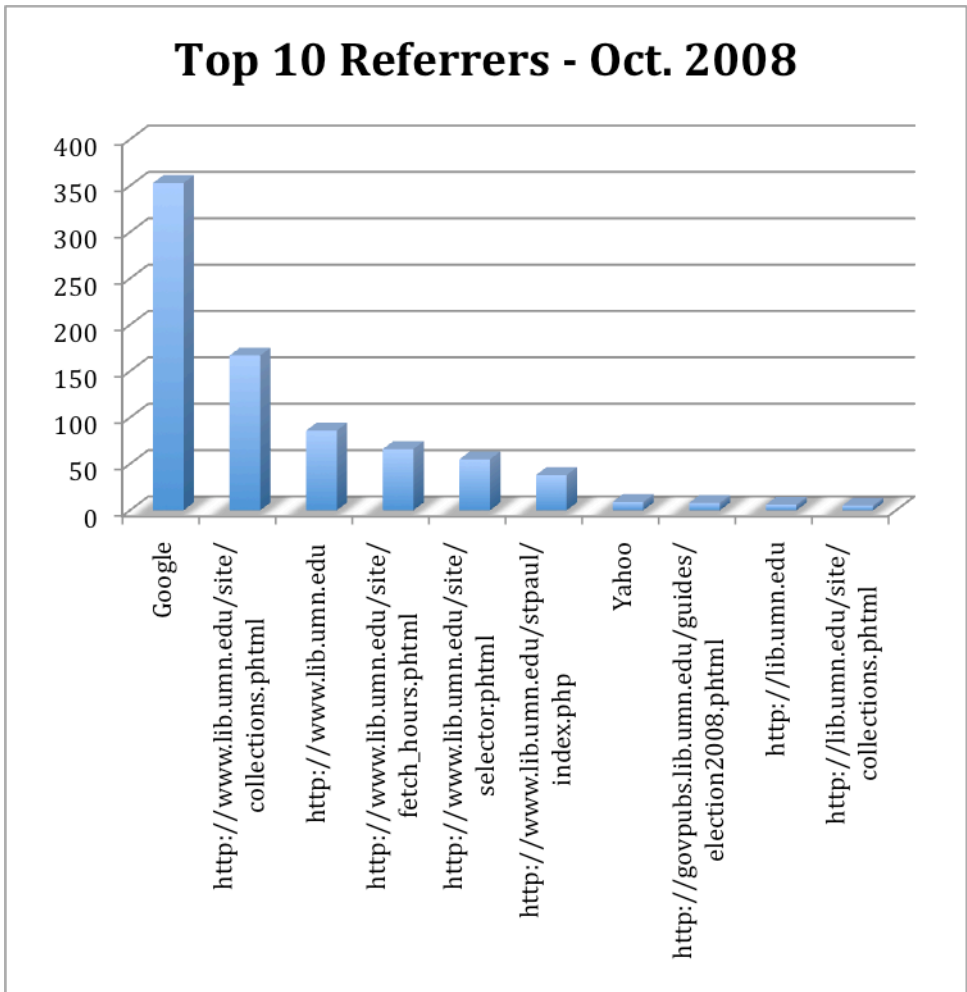
pushto	29	11 %
learning pashto	23	8.7 %
learning pushto	7	2.6 %
pashtun language	7	2.6 %
punjabi phonetics	6	2.2 %
pashto learning	6	2.2 %
smart learning center	6	2.2 %
walter smart learning commons	5	1.9 %
punjabi conversation	5	1.9 %
umn learning resource center	4	1.5 %
Other phrases	164	62.5 %

### *magrath.lib.umn.edu Website Statistics*

This set of statistics covers pages served on the magrath.lib.umn.edu sub domain, which contains pages related to Magrath Library on the St. Paul campus. Data was gathered from the Libraries' AWStats program, which ingests and parses Apache web server logs.

#### **Referrers**

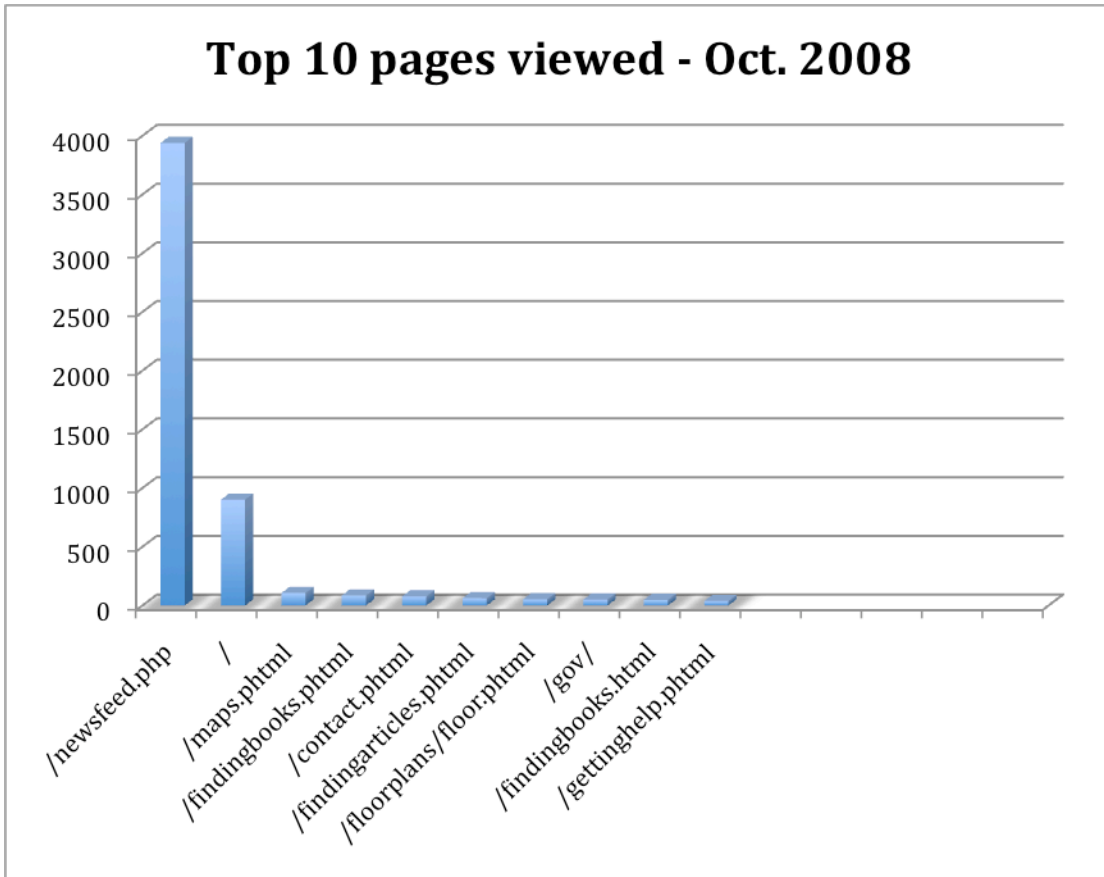
This chart shows referring domains for visits to all pages on the magrath.lib.umn.edu sub domain in October, 2008. In this chart I have combined search engine and web page referrals for those visits that log a referrer. As with many of our other sub domains, Google is the single most prominent referrer of traffic to this site.



**Page views**

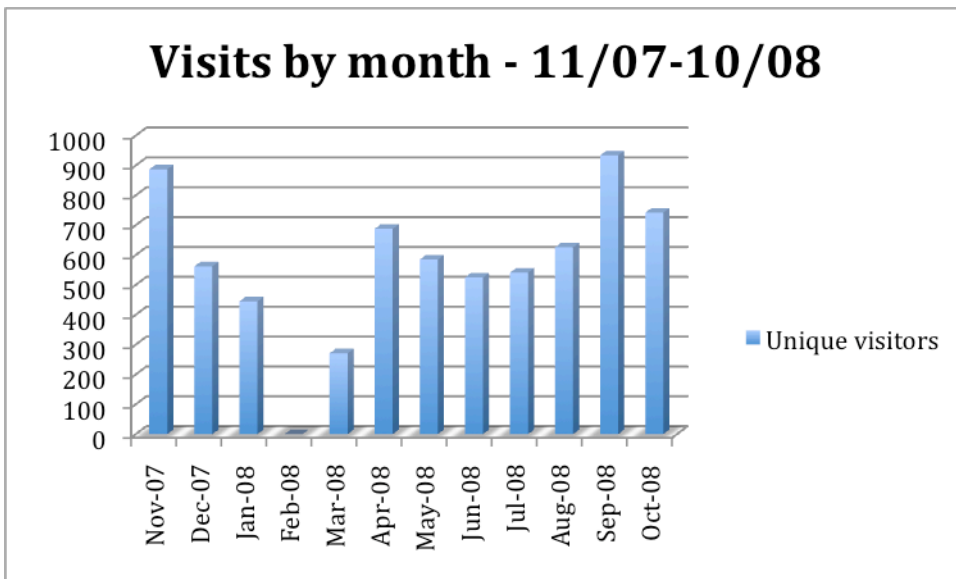
The following chart shows the ten most viewed pages on the magrath.lib.umn.edu sub domain. Magrath.lib has a newsfeed which receives the most traffic. Unlike other sites, the magrath.lib homepage is the second most popular page. The newsfeed is hit ~4,000 times per month and presents a nice model for other sites. However, after these two pages traffic drops off significantly.

## Top 10 pages viewed - Oct. 2008



### Monthly visitors

This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. This site follows a similar pattern to other subdomains in that traffic mirrors the rhythm of the school year.



