Using Web Metrics to Analyze Digital Libraries

Michael Khoo
The iSchool at Drexel University
Philadelphia, PA 19104-2875
+1.215.895.2474
michael.khoo@ischool.drexel.edu

Mimi Recker
Utah State University
Logan, UT 84322-2830
+1.435.797.2688
mimi.recker@usu.edu

Joe Pagano
The Library of Congress
Washington, D.C. 20540
+1.202.707.2488
jpag@loc.gov

Bart Palmer
Utah State University
Logan, UT 84322-2830
+1.435.797.2688
bart.palmer@usu.edu

Anne L. Washington
The Library of Congress
Washington, D.C. 20052
+1.202.707.5520
annew@crs.loc.gov

Robert A. Donahue
WGBH Educational Foundation
Boston, MA 02135
+1.617.300.3638
bob_donahue@wgbh.org

ABSTRACT
We discuss the use of web metrics at four digital libraries, the Instructional Architect, the Library of Congress, the National Science Digital Library, and WGBH Teachers’ Domain. We describe practical issues involved in implementing and using web metrics to track web site performance. We also report on the focused data mining of web metrics. Using session length as our example, we recommend that this metric, which was developed to track e-commerce, be reconsidered when applied in non-e-commerce settings, such as digital libraries. We conclude by discussing some of the limitations and possibilities in using web metrics to analyze and evaluate digital library use and impact.

Categories and Subject Descriptors
H.3.7 Digital Libraries: system issues, user issues

General Terms
Management, Measurement, Documentation, Performance, Design, Human Factors

Keywords
evaluation, usability, usage, web analytics, web metrics

1. INTRODUCTION
Web metrics (‘web metrics,’ ‘web analytics’) record and analyze visitor traffic to and through a web site. They can be used to estimate whether or not users’ goals are being achieved; to support usability studies and web design; and to provide feedback on web site use to developers, managers, and other stakeholders. Despite similarities in nomenclature, web metrics are distinct from webometrics (the analysis of the structure and links of the World Wide Web) [7], the system-based analysis of digital libraries [e.g. 5, 14], or social network, bibliometric, or other analyses [e.g. 8, 9].

Digital library web metrics track users’ interactions with a digital library web site [36]. As web metrics were initially developed as a way to understand the success of e-commerce web sites by tracking customers and purchases, and because educational and cultural institutions do not use purchasing as a success measure, useful web metrics for digital libraries need to be identified.

This paper describes web metrics research at the Instructional Architect (Recker, Palmer), the Library of Congress (Pagano, Washington), the National Science Digital Library (Khoo), and WGBH Teachers’ Domain (Donahue). The paper is structured as follows. Section 2 introduces technical factors associated with common web metrics tools and research issues; section 3 discusses organizational issues; section 4 presents case studies centered on interpreting session length metrics; and section 5 advances some preliminary conclusions about the appropriateness of session length metrics for digital libraries, as well as some wider lessons learned about implementing web metrics with digital libraries.

2. TECHNICAL FACTORS
We begin with an overview of some common web metrics technologies and units of analysis.

2.1 Web Metrics Tools
Web metrics tools differ widely in technology, functionality, complexity, utility, and cost. Three common tools and approaches are: (1), combinations of user panels and browser logging tools to track sample WWW user populations (e.g. Nielsen NetRatings: http://www.nielsen-netratings.com/); (2), collecting network traffic data directly from ISP servers (e.g Alexa, http://alexa.com/; Hitwise, http://hitwise.com/); and (3) using site-specific server log parsers or page tagging technologies to measure traffic through a particular site. All these methods record and report a wide range of data. The first two methods are useful for analyzing high-level traffic for commercial purposes, such as pricing Internet advertising on a particular site. The third set of methods is useful for generating fine-grained understandings of how a particular site is functioning. This paper focuses on the use of server log parsers and page tagging tools.

Server log parsers are established technologies which analyze the raw traffic data from server logs. These data can be viewed and manipulated in a desktop or browser window. Examples include AWStats (http://awstats.sourceforge.net/) and Webalizer (http://www.mrunix.net/webalizer/), and a range of commercial products [11]. A recent development has been tools which use javascript embedded in the HTML of every web page to report
user data to a database, often located on a third party server. Again, data can be viewed and analyzed in a browser window. Examples of page tagging tools include Urchin (http://www.onewebhosting.com/urchin.php), Google Analytics (http://analytics.google.com), and Omniture (http://www. omniture.com/). One central difference between server log and page tagging tools is that the latter only record data when the javascript is triggered by a web page being loaded in a browser window, presumably by a human. Server log analyzers, on the other hand, record all calls to a server, including those of machine (bot, spider) origin. While it is possible to configure server log tools to exclude much (although probably not all) non-human traffic, page tagging tools will thus often report lower levels of traffic than server log tools.

