Measuring law library catalog website usability

Rutgers University has made this article freely available. Please share how this access benefits you.
Your story matters. [https://rucore.libraries.rutgers.edu/rutgers-lib/59505/story/]

This work is the AUTHOR’S ORIGINAL (AO)
This is the author's original version of a work, which may or may not have been subsequently published. The author accepts full responsibility for the article. Content and layout is as set out by the author.


Terms of Use: Copyright for scholarly resources published in RUcore is retained by the copyright holder. By virtue of its appearance in this open access medium, you are free to use this resource, with proper attribution, in educational and other non-commercial settings. Other uses, such as reproduction or republication, may require the permission of the copyright holder.

Article begins on next page

SOAR is a service of RUcore, the Rutgers University Community Repository
RUcore is developed and maintained by Rutgers University Libraries
Measuring Law Library Catalog Website Usability: A Web Analytic Approach

Wei Fang
Digital Services Librarian
Rutgers-Newark Law Library for the Center of Law and Justice
Newark, New Jersey 07102

Marjorie E. Crawford
Head of Technical and Automated Services
Rutgers-Newark Law Library for the Center of Law and Justice
Newark, New Jersey 07102

1. Introduction

Cataloging is one of the core services provided by most libraries. With the advent of the Internet, libraries made a dramatic shift and began to offer Web-based catalog services. Consequently, using Websites to offer catalog content to the public was one of the hottest issues debated during the 1990s. The use of the online catalog by users to locate information is a paradise that has been lost.\(^1\) How to regain the popularity of the catalog Website is one of the hardest questions confronted by the management of libraries today, as more and more people use search engines, such as Google to find the information they need in order to complete research tasks.\(^2\)

In evaluating and measuring the input and output of the users upon the catalog, the law library recognized there was a critical need to learn more about how our patrons use this system. It is crucial to measure the usability of the catalog Website, in order to improve the user’s experience of the catalog. The traditional method is by using the transaction logs of the OPAC system. Our library catalog system provides a Web-based representation tool of the log file, which is called the Millennium Web Management Reports (MWMR). We have been using this tool for years and found that this tool is limited in terms of functionalities and accessibilities. While trying to find a better solution to overcome these limitations, we decided to exploit third-party Web analytics tools such as the Google Analytics.

The Rutgers University Law Library - Newark has used the Google Analytics to track its major services Websites, including the main Website and the catalog Website. We have successfully improved our main Website by analyzing data from the Google Analytics.\(^3\) This study used an experimental approach to measure the usability of our catalog by analyzing the transaction logs from the OPAC system and the results from the Google Analytics. The following contributions have been made: 1) Complementary to the OPAC transaction logs, we found that the Google Analytics as a Web analytic tool can provide extensive information on OPAC user’s navigational behaviors. 2) We found out the number of regular catalog Website users. 3) We discovered the most popular search method used by the users. 4) Suggestions were made to improve the user’s experience of the catalog Website. For instance, the catalog Website should be redesigned with more content and graphics to attract more users, since most users use computer systems with high resolution and color depth monitors, and high speed Internet connections to visit the catalog Website.

After we determined the way that data was being collected on the usability of the catalog, we tried to overcome some of the disadvantages associated with conducting surveys of patrons. In general, questions given in a survey can be open-ended or closed-ended: closed questions are considered more efficient and reliable, while open questions can help get unanticipated answers in respondents’ own words. Survey results can be dramatically affected by the way the questions are worded.\(^4\) Unlike paper-based surveys, online surveys overcome physical location limitations, but because of their subjective
nature accuracy of results cannot be guaranteed. Additionally, composing survey questions and conducting surveys are time-consuming and require a great deal of human input. In contrast, both the transaction logs of the catalog system and the Google Analytics provided statistics of the users’ search history in an automated way.

This article is organized as follows: background, literature review, research design, data collection, analysis, findings, discussion, recommendations, and conclusion.

1.2 Background
1.2.1 Rutgers-Newark Law Library’s Catalog System

The Rutgers-Newark Law Library (RNLL) is part of Rutgers School of Law-Newark. With more than half a million volumes, RNLL is the largest law library in New Jersey. Its collections include the statutes and court decisions of all 50 states, federal statutes and case law, federal and New Jersey regulations and administrative decisions, federal and New Jersey legislative history materials, the codes of ordinances for many New Jersey municipalities, Anglo-American legal periodicals, the primary materials of international law, extensive historical materials on English law, and a special collection of criminology and criminal justice materials.

