

This is the author's final, peer-reviewed manuscript as accepted for publication. The publisher-formatted version may be available through the publisher's web site or your institution's library.

Putting web analytics to use creating a data driven website

Tara Coleman, Joelle Pitts

How to cite this manuscript

If you make reference to this version of the manuscript, use the following information:

Coleman, T., & Pitts, J. (2013). Putting web analytics to use creating a data driven website. Retrieved from <http://krex.ksu.edu>

Published Version Information

Citation: Pitts, J., & Coleman, T. L. (2013). The right tools for the job: Using analytics to drive design. In T. Farney & N. McHale (Eds.). *Web analytics strategies for information professionals: A LITA guide* (pp.131-145). Chicago : ALA TechSource.

Copyright: ©2013 by the American Library Association.

Digital Object Identifier (DOI):

Publisher's Link: <http://www.alastore.ala.org/detail.aspx?ID=4258>

This item was retrieved from the K-State Research Exchange (K-REx), the institutional repository of Kansas State University. K-REx is available at <http://krex.ksu.edu>

Putting web analytics to use creating a data driven website

Tara Coleman, Joelle Pitts

Introduction

After nearly ten years and a major structural reorganization, Kansas State University Libraries determined a website redesign was an organizational priority during the Spring of 2011. This decision was a combination of several internal and external factors. Internally, the driving force for the redesign was an institutional reorganization focused on a data-driven, patron-based model rather than subject specific content and access. Other internal factors leading to the redesign included a catalog usability study in 2008 and two iterations of a LibQual survey, in which patrons demanded a more current, user-friendly website. In fact, the library website was identified as one of the most unsatisfactory aspects of library services as a whole. The site was also noticeably dated in terms of design and content.

Externally, the university developed a new strategic plan and “Vision 2025” seeking to elevate Kansas State University to the status of a top 50 public research institution by the year 2025. On the public face, Vision 2025 mandated the use of the official name, “Kansas State University” to strengthen national name recognition, official colors, and a desire for a standard look for all university web pages. This campus-wide research and branding effort served as the impetus to begin the redesign process in earnest. Thus, the combination of internal data and external stimulus provided a unique opportunity for the Web Services Librarian to embark on a data-driven redesign project.

At the time the redesign began, Kansas State University Libraries hosted a staggering 3,500 published web pages. While many of these pages were created with the intention of helping users find information, many more were either unlinked, buried, or were otherwise not providing helpful information to patrons. In addition, many of the pages were not necessarily intended for public view (internal documentation, etc.) or the content was written for print documents and not converted to web friendly language. It quickly became apparent that a large portion of the pages were created and uploaded during the previous 10 years by staff with changing responsibilities and web privileges, thus making many pages obsolete and in some cases, were forgotten altogether. Additionally, pages with current authors and content were not promoted or linked from appropriate locations.

The combined factors of the library reorganization, the vast collection of outdated, lost pages and the external stimulus from the Vision 2025 branding campaign dictated a change in overall layout, function and page content from a collection and subject centered site, to a patron and data-driven site. From early in the process, a data-driven

decision making process was envisioned to play a crucial role in the project in order to validate and assess if institutional and organizational objectives were indeed met.

Methods

Data-driven decision making relies on the accumulation and analysis of various types of data and sources of information pertaining to the project at hand. Data on usability, link navigation, and page views are crucial to understanding why a site must be redesigned in ways which meet user needs and organizational goals. Using existing data to plan, make critical decisions and legitimize decisions during a website redesign process is arguably the most efficient and user-centered approach to a project of this magnitude. While the data gathering process can be daunting and frustrating, the use of data to make and reinforce decisions informs the entire project, and facilitates each phase in the process. The next section will outline the process Kansas State University Libraries followed to gather and utilize data from various sources, and will discuss the usability testing planned prior to the final release.

Web Analytics

AWStats, an open source web analytic tool, was one of the first tools used to gather data about the Libraries' website. The tool allowed the redesign team to gather information on every web page publicly accessible online. A list was compiled of every published web page and AWStats was used to find out how many times each page had been visited. The team used the data from this tool to help sort web pages for use in an upcoming comprehensive content audit.