Another type of page tagging tool records users’ click locations on a web page, and represents these data in the form of a heat map. These tools provide useful visual summaries of large amounts of user interaction data (see 4.3). They tend to be one shot tools, although they can be used to support A-B testing of site designs; for instance, does a repositioned search box receive more or less attention from users?

Despite the apparent plug-and-play nature of many web metrics tools – a perception encouraged by tool vendors – in practice they can differ widely in cost, functionality, sophistication, and ease of use. As with other types of software and hardware, potential issues with web metrics tools include: limited functionality; limited documentation; limited vendor server capacity and bandwidth, and poor response times when viewing data; inability to download and archive data locally; poor technical support; and the user privacy and trust concerns that can arise when a piece of page code collects data on users and stores and/or transmits those data to third parties. Some sense of these issues can be gained from vendor web sites and representatives, reading web metrics blogs and discussion boards, and consulting colleagues. Note that while some tools offer free demo versions, these may have a limited capacity and functionality that may not repay the learning curve involved in implementing the tool in the first place.

### 2.2 Units of Analysis

Commonly reported web metrics include: the number of visitors to a site; the time and date of their visit; the geographical location of their IP address; whether they arrived via a search engine, bookmark, or link; the page(s) they enter and leave the site; the page(s) they viewed; time spent on individual pages; operating system; and monitor and browser configurations. There are no standard definitions for many metrics, although there are attempts (e.g. by the Web Analytics Association: http://www.webanalyticsassociation.org/) to develop these. Web metrics definitions can also change over time. For instance, Google Analytics’ calculation of the average visit time has, at different points, both included and excluded single page visits. Google’s rationale was that single page visits are often mistakes which result in short visits and skewed average session length data; and excluded these visits for a while, before being reinstating them following requests from users [12].

The more sophisticated web metrics tools can combine and manipulate data across many dimensions and levels of granularity. There are two basic strategies for making sense of large amounts of web metrics data. The first (discussed here) is to identify key metrics that can be reported on a regular basis. The second strategy (Section 4) is to mine web metrics data in a focused way, guided by one or more targeted research or evaluation questions.

#### 2.2.1 Visits, Unique Visits, Page Views, Hits

These terms are sometimes used interchangeably, but they measure different phenomena which can yield widely varying data. While a hit is often a vernacular term for web site traffic, it has not been considered a standard web metric for a number of years.

A visit is the sequential viewing of one or more pages on a web site by a visitor from the same IP address. A visit ends after no further activity from that IP address is detected for a certain period of time, often set at 30 minutes, after which the visitor is assumed to have closed the browser window or otherwise ceased interacting with that web site. A unique visit (as identified by IP address and/or persistent cookie) aggregates all visits made to a web site during a specified time-frame (day, week, month, year, etc.). For instance, the same visitor making three different visits over a 24-hour period is counted as one unique daily visitor; the same visitor making visits on a Monday, Thursday and Saturday is counted as one unique weekly visitor.

Page views count the number of times web pages on a web site are accessed during a visit, including repeat viewings of the same page. Issues have emerged with defining and tracking page views when AJAX (Asynchronous JavaScript And XML) technologies are present, as these permit in-page content refreshing without actual page refreshing, and constantly refreshed AJAX content on the same page will therefore count as only one page view.

A hit is a request for a page or page element (image, frame, movie, etc.). Each page view will usually generate more than one hit: for instance, a page with 15 embedded images will record a total of 16 hits. Despite the common practice of referring to website traffic in terms of hits, this metric is rarely used to report website traffic, as a single visitor can generate a large number of hits on just a few pages.

A web site will record more hits than page views (as a page may generate a number of hits), and more page views than visits (as a visitor may view more than one page). Further, while web metrics tools might be thought to be impartial measures of site traffic, different web metrics tools will measure hits, page views and visits in different ways. For instance, the NSDL runs both AWStats on its own servers and also uses Omniture to provide page tagging metrics, and AWStats consistently reports higher levels of traffic for the nsdl.org site than Omniture does. Here are the figures for May 2006, normalized to a nominal 1000 monthly visits to nsdl.org:

<table>
<thead>
<tr>
<th></th>
<th>AWStats</th>
<th>Omniture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits</td>
<td>37,317</td>
<td>n/a</td>
</tr>
<tr>
<td>Page views</td>
<td>14,401</td>
<td>4,165</td>
</tr>
<tr>
<td>Visits</td>
<td>1,875</td>
<td>1,000</td>
</tr>
</tbody>
</table>

The 37:1 ratio between hits recorded by AWStats and visits recorded by Omniture for the same time period illustrates the importance of specifying the tool and metric being used when web metrics are reported; and it follows that other sites’ published web metrics should be treated with caution, unless it is known how figure being reported was recorded and calculated.
2.2.2 Timeframes
Web site traffic fluctuates over time (Figure 1). For example, daily traffic the nsdl.org site of the NSDL peaks between 7:00 a.m. and 6:00 p.m. EST/EDT. Weekly traffic is lower at the weekend (suggesting that the site is used at or for work). Annual traffic peaks during fall and spring semesters and declines over summer (i.e. during the summer vacation), and during holidays such as Thanksgiving (in the United States) and Christmas.

