Chapter 1

Web Analytics Overview

Abstract

This chapter of Using Web Analytics in the Library *provides an overview of Web Analytics, the types of services available, and the type of data that they can provide.*

We live in an age of accountability. For libraries, that means it is no longer enough just to inform our stakeholders that the work we do is worthwhile, is good for the community, and is a responsible use of the taxpayers' (or owners') money. Instead, we must justify our services, from both financial and logistical points of view. Librarians must operate with a "return on investment" mindset, learning to demonstrate value offered for both public and private dollars.

At the same time, we have been shifting to a user-focused paradigm for the past decade. While our traditional way of conducting business has been one-directional, where we have made decisions for programs and services based on our own professional perspectives, we have recently become aware of the value of listening to patrons and of crafting programs and services based in large part on their input. With the growth of social networking tools being a major facilitator of societal change, users have become partners in numerous aspects of libraries' daily operations.

Nowhere is this more evident than in the library's digital branch,[1] or its Web presence. The library website has become vital as a community outreach tool and as an ongoing source of information to its users. However, rather than a one-directional, electronically posted version of a flyer or newsletter, the contemporary website is truly a place of community interaction among its users. The website can extend the services available at the brick and mortar library and offer a diverse set of interactive tools such as book discussions, live chat, print resources for download, video instruction and support for research, and blogs that invite user participation. All of these initiatives heighten engagement among the community.

Thus librarians are faced with the challenge of creating innovative tools in a changing society, while at the same documenting responsible use of our resources. In addition, we want to demonstrate within our own organizations that our website is constantly responding to trends and user needs. Finding tools to assist with these challenges is essential. Web analytics is just such a tool.

What Is Web Analytics?

Web analytics is a process through which statistics about website use are gathered and compiled electronically. An analytics program can be used as a tool to help you get to know your users—who they are, where they are coming from, and how they use your site. Ultimately, having access to information about your users helps you to make appropriate decisions about your site, whether those decisions apply to major redesigns of your site or to ongoing tweaking and minor changes reflective of shifts in customer usage or in your own current programs and services.

The origins of web analytics are from commercial website design, where tracking users' behaviors and actions directly relates to consumers' purchasing behavior. However, web analytics can be just as valuable for a nonprofit website as for an e-commerce site. One approach is to monetize your website goals, applying a dollar figure to certain activities such as user visits or online program registrations. Essentially,

however, librarians should simply translate the theories of profit-based customer usage into the goals and situations that are specific to your organization. No matter the kind of organization, the website should seek to understand the visitors' experience and be responsive to their needs.

Combining Quantitative with Qualitative Data

Perhaps the best known name in the field of web analytics is Avinash Kaushik. In addition to two books, Kaushik writes an excellent blog, Occam's Razor, and has numerous videos online regarding best practices of web analytics use. As mentioned previously as typical of web analytics literature, much of Kaushik's work focuses on e-commerce, as his own history in the field is from work with several major commercial companies. Today he is the Analytics Evangelist for Google.

Occam's Razor
www.kaushik.net/avinash

Kaushik's enthusiasm for web analytics is contagious. It's difficult to read his books, read his blog, or watch any of his numerous web instruction videos without also becoming passionate about web analytics. But one of Kaushik's overriding philosophies about web analytics is that the raw data tells only a portion of the story: the "what." To get at the "why," and even the "how," you must get into the mindset of the user. As Kaushik says, you must *ask them* (the users), using methods such as electronic surveys, focus groups, and user testing.

One method of asking for users' feedback is to automatically direct users to an online survey, using a tool such as SurveyMonkey or Zoomerang. Kaushik writes that three open-ended questions are enough:

1. "What was the purpose of your visit to our website today?"
2. "Were you able to complete your task today?"
3. "If . . . not, why not?"[2]

SurveyMonkey
www.surveymonkey.com

Zoomerang
www.zoomerang.com

If the visitors can compete your survey in just a few short minutes, they are much more likely to offer you feedback about their website visit. The answers to these questions will provide you with a wealth of information and will put your analytics data in context.

For a more intensive user feedback project, consider a formal usability study. Two books by Steve Krug are wonderfully helpful in introducing both the purpose and the execution of user studies: *Don't Make Me Think: A Common Sense Approach to Web Usability* and *Rocket Surgery Made Easy.*[3] Another excellent tool is *Usability Testing for Library Websites: A Hands-On Guide,* by Norlin and Winters.[4] Each of these tools can help you get familiar with a process whereby you observe actual users in real time trying to complete tasks on your website. The information you gather is invaluable in learning about user behavior, and the observations (often surprising) provide critical information for web designers.