## **Conclusions**

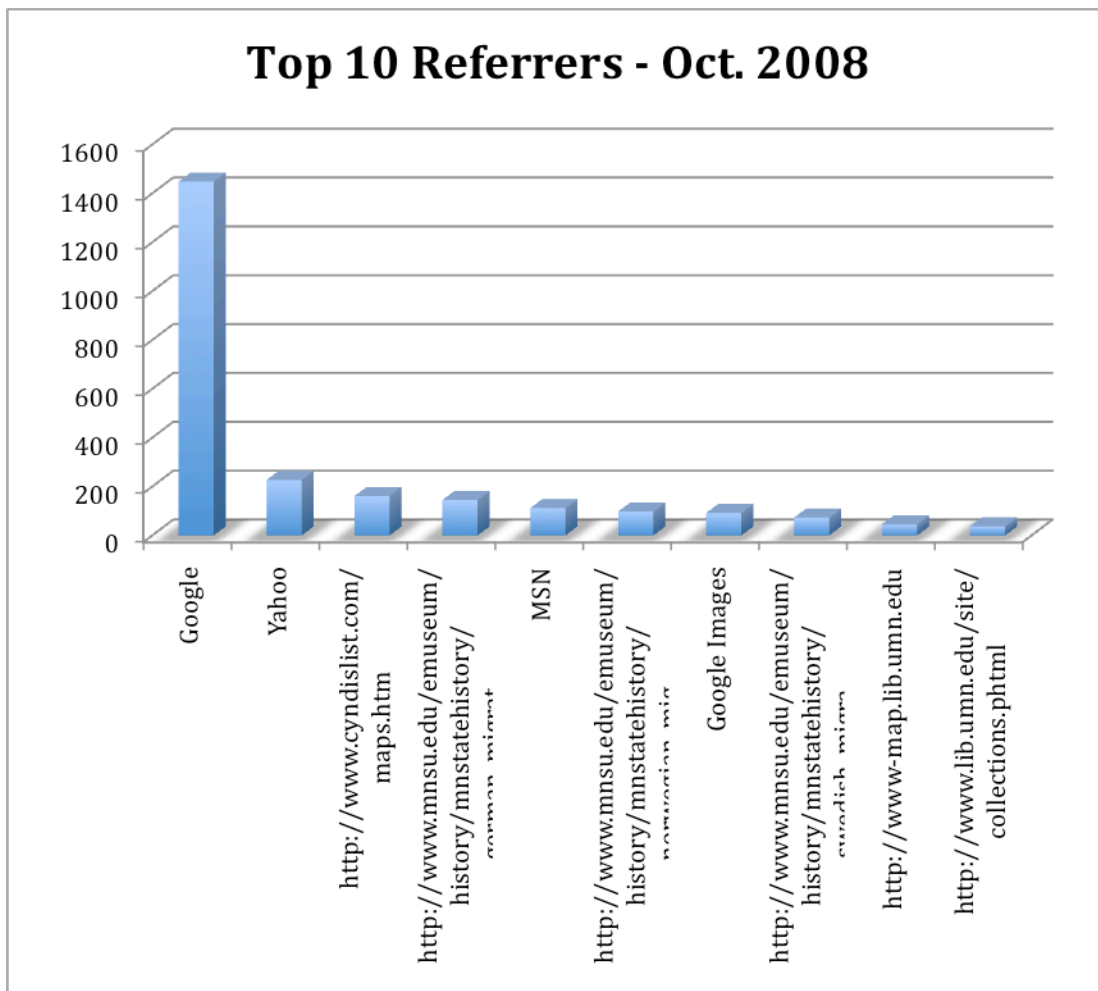
Traffic to the magrath.lib.umn.edu sub domain is largely search-engine driven, and generally restricted to the site's newsfeed and home page. Visits to pages within this site are infrequent. However, the newsfeed does present a dynamic and trafficked resource that bucks the trend shown by most of the other library subdomains.

## ***map.lib.umn.edu Website Statistics***

This set of statistics covers pages served on the map.lib.umn.edu subdomain, which primarily contains pages related to the Borchert Map Library. Data was gathered from the Libraries' AWStats program, which ingests and parses Apache web server logs.

## **Referrers**

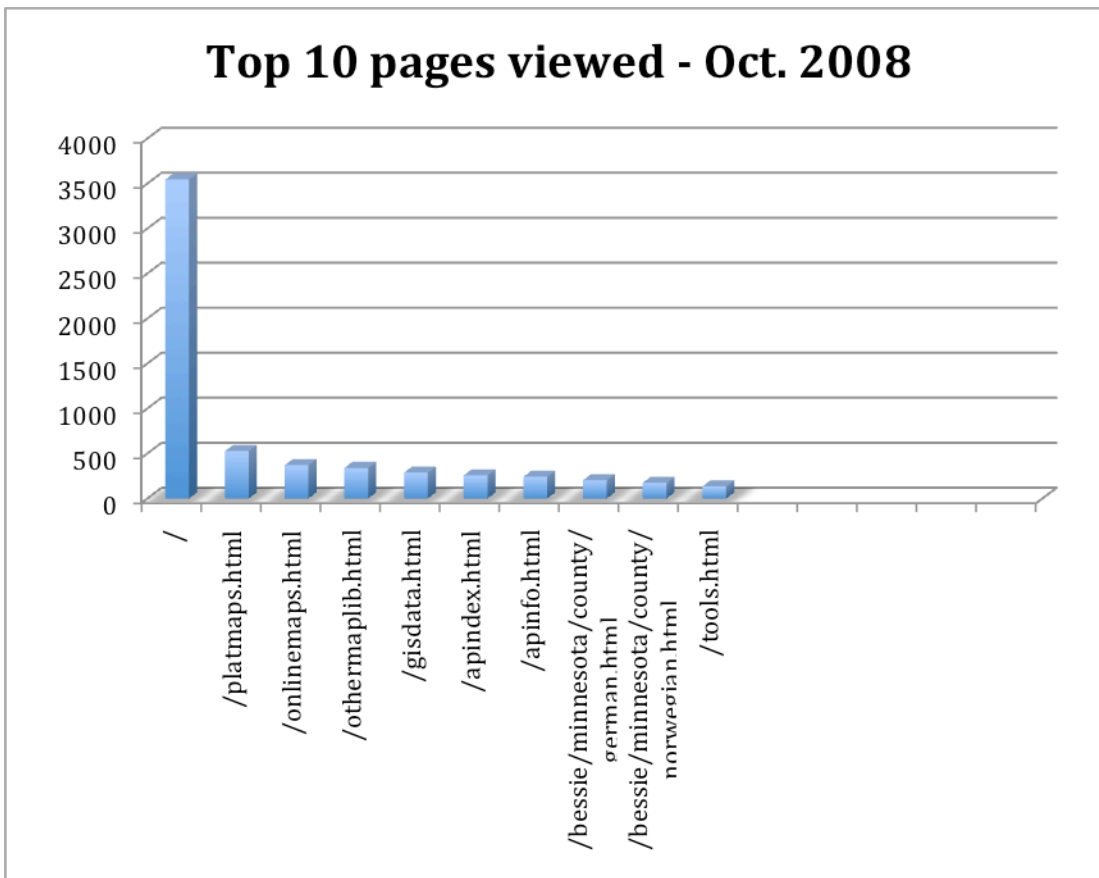
This chart shows referring domains for visits to all pages on the map.lib.umn.edu subdomain in October, 2008. In this chart I have combined search engine and web page referrals for those visits that log a referrer. As with many of our other subdomains, Google is the single most prominent referrer of traffic to this site.





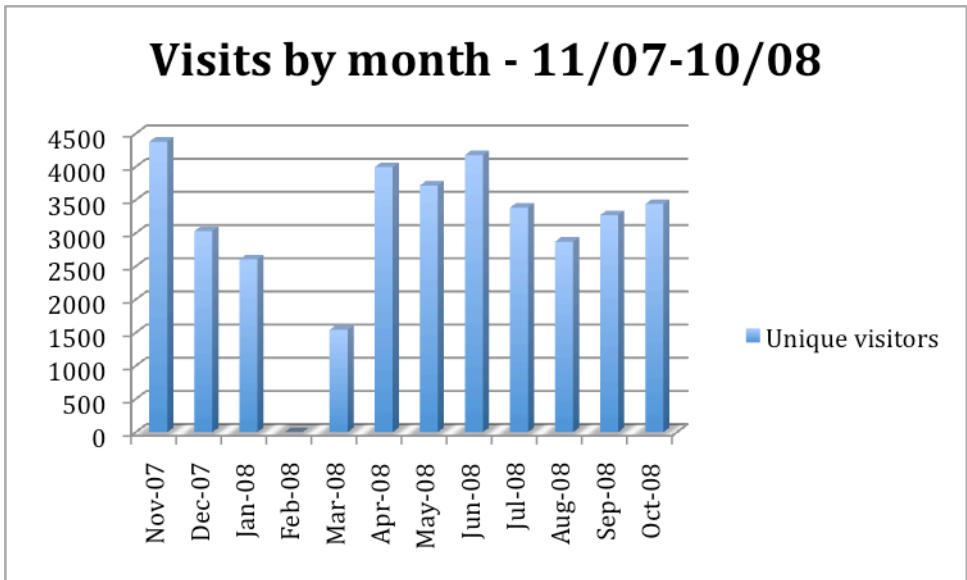
### Page views

The following chart shows the ten most viewed pages on the map.lib.umn.edu subdomain. As with many of our subdomains, the site's home page receives many more views than any other page on the site.



### Monthly visitors

This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. Map.lib receives traffic year round and site load doesn't seem as centered around the school year as is the case in many other library subdomains.