Sampling time-frames have a major influence on how web metrics are interpreted. For example, if traffic declines in December, is this good or bad news? In the case of the NSDL, such a decline happens every year during the school vacation; it is expected and not necessarily a cause for concern. On the other hand, the same decline would be more worrying for an online retailer hoping for sales over the holiday season. Making sense of temporal fluctuations requires longitudinal comparative data, ideally at least thirteen months’ worth, at which point annual fluctuations can be identified.

2.2.3 Referrals, Geosegmentation, Clickstreams, Bounce and Exit Rates
Referral metrics track how visitors arrive at a web site. This is usually by clicking a link on a web page, a link on a page of search engine results, loading a bookmark, or entering the site URL into a browser address bar. Referral data can identify significant referring sites and TLDs (e.g. .com versus .edu), etc., as can also give an indication of a site’s search engine visibility. Geosegmentation data can provide detailed information of the location of a visitor’s IP address (e.g. country, state, city). Clickstream data visitors’ landing pages, their path through the site (indicating whether visitors appear to be getting lost), and their exit page. Finally, bounce rates measure the percentage of visitors who leave after viewing one page; and page exit rates measure the percent of visitors who exit on a particular page after visiting more than one page.

2.2.4 Session Length
Session length metrics measure the amount of time spent on a page (page session length) or on a visit (visit session length). Session metrics are becoming more popular, on the assumption that they offer more accurate data than page views [26]. However, it is still difficult to draw any direct conclusions solely from session length data. It is impossible to tell whether a user interacted with a web site for entire recorded length of the session: they could have been looking at another page or site (e.g. by using browser tabs [35]), or engaged in other behavior (e.g. taking a phone call).

2.3 Summary
Web metrics tools come in a wide variety of forms and provide many units of analysis. They can be used for regular reporting or for focusing on specific critical questions. However, there is little standardization across tools, and care must be taken when comparing metrics across different tools and sites.

3. ORGANIZATIONAL FACTORS
Web metrics implementation involves a range of organizational issues. Digital libraries are sociotechnical systems, complex mixtures of people, organizations, technologies, practices, policies, and other phenomena [6, 10]. As has been described, web metrics tools vary widely in form, function, and definition.
(STEM) educational resources for educators and learners of all ages [23]. NSDL has worked for a number of years to develop a web metrics strategy [17]. As of 2006, NSDL projects used a range of tools on their own web sites, with the attendant problems of aggregation and consistency. To standardize its web metrics and to report aggregated program metrics to NSF, NSDL began to implement Omniture web metrics on selected NSDL project sites. Addressing privacy concerns, Omniture was configured so that after a user session ended, session cookies expired from the visitor’s computer, and any IP address information was deleted by Omniture, leaving just an anonymous record of the path a visitor took through the web site. Finally, new recommended wording was developed for NSDL privacy policies that described the Omniture tracking technologies and their data and privacy implications.

The Omniture implementation proceeded slower than expected. This was often as a result of a lack of resources at the individual project level. Some projects relied on part-time student assistants who may not have had the time or expertise to implement Omniture. Again, Omniture implementation requires access to site code, and projects with external web developers had trouble accessing the project servers in order to implement the code. (In one case the contracted developers were so unreliable that not only was Omniture not implemented, neither were significant portions of the project site itself). Project managers are often university faculty working part-time on the projects who may not have web metrics experience. In general, new projects often have more immediate priorities than implementing web metrics, such as getting their site up and running in the first place. These findings generally support Lagoze et al.’s observations that successful technology implementation in a distributed digital library organizational setting requires the presence of a minimum set of skills and resources – time, server access, technical skills, etc. – at each project, with the absence of any one skill hampering implementation, sometimes severely [21].

4. SESSION LENGTH DATA

In this section we present a series of short case studies focusing on visitor session length data. Session length metrics can measure both the amount of time spent on a web site (visit session length), and the amount of time spent between opening and leaving a web page (page session length) (see 2.2.4). They are gaining attention as indicators of web site use. In July 2007, Nielsen/NetRatings added session length as a primary metric, partly to address concerns with the reliability of page views (recall that high page view counts might be generated from poor/convoluted navigation, as well as by visitor engagement) [26]. This move precipitated considerable discussion in web metrics communities [e.g. 37]; and its immediate effect was to raise video sites, like YouTube much higher in overall ratings, over search sites like Yahoo or Google.