Website Goals and Implementing Web Analytics

Do you have goals for your website? Before you ever start to collect and review data, you should have key performance goals in mind for your digital branch. Do you have a strategy for your website and its role in your overall strategic plan and your mission? First, establish your website goals, based on your library's mission and strategic plan. Build the website to reflect that mission and that plan. With clear goals in mind, you can use the data to help you make decisions about your website. Avinash Kaushik calls these "actionable insights."[5]

Kaushik sees no reason to collect data if there is no direct match to a potential decision for improving your website. He also stresses the 10/90 rule, which states that 10 percent of an organization's cost for web analytics is for the tool itself and its setup, and 90 percent of the expenditure is for a person (or team of persons) to analyze the data for actionable insights. It's not just about the numbers; *analysis* is key, and that analysis is the most challenging part of the process. As you consider implementing web analytics, it is important that everyone has a firm understanding of the critical role of savvy personnel in regards to data analysis for your specific organization. A strong culture of data will include both the numbers and the people-generated actionable insights.

Another caveat about data collected from web analytics tools is that the numbers themselves are never 100 percent accurate. Much of this has to do with how the analytics tools work, which will be explained at various points in this report. Whatever the source of the odd inaccuracies, they are ultimately not particularly significant. What are important are trends over time (six months to two years), patterns of usage (such as seasonal spikes or dips), and insights you can draw through advanced tools such as user segmentation (for

example, finding out what your Polish language users are seeking). Slight discrepancies in the numbers are normal and should not be considered worrisome.

Analytics Data and Insights

Are your visitors doing what you expect them to do, or what you want them to do? Are they following the path you thought they would follow when you designed your menu system? Are visitors to your Digital Library page finding the link to historical photos of your city or the university's archival images? These are some of the questions you should be able to answer by using an analytics program. You can follow a user's clickstream, find out which pages get the most visits, gather numbers about how long users stay on each page and where they come from. You could in theory track every visitor to your site, and each click of each of those visitors. Kaushik refers to "the paradox of data": there is so much data, but so many barriers to good insights.[6]

With so many metrics available, what should you be looking for, and what conclusions can you draw from those metrics? These questions will be investigated in more detail throughout this report (see especially chapter 2 for a longer glossary), but for now I will offer a just a few examples. As you look at these examples, consider how you could use these statistics to make changes in your website content and design, and how you could also use the statistics to justify website personnel and support.

- **Page views:** This is just what it sounds like: number of times an individual page is viewed. You can watch the numbers in this category for trends over time, seasonal spikes and dips, and comparisons with other pages in your website. You can tell from the various page view rates within your website which pages are most popular, and which pages are rarely viewed. Another way to use these figures is to compare your total page views per month with your walk-in numbers. Thus, what percentage of your total use is from your digital branch? This comparison will of course be an estimate, since many customers use the website and also visit in person, but it is nevertheless an extremely useful comparison.
- **Time on page:** You can study this metric in conjunction with the purpose of the page. For example, you may not want a visitor to spend much time on an opening page with basic information and descriptive links. A short time on page average could indicate that people are finding what they want quickly. However, another page may offer richer content, such as bibliographic citation helps or blog posts. In such cases, you would prefer a longer time on page.
- **Unique visits and returning visits:** The unique visits metric indicates the number of visits to your site by different users, while returning visits indicates users who return from at least one previous visit. Since return visits are measured by cookies, users who turn off cookies will be counted as unique rather than return visitors. This is an example of how the metrics numbers may not be 100 percent accurate.

  Obviously you want to have an increasing number of visitors over time, but in general website visits will trend upward, and this fact alone will not give you a lot of actionable insights. Partner this metric with other information, such as increasing numbers of return visits to the YA blog when the YA librarian makes more frequent entries, demonstrating increased engagement with the target community.
- **Bounce rate:** This metric counts visitors who left your site after visiting only one page. A high bounce rate is often considered undesirable, and you will see this assumption in the literature. However, bounce rate is another example of a metric that should be compared to the actual page. If visitors are finding exactly what they want when they land on a page in your site, either at the opening page or some other file within the site, they may indeed leave after viewing only that individual page. A high bounce rate for e-commerce would be considered bad since the customer did not continue to browse and eventually make a purchase. In information websites, analysis of bounce rate is more complex.
- **Visitor information:** Many metrics can give you insights about your users, including geographic location, browser used, operating system used, and what path they took through your site. The more you know about your visitors, the more you can adapt your site to meet their needs. For example, if you find that many people are using mobile devices to view your site, you will know that designing for mobile interfaces should be a priority for your web development team.