**Conclusions**

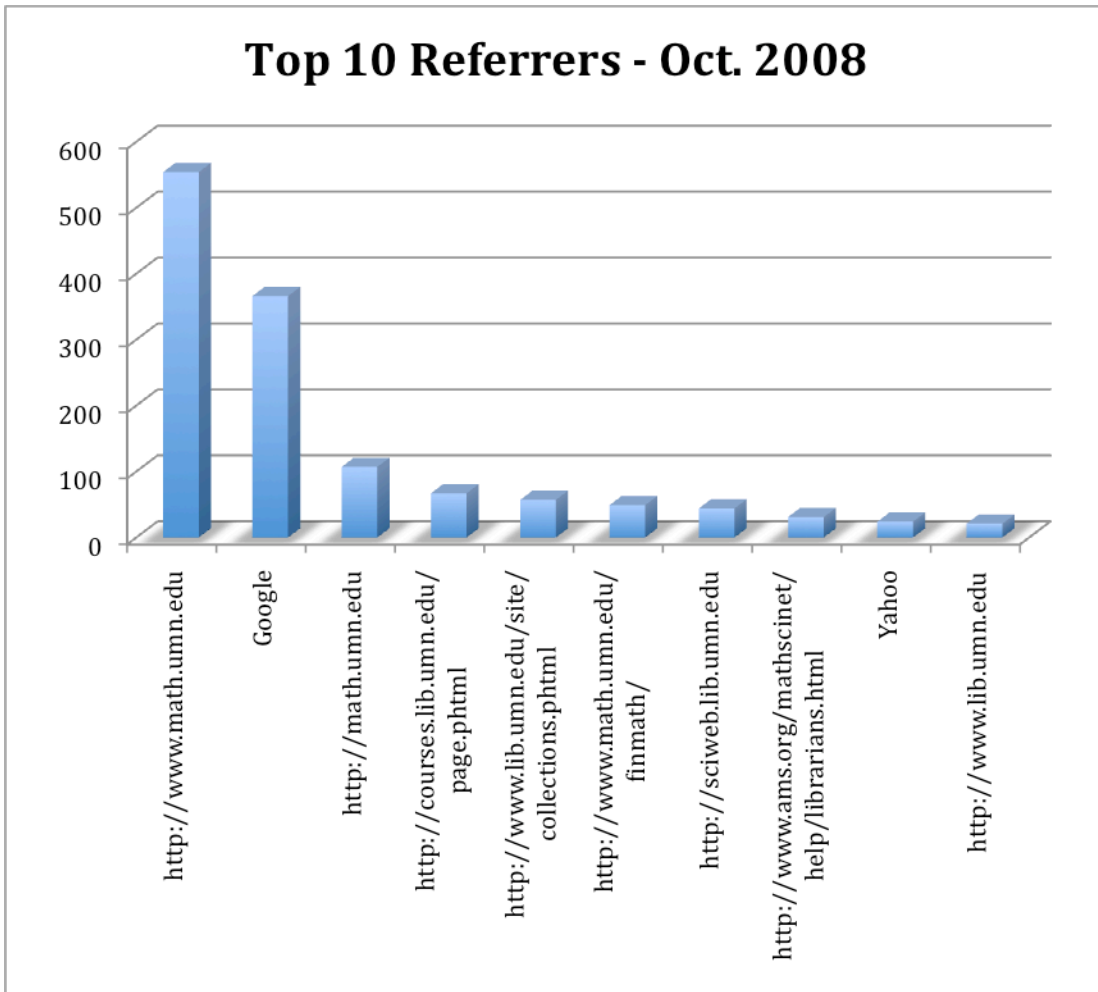
Traffic to the map.lib.umn.edu subdomain is largely search-engine driven, and generally restricted to the site’s home page. The site is used year round.

*math.lib.umn.edu Website Statistics*

This set of statistics covers pages served on the math.lib.umn.edu subdomain, which contains pages related Mathematics Library. The data was gathered from the Libraries’ AWStats program, which ingests and parses Apache web server logs.

**Referrers**

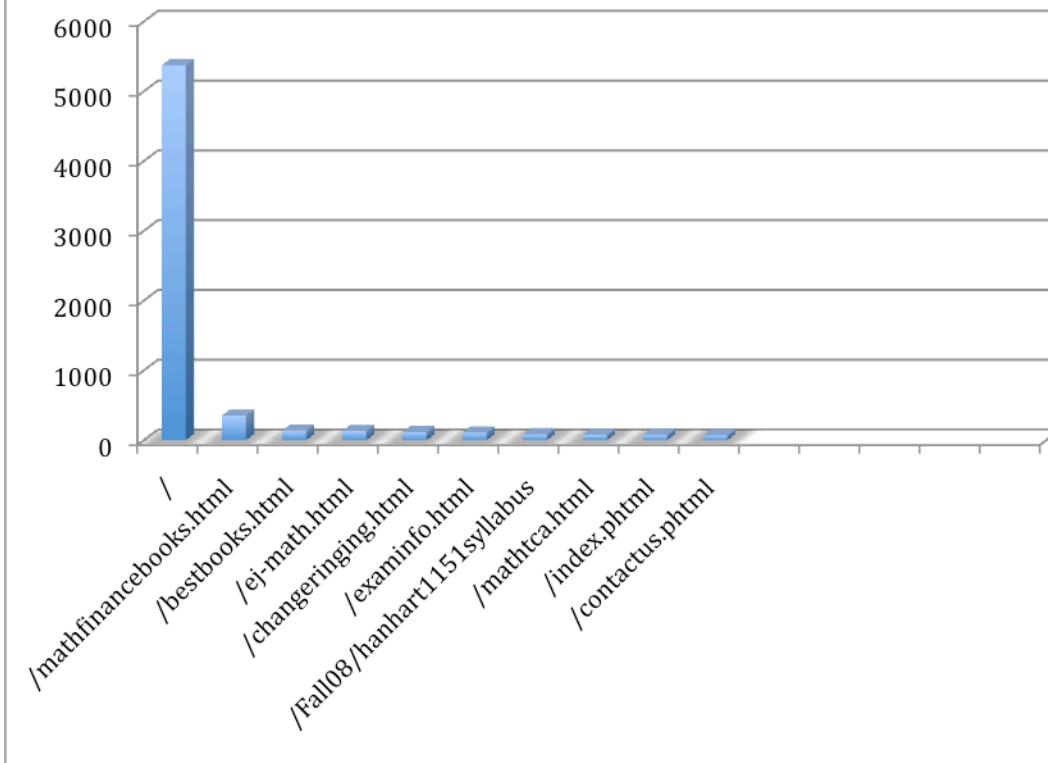
This chart shows referring domains for visits to all pages on the math.lib.umn.edu subdomain in October, 2008. In this chart I have combined search engine and web page referrals for those visits that log a referrer. This math.lib site bucks the trend and Google comes in second as a referrer. The largest driver of traffic to the site is the main University of Minnesota Math Department page. There is a sharp drop-off after these two referrers.



**Page views**

The following chart shows the ten most viewed pages on the math.lib.umn.edu subdomain. As with many of our subdomains, the site’s home page receives nearly all the views on the site.

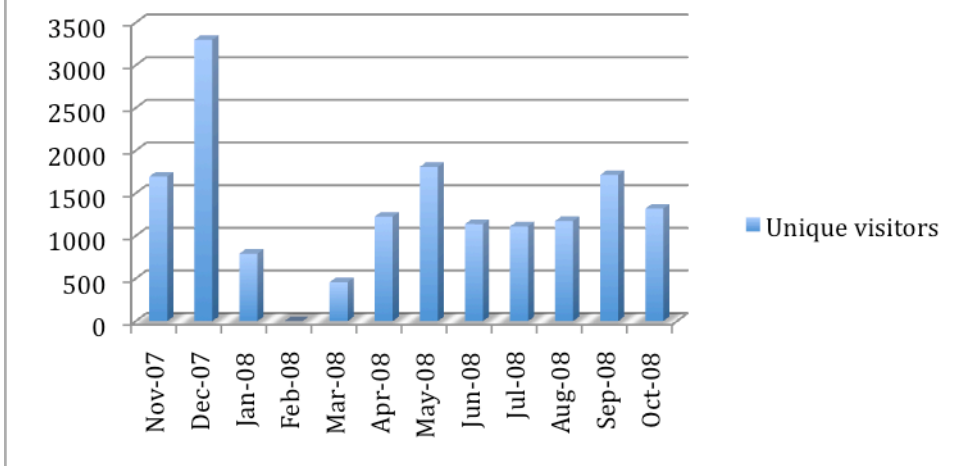
## Top 10 pages viewed - Oct. 2008



### Monthly visitors

This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. This site follows a similar pattern to other subdomains in that traffic mirrors the rhythm of the school year.

## Visits by month - 11/07-10/08



## **Conclusions**

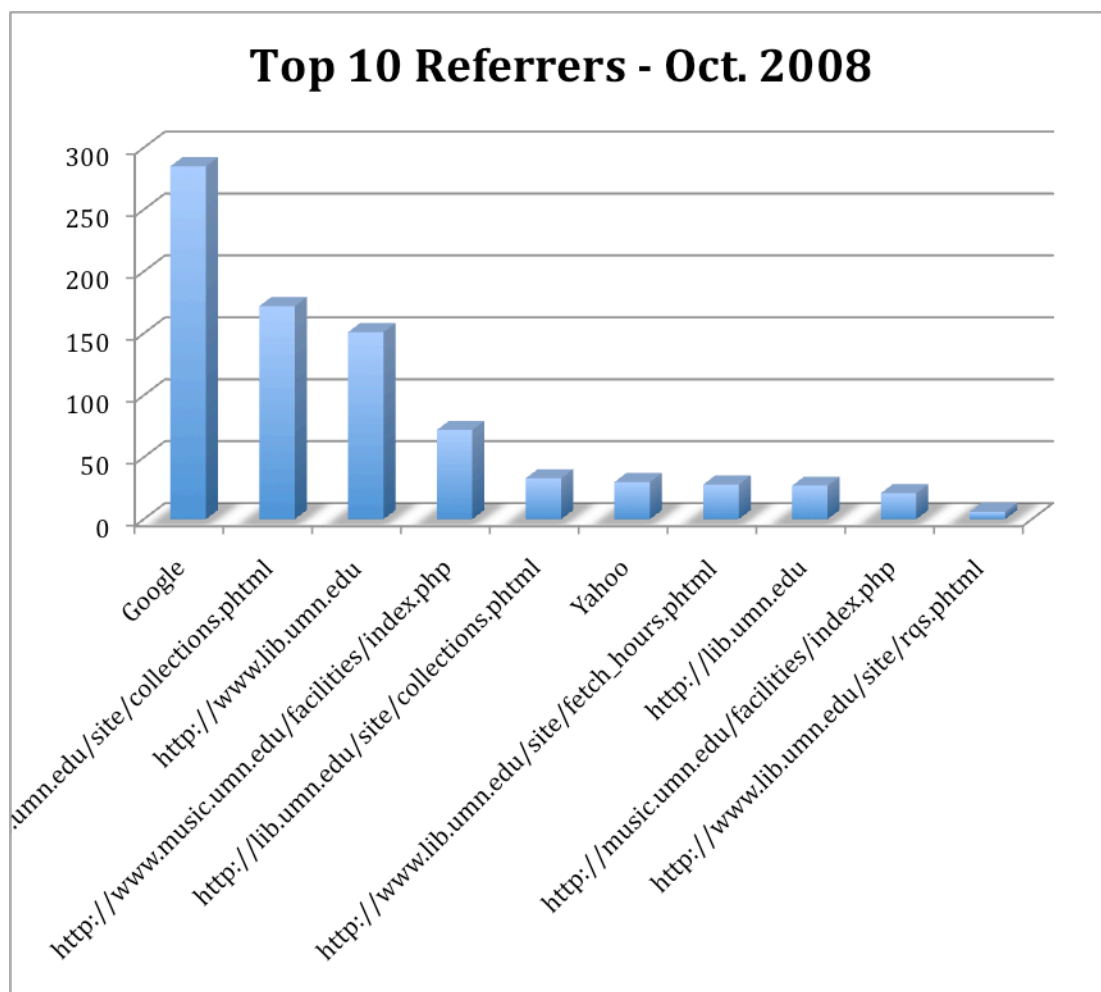
Traffic to the math.lib.umn.edu subdomain is search-engine driven, and generally restricted to the site's home page. However University department and course pages do push traffic to the site.