In this section we consider four case studies, from the Library of Congress, the National Science Digital Library, WGBH Teachers’ Domain, and the Instructional Architect, which explore various ways that session length has been used a metric.

4.1 The Library of Congress

The Library of Congress is the United States’ oldest federal cultural institution (http://www.loc.gov), and maintains numerous online collections. The LoC studied session times from pages in the "Selected Civil War Photographs" collection of the American Memory series (http://memory.loc.gov/ammem/cwphtml/). The aim was not to come to any positive or negative conclusions about time spent on a particular page, but to find a way to interpret session length beyond the data produced by most standard metrics tools. In particular, we asked if it were possible to determine if there was a way to evaluate an optimum session time for a specific page, that is, the amount of time most users would be expected to spend on a page of content based on looking at the range of actual times spent on a page. The premise here is that pages should satisfy certain specific needs of users. A page trying to meet multiple needs, will probably meet few needs and optimum session time will be more difficult to determine. A page meeting a few needs well will show patterns of optimal session time because users will be engaging with the content in similar ways. There may be more than one optimal time for a page, for example a user reading a page on line, versus printing the page to read later, but in both cases a cluster should be observed around a certain amount of time.

4.1.1 Data Cleaning and Page Selection

Data were captured for one-week in late 2007 and imported into a database. Session time is associated in a single record with a particular page and the page’s related views. The data had to be reformatted for calculations associated with an individual page (record) with multiple session times and page views. Two options to reformat the data were to create a relational database, or to associate all data within records of the same page together; the latter was chosen.

The data warehouse supplied the following fields: URL, session time, views, and visits, further parsed to include directories and file names. A unique ID field was added to each record, and a further field was added for the estimated number of words in the page. The modified database thus contained a unique identifier, views, visits and session time fields. Each page had multiple records for every session time category. These data were used to calculate the standard deviations of page views related to the ranges of session time.

Records dealing with search were eliminated, as were pages where most of the session times were limited to 30 seconds or less. Session lengths were compared across different types of pages and across several slices of time, and several high traffic pages from the digital library collections were selected.

Session length may appear as an actual value or as a range. The former may be a challenge because of aggregation, while the latter may be difficult to work on mathematically. The session length data in this study were in the following bins: 0-15, 15-30, and 30-60 seconds; and 1-3, 3-5, 5-10, 10-15, 15-20, 20-30, and over 30 minutes. Units in seconds were converted to minute fractions. In order to aggregate information from a single page, the ranges were treated as categorical data. For the calculations, it was assumed that the session time was the mean amount of time in the range; for instance, 2 minutes was the assumed mean for the 1-3 minute range. Session time values were reported in hours, and the maximum time spent on page was calculated as the number of pages views times the mean range time.

Time ranges often approach some sort of logarithmic scale, which complicates computations. One approach would be to take the interval associated with the most page views; however that discounts the fact that the distribution may be uneven with two
very high values bookmarking very low values. To avoid this, five time segments ranging from 0.25 minutes to 20 minutes were analyzed.

### 4.1.2 Interpreting Session Time

The analysis was conducted on a sample data set of pages from one collection. To explore the optimum session time for an individual page, we looked at the standard deviations for all seven-session ranges, based on views associated with each session range. We then reviewed the standard deviation for every group of three consecutive ranges. This gave us five ranges to compare. We looked for variation in views associated with time spent on a page, and the actual aggregate time spent on the page (Figure 2). The total views for the three pages selected were 13K, 11K, and 25K. To take into consideration bounces and inactive session, we looked only at the scope of seven session ranges between .25 and 20 minutes and eliminated the data at less than 30 seconds and more than 20 minutes. The lowest value was eliminated because it likely included a large number of false drops, and the highest was eliminated because of the likelihood that the user had abandoned the page. After excluding the lowest and highest values, the total views number dropped to 3200, 3700, and 13,000 respectively mostly due to elimination of views in the 30 second or less range.

The analysis revealed large differences in the standard deviation of different session time ranges. While it is not possible to guess why these differences occur, it is useful to think about the relative differences in how time is spent on a page. For instance, a page with views equally distributed across all time segments is likely a confusing page for users. While comparing standard deviations between pages which had little value, there was value in comparing standard deviations of views to the total session length. If a guess for an optimum time could be made, it may be associated with the range of three values that correlate to a low standard deviation, relative to the total time spent on the page.

### 4.1.3 Summary

The study identified a method for understanding how session length may be used to understand collection use. Additional analysis might take into consideration the type of page being viewed by analyzing the number of links or the number of words on the page. Future work will apply this methodology for session length to several collections.

### 4.2 Teachers’ Domain

This section is based on web metrics reports produced by Teachers’ Domain for NSDL in 2006. These are available at: http://eval.comm.nsdl.org/reports.html

Teachers’ Domain (TD: http://www.teachersdomain.org/) is a digital library based at WGBH in Boston that provides teachers with pedagogically sound media-rich resources such as video clips, interactive activities, and images from educational public television shows such as NOVA and A Science Odyssey. TD curricula are aligned and referenced to state science standards. Registration is required to use the site and resources [25].