When getting started, concentrate on a few effective metrics for your own site and follow those. As you build experience and confidence with those metrics, you will begin to add more dimension to your analysis. Start small, as the amount of data can be absolutely overwhelming unless approached with planning. Web analytics need not be an in-house example of information overload.

Web Analytics Tools

There has been an exponential growth in the web

analytics field in the last decade. Google launched its free tool, Google Analytics, in 2006, and the field has continued to boom since then. Many commercial tools exist, including Coremetrics, Adobe's Omniture, and WebTrends.[7] Open source tools are also available, including Piwik, which is billed as an open source alternative to Google Analytics.

Google Analytics has become an extremely popular tool among libraries, due not only to its tremendous power, but also to its free availability, its ease of use, its flexibility, and its clear reporting mechanisms. This report will use Google Analytics as a model for library implementation. The various chapters of the report will look at details of the Google Analytics program, how to install and set up the program for your own library, and how to interpret the reports. Individual case studies from libraries will show real-world examples of Google Analytics in use, where librarians have used Google Analytics to help them better understand their users' website behaviors and understand how they can continually improve their customers' website experience.

Piwik
http://piwik.org

A Note About User Privacy

As you begin to learn more about web analytics and its capabilities, you will see that a tremendous amount of data can be collected about your end users. This behind-the-scenes data collection may make many librarians uncomfortable. However, the reality is that virtually all websites collect some user data through the operation of server logs. While completely eliminating data capture is an unrealistic goal, intentionally adding a tracking tool such as Google Analytics to collect personal information about the library's website visitors may seem to be the antithesis of our service philosophy. How can we reconcile the priority of personal privacy with the organizational need to examine website usage and statistics?

Essentially, what is important to libraries is that our users operate with complete anonymity at all times and that they also maintain ultimate control of their data. The ability to opt out of any data collection is key and should be clearly offered as an option to our website visitors. In addition, we must be thoughtful about what data to collect and how it will be used. Libraries must strike a balance between user privacy and organizational effectiveness, with the scales always being tipped in favor of user choice.

The excellent TechSoup article "Site Statistics and User Privacy for Nonprofit Websites," by Elliot Harmon, discusses at length the issue of nonprofit organizations' struggle with website end-users' privacy. Ultimately Harmon concludes that clear communication with the site's users from the outset is the best approach. When TechSoup's own website added Google Analytics, the leaders posted an entry to their blog explaining the role of analytics in their overall communication strategy, the potential implications of adding the analytics tool, and clear reasons for their decision. Harmon quotes library consultant Jessamyn West as saying, "Have good reasons [for using web analytics]. If you don't have good reasons, that sends the message that you really haven't thought about the implications of these tools."[8]

If you don't already have a privacy policy for your website, beginning to use an analytics program is an excellent time to move this to the top of your priority list. A privacy statement should include information for your visitors about what personal data they may be asked to provide, either as registered users or as guests, what information may be collected without their explicit knowledge (such as their geographic location), how you will store and use the information, and what third parties (such as the analytics program host) may have access to it.[9] Finally, you will also want to offer your users clear instructions of how they might opt out of as much as the data collection as possible.

Google's Approach to Privacy with Google Analytics

The approach to privacy at the other end, the host company for the analytics tool itself, is a key part of the decision process in regards to using web analytics. (Some librarians prefer open source analytics programs run from their own local servers for just this reason.) However, just as libraries choose automated circulation systems and online public access catalogs with patron privacy built in, so too can we select analytics tools that build in sensitivity to user anonymity. Google Analytics (GA) is a good model in this respect, as the company is clear, direct, and intentional in its privacy policies. GA has its own internal policies that strongly support end-user privacy. It actually includes a privacy statement in its licensing agreement, and encourages full disclosure and individual privacy statements for all users of their tools.

Brian Clifton, a GA expert and author of *Advanced Web Metrics with Google Analytics,* says Google made an intentional decision not to collect user information from the beginning of its analytics program development. GA works through the use of cookies, a prevalent tool in web analytics. When you install GA, you add a string of code to each webpage for which you

will be gathering statistics. That code in turn communicates with the user's page by sending it a cookie, or tracking code, when it visits your site. (See chapter 2 for more technical information about how GA works.) All data collected and then reported with GA, however, is in aggregate and anonymous form.[10]

In addition, as mentioned above, the company requires adherence to its own privacy policy for all clients of its analytics tool. The privacy statement is issued as part of its Terms of Service agreement and imposes strict limits on GA clients in regards to third-party access to collected data. GA clients must also agree to post and abide by a clear privacy policy of their own, informing their website visitors of the presence of the analytics tracing program and use of cookies.[11] See the GA Terms of Service for more information. In addition, Clifton offers a sample "Best Practice Privacy Statement When Using Google Analytics."[12] Google itself offers a user opt-out tool for some browsers (see http://tools.google.com/dlpage/gaoptout). The link to this opt-out tool should be posted on your website or within your privacy statement.