The library has been using the Integrated Library System (ILS) from Innovative Interfaces Incorporated (III) since 1992. The catalog Website, WebOPAC (Web Online Public Access Catalog), is the Web-based interface of the catalog system. The WebOPAC is just a small portion of the entire III system. Yet, it is the most important platform that our patrons use to access our own collections. The server for this system is running the Millennium software for acquisitions, serials, cataloging, interlibrary loan, and circulation. This catalog system contains 114,668 bibliographical records and course materials for 34 courses. It has 1,615 registered patrons.

The usage of the catalog system is tracked by its built-in logging program and recorded in the transaction logs. The Millennium Web Management Reports (MWMR) functionality is a Web-based tool used to access the transaction logs of the OPAC.

1.2.2 Google Analytics Background

In March 2005, Google acquired a Web analytics firm called Urchin Software. Thousands of popular Websites and marketers used to use software solutions from Urchin to better understand the user’s experience as well as to optimize content. In November 2005, Google released the online version of Urchin, named Google Analytics. Unlike the original Urchin, which was priced from $899 to $4,995, Google offers this hosted service for free. Due to the popularity of the service, Google placed new applicants on a waiting list until Google Analytics became generally available to the public in mid-August 2006.

All of the functions that Google Analytics offered are free in our case, but high traffic Websites are subjects to fees. According to Google Analytics’ Terms of Service, “the Service is provided without charge to you for up to 5 million pageviews per month per account, and if you have an active Adwords campaign in good standing, the Service is provided without charge to you without a pageview limitation”. Privacy could be an issue. As Google Analytics states on its Website, it generates “aggregated non-personal information” to share with third parties. Thus, high-security Websites are recommended not to use this service.

In May 2007, Google Analytics redesigned its user interface in hope of giving its users a better
feel and more functionality. In this study, we used the newest interface and functionalities of Google Analytics.

1.3 Literature Review

The traditional catalog changed with the introduction of the online catalog and offered users some “powerful capabilities”.[10] Many of the early researchers proposed technical enhancements to the online catalogs, but these enhancements were never implemented in the OPACs. However, many of these enhancements are now being used by “Web search engines and online book stores”. [11]

User goals on online catalogs have been of interest to researchers since the 1980s.[12] According to Yu and Young,[11] the early studies investigated the following: user demographics, search behaviors, knowledge of system, knowledge of subject matter, library settings, search strategies, and OPAC systems. For their OPAC studies, Yu and Young employed data-collection methodologies that included experiments, interviews, questionnaires, observations, and transaction logs.

The library literature reflects that transaction logs have been widely used to study and monitor the use of OPACs.[13] Each time a user connects to a Web catalog using a browser and places a query, the system writes a record and affixes a date and time stamp. This session file is, then, stored in a transaction history file or transaction log on the local system.[14] Studies on monitoring, analyzing, and evaluating transaction logs span more than thirty years.[15] While transaction logs are useful tools, they have some limitations. For example, it is not possible to tell why users changed their search terms or if the user had the expertise to properly evaluate the search hits.[16]

Catalog studies showed the first and second generation online catalogs provided very little help to searchers of these systems. During the last decade, libraries have simplified the search interface to facilitate a search strategy that is similar to the Google search experience. Karen Markey reviewed twenty-five years of user searching literature and concluded in her findings that users enter a few brief terms into the online catalog during an IR (Information Retrieval) session.[17] In addition to the research on user interface, the researchers showed that using more graphics on the Website could attract more people, extend their overall visit length,[18] and provide more interactions for the users.[19]

An emerging technology on the horizon is the use of third party Web analytic services. According to the Web Analytics Association,[20] “Web Analytics is the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage.” With Web analytics, one does not need to worry about location-based problems inherent in paper-based surveys or about receiving inaccurate information. Plus, all the data is collected automatically with high accuracy. Examples of available Web analytics tools include VisiStat, StatCounter, ClickTracks, and Google Analytics. By far the most sophisticated Web analytics tool is Google Analytics.[21] It provides hosted service for Web analytics, through which collecting and analyzing Web usage data can be done in a finger-snap. It is a valuable tool for those who need to determine their Website's performance in a fast and reliable way.[12] There are some articles that discussed using Google Analytics to measure library Website usability.[3] However, to our knowledge, there is no reported literature on the use of Google Analytics to track a catalog system in combination with the catalog transaction logs.