## ***music.lib.umn.edu Website Statistics***

This set of statistics covers pages served on the music.lib.umn.edu subdomain, which contains pages related to the Music Library.

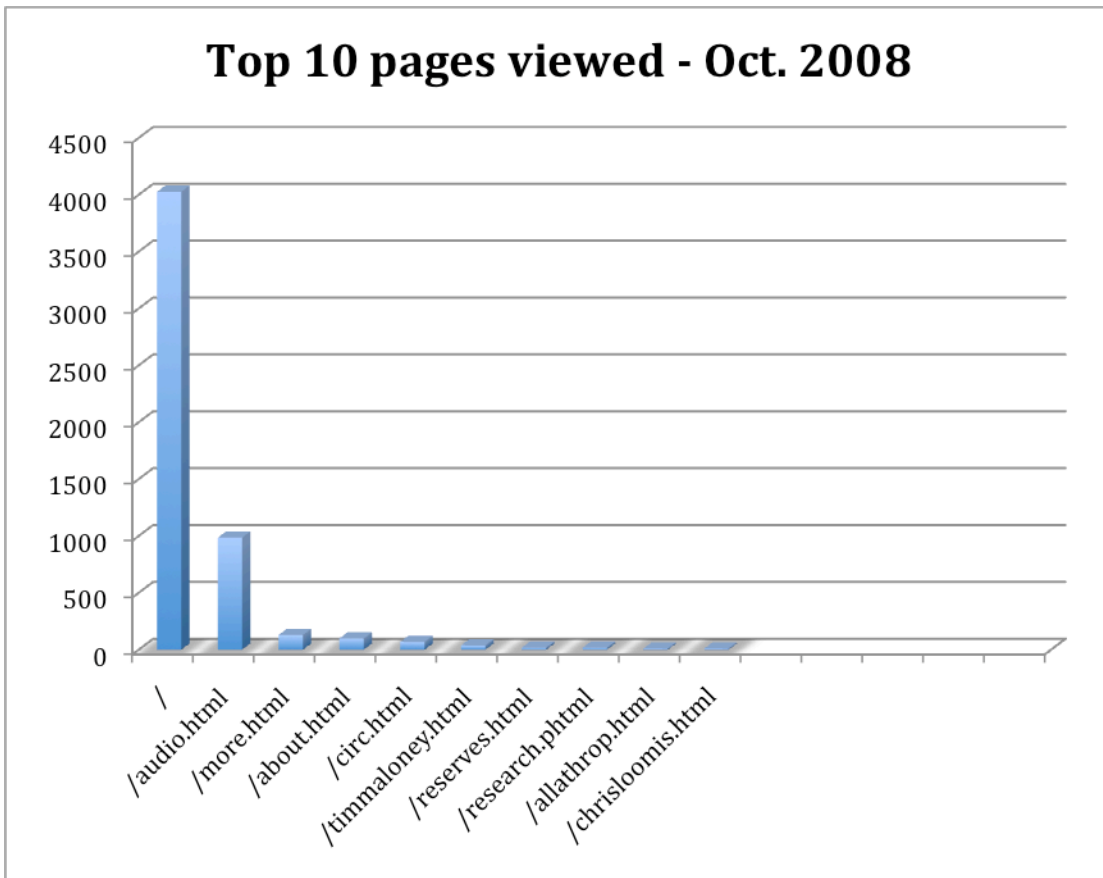
## **Referrers**

This chart shows referring domains for visits to all pages on the music.lib.umn.edu subdomain in October, 2008. In this chart I have combined search engine and web page referrals for those visits that log a referrer. As with many of our other subdomains, Google is the single most prominent referrer of traffic to this site.



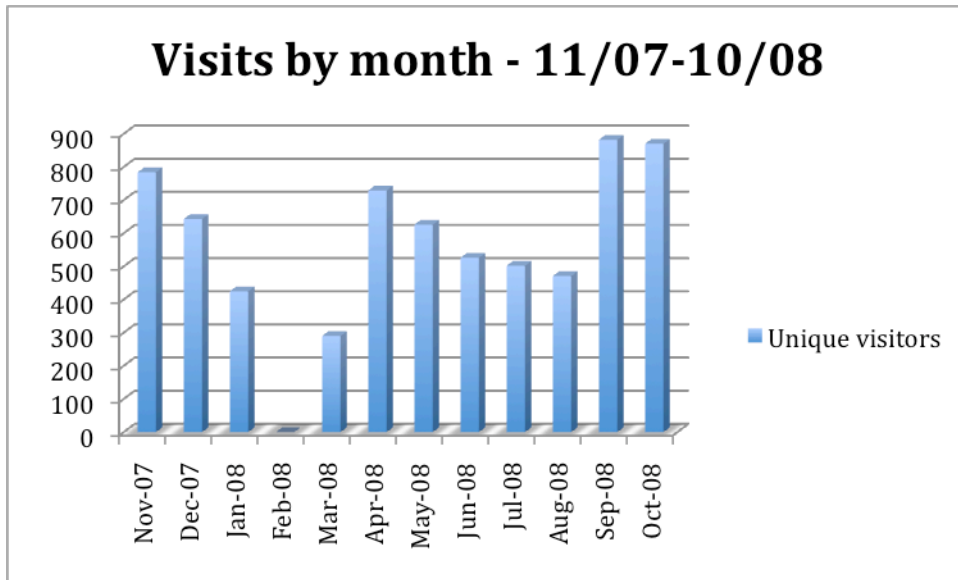
### **Page views**

The following chart shows the ten most viewed pages on the music.lib.umn.edu subdomain. As with many of our subdomains, the site's home page receives many more views than any other page on the site.



### **Monthly visitors**

This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. This site follows a similar pattern to other subdomains in that traffic mirrors the rhythm of the school year. There is some missing data in February and March, 2008.



### **Conclusions**

Traffic to the music.lib.umn.edu subdomain is largely search-engine driven, and generally restricted to the site’s home page. Visits to pages within this site are infrequent, excepting the audio page.

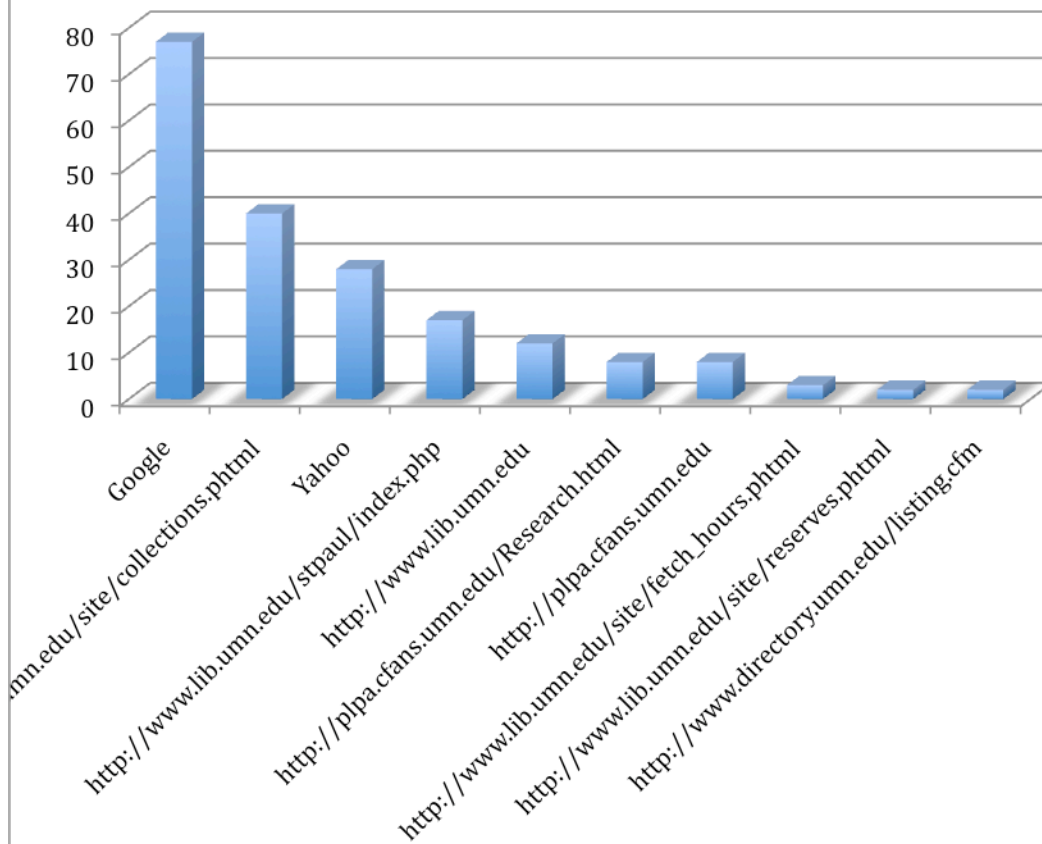
### *plant.lib.umn.edu Website Statistics*

This set of statistics covers pages served on the plant.lib.umn.edu subdomain, which primarily contains pages related to the Plant Pathology Library.

### **Referrers**

This chart shows referring domains for visits to all pages on the plant.lib.umn.edu subdomain in October, 2008. In this chart I have combined search engine and web page referrals for those visits that log a referrer. As with many of our other subdomains, Google is the single most prominent referrer of traffic to this site.