Google Analytics Terms of Service
www.google.com/analytics/tos.html

Privacy and Service

Ultimately, citizens make constant decisions about their own comfort level with sharing personal data online. Individuals may choose to participate—or not—in social networking sites, online banking, online credit card use for purchases, online surveys, and casual Internet browsing for personal interest. What is important for libraries is that our users are fully informed of the parameters of their relationship with our digital branch, including what information we collect about their visits, how that information is collected, and how it is used. Our users should find those details clearly posted on our websites, and information about how they can take personal control of the tracking mechanism should be easy to find. Creating a thoughtful rationale that considers both patron privacy and organizational effectiveness will help end users and librarians alike feel comfortable with website analytics.

Resources

Clifton, Brian. *Advanced Web Metrics with Google Analytics,* 2nd ed. Indianapolis, IN: Wiley Publishing, 2010.

Electronic Frontier Foundation. www.eff.org.

Google Analytics Opt-Out Browser Add-On (Beta as of March 31, 2011). http://tools.google.com/dlpage/gaoptout.

Google Analytics Terms of Service. www.google.com/analytics/tos.html.

Harmon, Elliot. "Site Statistics and User Privacy for Nonprofit Websites: Learn the Facts Before You Install Analytics Tools." TechSoup, Oct. 26, 2009. Available online at www.techsoup.org/learningcenter/webbuilding/page12238.cfm. Accessed March 31, 2011.

Kaushik, Avinash. Occam's Razor. www.kaushik.net/avinash.

———. *Web Analytics: An Hour a Day.* Indianapolis, IN: Wiley Publishing, 2007.

———. *Web Analytics 2.0.* Indianapolis, IN: Wiley Publishing, 2010.

King, David Lee. "Building the Digital Branch: Guidelines to Transform your Website." *Library Technology Reports* 45, no. 6 (Aug./Sept. 2009).

Morgan, Joe. *Google Analytics for Libraries.* Available online at http://josephsandersmorgan.com/home/web-analytics. Accessed March 31, 2011.

Quinn, Laura S. "A Few Good Web Analytics Tools." TechSoup, April 29, 2007. Available online at www.techsoup.org/learningcenter/internet/page6760.cfm. Accessed March 11, 2011.

Web Analytics Association. www.webanalyticsassociation.org.

Notes

1. David Lee King, "Building the Digital Branch: Guidelines to Transform Your Website," *Library Technology Reports* 45, no. 6 (Aug./Sept. 2009).
2. Avinash Kaushik, "The Three Greatest Survey Questions Ever." Occam's Razor, April 10, 2007, www.kaushik.net/avinash/2007/04/the-three-greatest-survey-questions-ever.html (accessed March 25, 2011).
3. Steve Krug, *Don't Make Me Think: A Common Sense Approach to Web Usability,* 2nd ed. (Berkeley, CA: New Riders, 2006); Steve Krug, *Rocket Surgery Made Easy* (Berkeley, CA: New Riders. 2010).
4. Elaina Norlin and C. M. Winters, *Usability Testing for Library Websites: A Hands-On Guide* (Chicago: ALA, 2002).
5. Avinash Kaushik, *Web Analytics: An Hour a Day* (Indianapolis, IN: Wiley Publishing, 2007), 15.
6. Avinash Kaushik, *Web Analytics 2.0* (Indianapolis, IN: Wiley Publishing, 2010), 4.
7. Kaushik, *Web Analytics: An Hour a Day,* 5.

8. Elliot Harmon, "Site Statistics and User Privacy for Nonprofit Websites," TechSoup, Oct. 29, 2009, www.techsoup.org/learningcenter/webbuilding/page12238.cfm (accessed March 12, 2011).
9. Ibid.
10. Brian Clifton, *Advanced Web Metrics with Google Analytics,* 2nd ed. (Indianapolis, IN: Wiley Publishing, 2010), 61.
11. Clifton, *Advanced Web Metrics,* 60–62.
12. Ibid., 63.