AN INTRODUCTION TO NSDL WEBMETRICS[1]

Version: August 2006

Contents

1 Introduction: NSDL and Omniture

'Webmetrics' is a general term used to describe various methods for collecting and analysing traffic to and through a web site. Webmetrics are often promoted as being of interest to commercial sites seeking to maximize sales and advertising revenue, but they are also useful for non-commercial and not-for-profit sites, where they can be used to supplement web site usability studies, outline user audience demographics, and track trends in site usage.

During the early years of the NSDL program, NSDL operated without any central guidelines for webmetrics tools and analysis. Core Integration itself used a free, open source utility (AWStats) to measure traffic on nsdl.org; from CI's point of view, however, it was unknown what forms of webmetrics existed among other NSDL projects. More recently, CI has emphasized the importance of consistent program-wide webmetrics, and has developed guidelines for webmetrics standards that are included in the Memorandum of Understanding established between CI and the NSDL Pathways Projects. The MoU guidelines recommend that Pathways projects implement a third-party webmetrics service provided by Omniture (omniture.com). CI pays for Omniture implementation across all projects, and running Omniture incurs no financial burden for Pathways projects. CI does not prevent Pathways from running their own webmetrics tools in parallel to Omniture.

CI began implementation of Omniture on nsdl.org at the end of 2004. Since April 2005, CI has been including other Pathways projects, and DLESE, on the Omniture contract. As of July 2006, the following projects had been implemented, were in the process of being implemented, or had been contacted about implementation:

- National Science Digital Library – nsdl.org
- Applied Math and Science Education Repository (AMSER) – amser.org
- Bioscience Education Network (BEN) – biosciednet.org
- Communities for Physics and Astronomy Digital Resources in Education – compadre.org
- Computer Science Educational Reference Desk (CSERD) – shodor.org
- Digital Library for Earth System Education (DLESE) – dlese.org
- Materials Digital Library (MatDL) – matdl.org
- Math Gateway – maa.org
- Middle School Portal – msteacher.org
- National Engineering Education Delivery System (NEEDS) – needs.org
- Science Education Resource Center (SERC) – serc.carleton.edu
- Teachers' Domain at WGBH – teachersdomain.org

[1] For further information, please contact: Michael Khoo, NSDL-Core Integration: mjkhoo@ucar.edu

Omniture works through a combination of javascript (placed on each page of a web site) and non-persistent cookies. Omniture records data such as the page being viewed, time of visit, browser type, previous page and next page visited, etc. These data are stored on and can be accessed and analyzed remotely at the Omniture web site. The particular configuration of Omniture services chosen by NSDL means that after a user session has ended (determined by a visit going for 30 minutes without initiating any further action), the session cookie expires from the visitor's computer, any IP address information is deleted by Omniture; and all that remains is an anonymous record of the path a visitor took through the web site in question.

The Omniture contract provides CI with centralized access to webmetrics for each of the instrumented projects, and to a high-level, coarse-grained 'roll-up suite' that aggregates basic metrics across all of instrumented projects. Individual projects have access to the metrics for its own site, and also to the roll-up suite, but are unable to view the metrics for any other project. Using Omniture is advantageous for CI, as the same tool is implemented on all project sites, leading to consistency in cross-project analyses. The tool also provides access to Pathways project webmetrics, without having to rely on the projects themselves to supply them (although CI does rely on Pathways staff to implement and maintain the Omniture code on each site).

2 Choosing Tools and Metrics

Webmetrics tools differ widely in cost, functionality, and ease of operation. They can be free or proprietary, implemented on project servers or operated remotely from third party servers, and can can require varying levels of expertise and server access for their operation.

Different webmetrics tools measure website traffic in different ways, and statistics can differ considerably from tool to tool. Other factors, such as whether or not visits from site developers excluded, whether or not visits from bots and crawlers are excluded, and so on, can also affect webmetric reports.

Visits, page views, and hits

Website traffic can be measured in a number of different ways, including 'visits,' 'page views,' and 'hits.'

A *visit* is defined as the sequential viewing of one or more pages by the same visitor, within a specified time frame. Omniture defines a visit as:

> The number of times a visitor browses to your site. A visit begins when a person first views a page on your company's Web site, and lasts until that person stops all activity on the site for a period of 30 minutes (Web analytics general industry standard). The industry assumes that, given the length of time, the person likely closed their browser and reinitiated a separate "visit."