## Top 10 Referrers - Oct. 2008

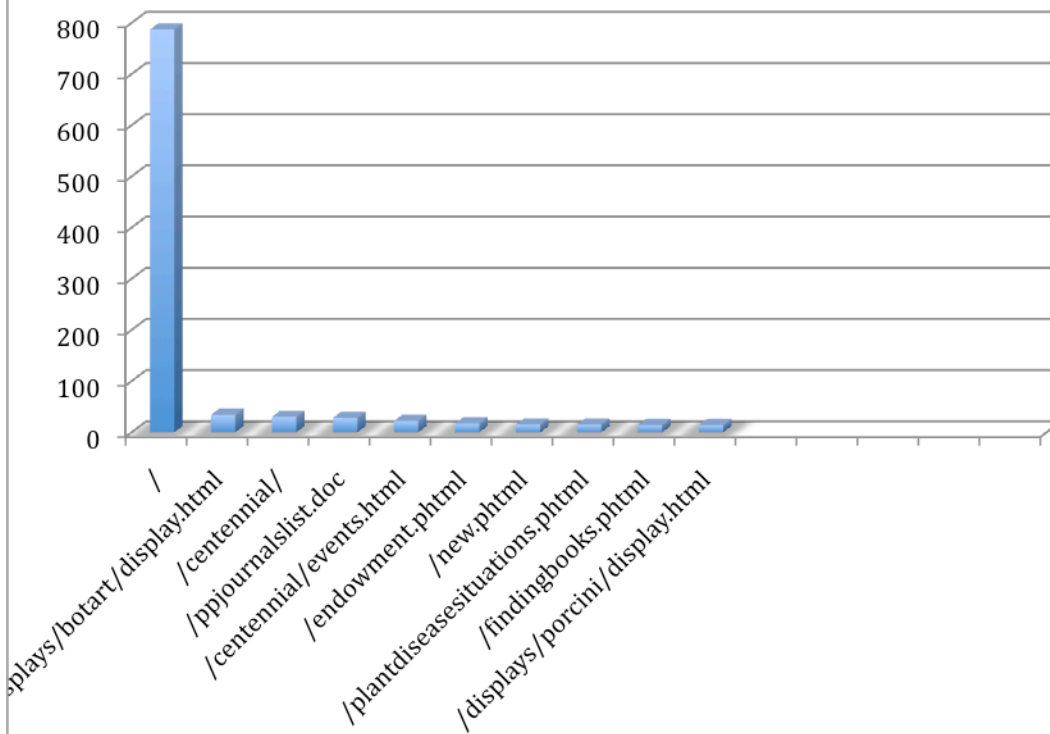


### Page views

The following chart shows the ten most viewed pages on the plant.lib.umn.edu subdomain. As with many of our subdomains, the site's home page receives the vast majority of traffic to the site with a stunning drop-off in the secondary pages.



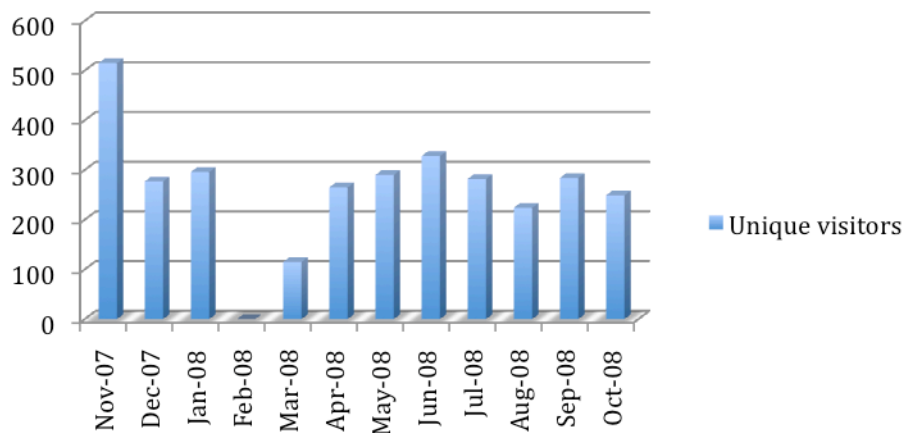
## Top 10 pages viewed - Oct. 2008



## Monthly visitors

This chart shows unique visitors per month from November, 2007 through October, 2008. Data from February and a portion of March are missing. This site follows a similar pattern to other subdomains in that traffic mirrors the rhythm of the school year.

## Visits by month - 11/07-10/08



## **Conclusions**

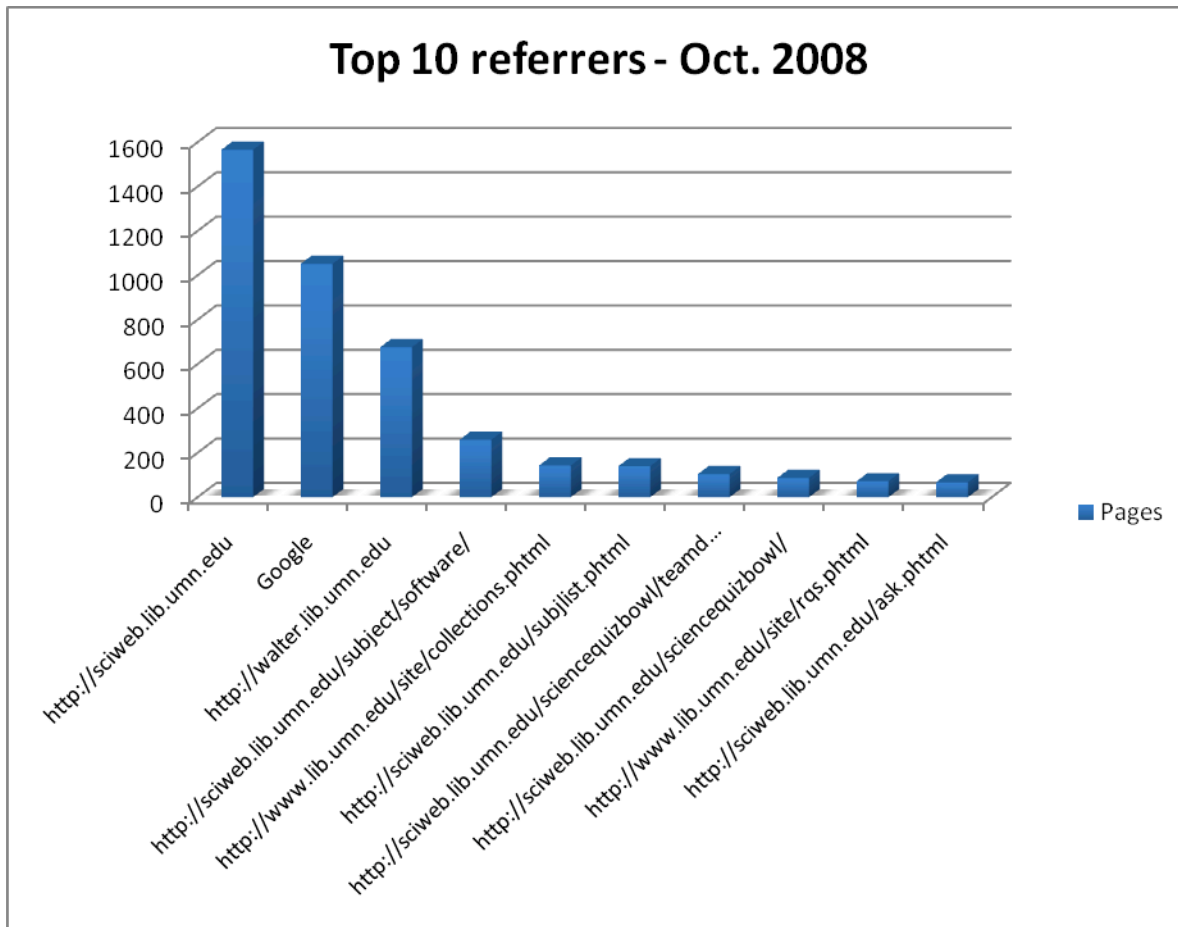
Traffic to the plant.lib.umn.edu subdomain is largely search-engine driven, and generally restricted to the site's home page.

### *sciweb.lib.umn.edu Website Statistics*

This set of statistics covers the Science/Engineering subdomain (sciweb.lib.umn.edu). Statistics reviewed here are for the month of October 2008 and come from the Libraries' AWStats program, which ingests and parses Apache web server logs.

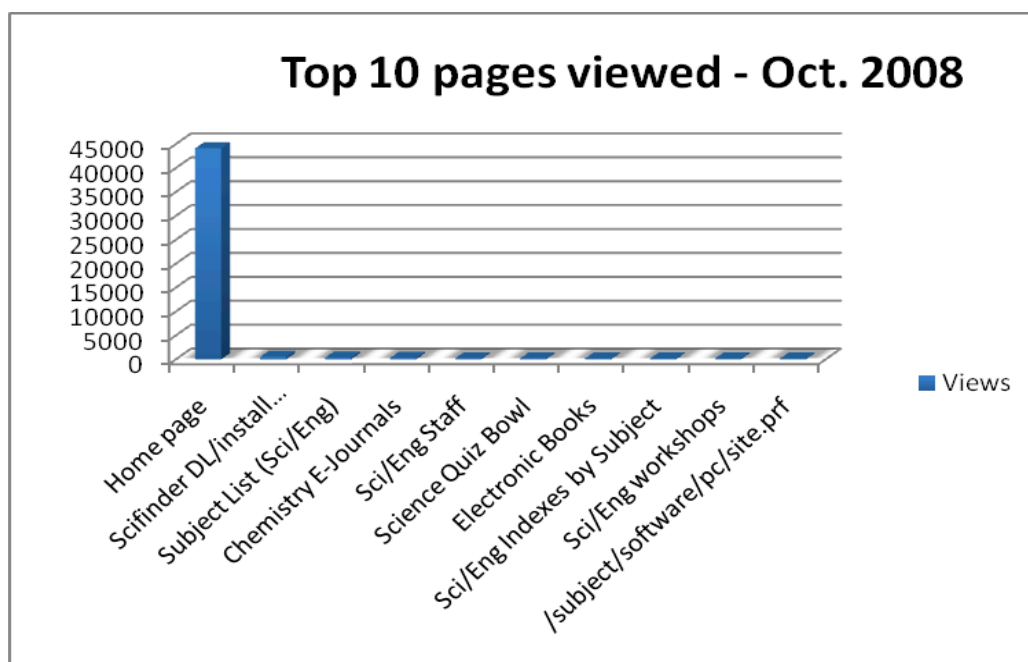
## **Referrers**

Most users arrive via a direct link. Google is the next highest referrer, followed by the Walter Library home page. There is a sharp drop off in the number of referrals following those three. The remainder of the top ten referrers are an assortment of pages within the sciweb subdomain, along with a couple of lib.umn.edu pages.