2. Research Design

This study will address the following questions:
1) What usability statistics are provided by the MWMR and the Google Analytics respectively?
2) By tracking the usage of the catalog Website using both tools, what can we find about the user’s
search behavior? For instance, what is the most frequently used search method?
3) Based on the findings, how can the Web redesign team determine the effectiveness of the catalog Website and further propose redesign ideas and content to better fit our visitors' information seeking needs?
4) What are the pros and cons of using the MWMR and the Google Analytics as tools to measure library catalog Website usability?

2.1 Data Collection
We have been tracking our catalog Website using both the Google Analytics and the MWMR tools since December 18, 2006. The results reported in this paper are based on the data from December 18, 2006 to December 18, 2007. In order to make a comparison between the MWMR and the Google Analytics functionalities, two sets of data from each system were collected by the Digital Services Librarian.

Anyone with a Google account can use the Google Analytics service. Once a Google account holder signs up for the Google Analytics service, Google sends a confirmation email and provides codes to insert into each Webpage to be tracked. The code has to be inserted right before the </body> line in the HTML code of each page to be analyzed. Since our Web pages are generated dynamically from some templates, we simply inserted the tracking code in the templates, including all srchelp HTML pages and OPAC menu page. All the pages based on these templates were thus tracked, so our whole installation procedure was done within 20 minutes. The Google Analytics service can be easily deployed on multiple Websites simultaneously.23 The Google Analytics tool will start tracking as soon as coded Web pages are online. However, reports offered by the Google Analytics service averaged a two-hour delay. For instance, results for 9:00 a.m. show up around 11:00 a.m., meaning that visitors’ activities cannot be tracked in real time.

The MWMR can collect data automatically as well. However, training is required to set the parameters of the system correctly or the results could be dramatically different. Also, the data that the MWMR collected will be automatically summed up after ten days. Thus, data for each day has to be collected manually if a long-term detailed report is needed.

Since the MWMR is an integral component of the catalog system, it maintains a history file of search transactions, a.k.a the transaction logs. As a third-party service hosted by Google, the Google Analytics report provides the following data: the number of visitors to each Webpage, the site’s overall benchmarks, the summary of the visitor’s profiles, and the navigational information about visitors who browse the site. There is overlap in data offered in reports by the Google Analytics and the MWMR in only a few categories. Each tool offers something the other does not.

In general, the following data elements are available from the Google Analytics service but cannot be found in the MWMR:
1) Visitors’ system configurations
2) Visitors’ geographic locations
3) Networks and net speed the visitors used
4) Detailed navigational information, such as landing page and exit page
5) Referring sites information
6) Visualized summaries of the site’s activities

On the other hand, the MWMR provides the following data elements which cannot be found in
Google Analytics:

1) Number of User Keyed Searches
2) Number of System Suggested Searches
3) Number of Records Retrieved
4) Number of Searches Limited
5) Number of Exported Records
6) Number of Displays Invoked
7) Current Searches

As a consequence, we collected two data sets from each system to find the common shared data elements, in order to compare and analyze the differences between them. We paid close attention to the following data elements:

1) Number of total visits
2) Number of Pageviews
3) Number of Records Retrieved
4) Bounce Rate
5) Average time spent on site
6) Percentages of new visitors and returning visitors
7) Users’ system configurations, such as screen resolution and color depth

Again, both systems can collect data and generate reports automatically; therefore, these data sets can be relied on to provide valid results if both systems are installed and configured properly.

2.2 Data Analysis

We use functions from both systems to generate full-scale user activity reports. However, our staff members have to interpret the reports and summarize them for administrative members and the redesign team. The following two subsections describe the functions we used for the reports and the information we derived from the reports.

2.2.1 Data Analysis Functions used in Google Analytics

The Digital Services Librarian took advantage of the following Google Analytics features: Dashboard (visualized summaries), Content by Title, Site Overlay, Traffic Sources Overview, Visitor Segmentation, Bounce Rate, Entrance Paths, Filtering, Data Export and Email and Add Other Users. We will discuss these features in more details below.