Overall visitor session times for TD site are calculated from log transactions beginning when users log in to the site. They average about 10 minutes (with the caveat that the length of the last page view cannot be calculated and included). This is in line with expectations, based upon the average length of TD videos and data from focus groups, that suggests a typical session involves a teacher visiting the site to use a resource (e.g. a video) in the classroom, with showing the clip, answering questions, etc., and takes about 10 minutes.

Figure 3a shows the results for four different months from 2005 and 2006. The curves change only slightly from month to month, suggesting that these monthly samples are representative.
Logarithmic units are used to compress the time scale and binning, and also because humans tend to characterize things like session in length in terms of time scales rather than the actual values themselves, e.g.: seconds (a “short” visit) vs. minutes (a “regular” visit) vs. hours (a “long” visit).

Figure 3a is clearly not a simple Gaussian curve. The jaggedness on the left side of the distribution (1 minute and shorter) is due to the binning of the data. Upwards of 1 minute the distribution evens out and does have a Gaussian-like appearance that one might expect from a single population. Note that the peak of the long “hump” is just short of 10 minutes (although all values in this distribution are underestimated, as the time spent on the exit page has not been included). These are likely the users that TD had in mind when developing the site, who come to the site and browse and select from the resource offerings. The distribution suggests that there is more than one population of TD user – something not considered when the site was planned – and that this new group is significant in size. Manual checking of a small sample of short session length visits showed that they tended to be referred from very specific search referrals (e.g. ‘carbon cycle diagram’) that guided them directly to the TD resource on that topic. It is probable that, as opposed to the browsers, these are users with specific search needs, who want to locate something, evaluate it, and bookmark, copy, or use it and move on to the next thing on their to-do list.

A second data set looked at the session login frequency of registered users, including those who visited the site multiple times, by recording the time and date of the beginning of the session, and the time between each pair successive logins by the same user, counted individually across all users. The distribution breaks down into three regions (Figure 3b).

1. Several “spikes” starting at 1 day, at one day intervals (if the diagram is enlarged, this pattern can be followed out to nearly a month). These spikes suggest that when many users access TD, they do so around the same time of day, although not every day.
2. A broad “hump” between a few minutes and ~12-18 hours. These are the situations where people log in two or more times a day and the lack of finer structure suggests that these inter-access intervals are more random in nature.
3. Below a few minutes, a large “tail” down to the 1-second level.

The short-frequency tail suggests that many users were re-logging on every few seconds. It was possible that a bug was regularly erasing session IDs. However, further investigation of the login data showed that the same user’s account would begin a session approximately 20-30 times within the same 1–2 minutes on the same day. As a result, the distribution from the top 20 users (representing a mere 0.028% of the entire user sample) accounted for almost 50% of the observed short frequency tail.

In summary, the TD users didn’t behave the way the TD site design expected them too, and these results affected the analysis and interpretation of other TD datasets. For example, student users are underestimated (as they log in with teacher IDs); data concerning ‘top’ users (based upon number of logins) is heavily corrupted by these other usage patterns; and site protection and personalization features based on user ‘type’ (e.g., teachers vs. students) may not be as effective as hoped for, if teachers permit students use their accounts. There can therefore be more than one population of users, whose definitions of ‘success’ on the site might not be the same as the model used in its construction: something to consider when contemplating future development to incorporate different needs. One plausible explanation for these data is that teachers were giving out their TD password to students in a computer lab to do work on the site, and that what the logs were showing was 20 or 30 students logging in simultaneously. This also might account for some of the short-length sessions, if the instructions given to the students required them go to a specific URL to view a single resource on the TD site.

4.3 The National Science Digital Library

The NSDL maintains a ‘portal’ site at nsdl.org that supports a range of user activities, including: search across various NSDL collections; browsing and visualization tools; ‘community areas’ for K-12 teachers, college faculty, librarians, and others; information for NSDL developers and community members; access to blogs; and other services. Session data for nsdl.org suggest that most common action of visitors was to proceed directly to the search results page (presumably by initiating a search), and then to leave the site. This accounted for 55-60% of all traffic that passed through the nsdl.org front page. Other sections of the site were accessed less frequently.

This pattern was triangulated with a heat map visualization of one month’s user activity on the nsdl.org front page using ClickDensity (http://www.clickdensity.com/), a javascript page tagging tool that records the location of each user’s click on a page, and displays the data in the form of a heat map (see 2.1) (Figure 4). The data showed that users clicked mainly on the search box (presumably to enter a search term) and also on the search button (presumably to initiate a search). Data which tracked the ‘time to click’ for each page element showed that while the mean time to click on the search box was 25.8 seconds, the mode was only 1 second, suggesting that many users clicked straight into the search box once the front page had been loaded. (The presence of clicks apparently ‘outside’ the front page in Figure 4 is explained by the fact that the nsdl.org layout at the time used a centered fixed-width HTML table, which left gaps in on either side when loaded in a wide browser window.)