## Pages Viewed

By far, the most viewed page on the Sci/Eng subdomain is the home page. All other pages in the subdomain have a very small number of views compared to the home page.



Page	Views
Home page	44145
Scifinder DL/install instructions	535
Subject List (Sci/Eng)	375
Chemistry E-Journals	299
Sci/Eng Staff	218
Science Quiz Bowl	200
Electronic Books	185
Sci/Eng Indexes by Subject	183
Sci/Eng workshops	174
/subject/software/pc/site.prif	163

## Keywords/Key Phrases

Most searches that lead to the Sci/Eng subdomain include the keywords “walter library.” Other keywords and phrases occurring with some frequency include “scifinder” and “science quiz bowl,” which is consistent with the place of the pages for Scifinder and Science Quiz Bowl among the top ten in this subdomain.

565 different keyphrases	Search	Percent
walter library	129	11.2%

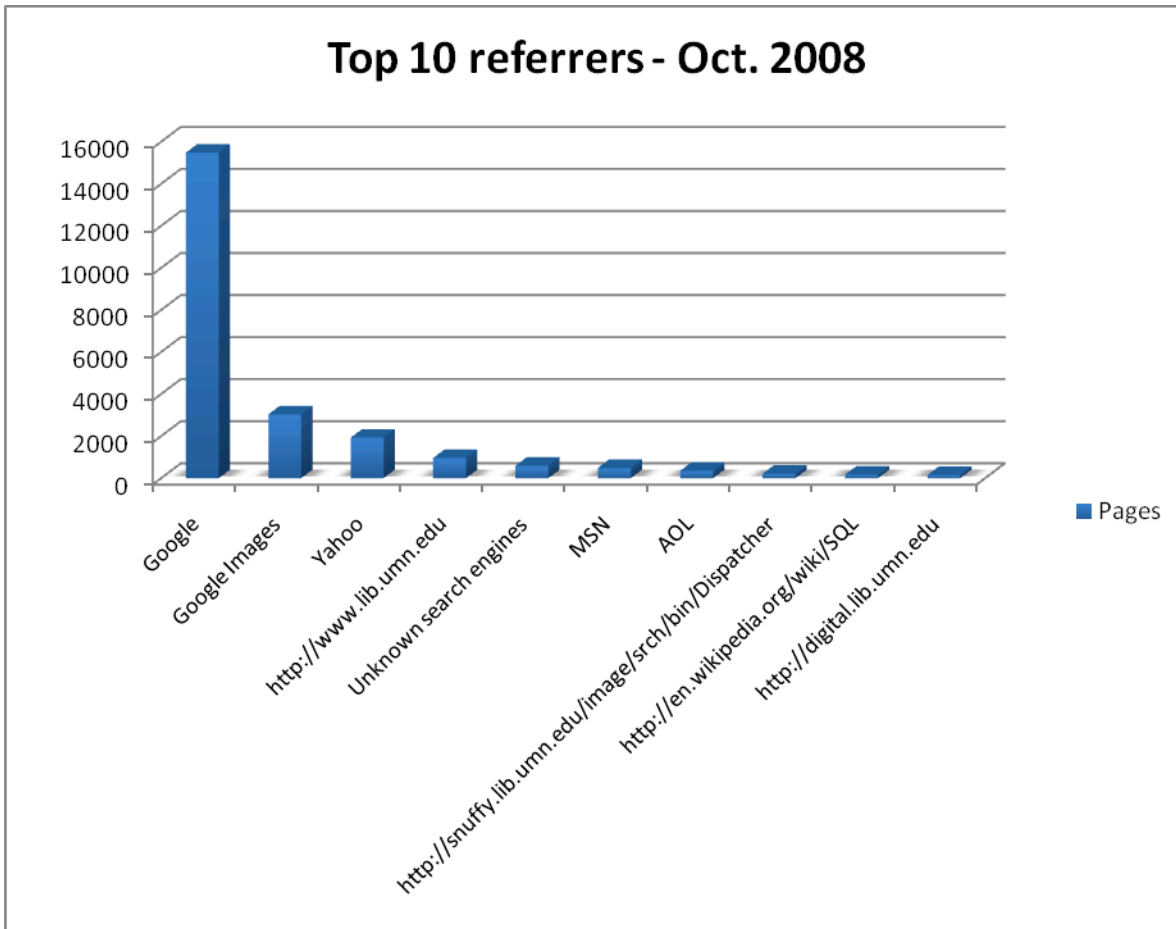
walter library umn	97	8.4%
walter library university of minnesota	60	5.2%
umn walter library	32	2.7%
scifinder	31	2.7%
science quiz	18	1.5%
science quiz bowl	13	1.1%
science quiz bowl questions	12	1%
Scifinder scholar	11	0.9%

### *special.lib.umn.edu Website Statistics*

This set of statistics covers the Archives & Special Collections subdomain (special.lib.umn.edu). Statistics reviewed here are for the month of October 2008 and come from the Libraries' AWStats program, which ingests and parses Apache web server logs.

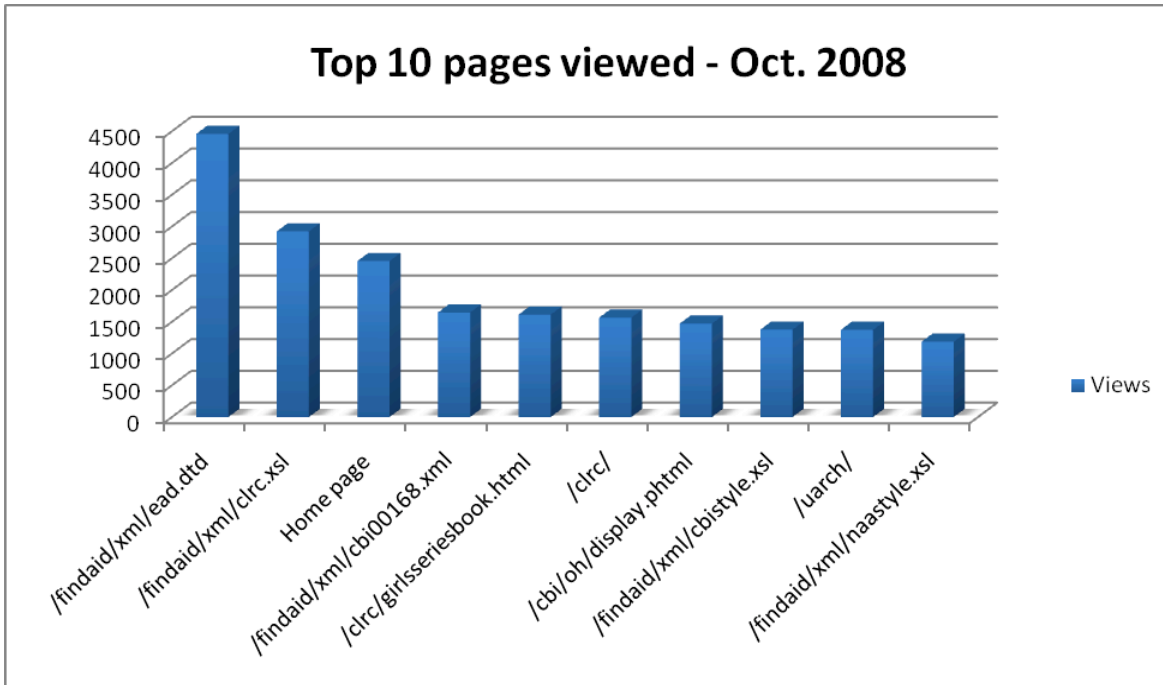
#### **Referrers**

The vast majority of users arrive via Google, which has more than three times the number of referrals as the number two referrer, Google Images. Several other search engines are also represented in the top ten. The Libraries' home page is fourth on the list. IMAGES and the Digital Collections home page also make the top ten, though with relatively small numbers. Interestingly, a link to a finding aid in the Wikipedia article on SQL is the ninth most popular referrer (the article links to the finding aid for ANSI X3H2 Records, 1978-1995, which was also the fourth most viewed page during October 2008).



**Pages Viewed**

The most frequently viewed page during October 2008 was the EAD DTD. This indicates that archival finding aids were the most viewed pages, since the DTD file must be accessed any time a finding aid is viewed. Several XML style sheets also appear in the top 10 pages viewed. These stylesheet files are specific to each archival unit and called upon every time a finding aid from that unit is viewed. Thus that statistic indicates that CLRC, CBI, and Northwest Architectural Archives finding aids are most often viewed. The main Archives and Special Collections home page had the third most views, while the home pages for the CLRC and University Archives also landed in the top ten.



### **Keywords/Key Phrases**

Of note here are the number and variety of key phrases that resulted in a link to the Archives and Special Collections subdomain. No single phrase or combination of similar phrases appears to dominate searches, with the top phrase “kerlan collection” only comprising 0.6% of searches during October 2008.

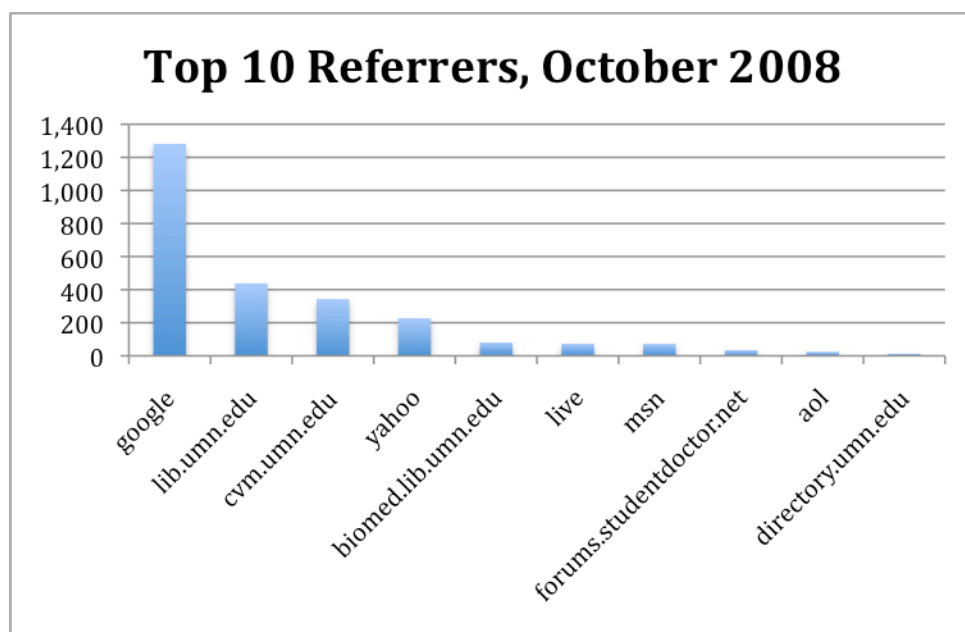
14720 different keyphrases		
	Search	Percent
kerlan collection	116	0.6%
university of minnesota archives	64	0.3%
ymca history	61	0.3%
ymca logos	60	0.3%
hygiene posters	58	0.3%
hygiene	57	0.3%
felice holman	57	0.3%
social welfare history archives	48	0.2%
archives	47	0.2%
ymca archives	40	0.2%
<i>Other phrases</i>	<i>17972</i>	<i>96.7%</i>

## *vetmed.lib.umn.edu Website Statistics*

This set of statistics covers pages served on the vetmed.lib.umn.edu subdomain, the site for the University's Veterinary Medical Library. The data here is from October 2008 and is drawn from the Health Science Libraries' web statistics as gathered using Google Analytics. This site and biomed.lib.umn.edu are the only of the Libraries' sites that currently use Google Analytics.