Dashboard (visualized summaries)

This is the feature we like the most. It is literally a nutshell of the Website log information. The dashboard offers a visual summary of the Website activities in a time range that we can customize. For instance, for a given time period, we can determine how many users have visited the Website, how many pages have been viewed, how long on average a patron stayed on the Website, which content was the most popular, and which Website the patrons navigated from before visiting our site (see Figure 1).

Content by Title

Content by Title presents a list of the most popular items on our Website. By analyzing data from this feature, we figured out what content was attracting visitors. For instance, we learned that our top hit between December 18, 2006 and December 18, 2007 was the Title Search page (see Figure 2).

Site Overlay
Site Overlay shows clicking summaries laid over an actual Webpage. This feature gives us a direct way to find out if a link had been clicked, as well as the number of clicks on each link.

Traffic Sources Overview
This feature summarizes the methods that visitors used to find our Website. For instance, it shows how many users used direct link to visit our Website and how many users were redirected to our site by another Website (see Figure 3).

Visitor Segmentation
The Visitor Segmentation feature adds 18 more predefined segments for further drill-down into any of 80 Google Analytics reports (see Figure 4). By employing this feature, we could combine any Google Analytics report with other information, such as country, region, and keyword, to generate a new report that presents detailed information regarding our visitors. For instance, we could see detailed information about visitors who viewed the Search by Title page and where they were coming from—that is, visitor segmentation based on region.

Bounce Rate
Bounce Rate is a term used in Web analytics. It can refer to two different numbers for a Webpage or Website. The Bounce Rate for a single Webpage is the number of visitors to the Webpage who leave within the specified timeout period without viewing another page, divided by the total number of visitors to the page. The Bounce Rate for a Website is the number of Website visits which included only a single page of a Website per session divided by the total number of Website visits. Bounce Rates can be used to help determine the effectiveness or performance of an entry page.24

Entrance Paths
Entrance Paths show the navigational path of the visitors: the first pages the visitors started with, then which pages they subsequently viewed, and eventually the last pages viewed during their visit. We used this function to track where the users visited after they landed on the Search by Title page (See Figure 5).

Filtering
Filtering is a very powerful feature. Data collected by the Google Analytics can be customized in order to provide accurate reports. For instance, we can exclude all traffic from a domain or IP address (Internet Protocol). We used this feature to exclude all internal access generated by our staff members so that all the reports show only our patrons’ activities.

Data Export
The Google Analytics tool allows users to export report data in text, XML, and MS Excel formats. This feature is powerful because it generates data that can be analyzed with other statistical programs. It could be imported to MS Excel or other statistical software for further analysis.

Email and Add Other Users
By changing settings of the Google Analytics service, an admin user can decide if adding or deleting users is needed. A user must have a Google Gmail account in order to be added to this service. Each user can be assigned different privileges in order to view various reports and profiles. The Google Analytics service also allows admin users to define a list of emails and reports to be sent immediately or
on a scheduled basis.

2.2.2 Use of the MWMR Reports
   The library catalog system comes with a transaction log reporting function, which could be accessed by staff members using DOS-like terminals or using the MWMR (see Figure 6). Both methods access the same set of data collected by the catalog system automatically. The system administrator can assign parameters for different data collection configurations. For instance, how long the server should keep the transaction logs can be configured.

   Since the DOS-like terminals are very inefficient and very difficult to use, we only used the MWMR features to collect the transaction logs of the OPAC system in this study.

   The following reports from MWMR were used in our study:

   **Search Statistics**
   When this function is called, the system displays the report on search activities. The Search Statistics report shows an analysis of daily system usage for the past ten days. This report includes the number of user keyed searches, system suggested searches, records retrieved, searches limited, exported records, and displays invoked.

   **Indexes Used for Search**
   The Indexes Used for Search Results report shows the total number of searches and the percentage of total searches performed using each index defined in the database.

   **View Current Searches**
   The View Current Searches report displays searches that are currently being performed on the OPAC system.