Using the filtering capability of AWStats, the redesign team compiled a list of every web page in the www.lib.k-state.edu domain and determined the number of page visits during the year 2010. Once compiled, department and subject-related lists of web pages and statistics were created to allow library staff to review the information and decide if the pages should remain online and in their current state.

AWStats also allowed the Web Services Librarian to conduct a "toxic waste" experiment. A blank page was uploaded and buried deep on the website. Published years ago, this page shows how often a page is hit by a web crawler or bot. The number of page visits on the toxic waste page helped the team determine the number of page visits to disregard when analyzing the statistics (539 crawler hits in 2010). Unexpectedly, the toxic waste page had many more hits than some of the Libraries'

legitimate, linked content pages, which further convinced the redesign team that a total overhaul of the site was necessary.

The data gathered from AWStats allowed the redesign team to sort information to see what was used and what was not, though without as much context as needed. For example, AWStats identified the homepages of the Libraries' three small branches as all in the top ten most visited library pages in the domain. Out of context this would seem to show that the branch libraries' web pages are highly used, when in reality they are the default homepage for all internet browsers loaded on computers in those buildings. The main library homepage is not the default homepage for all computers in the main library building, thus skewing the statistical results of the program. Not surprisingly, other pages that floated to the top of the most visited list were pages that were well designed, frequently updated, promoted heavily by the content owner and highlighted during library instruction sessions.

One benefit of using the AWStats analytical tool is that it allowed the team to quickly get the number of page visits into a single document and sort them. The major drawback was the lack of context provided with the page visit data; without knowing how patrons got to the page and where they went once they left, results can be skewed during analysis and lead to ill-informed decisions.

Google Analytics was the second major web analytics tool used to analyze data. It told the story behind the numbers found using AWStats and that story helped guide the team's decision making, especially when the redesign team decided to remove links from the homepage. Using Google Analytics the team determined how often page links were clicked on, how long patrons stayed on a page, how they navigated to a page, and where they went after.

The information gathered from Google Analytics was at times very surprising. At the beginning of the redesign, the library homepage had 23 links to other library web pages, one link to another campus department housed in the main library, and four links to social media. While the team was unable to use Google Analytics to gather information on the social media or the other campus department, the information gathered from the other 23 links was very helpful.

Of the 23 links on the homepage, several were repeated links. There were three links to the "Ask a Librarian" page. The one clicked on the most was an icon with a question mark. There were two links to the catalog and two links to the database page. The links in the global navigation were used more than the links further down on the page. This information helped verify that it is not necessary to repeat links as often in order to drive

patrons to the the desired page and that graphics can help guide patrons to the appropriate links.

Other Analytical Tools

Although not a web analytics tool per se, the LibQual Lite survey administered to the Kansas State University community in the Spring of 2011 served as a baseline for an analysis of the Libraries' website. LibQual responses and comments are based on user opinion and desired level of service. After a quantitative and qualitative analysis of the data, the redesign team found that the Libraries' website was the most unsatisfactory aspect of library service for every patron group (undergraduate, graduate, faculty, staff and community member). There were several negative comments associated with the library website, varying in degree from mild annoyance to outright contempt. In many cases, patrons confused the OPAC system, Voyager, with the library website as a whole, negatively skewing the results. However, the website was eventually identified as one of the most important aspects of library service to focus on in order to increase overall patron satisfaction with the Libraries.

LibStats, an application used to track reference questions, was also used as an analytical tool during the redesign decision making process. LibStats is not a web-based analytical tool, but one that informed time sensitive content decisions after overall content and style decisions were made. The application allowed the team to determine which reference questions are asked most frequently during different time periods in the academic year. Though many questions are common throughout the year, such as printing and interlibrary loan questions, many occur most frequently during specific time frames allowing the group to decide which content should utilize valuable homepage real estate during different times in the semester, and which content would always be featured.

