

Getting to Know Web Analytics

Abstract

This chapter of Using Web Analytics in the Library provides guidelines for selecting and implementing web analytics tools in a library context. The author examines different variables that determine which tool might be appropriate and offers suggestions determining which tools meet different needs.

One of the first things you will need to do as you consider implementing web analytics is to select the appropriate tool for your library. In this report, I will focus specifically on Google Analytics (GA) to illustrate various aspects of web analytics. But before we get into the basics of any one program, it is useful to have some foundational knowledge of how the tools work, what programs are out there and how to choose one, and what the most used standard metrics are.

Web Tracking Basics: Data Collection Mechanisms

It's ALL clickstream data. . . .

When we typically hear about collecting a Web user's *clickstream data*, this term can actually refer to different mechanisms used to track and store users' activities while on the Web. There are a variety of tools used to capture user information, with the two main approaches being log file data capture and page tagging.

Log File Data

Web log files were the original method of capturing and storing information about visitors to individual

websites. Typically a request for your website comes to your server, and the server creates an electronic file entry in the log for that request. Web logs capture information such as the page name, IP address and browser of the visitor, and date and time stamps.¹ Web servers collect data and create logs as part of their regular activity, independent of the user's browser, which makes the data readily available.

Relying on this method of data collection for user analysis has several disadvantages. For one thing, the data may be hard to access. If your library contracts with an external web hosting service for your website, you must work with the service to access server log file information. If your server is controlled locally, log files will probably be maintained by the information technology (IT) department rather than the website design department. Analysis and use of this information would require close collaboration with IT. In addition, log files are primarily intended for the capture of technical information. While this information is useful in the overall analysis of information technology resources, web logs are not the most effective way to capture and analyze website visitor behaviors.

Ultimately, server log information can help you evaluate traffic numbers in regards to your server load and capacity, but it tells you very little about your users or the effectiveness of your site in relation to your goals.

Page Tagging

The other common method of web analytics used today is often referred to as *page tagging*. This method of collecting data involves inserting tags, or lines of JavaScript code provided by an analytics program, into the source code of a webpage (mylibrary.org).

The tag code collects data from the visitor's browser and sends it to the analytics program's remote host computer, where reports are available to the mylibrary.org owners.

Page tagging has become quite popular, as both implementation and management of the analytics tool are much easier than through using the log file method. Many of the analytics tools available today are a service of a third-party vendor, which frees the mylibrary.org owners from having to develop an internal technical analytics infrastructure.

Advantages of Page Tagging as An Analytics Tool

In his book *Web Analytics: An Hour a Day*, Avinash Kaushik points out a number of advantages to using page tagging:²

- Tagging is extremely easy to implement. Once you sign up with an analytics program vendor, you add a few lines of code to the <head> section of each HTML page. The analytics program immediately begins to capture vast amounts of data from its visitors.
- A tagging solution is possible for those without access to their servers. Analytics data gathered through JavaScript tags is made available through the third-party analytics program host, and thus you do not need to rely on server access.
- As a webmaster or a design team, you have significant local control over data review and analysis. While typically a tremendous amount of data is available, you can easily select the data that is most relevant to your own specific goals and outcomes.
- Similarly, you can create administrative reports that clearly highlight data relevant to your library's priorities.
- Innovation in web analytics is focused on page tagging. Using this data collection mechanism keeps you closer to ongoing upgrades and revisions in the field.

There are, however, some concerns with page tagging as an analytics tool. The tagging process involves the use of cookies, and some web users turn off both cookies and JavaScript. These visitors will be invisible to you. In addition, as mentioned in chapter 1, the use of cookies raises some privacy issues for our customers who do not turn off cookies in their Web use. These issues should be clearly discussed both within the library and with our website visitors.

Cookie Basics

All cookies are small chunks of data that are sent from a website you visit to your hard drive so that your

computer can provide information to your browser as you use the Internet. There are various distinctions to keep in mind: *session cookies* versus *persistent cookies*, and *first-party cookies* versus *third-party cookies*.