These data were further triangulated with 2006 usability testing of the nsdl.org search results page – including paper prototyping.
interviews, and surveys – in which test subjects (K-12 educators and university TAs) described how they wanted to use NSDL quickly and efficiently for ‘just in time’ professional development, searching for lesson plans, information, images and so on for quick integration into their existing pedagogical activity. Conversely, they expressed frustration with elements that might impede their rapid use of NSDL, such as poor navigation and distracting graphics (see also [34]). The usability data therefore pointed towards users wanting to use the site quickly for ‘one stop shopping’ (c.f. the findings from the Teachers’ Domain and IA, this section).

Taken together, all these data suggest that NSDL users who are shopping’ (c.f. the findings from the Teachers’ Domain and IA, pointed towards users wanting to use the site quickly for ‘one stop shopping’ (c.f. the findings from the Teachers’ Domain and IA, this section). Conversely, they expressed frustration with elements that might impede their rapid use of NSDL, such as poor navigation and distracting graphics (see also [34]). The usability data therefore pointed towards users wanting to use the site quickly for ‘one stop shopping’ (c.f. the findings from the Teachers’ Domain and IA, this section).

In this inquiry, the usability data and both sets of web metrics were mutually supporting. The former provided context for interpreting the web metrics, and the latter countered criticisms that the usability subjects might have been making false reports.

### 4.4 The Instructional Architect

The Instructional Architect (http://IA.usu.edu) is an end-user authoring service designed to support the instructional use of resources in the National Science Digital Library and, more generally, on the Web. The IA increases the utility of online learning resources for classroom educators by supporting users (particularly teachers) to discover, select, sequence, annotate, and reuse online learning resources stored in digital libraries, in order to create instructional objects (e.g., lesson plans, study aids, homework – collectively called projects) [26]. Many teacher-created IA projects are fairly simple: teachers are not web professionals attempting efficiently to address classroom and learning needs. To support their needs, much of IA is similar to blog software coupled with a social bookmarking system; and a user-centered design process has ensured that the system meets the basic requirements of teachers who wish to use digital library content to quickly and easily meet classroom demands.

From the home page of the Instructional Architect, users can:

**Browse.** Users can access projects by keyword searches or by browsing by subject area, grade level, author’s last name, or title.

**Register.** Users can create a free account, which provides them exclusive access to pointers to their saved online resources and IA projects.

**Login.** Users can log in to their personal accounts, and take advantage of the IA tool set and features. Guests can log in, although with greatly reduced functionality.

The IA offers three major usage modes for registered users. First, the My Resources tool lets users search for resources in the NSDL, with results and metadata being displayed in abbreviated form. Users can select resources for further use by adding them to a folder, and can also add Web resources by entering URLs (although Web resources will not have associated metadata).

Second, the My Projects tool allows users to create web pages in which they sequence and annotate their selected resources in order to create instructional projects. Users can add basic project metadata such as subject area, grade level, and core curriculum standard, and these metadata are used to support project search and browse. Third, users can Publish and share their projects, setting permissions for who can view them (user-only; student-view via a user-created password; and public view, i.e. anyone browsing the IA site).

#### 4.4.1 Web metrics Strategies

IA development began in 2001 as an NSF sponsored project. In the beginning, very little user tracking was implemented, as the focus was on system functionality. IA gradually began logging various aspects of online user activity in a custom back-end PostgreSQL database. Since IA users register, activity can be logged on a per user basis, rather than being inferred. In June 2007, IA implemented Google Analytics (GA). Discrepancies between custom and GA tracking were discovered (and expected): for example, from June-December 2007, GA reported 65,511 project views, while IA tracking reported 112,261 views.

IA’s initial web metrics approach was driven by marketing questions: how many users does IA have and how often do they visit? There was particular interest in tracking the impact of teacher professional development workshops. Analysis of repeat user site activity showed that the IA was ‘sticky’ for about 10% of its users, perhaps representing the ‘early adopters’ [27]. Since free sites that offer user registration seldom report this number, it is difficult to know how this metric compares with other sites.

User data logging was subsequently expanded, and the content of IA projects was quantitatively analyzed. These analyses required complex SQL queries, data scrubbing and transformation, and revealed that users appeared to prefer smaller granularity digital library resources, and that these require more contextualization that larger resources [28]. While these analyses were very useful, they were not easily done with traditional web metrics data. Custom tracking continues to be important for measuring outcomes unavailable from third party tools. For example, from 2002 to January 2008, over 2,700 users registered, 5,400 projects were created, and 20,500 online resources saved in 6,700 folders; and since August 2006, IA projects have been viewed over 258,000 times. These numbers, by Web standards, are small; however, from a quantitative point of view, analysis and meaningful interpretation is daunting.