A related concept is that of the *unique visitor*, who is identified by IP address and/or persistent cookie, and who can make multiple visits during a day, month, or year, and still be counted as one unique visitor (for instance, the same visitor making three different visits over a 24-hour period could be counted as one unique visitor). Unique visitors are ignored in NSDL webmetrics, as in compliance with NSDL's privacy guidelines, Omniture does not store users' IP addresses.

Each visit will include one or more *page views*. Omniture defines a page view as:

> One complete Web page load in a user's browser. Page Views count the number of times an entire page has been accessed, rather than individual site elements.

Finally, a *hit* measures a request either for a page or for any of the elements in that page (such as an embedded images). According to Omniture:

> Every time a server is called to retrieve an image, text, frame, or other information for a Web site, a hit is registered. Depending on site content, you could have 10 hits or more for one full-page view.

Despite the common practice of referring to website traffic in terms of hits, this metric is rarely used to measure website traffic, as a single visitor can generate a large number of hits.

A web site will record more hits than page views (as a web page may generate a number of hits), and more page views than visits (as a visitor may view more than one page). Different webmetrics tools will measure these hits, page views and visits in different ways. For instance, NSDL runs webmetrics software called AWStats on its own servers, and the AWStats webmetrics consistently report higher levels of traffic for nsdl.org site than Omniture does. In May 2006, for example, AWStats reported over 1/2 million hits on nsdl.org, 197,916 page views (compared to 57,233 reported by Omniture), and 25,773 visits (compared to 13,743 reported by Omniture). The extreme difference between the 512,841 *hits* reported by AWStats for the month of May, and the 13,743 *visits* reported by Omniture, demonstrates how important it is to specify the tool, the technology, and the metric being used when webmetrics are being reported. It follows therefore that webmetrics reported by other sites should also be treated with caution, unless it is known exactly what figure is being reported, and how it is being calculated.

Webmetrics nsdl.org for May 2005	AWStats	Omniture
Hits	512,841	n/a
Page views	197,916	57,233
Visits	25,773	13,743

In the case of NSDL, the metric chosen for reporting of website traffic is that of the visit ("A visit begins when a person first views a page on your company's Web site, and lasts until that person stops all activity on the site for a period of 30 minutes"). The underlying assumption behind this choice was that each visit represents an attempt, by a user, to complete a specific task on nsdl.org.

3 Choosing timeframes

Sampling timeframes and periods have a major influence on webmetrics, and nsdl.org is no exception; and traffic to nsdl.org exhibits a number of temporal rhythms. These include daily peaks between approximately 7:00 a.m. and 6:00 p.m. (which indicates that most users come from the same time zones as the continental USA) (figure 1), and lower weekend traffic (indicating that most people use the site at work) (shaded areas, figures 2 and 3). An annual rhythm shows monthly peaks during the school year and declines during the school vacations (figure 4). These and other temporal rhythms requires averaging and smoothing across various time scales in order to identify specific trends. For instance, in the case of Omniture, the Omniture metrics did not become reliable indicators of annual trends in nsdl.org traffic until after the second year's statistics had begun to be collected (i.e., after May 2006).

4 Filters

A final variable affecting nsdl.org webmetrics involves the contribution of non-users (e.g. developers) to overall traffic in the reported statistics.[2] Including developer traffic in overall metrics

[2] Omniture automatically excludes search engine traffic, such as bots, crawlers, and spiders, from the metrics.

can cause the metrics to report a higher level of use than is actually occurring. To address this situation, developer computers at NSDL Core Integration projects have, as far as possible, been excluded from the Omniture statistics through the use of IP filters and cookies (although for various technical and organizational reasons, these methods are not foolproof).

5 Uses for webmetrics

Sophisticated webmetrics packages such as Omniture can provide rich and highly detailed views of traffic through a website. They allow website data to be parsed in many different ways, and at many different levels of granularity. However, what do these data tell us about a particular website? They do provide statistics, but inferring from webmetrics data to hypothesized user behaviour(s) is tricky. Such inferences will often *remain* hypothetical until they are triangulated with observed/reported user behaviours. In brief, it is impossible to infer any particular user behaviour with 100% certainty solely from webmetrics.