Session Cookies and Persistent Cookies

A session cookie is stored only as long as your visit lasts within a particular website, stores your browsing history for as long as you stay at that website, and is erased when you close your browser. A persistent cookie, on the other hand, stays stored on your hard drive as long as the file is programmed to stay there, or until you manually remove it.³

First-Party and Third-Party Cookies

Most web analytics programs, including Google Analytics (GA), use only first-party cookies. First-party cookies are set by the host website itself and are used by that host website to store a user's data so that person can return to the site without starting all over as a new customer. In addition, the first-party cookies used by GA allow GA to analyze things such as which keywords and referring sites bring visitors to your site.⁴

Third-party cookies are set from a domain outside the one shown on the user's address bar, usually without the knowledge of the user, and are typically used by advertisers to collect and store an individual's browsing habits. These cookies are much more likely to be turned off by web users, as they are considered more invasive and thus more objectionable.

Other Web Data Capture Methods

Additional methods of data capture include web beacons and packet sniffing. Web beacons are 1 × 1 pixel GIF images placed in webpages, usually hosted by a third-party server. Packet sniffing captures all of the user's data, including passwords. Both methods are typically used with commercial websites so that advertisers can track the effectiveness of their ads through number of views, user behavior, and so forth.

To a great extent these methods have fallen out practice, due mostly to privacy concerns and the fact that a vast number of users (some estimates as high as 40 percent) turn off third-party cookies. Many fewer people turn off first-party cookies, as they are much less invasive, they collect only anonymous data, and it is very difficult to browse the Web if you turn them off.

Choosing a Program

Shopping for a web analytics program is quite similar to shopping in general: you thoughtfully consider what is important to you, and then you examine the options within your budget. Analytics considerations

include the cost, the level of sophistication for data collection and analysis, the reporting options, and the program's overall accessibility. You may also consider advanced options such as segmentation of users, depth of analysis for the individual metrics, and host-level support. GA is an excellent choice for most libraries, as it offers tremendous data collection, ease of use and customization, and support. And there is no cost for its implementation.

GA, as a free service, has been so successful that the commercial services have had to continually refine and specialize their products. Examples of commercial products include Omniture, WebTrends, ClickTracks, and CoreMetrics. These programs can cost thousands of dollars per year, but usually include a range of services along with the analytics program and may be good alternatives for the largest library systems. For a thorough discussion of what to look for in a commercial analytics vendor, see Avinash Kaushik's blog post "Web Analytics Tool Selection: Three Questions to Ask Yourself" from January 30, 2007, and chapter 2 in his book *Web Analytics 2.0*.

*Web Analytics Tool Selection:
Three Questions to Ask Yourself*

www.kaushik.net/avinash/2007/01/web-analytics-tool-selection-three-questions-to-ask-yourself.html

At the other end of the spectrum are open source alternatives. Piwik and OWA, or Open Web Analytics, are GPL-licensed, downloadable systems that use page tagging. AWStats is a log file-based GPL-licensed product. There are quite a few comparison sites available via a simple web search for "open source web analytics."

Piwik

<http://piwik.org>

Open Web Analytics

www.openwebanalytics.com

AWStats

<http://awstats.sourceforge.net>

Although there are various reasons for an individual organization's ultimate choice of a web analytics program, for the purposes of clarity I will use a single program for illustration and description throughout the rest of this report. While not specifically endorsing GA or rejecting other programs, I will use GA as a frame for library-specific descriptions of how to use website user data to analyze your website goals and customer satisfaction.

GA Basics

Like the page tagging analytics programs described earlier in this chapter, GA begins with the insertion of JavaScript code tags in your webpages. The GA program is available to anyone who has a basic Google account. Once you are a registered Google account holder, you can very easily sign up for the analytics program. Details of the step-by-step process are outlined in chapter 3.