The more recent addition of voluminous GA data has permitted more detailed research questions about patterns of activity for IA’s two distinct user populations: teachers, and their students. Do teachers and student usage patterns differ? If so, what is each group doing? The following data were particularly illuminating in this respect: traffic sources, page popularity, and bounce/exit rates. (Unless otherwise specified, all GA data reported were collected between 1 June and 31 December 2007.) First, like the NSDL, GA timeframe data showed access ebbs during summer months and weekends, indicating primarily school use. Second, GA provides reports about users’ geographical location and how users access the IA. These analyses showed greatest use in areas where IA conducts the most teacher workshops. They show that traffic sources are largely direct access (60%; bookmarks and typed URLs), compared to 27% from search engines and 13% for
links from other web pages. This suggests a largely purposeful user base, consisting of logged-in teachers and students accessing instructional resources. This interpretation is supported by the following two results.

GA also reported that the ratio of student login paths to teacher login paths was nearly constant, at about 6:1. This suggests that the site is used as intended: teachers design IA projects then send their students to them. Recall that IA projects consist of URLs to digital library or Web resources with accompanying teacher instructional annotations (somewhat like blogs). Thus, teachers want students to click on their selected links and exit the site. GA data show that for all IA project pages, the bounce rate is 86%, and the exit rate is 70%, suggesting that most IA projects are either the only or final page viewed. In contrast, pages available only to registered users have a bounce rate of 12% and an exit rate of 3%. Finally, GA reports session length as time on site and pages viewed per visit. Unfortunately, due to the way these metrics are reported in GA, there is no way to differentiate between logged in teachers and their students. However, a comparison of the time on page for registered-only pages and IA project pages indicates that project viewing (on average) is two and a half minutes longer than the mean for registered-only pages. This shows that project viewing is longer than editing usage (on a per-page basis), again suggesting differences between student and teacher use. IA is currently implementing tracking code to help us make better sense of session length data.

4.4.2 Summary
Web metrics have provided valuable insight about use and users of the IA. While it is often challenging to triangulate between the two sources of tracking data, as well as with other forms of data (e.g., user surveys), it is only by considering all sources that impact on users’ knowledge, attitudes, and behaviors can be understood.

5. DISCUSSION
Each digital library has its own technical and organizational concerns. Despite these differences, two main areas of findings emerge from this exploratory work: theoretical questions concerning the dynamics of session length metrics, and practical questions driving different webmetric implementation and analysis strategies, as well as the resources needed for implementation and analysis.

5.1 Session Length
The results indicate that some of the basic assumptions in considering session length as a key metric — namely, (a) that session length is a viable alternate metric to page views, and (b) that longer session lengths are preferable to shorter session lengths — are problematic, at least in the context of digital libraries. We suggest that this is partly because this putative new industry standard implicitly posits an absolute relationship between session length and web site quality, whereas in fact this is a contextual relationship. For some sites, short sessions might be indicators of quality; for other sites long sessions might be indicators of quality.

The contextual nature of session length as a metric is apparent in all the studies. The NSDL study provides evidence that shorter session lengths are desirable for users who are attempting particular and/or focused tasks based around the use of the nsdl.org search engine. (It should be noted that the search tool is by no means the only one available to NSDL users. For instance, NSDL has recently launched a series of blogs called ‘Expert Voices,’ http://expertvoices.nsdl.org/ where usage patterns indicative of user satisfaction, such as session times, may be different from those found in the use of the search tool.) The Library of Congress study suggested that the notion of relative session time is useful; the absolute measure of session time cannot be compared between pages, but is useful to see within the same page over time. In the case of Teachers’ Domain, the sessions were longer, but it was assumed in this case that session length was driven by the length of the TD video resources being accessed; and again, that having achieved this task, that the TD user would then want to exit the site. In the case of IA, analyses of session length has shown differences in average session length between users undertaking different types of tasks, again showing that session length is not a ‘one size fits all’ concept.

The desire for long session times may be understandable from the point of view of Web content providers interested in ‘sticky’ sites that can be used as platforms for generating advertising revenue, but from the point of view of educational content providers attempting to address the ‘just in time’ professional development needs of a wide range of educators, long sessions may not necessarily be desirable. Rather, ‘efficiency,’ ‘ease of use,’ and other attributes highlighted in the studies, would seem to be preferable. For instance, research into educators’ perceptions of quality in online educational resources suggests that anything that is distracting from the task at hand contributes to perceptions of poor quality [34]. If we equate distraction with time wasted, and therefore more time spent on site, the argument can be made that longer session times could, in some circumstances, represent poorer quality interaction, and more distracted and less satisfied users, than shorter session times do.