General webmetrics

There are some general webmetrics that are useful without necessarily being triangulated with user behaviour, such as: How many people are using your site? Is usage trending upwards, or downwards? and so on. Using even these data involves however paying attention to the issues outlined in previous sections (e.g. selecting an appropriate metric, selecting an appropriate timeframe, averaging data out over specified time frames, excluding developer traffic, and so on). In the case of CI, developing a protocol to guide general usage data collection took a number of months. Issues included: deciding upon an appropriate and useful metric; identifying and excluding particular ranges of IP addresses form the data; addressing the privacy concerns of various Pathways projects with regard to the data collected by Omniture (it turns out that all Omniture data are ultimately owned by NSDL); supporting Pathways projects correctly to implement the Omniture tools; and so on.

The chosen metric for nsdl.org (as described above) was that of the visit. NSDL assumes that each visit represents an attempt, by a user, to complete a specific task on nsdl.org. Even this apparently simple assumption is however open to qualifications. For instance, from the Omniture metrics alone, we have no way of telling whether two visits from the *same* IP address 30 minutes apart represent two visits by the same person, or two visits by two different people using the same computer (e.g. in a school library); two different users accessing nsdl.org from the same computer in a library will, from the point of view of Omniture, count as the same visit, user, and task. Similarly, we have no way of knowing whether or not two visits from *different* IP addresses on the same day may be two different people using two different computers, or the same person using a work and a home computer; for instance, a teacher working using nsdl.org their school office computer, who then goes home and finishes their task on their home computer, will count as two separate users accomplishing two separate tasks.

Using webmetrics to answer specific questions

Webmetrics come into their own when they are used to answer specific research questions about a website. For instance, Omniture is being used to inform decisions related to nsdl.org website design. The Omniture statistics tell us that the most common monitor resolution amongst nsdl.org users is 1024x768 pixels, and that use of this monitor resolution is increasing; and this has informed how nsdl.org web pages is laid out. Again, Omniture showed that visitors to the search results page rarely used the left navigation column, suggesting that the left navigation should be removed from this page (to leave more room for the search results). Finally, Omniture pathflow analyses show that the

majority of visitors to nsdl.org visit the front page, then go to the search results page, then leave the site; which suggests that most users wish to search NSDL for resources, select their search results and then leave, and not spend time exploring the site (this observation has been confirmed in interviews and user testing). This last finding runs counter to webmetrics notions of site 'stickiness,' which suggest that it is preferable that users spend more rather than less time on a website.

Identifying research questions from webmetrics

Interesting patterns in webmetrics data can also point to further research questions. For instance, webmetrics for nsdl.org show an unusual pattern of referrers. Omniture defines referrers as:

> Locations from which visitors to your site come. They can be Web sites with search engines or advertisements, e-mail messages, Web sites with links to your site, or a shortcut on your own computer. For example, if a visitor clicks on an ad banner or link of any kind from site A and arrives at site B, site A 'refers' you to site B. Site B's SiteCatalyst account will register site A as a referrer and log one page view for site B.

Omniture records referrals to nsdl.org as follows: links from search engine pages (20% of all referrers), links from other web pages (40%), and typed and/or bookmarked URLs (40%). These proportions (20:40:40) differ from those in other NSDL projects, and in comparison nsdl.org has:

- a higher proportion of referrals from links in web pages
- a higher proportion of referrals from typed/bookmarked URLs
- a lower proportion of referrals from search engines

While many referrals to nsdl.org are from Google, Yahoo, and other search engines, cumulatively, many more referrals to nsdl.org come from links embedded in non-search engine pages. However, webmetrics do not tell us how or why these links were created in the first place, and to understand this phenomenon further, additional non-webmetrics research is needed.

Summary

Webmetrics allow NSDL to gain a rich understanding of use of the nsdl.org site and NSDL as a library. Care however has to be taken in specifying which tools and metrics are used, and also in distinguishing between what webmetrics tell us about website traffic (which they can do in great detail), and what webmetrics tell us about users' behaviours (a much more difficult area to understand).