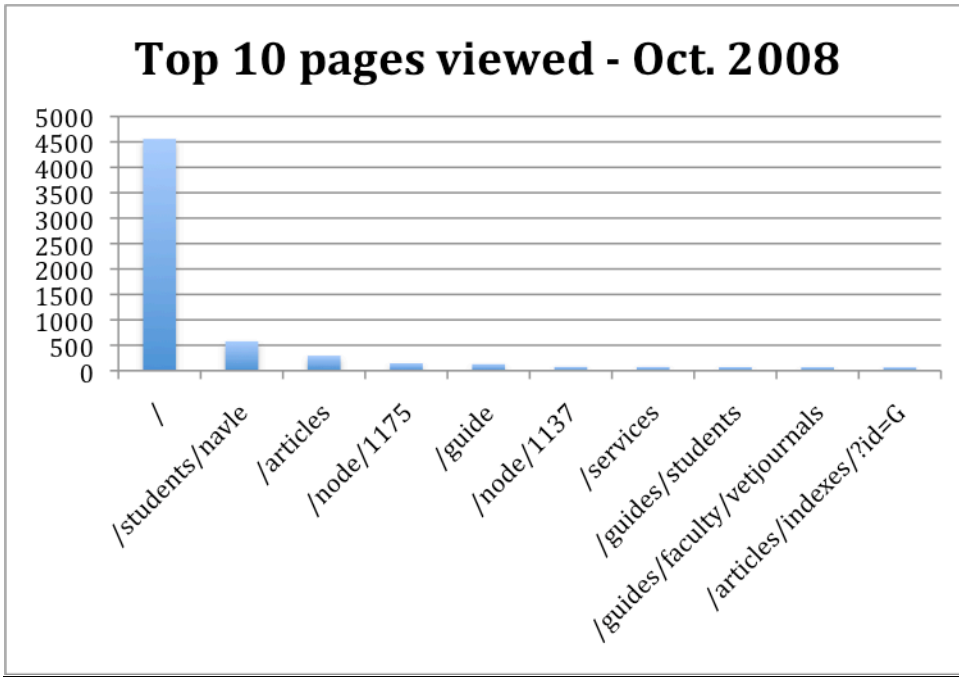
### **Referrers**

The chart below shows the ten sites that referred the most visitors to the VetMed site in October 2008. As with so many of our sites, Google is responsible for directing more users to this site than any other referrer. As with many of our branches, the Libraries' main web site is also a major source of traffic. It is gratifying to note that the third-place referrer during this sample period was the College of Veterinary Medicine's website, which likely indicates a strong library presence on that site.



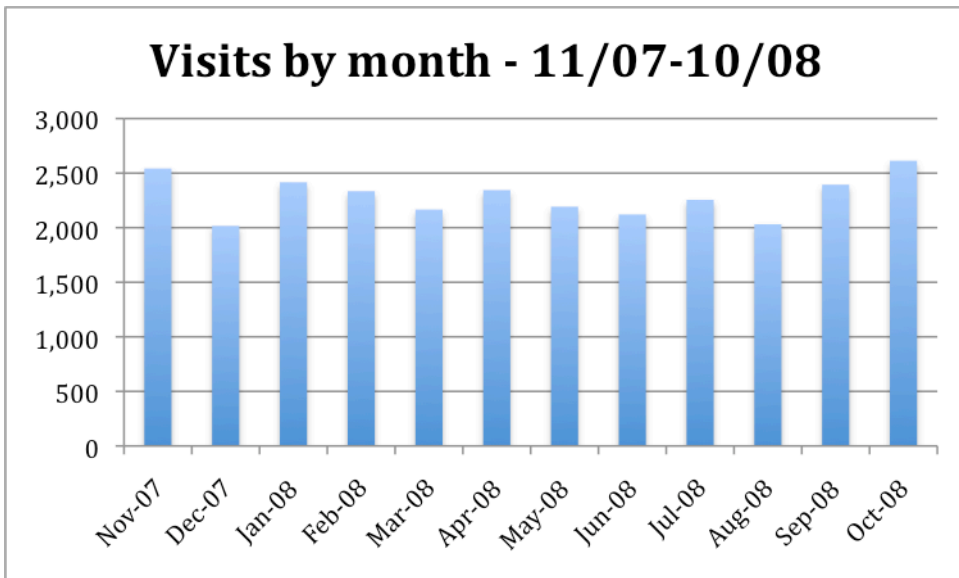
### **Pageviews**

The following chart shows the number of times the ten most frequently viewed pages on the VetMed site were accessed by users in October 2008. The site's home page was viewed nearly eight times as frequently as any other page. Two of the pages on this chart show the native URL structure of the Drupal content management system, "/node/1xxx". The page listed here as /node/1175 is a page that collects links to instructions for authors from a number of journals. The page listed here as /node/1137 has information about a recall of tainted pet food.



**Monthly visitors**

The chart below shows the number of visitors to the VetMed site each month for the year ending October 2008. Like many of our sites, the month-to-month fluctuations of traffic on the VetMed site roughly follow the contour of the school year. However, the VetMed site is unique in that these fluctuations are slight. Traffic is relatively steady all year. It would be interesting to learn if this reflects the academic calendar of the College of Veterinary Medicine, and if there are more students in summer classes than in other colleges.





## **Future explorations**

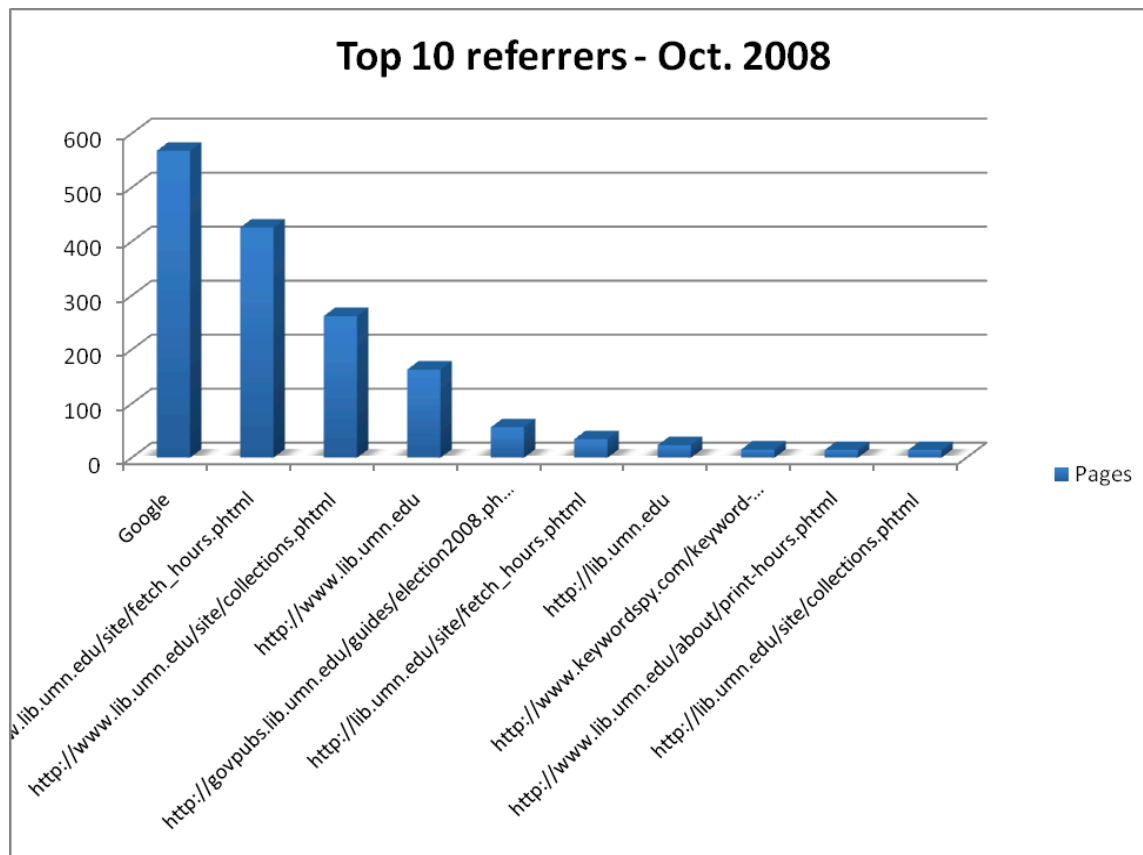
As with the Bio-Medical Library's site, the VetMed site runs on the Drupal web content management system, and usage of the site is tracked using Google Analytics. As we prepare to move the rest of our sites to Drupal and consider options for new web statistics-gathering tools, we would be well advised to look to the HSL sites as a potential model.

## ***walter.lib.umn.edu Website Statistics***

This set of statistics covers the Walter subdomain (walter.lib.umn.edu). Statistics reviewed here are for the month of October 2008 and come from the Libraries' AWStats program, which ingests and parses Apache web server logs.