Once a comprehensive list of the Libraries' web pages and statistics was created, the content audit began. The content audit plans were framed around the article "Spring Cleaning: Finding your lost sock. The guide to content audits." published by Nick DeNaardis on the website .eduGuru. His article advised inventorying every page on a site and compiling a document with components such as: page title, owner, last updated, visits, and currency. DeNaardis advised that once the data was compiled, the team should inspect the data, remove outdated content and edit. The original goal of the Libraries' audit was to identify which pages were online, remove out-of-date and irrelevant information, and update pages determined of value to the organization. The article provided sound advice, but the Web Services Librarian quickly discovered the content audit was going to be a challenge. This was the first content audit the Libraries undertook. Over 3,500 published pages were discovered, more than the number of

freshman students admitted to the university every fall. Not only were there a lot of pages, many of them were unlinked and forgotten until the audit began. Because they were forgotten, many of them were out of date.

Due to the size of the organization, the specialization of the information online, and the large number of published pages, the content audit was distributed amongst library staff. Pages were grouped together by department or content owner/expert. Pages that had no obvious content owner were reviewed by the web services librarian. Each content owner was sent an excel spreadsheet of links and statistics and asked to review each page listed on the spreadsheet. Content owners were asked to decide if pages should be kept online with the understanding that they would either be updated if necessary, or archived and deleted. In other words, the pages were taken offline but a copy of the local file was kept for archiving purposes.

On first glance, the content audit seemed an easy first step in the overall redesign process, but once begun, the challenges of the work became apparent. The review of over 3,500 web pages was overwhelming for some. The distribution of the web pages was also uneven. The review of department and unit web pages was, of course, not a high priority for many content owners, making the review process time consuming. Additionally, while many of the pages had low usage statistics, once discovered, many content owners wanted to keep pages they did not realize were published, a decision making process made more difficult by the fact that no previous standards or benchmark for deleting old pages was in existence. These factors contributed to a very lengthy, frustrating process which prevented the website redesign team from moving forward on schedule.

With the initial data in hand, the website redesign team was charged to act. The composition of the team focused on the skills needed to get the job done: programming, web design, graphic design, data interpretation, and experience working directly with users. A viable record of the project was also deemed a priority, as was developing a plan for archiving old pages, so the team included an archivist as well.

Towards the end of the redesign timeline, a final data gathering activity in the form of a usability study was conducted on the redesigned site to reinforce the analytic data gathered over the course of the project. This information, combined with the classic web analytics from Google and AWStats and library data gathered from LibStats and LibQual, provided a holistic, complete and positive analysis of the essential content and design elements included in the redesign. This information package was delivered to library staff and stakeholders prior to release of the new site and allowed the redesign team to confidently move into the post-evaluation stage of the project.

Project Management

The redesign was formally outlined under a project charge. Within the charge, each member of the team was asked to follow a simple decision making process:

- Analyze and interpret the data
- Consider outside factors and opinions
- Examine the precedent set by peer institution libraries
- Return to analysis for final determination

All aspects of project management centered on this data driven decision making model. This model allowed the redesign team to make decisions efficiently while legitimizing the decision making process in the view of library staff and stakeholders. Though certainly helpful, this data-driven decision making model posed several challenges, including overcoming historical precedent, library and campus politics, and campus-level web standards and expectations.

Challenges/data integration, communication, solutions and communication

During the planning of the website redesign, there were links and content for which data was unavailable and circumstances wherein the data gathered were not robust enough to use in decision making. One such example is the Libraries integration of a Media Wiki in the year 2008. The original intent of the wiki was to allow for easy internal collaboration and information sharing, but quickly fell victim to scope creep. Soon the wiki became home to internal and sometimes sensitive data as well as public data such as class and subject guides and public room policies. The wiki pages created by staff were linked from public as well as internal library web pages, however, all wiki information was available for public view. One challenge of gathering statistics for the wiki was that each time a page was edited, the “number of times accessed” page hits would increase, making it almost impossible to determine the number of patron views and staff edits making up the total number. The solution was to find a new tool that filled the same need for easily editable web pages. The LibGuides content management system was purchased its scope limited to only class and subject research guides.