Figures: Examples of daily, weekly, monthly, and annual variations in nsdl.org traffic

Figure 1
Unit:
Weekday

Sub-unit:
Hour

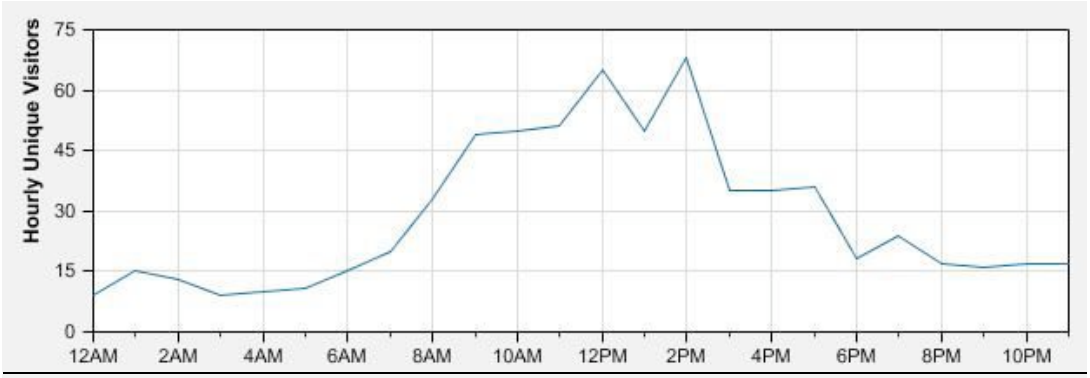

Figure 2
Unit:
Week

Sub-unit:
Hour

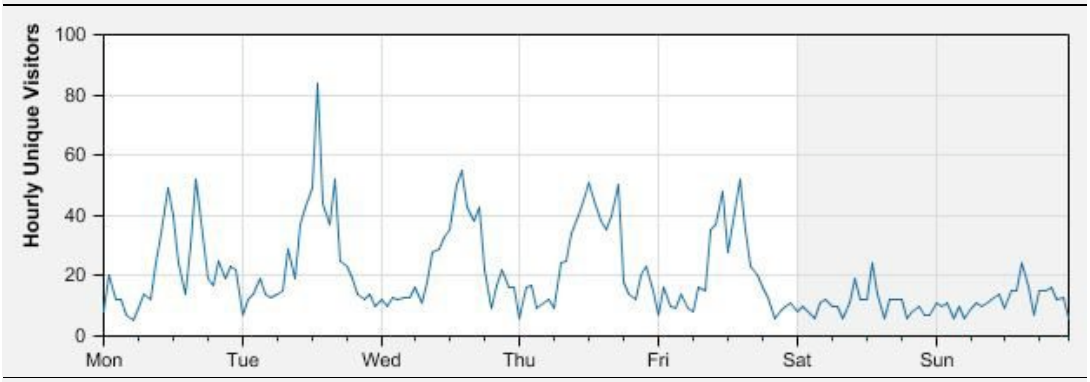

Figure 3
Unit:
Month

Sub-unit:
Day

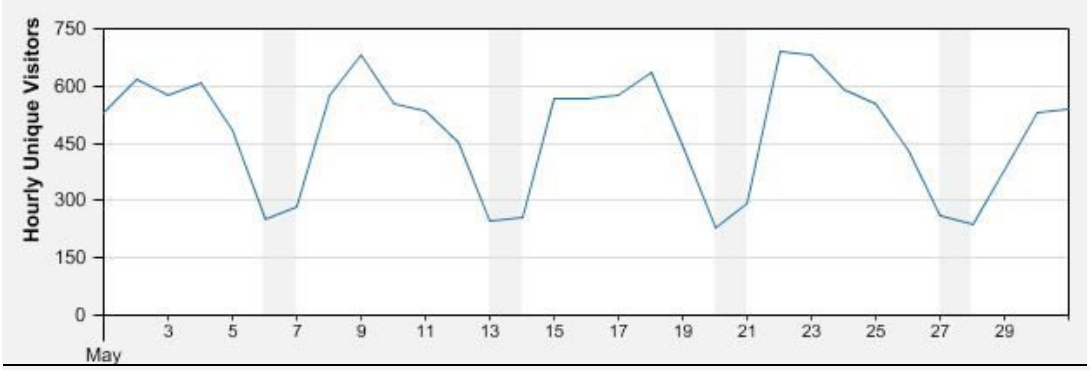

Figure 4
Unit:
Year

Sub-unit:
Week