But what will the program measure? How will you use the various metrics to help you understand more about your website customers, and whether your website is successful? This section will outline some of the basic analytics terms as defined by the Web Analytics Association, along with some comments regarding their relevance to library usage.

Definitions from the Web Analytics Association are noted with the (WAA) designation, and are taken as posted in Occam's Razor.⁵

- **Page:** "A page is an analyst definable unit of content" (WAA). This is an individual HTML file within your site, with page tagging in the <head> section of the source code. An example is mylibrary.org/youthservices.html.
- **Page views:** "The number of times a page (an analyst-definable unit of content) was viewed" (WAA). As mentioned in chapter 1, important things to watch are trends over time (six months to two years), dips and spikes according to your own program calendar, unexpected dips and spikes that may need observation and reflection, comparisons within your site among pages, and overall total views as compared to your total walk-in numbers.
- **Visits/sessions:** "A visit is an interaction, by an individual, with a website consisting of one or more requests for an analyst-definable unit of content (i.e. "page view")" (WAA). Again, you can watch for trends over time and unexpected changes.
- **Unique visitors:** "The number of inferred individual people (filtered for spiders and robots), within a designated reporting timeframe, with activity consisting of one or more visits to a site. Each individual is counted only once in the unique visitor measure for the reporting period" (WAA).
- **New visitor:** "The number of Unique Visitors with activity including a first-ever Visit to a site during a reporting period" (WAA). Don't forget that if a user has turned off cookies, she will be counted as a new visitor whether she has been to your site before or not. Nevertheless, you can watch for trends over time to see if your site is gaining ground in your target market. You can also use this metric to watch for upward jumps after a

significant redesign, a large marketing program, or other library events where increasing your Web presence has been a goal or turns out to be an organic but unplanned side effect.

If this increase in new visitors happens as a pattern in parallel with significant library events, you can begin to plan for the potential incoming website traffic by intentionally updating your site to coordinate with those events. For example, make sure your content is fresh, add notices about other upcoming events, rev up activity within your social networking sites, and so on.

- **Repeat visitor:** “The number of Unique Visitors with activity consisting of two or more Visits to a site during a reporting period” (WAA). Again, partner this information with library events, such as summer reading for public libraries or research paper time in an academic library. Another tool that can be extremely useful is segmenting users, or looking deeper (drilling down) into an individual segment of your user population. I will include examples of segmentation in chapter 4.
- **Return visitor:** “The number of Unique Visitors with activity consisting of a Visit to a site during a reporting period and where the Unique Visitor also Visited the site prior to the reporting period” (WAA). The return visitor has been to your site before, but has not been there as often as the repeat visitor. Again, user segmentation would be beneficial to understand more about these customers and what they are looking for, and then hopefully to improve their website experience and their success rate with the library’s online resources.
- **Entry page:** “The first page of a visit” (WAA). With analytics tools, you can tell where the website visitor comes from on the Web, such as through a search engine, a specific web address (URL), or a link from an outside blog. (See also *referrer*, below.) You may assume that your homepage will have the highest count as the customers’ entry page, but it would be interesting to see if this is the case. For example, one of your departments may attract quite a few customers directly to its own page, or a construction update page for a renovation or new building project may be extremely popular. As with the other metrics, noticing these figures helps you to gauge your customers’ interests and information needs—and in general, helps you to get to know them better.
- **Landing page:** “A page intended to identify the beginning of the user experience resulting from a defined marketing effort” (WAA). Although some literature does use this term almost interchangeably with *entry page*, the WAA defines it separately. When used in e-commerce, the landing page’s ultimate purpose is to easily facilitate a purchase. The customer would come in at the

entry page, and then “land” when he is ready to take action. If using the term within this framework, a landing page for libraries might be a page where a user can register for a program, join a particular online discussion group, or RSVP for a library instruction session.