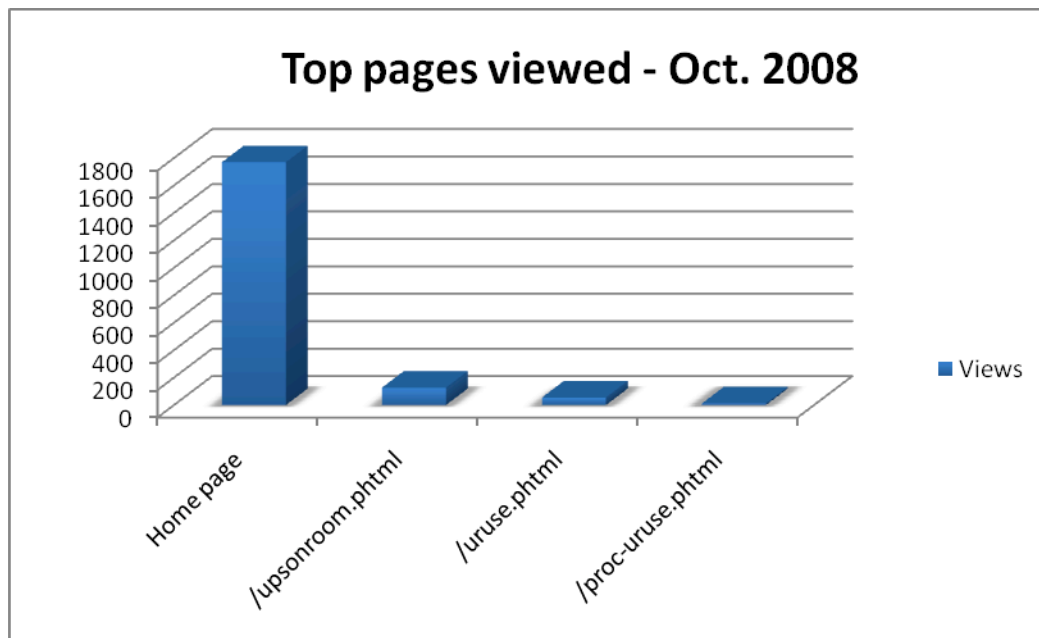
### **Referrers**

The most popular gateway to the Walter subdomain is Google. The most common referrers after Google comprise mostly pages within www.lib.umn.edu, including the hours page, the collections page, and the Libraries homepage.



## Pages Viewed

There are very few pages in the walter.lib.umn.edu subdomain, so only four pages appear on the list of pages viewed. The home page is by far most frequently viewed page. The other pages viewed provide details about the Upson Room and its usage. The Walter subdomain appears to serve primarily as a gateway to the sciweb subdomain, which can be seen in the referrer statistics for that subdomain.



## Keywords/Key Phrases

The phrases most often searched unsurprisingly include some permutation of the keywords “walter library,” with “upson room” also appearing occasionally.

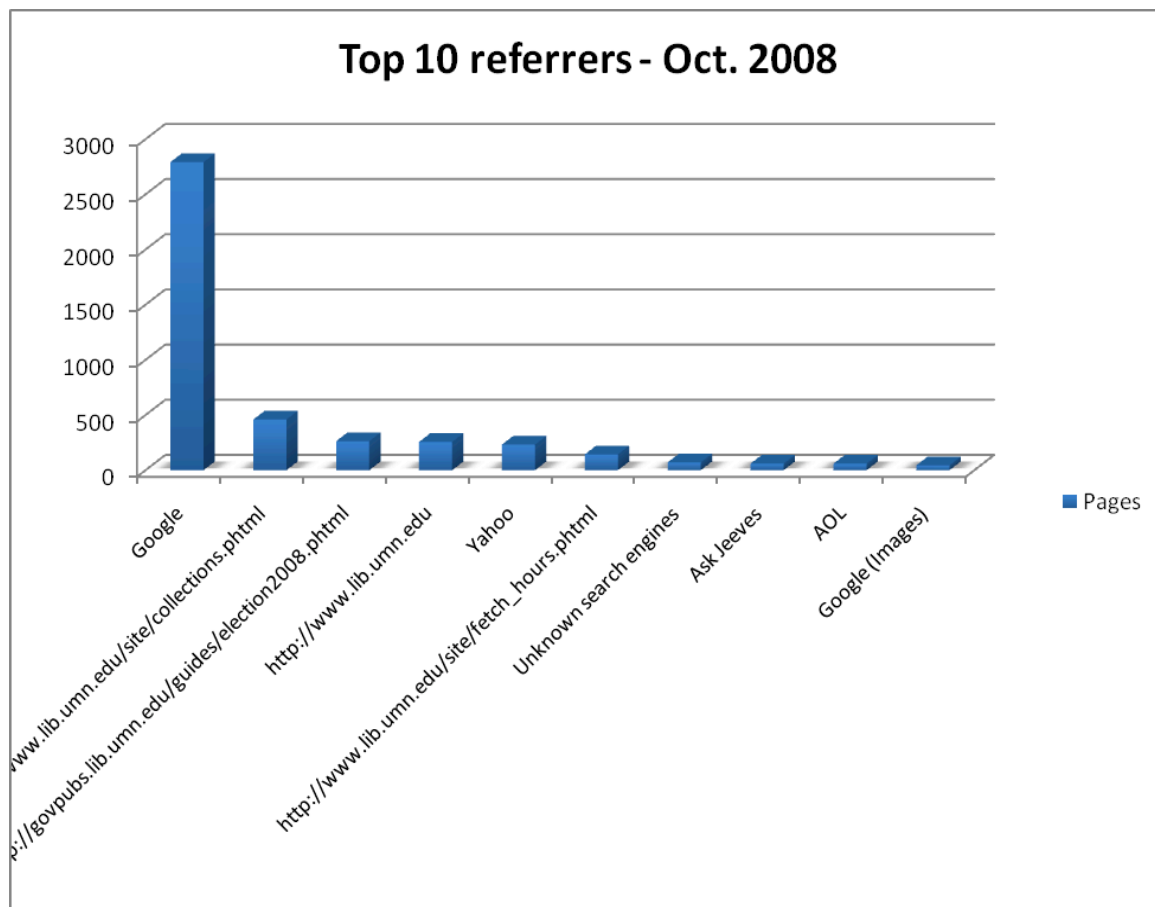
102 different keyphrases	Search	Percent
walter library	317	54.9%
walter	47	8.1%
walter library hours	29	5%
university of minnesota walter library	13	2.2%
walter library university of minnesota	12	2%
walter umn	8	1.3%
upson room	7	1.2%
umn walter library	6	1%
walter library umn	6	1%
walter library computer lab	5	0.8%

## wilson.lib.umn.edu Website Statistics

This set of statistics covers the Wilson subdomain (wilson.lib.umn.edu). Statistics reviewed here are for the month of October 2008 and come from the Libraries' AWStats program, which ingests and parses Apache web server logs.

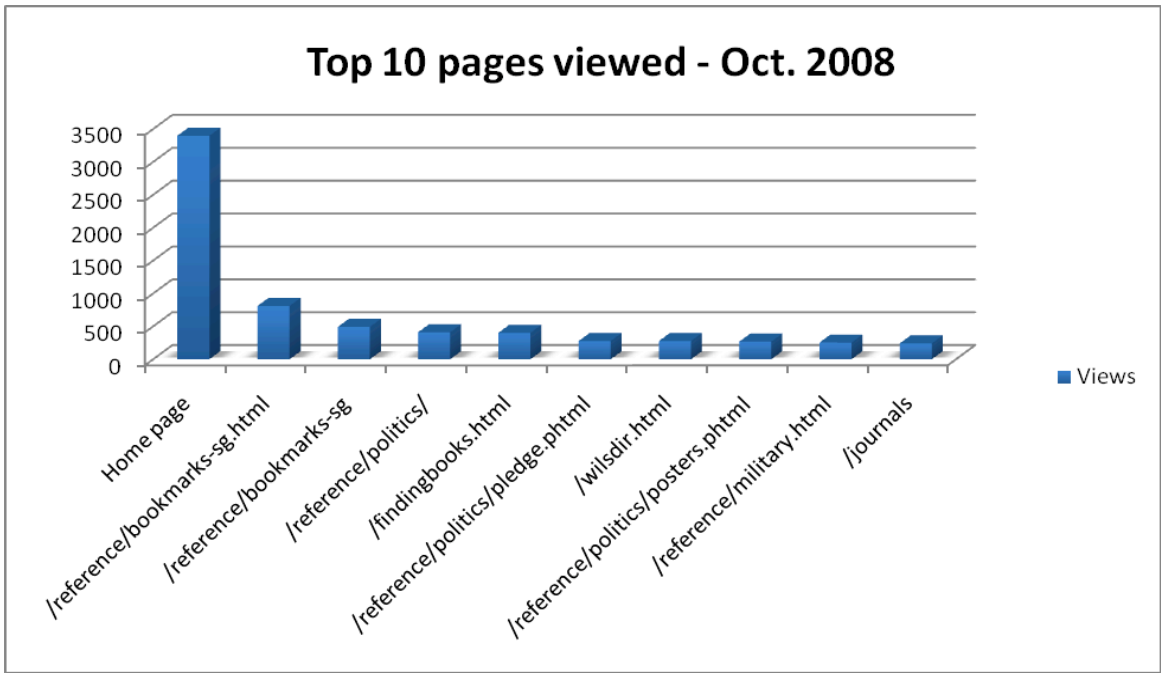
### Referrers

By far, the most popular gateway to the Wilson subdomain is Google. Other search engines are also represented, but none have anywhere near the number of referrals that Google has. Also appearing in the top ten referrers are the Libraries' collections and hours pages.



### Pages Viewed

The Wilson home page is by far most frequently viewed page in the subdomain. Other popular pages in October 2008 included pages on politics (presumably due to the impending election), bookmarks for reference staff, and pages detailing how to get to Wilson Library and how to locate materials in the Wilson building.



**Keywords/Key Phrases**

The keywords “wilson library” are very commonly used. “Wilson” appears in every one of the top 10 phrases.

1600 different keyphrases		Search	Percent
wilson library	621	19.6%	
wilson library university of minnesota	113	3.5%	
wilson library umn	106	3.3%	
university of minnesota wilson library	57	1.8%	
wilson	42	1.3%	
wilson library u of m	42	1.3%	
tc wilson library	37	1.1%	
u of m wilson library	37	1.1%	
wilson library map	24	0.7%	
wilson library minnesota	22	0.6%	