One major challenge the redesign team faced was the absence of data or analytics on the usage of library social media and external applications. As mentioned above, the analytical tools utilized for this project included Google Analytics, AWStats, LibStats and other internal analytics. These tools do not track user navigation outside of the Libraries’ domain, thus the decision making process was stalled when such sites and applications were encountered during the redesign. For example, a Twitter feed is synced directly onto the Libraries’ homepage and was originally envisioned as the avenue used to alert patrons (both subscribers and homepage visitors) and staff of current events and

building issues, such as fire alarms. Unfortunately, the analytical tools in place during the website redesign did not allow the redesign team to gather usage or click data on the Twitter feed. The team decided that because social media has been identified as a major promotional avenue and because the statistics could never account for the number of patrons who read the feed on the homepage but never clicked on it, the Twitter feed would remain live on the homepage and would be included in ongoing usability studies.

Similar issues stalled the redesign process in the form of entities and applications not hosted within the Libraries' domain. RefWorks, Google Calendars, Zoho and digital collections site (hosted in D-Space and CONTENTdm) are examples. While the team was able to gather some data on the digital collections site using Google Analytics, the data was not robust enough to implement a data-driven decision making model in those cases.

Fortunately, the redesign allowed the creation of a more navigable homepage infrastructure so that tools and applications like RefWorks need not necessarily be linked from the homepage. In this case, as in the Twitter feed issue, external applications will be included in ongoing usability testing to determine empirically which links should appear in the top tier and which should be linked from elsewhere on the site.

Campus and library politics challenged the redesign implementation at times. Kansas State University is unified under a common "Vision 2025" aimed at elevating the university to a ranking in the top 50 public research institutions. While campus units, departments and colleges are each working towards that goal, in some cases, efforts became entangled. The introduction of new university web standards three weeks prior to the proposed launch date for the redesigned site was such a challenge. The redesign, of course, did not meet those standards and the implementation was delayed by several months.

In order to more effectively and fairly communicate with library stakeholders, web analytics were used to justify decisions made and develop creative solutions for content and design elements throughout the project. In this way, the transition and acceptance of old site to new was made easier. Communication was a crucial aspect of the project charge and the team agreed on incorporating it into each phase of the project.

Communication was accomplished in many ways. Each team member was responsible for updating assigned departments and units when appropriate. In addition there were weekly updates to library administration, and monthly in person updates at library-wide

staff meetings. The redesign team also hosted several open forums for library staff to voice their opinions on different aspects of the redesign as the project unfolded. This allowed the organization as a whole to voice their opinions on the public face of the Libraries, while simultaneously providing a platform to educate the staff regarding the data-driven decision making process. However, despite the team's broad communication plan, many individuals expressed concern regarding links, and content which would be affected by the redesign process. It was important to communicate that decisions were prompted by data gathered from analytical tools, not the relative importance of the content or link; a process made much easier with the data in hand.

Best Practices

Before starting on a redesign, it is important to gather all of the data first. This may seem obvious, but it is surprising how many decisions are made without this critical information. Having the data at hand will also help solidify a case when recommendations or decisions are made that displease others (ie: removing pages, highlighting one page over another, etc.). Quantitative numbers are often fairly easy to come by; those can be gathered from web analytics tools such as AWStats and Google Analytics. K-State Libraries' relies heavily on social media as a communication tool. Tracking how many people follow The Libraries on Twitter or click on one a Libraries' blog links is easy, but it is harder to track how many people read the tweets on the homepage or follow the blogs via a RSS feed reader. While quantitative data is always important, qualitative data is valuable as well, and the two data types work in tandem to provide a redesign team with a more holistic analysis of patrons' interactions with the site. Sources of qualitative data may include reference statistics, quality control surveys or usability studies.