- **Exit page:** “The last page on a site accessed during a visit, signifying the end of a visit/session” (WAA). Kaushik says this is where you can look for “leakage” in your site, determining where the highest percentage of users leave your website.⁶ However, Kaushik also states that the metric actually tells you very little about the success or failure of your site or the individual page. For example, 52 percent of your visitors may leave your site from the Upcoming Events page, but this could be either good (they are finding just what they want) or bad (they are annoyed that there is nothing for them)—or some of both! It’s extremely difficult to make judgments about the exit page by using the numerical result alone. Rather, combine it with surveys or other mechanisms for user feedback.
- **Visit duration:** “The length of time in a session. Calculation is typically the timestamp of the last activity in the session minus the timestamp of the first activity of the session” (WAA). Normally you want your website to be engaging enough to hold customers’ interest for at least several minutes. A longer visit duration is considered desirable in e-commerce sites where the goal is to have the customer find multiple products for purchase. However, library websites will have some pages where a speedy result is desirable (a successful in-and-out catalog search for a specific title), and some situations where a longer visit is desirable (engagement with a digital image collection). Don’t attempt to lump this result as one individual metric across your site and make any assessments based on the single figure.

Visit duration is also categorized as *time on site* and *time on page*. One thing to keep in mind when reviewing these particular metrics is that the time being measured is the number of minutes the site or page is *open*, which does not necessarily equal the amount of time the visitor is actually engaged with your content. The clock will time out after some inactivity, frequently 29 minutes.

- **Referrer:** “The referrer is the page URL that originally generated the request for the current page view or object” (WAA). Another way to say this is that the referrer is the URL from which your visitor clicks to get to your site. The referrer can be *internal*, coming from within your website, or *external*, coming from outside your site. Related terms are *search referrer*, which indicates whether the visitor came via a search engine, and *source of visit*, which is the origin of the referrer,

including specific search engines or another specific website.⁷

- **Conversion:** “A visitor completing a target action” (WAA). This term originated with e-commerce, and in that environment, it usually means a visitor completes a purchase. For libraries, a conversion can be much more general, and the WAA definition is in keeping with the broader interpretation. Examples for libraries could be any goal or target action you designate in your analytics configuration, such as placing a hold on a catalog item, adding a comment to a blog, or clicking on a link to a special collection page.

Here are some useful terms in addition to those defined by the WAA:

- **Google Analytics Tracking Code (GATC):** The GATC is the snippet of code that is added to each of the webpages using GA. This code, or the *page tag*, is what enables the data tracking and collection.
- **Bounce rate:** As described in chapter 1, the bounce rate indicates the number of visitors who left your site after visiting only one page. A high bounce rate is generally considered undesirable, since e-commerce sites prefer visitors to browse and ultimately buy. However, just as with visit duration, bounce rate can be either good or bad depending on the purpose of an individual page and your user’s expectations.
- **Click density:** *Click density* refers to the number of times a link was clicked by a visitor. You can view this and view the typical path of the customers by using the analytics tool’s site overlay feature. As Avinash Kaushik says, this feature enables you to walk in the shoes of the visitors, assessing whether they follow the path you’d like them to follow or indeed the path you think they will follow.⁸ For libraries this might mean that visitors rarely use your departments’ named links, but instead rely heavily on the search bar. Are people clicking on the link to the director’s blog? If not, you might want to find a more prominent placement for that link. Click density analysis is considered one of the most powerful tools for actionable insights, as the site overlay literally lets you look at your website and your actual customers’ use.
- **Funnel:** The funnel is the set path a visitor takes before a goal conversion. In e-commerce, this would be a path a user takes to the checkout process, but a funnel can also be a path toward a different kind of goal. For example, we can envision a specific funnel, or visitor path, for a catalog search or an interlibrary loan request.
- **Key Performance Indicators (KPIs):** KPIs are key factors, specific to your organization, that measure success.⁹ These goals will be defined by

your library’s mission and thus the mission of your digital branch. Do you have a percentage of your user population that you want to have visit each month? Do you want to see an increase in program attendance based on website advertisement? You could define goals based on usage trends, participation in social networking tools, and any conversion rates you define.