3. Findings
   Based on the data sets we collected from both the MWMR and the Google Analytics tools over the past year, we made the following discoveries:

   **User Profiles**
   2) About 58 percent of our users visited the Website regularly.
   3) On average, each visitor visited three pages.
   4) Most of our users came from the New York/New Jersey region.
   5) Some 21 percent of users visited only the first page.
   6) About 66 percent of users used Firefox as their Web browser and Internet Explorer was used by about 30 percent users.
   7) Only 1.47 percent of users are still using the dialup connection.
   8) 46 percent of our users visited the Website via Rutgers University’s computers. These users spent 2 minutes on average on the site.
   9) High speed Internet connection users spent more time (4 minutes on average) than those users on the Rutgers University’s computers.
   10) Returning visitors stayed almost twice as long (2:31 mins vs. 1:50 mins) as the new visitors on the site.
   11) About 49 percent of the users used monitors with a 1024 by 768 resolution and 19 percent of users used monitors with a 800 by 600 resolution.
12) 96 percent of users used monitors that supported at least 24-bit of colors.
13) 11 percent of users visited the Website directly as their favorites link (Direct Traffic).

**Site’s Performance**
1) Users visited our catalog website 7,037 times and loaded 21,551 pages from 24 countries.
2) The Bounce Rate for the entry page was surprisingly lower than other pages on the site (7.54% vs. 21.54%).

**Catalog Search Method**
Based on the results from Entrance Paths, we found the most popular search method was Search by Title (37%), followed by Advanced Keyword Search (31%), Search by Author (18%), Search by Subject (17%), Search by Call Number (3%) and Z39.50 (2.7%). About 75.42% of people used the page for Search by Title, these users found and used also the Search by Keywords advanced functionality. The page for Search by Title had a very high Bounce Rate (67.27%), and 52.22% of users exited the site from this page.

**Navigational Analysis**
Based on the results from Site Overlay and Visitor Segmentation, we found that:
1) Our catalog menu is not efficient. For instance, in order to Search by Title (the most frequently used search method), the user has to click 3 times from the catalog menu to get the search box. If the user starts at the library’s main Website, he/she has to click as many as 6 times.
2) The color of the side menu bar is too distractive in terms of color and text. When we used the Site Overlay function, we could not really see the details of the Website. This gave us a feeling that our users would have the same experience.
3) Only the law library main Website has links to the WebOPAC site.

**Search Engines Cannot Find the Catalog.**
One thing we found surprising was that the catalog Website is not indexed by any third-party Web search engine, such as Google and Yahoo. Because neither the Digital Services Librarian nor the Head of Technical and Automated Services could access the server to read robot.txt. This robot.txt file is not located in the liveftp folder. Consequently, the Digital Services Librarian used Google Webmaster Tools to locate the robot.txt file. We discovered that almost all the directories on the server have been blocked from Web search engines.

**4. Discussion**

**4.1 Comparison of the MWMR and the Google Analytics as Web Usability Measurement Tools**
The catalog system has a built-in statistical report function that can be accessed using a terminals and the MWMR. However, the numbers in the statistical function reports contain activities generated by both the WebOPAC and the Millennium terminals, such as the catalog and the circulation terminals. As we discuss below, the built-in statistical report functions are not practical enough for generating Web usage information and, thus, difficult to measure the performance. The WebOPAC system includes a limited functionality for collecting statistics, and the law library is forced to accept only the parameters offered by the vendors. Although the vendor offers an enhancement Web management statistical product, the law library is unable to purchase this option because of budgetary constraints. Based on the system configurations and modules purchased, some schools may have better experiences with using the add-on functions. Yet, with a standard configuration, as we are using, it has the following limitations:

1. Customized reports have to be built by staff members.
2. Only one admin user is allowed to log into the system at a time to run the statistical function and view the reports.
3. Each report may have a size limitation on the numbers of records that can be used. The maximum number of records for each report is limited to 60,000.
4. Reports are not cross-referenced.
5. Report is not visual enough to give admin users instant feedbacks.
6. Running report scripts may slow down the system performance.
7. Report data has to be analyzed by experienced personnel.
8. It is difficult to export data out of the MWMR since only one proprietary file format is offered.

On the other hand, we found that the built-in statistical functions are flexible and offer information in a more specific way compared with Google Analytics. For example, the MWMR can provide the number of records retrieved for each search from the catalog. The Google Analytics can only tell which search page was loaded, but cannot tell how many records were displayed on that result page.