Once the data is compiled, it is important to tell the story behind the numbers. What does the data mean and what will be done about it? When analyzing, keep in mind that the information may not mean what it appears to mean. For example high statistics on a page could have resulted from web crawlers or because the page is the homepage for the computer lab browsers. Tools such as Google Analytics may reveal that a large number of people click on a link, but do not stay there. Once context is given to the numbers, it becomes easier to make decisions.

Web analytics should be the foundation for any web redesign project, but it is important to begin the project with an understanding of the type of data will be collected. Different analytical tools provide different types of data, with different levels of granularity. And even the best analytical tools will not be able to provide data on certain aspects of a website, especially those elements which are hosted in other domains. Tools like AWStats and Google Analytics are wonderful for providing snapshots of web traffic on

various pages and how users navigate through them. However, these tools do not share how useful the information on a given page was, or how easy it was for the patron to find the information they were seeking. In cases where no data is available or when the data gathered is not robust enough to use in the decision making process, it is important to combine web analytics with usability testing, surveys and other types of qualitative data to ensure that the picture is accurate of how users interact with a site. Once the data is gathered and analyzed it, it is helpful to distribute or communicate the analysis to library stakeholders and staff. Seeing the data and knowing about the data-driven decision making process provides a legitimate basis for the decisions made by the project team and allows the decision making process to advance more smoothly. Empowering staff to improve or promote the pages that have low stats can also be helpful. When communicating, it is useful to show a visual representation (web page mock up) of the redesign plans. Provide a visual representation of the analysis, as stakeholders will be more likely to accept changes or modifications made to the website.

The first task of any web related project, especially a redesign project, should be the creation of a project charge which outlines various aspects of the project as they relate to the ultimate goals and objectives. It should include: an introduction to the project, a statement of the problem, and the purpose of the group's activities to address the problem.

The team is the heart of any project. It is critical that an effective team with the right skills be chosen. In the case of the Kansas State University Libraries website redesign, the project team included the Web Services Librarian, Systems Analyst and Programmer, Graphic Designer, Archivist, Service Quality Librarian and two public services librarians. Each individual contributed to the overall design process in a unique way, representing their departments' outlook and perspective.

Finally, website redesigns should not take place in isolation. It is imperative that the project team analyze design and content elements at peer institutions in order to stay current with trends and issues faced by similar populations.

Conclusion

A functional, effective and timely website redesign is possible through the use of web analytics and effective project management. The Kansas State University Libraries redesign team formed under a project charge utilizing the skills and energy of a unique project team. The project took place under a data-driven design impetus utilizing web analytics to legitimize decisions made.

Both qualitative and quantitative tools were used to gather the data necessary to redesign the website. Web analytics were used to legitimize the decisions made at every step in the process. AWStats provided raw data regarding hits to the Apache server. It provided the team a much more accurate picture in the form of specific calls for each document. Google Analytics showed where in each page patrons clicked, how many times, where patrons came from and where they went from the page in question. It also showed how long patrons lingered on each page which helped to justify content decisions as pages were rearranged and deleted. These analytic tools, coupled with qualitative measures such as the LibQual survey, LibStats and usability testing, allowed the Libraries' redesign team to create a holistic view of the site prior to any design or content decisions. This data driven decision making process was combined with a sound project management process resulting in an attractive, usable, patron-centered website in under one year.

The following represent the best practices gleaned from the website redesign process at Kansas State University Libraries:

- Conduct a content audit to determine the number and content of the pages published to the web
- Gather qualitative and quantitative data for a holistic view of the current site
- Develop a data-driven decision making plan, including how to make decisions when data is not present or robust
- Develop an effective project charge which outlines the management of the project (see appendix A for outline of the project charge)
- Recruit an effective team. Analytics go only so far unless the redesign team has talent and time enough to incorporate the data into a functional website
- Use analytics to legitimize the redesign decisions to library and campus stakeholders
- Plan for ongoing evaluation utilizing web analytics and qualitative measures like usability testing

Though every institution and redesign project will be different, these practices will guide other information professionals in large-scale website redesign projects towards a practical, effective, data-driven design system.