For more information, see a complete list of terms in the Google Analytics Glossary.

Google Analytics Glossary

www.google.com/support/googleanalytics/bin/topic.py?topic=11285

Possible Actions

Looking at these metrics, you can begin to see some possible actions libraries could take as a result of an analysis of their website analytics data. Here are just a few examples:

- If you find low page view:
 - Remove the page and free that space or personnel time for something else.
 - Relocate the link so more people can find the content, particularly if you think it is important.
 - Redesign or rewrite the page itself to make the content more appealing.
 - Perform a user study to make sure the page works the way you think it does, such as digital images that should be easily found and used.
- If you find unexpected click density:
 - Reorganize your menu structure
 - Relocate your menu, making sure its placement is consistent throughout your site.
 - Change the names of your links. (Frequently libraries rely on professional jargon for link names, and users don’t understand those terms. Use your users’ vocabulary, not your professional jargon.)
- If you find unexpectedly high traffic in one department:
 - Drill down further in the reports to analyze use; expand content and services in response to user demand.
- If you find source of visit data that shows a consistent increase in visitors using mobile devices:
 - Put more effort into designing pages for the mobile systems.

Chapter 3 will provide specific information about setting up GA and will begin to show how you can identify these metrics from the GA graphs and reports.

Resources

All About Cookies.org. "About Cookies: Are All Cookies the Same?" www.allaboutcookies.org/cookies/cookies-the-same.html. Accessed March 14, 2011.

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

Google Analytics. "Glossary." www.google.com/support/goanalytics/bin/topic.py?topic=11285. Accessed March 31, 2011.

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

Kaushik, Avinash. "Web Analytics Standards: 26 New Metrics Definitions." *Occam's Razor*. Aug. 23, 2007. www.kaushik.net/avinash/2007/08/web-analytics-standards-26-new-metrics-definitions.html. Accessed March 31, 2011.

Tonkin, Sebastian. "Top Ten Myths About Google Analytics." *Google Analytics blog*. May 28, 2009. <http://analytics.blogspot.com/2009/05/top-ten-myths-about-google-analytics.html>. Accessed March 31, 2011.

Web Analytics Association. "Standards Committee Deliverables." www.webanalyticsassociation.org/?page=standards. Accessed March 31, 2011.

Notes

1. Avinash Kaushik, *Web Analytics: An Hour a Day* (Indianapolis, IN: Wiley Publishing, 2007), 26.
2. Kaushik, *Web Analytics*, 32.
3. "About Cookies: Are All Cookies the Same?" www.allaboutcookies.org/cookies/cookies-the-same.html, All About Cookies.org website (accessed March 14, 2011).
4. Sebastian Tonkin, "Top Ten Myths about Google Analytics," May 28, 2009, *Google Analytics blog*, <http://analytics.blogspot.com/2009/05/top-ten-myths-about-google-analytics.html> (accessed March 13, 2011).
5. Jason Burby, Angie Brown, and WAA Standards Committee, *Web Analytics Definitions* (Washington DC: WAA, Aug. 16, 2007), as quoted in Avinash Kaushik, "Web Analytics Standards: 26 New Metrics Definitions." *Occam's Razor*, Aug. 23, 2007, www.kaushik.net/avinash/2007/08/web-analytics-standards-26-new-metrics-definitions.html (accessed March 18, 2011). See also Web Analytics Association, "Standards Committee Deliverables," www.webanalyticsassociation.org/?page=standards.
6. Avinash Kaushik, "Standard Metrics Revisited: #2: Top Exit Pages," *Occam's Razor*, Dec. 27, 2006, www.kaushik.net/avinash/2006/12/standard-metrics-revisited-top-exit-pages.html (accessed March 18, 2011).
7. Brian Clifton, *Advanced Web Metrics with Google Analytics*, 2nd ed. (Indianapolis, IN: Wiley Publishing, 2010), 6.
8. Kaushik, *Web Analytics*, 10.
9. Clifton, *Advanced Web Metrics*, 11.