This in turn raises the question: What is, if any, an appropriate session length metric for a digital library? Here, we argue that the parameters of such a metric can only be determined through the triangulation of web metrics with other forms of data, such as usability testing, interviews, and surveys. Only when users’ needs and tasks are understood in the context of the overall design of the site in question, can predictions be made as to the desirable length(s) of session time(s) for a particular site.

5.2 Practical Web Metrics Issues
In addition to these specific findings, a number of wider issues associated with web metrics implementation and management in digital library projects were identified.

First, it is important to provide adequate project resources in such areas as technical expertise, server access, staff time, etc. As the Library of Congress case shows, even obtaining, cleaning, and binning data for initial analysis can be a laborious process. In the case of NSDL, distributed web metrics implementation was found to depend heavily on the resources available at the local project level.

Second, when analyzing web metrics, it is important to remember that no firm inferences regarding users’ intentions can be made solely from web metrics [16, 36]. Web metrics record the actions derived from users’ thought processes and intentions, not the thought processes and intentions themselves; in psychological
terms, they are a measure of user behavior, not of knowledge or attitude. For instance, there is no way to tell whether two visits from the same IP address are from the same person, or two or more people using the same computer at different times (e.g. in a school library), or a group of students using one computer at the same time (as in the case of IA). Similarly, in the case of session length data, we do not know just from the data whether or not short or long visits are satisfactory or unsatisfactory. One way to address these inference issues was to triangulate web metrics data with other data, such as usability work, field studies, user panels, interviews and focus groups, etc. [e.g. 16] Evidence for web metrics inferences and interpretations can be found in other data sources, and web metrics can also be used to support inferences and triangulations from those other data sources.

Third, evaluators are often asked to provide succinct summaries of ‘what is going on’ to developers, managers, sponsors, and others; however, web metrics rarely (if ever) provide unambiguous information. Successful internal and external communication of web metrics should therefore describe the provenance and relevance of the data, the nature of any triangulation, the tool(s) and measurements used, the possibility that other tools might provide different data, the difficulty of drawing inferences, and possible sources of and margins of error. Here, one common potentially misleading audience mental model of web metrics is that they are typified by Gaussian distributions (i.e. with coincident means, modes and medians). However, Web sites, as Internet nodes, exhibit many of the power law distributions typical of the Internet, characterized by a small number of data with high frequency counts at one end of the distribution, and a large number of data with low frequency counts at the other end [c.f. Pareto, Bradford, Zipf, and other distributions: e.g. 2, 3, 4, 29, 31]. For example, most referrals will usually come from a few well-connected sites, while many other sites will contribute just a few referrals each; again, a few pages on a site will be viewed many times, whereas most pages will be viewed very few times (contrary perhaps to the expectation that all parts of a site should prove attractive to visitors). In communicating metrics it should be explained that these highly skewed distributions are not bad news, but are expected, and that there may be ‘long tail’ arguments that could be made regarding the value of data points that lie far out along the distribution [1].

6. CONCLUSION
The web metrics projects described in this paper have covered a wide range of project types, tools, data, research questions, and findings. Across this complexity, a number of common themes have emerged. These include the importance of:

1. Identifying the appropriate tools for the job (for instance, is page-tagging an appropriate solution if a project does not have easy access to its servers?)
2. Supporting web metrics implementations with sufficient and appropriate resources at the project level
3. Identifying and where possible reconciling different definitions for the same web metrics
4. Identifying and tracking key metrics longitudinally in order to begin identifying and accounting for temporal fluctuations
5. Triangulating web metrics with other research (interviews, surveys, etc.)
6. Analyzing and reporting web metrics in clear and comprehensive ways that address possible audience misconceptions
7. Integrating web metrics research into project evaluation and sustainability goals

In addition, we have discussed and demonstrated some of the uses of session length metrics, and identified some of the potential pitfalls involved in moving to considering session length as reliable indicators of web site use, particularly the question of whether, in terms of session length, ‘more’ is necessarily ‘better’; and we have proposed that for session length metrics to be useful, they have to be triangulated with other data associated with the context of use. These findings are useful both for practitioners engaged in regular web metrics reporting, and for researchers answering wider digital library research questions.

The limitations of this work so far entail the work needed to obtain generalizable results. Much discussion was required to align the research questions and units of analysis presented in this paper. However our findings suggest that this is a fruitful area for ongoing research, and we hope in future work to build on these models and results, in order to work towards more generalizable web metrics models of digital library use.

7. ACKNOWLEDGEMENTS
The IA work was supported by grants from the National Science Foundation (ESI-0544440 and DUE-0333818). IA thanks the participating teachers. The NSDL and TD work was supported by a grant from the National Science Foundation (DUE-0424671). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

8. REFERENCES