In contrast, the above limitations in the MWMR were not found in the Google Analytics service. Web Analytics tools, such as the Google Analytics service, offer objective and multi-faceted statistical data for Webmasters to better understand the interaction between their visitors and their Websites. It offers a Web-based interactive, cross-referenced, graphics-rich interface in a visual way. No special training is required to use the Google Analytics service. It is much easier to convince the library administration with easy-to-understand graphic charts. The Google Analytics tool collects and stores data as patrons browse. Additionally, the Google Analytics service offers some information that the catalog system is unable to provide, such as what kind of browser a patron has used and which Website has brought the patron to our catalog Website. The Google Analytics tool also allows users to define filters to eliminate unwanted counts. For instance, a user can define a filter to take out accesses that were generated by an IP address or a domain name.

Using both the MWMR and the Google Analytics Web tools, the authors were unable to completely explain users’ behaviors on the OPAC on a full scale. For instance, both approaches did not provide data of the users’ cursor movements. The OPAC researchers are trying to use different technologies to determine all the moves that patrons make as they browse through a site, since this kind of data can provide patrons’ information seeking behaviors in a more in-depth level.²⁵

Collecting and analyzing data using the MWMR effectively was a major concern of the library administration. In assessing data collected on services delivered to patrons, the library management was concerned about several major issues. For example, there were concerns as to whether the data collected would be reliable and valid.²⁶ We compared access information from the MWMR and the Google Analytics reports and found that the Google Analytics report revealed fewer hits than the MWMR. As we described above, the MWMR showed both internal and external access activities. Knowing that we have an active Cataloging Unit that keeps inputting and verifying records into the system, the difference is justified.

4.2. RNLL’s Catalog Website Redesign Recommendations

We posit that redesigning the catalog interface would improve the search experience and encourage researchers to use the catalog Website more frequently. Both the MWMR and the Google Analytics reported facts about the catalog Website in a visual presentation, but these tools were not
smart enough to make suggestions as to how to improve the Website performance and the user interface. We will need to get the input of librarians and ideas from the literature in order to make effective changes. For the Website redesign project, the Digital Services Librarian, the Serials Librarian and the Head of Technical and Automated Services were involved in the decision-making process. The process for our Website redesign project was as follows: The Digital Services Librarian received and interpreted the reports from the Google Analytics and the MWMR, and then distributed the interpretation of the reports to the Serials Librarian and the Technical Services Librarian. Based on feedback from these two librarians, the Digital Services Librarian developed new design suggestions that in turn received further comments. Final decisions about the redesign of the Website will be made by library administrators. All accepted changes will be ultimately implemented by the Digital Services Librarian, who continuously monitors the Google Analytics and the MWMR data and repeats the above process as necessary.

Based on our findings from analyzing the Google Analytics and the MWMR data, the following recommendations have been made on redesigning the catalog Website:

1) Use 800 by 600 as the resolution for the new Website templates.
2) Reorganize the search menu in the order of how frequently they are used by our patrons.
3) Use more graphics, mainly because the computer monitors of most users accept high depth of colors and use high-speed Internet connections.
4) Give proper names to all page titles in the HTML code, since the Content by Title shows popular items by page titles.
5) Create a separate page for the content in the current left menu bar.
6) Redesign the content: On average, each user visited three Web pages, which indicated the users stopped at the search page and then ended their session. The librarians can contribute additional content to attract more visitors.
7) Rewrite the robot.txt file so that public search engines can find the catalog Website and bring in more visitors.
8) Add more links on the other law library Websites, such as New Jersey Legal Library site to redirect visitors to the OPAC Website.

5. Conclusion
The purpose of this study was to measure the usability of the catalog Website. Analysis of the data collection of the searches performed on the OPAC has provided the administration with the strength and weakness of the catalog Website. Furthermore, we have shown the Google Analytics service can provide additional useful information and should be used in combination with the transaction logs within a proprietary catalog system. The findings described in this study have been reported to our administration to be used as performance benchmarks and our recommendations will be used to make changes to our catalog Website. After redesign changes have been implemented, the authors will continue to measure the usability and report upon the effectiveness of such changes in our future work.
Notes


