# A Survey of Web Log Data and Their Application in Use-Based Design

Mary C. Burton
*IBM*
*mburton@us.ibm.com*

Joseph B. Walther
*Rensselaer Polytechnic Institute*
*walthj@rpi.edu*

## Abstract

*Web-based logs contain potentially useful data with which designers can assess the usability and effectiveness of their choices. Most guides to World Wide Web (Web) design derived from artistic or usability principles feature no empirical validation, while empirical studies of Web use typically rely on observer ratings. Several sources of unobtrusive usage data are available to Web designers, including Web server logs, client-side logs, and other data. The naturally-occurring traces recorded in these logs offer a rich data source, amenable to normative use assessments and to experimental research comparing alternative Web designs. Identification of types of Web server logs, client logs, types and uses of log data, and issues associated with the validity these data, are enumerated. Finally, frameworks that outline how sources of use-based data can be triangulated to assess Web design are illustrated, and an approach to experimentation that overcomes many log data validity issues is presented.*

## 1. Introduction

Numerous sources offer rules and guidelines for optimal Web page design, derived from rhetoric, visual communication, and usability studies. Yet few if any have verified these recommendations using large-scale empirical methods. The ways in which users interact with a World Wide Web site provide potentially valuable data on the usefulness and effectiveness of Web design elements and content. The log files recorded by Web servers, and client logs, offer potentially useful data about users' Website interactions, which may be studied to generate inferences about Website design. At the same time, aspects of Internet connectivity such as floating Internet Protocol (IP) numbers, shared computers, and other issues, limit the validity of normative interpretations about Web usage based on server log data alone.

While information pertaining to the previous observations has appeared in a variety of information sources, a synthesis is provided here, and is keyed to the features of commercially available logfile analysis programs. Moreover, we develop frameworks which extend the ability of analysts to utilize these data: Triangulating server data with client-side log data and other sources of information, or collecting server data in the context of formal experiments, overcome many of the problems described elsewhere. This paper explores the use of Website user interaction data, or "use-based" data, in the context of Website evaluation and design.

### 1.1. Assessing the impact of Websites

Korgaonkar and Wolin [14] estimated that 55 million people surf the Web, and that on-line traffic has been doubling every 100 days; yet many companies have been disappointed with the commercial potential of the Web, as Web user purchases are less than expected. In order to remedy disappointing commercial results on the Web, they propose, a better understanding of Web users is needed. Korgaonkar and Wolin's original research methods offered a fine-grained analysis of Web users' motivations, concerns, and demographics in the context of three uses of the Web: the number of hours spent on the Web; percent of time spent on the Web for personal versus business use; purchases made on the Web, and the approximate number of orders placed on the Web. Such analyses begin to provide a use-based glimpse into the behavior of Web users.

A frequent goal for Website designers is to keep as many Web users as possible at their site for long periods of time and to have users navigate to numerous Web pages within their site. This is referred to as the "stickiness" of a Website. Media Metrix tracks "sticky traffic," which is the Web equivalent to Nielsen Media Research's television ratings. Media Metrix tracks 21,000 Websites, and generates monthly reports based on the number of "unique visitors" at each site for that period [15]. Website stickiness is one indication of Websites' popularity and usefulness. Yet stickiness and other characteristics need not only inform gross levels of popularity. The data from Web traffic can also inform design and evaluation in more detailed ways, and are available to anyone with access to servers' internal files.

### 1.2. Web design guidelines: Rhetorical, empirical, and "click" studies

What are the possible criteria, and sources of guidance to achieve those criteria, that pertain to the performance and design of Websites? Good design requires an understanding of Website users' tasks, needs, and expectations, and many Web guidelines attempt to address these concerns. While this work may be considered a sub-field of usability, it tends to differ in two respects. First, much Web design guidance tends not to be presented in the context of use goals, but as universals. Second, there is often a less apparent basis in research or a disciplinary foundation than in other usability sub-fields. Hence, much such advice is questionably based in traditional rather than empirical knowledge. Guidance on Web design—from books, Websites, or corporate guidelines—is often aimed at increasing consistency, predictability, and ease of use of Web interfaces. As often as not, guidelines emerge through commonly used graphical user interface elements that become standard, or are based on designer and programmer preferences.

Among the most popular of such sets of guidelines, Jakob Nielsen, a highly respected Web user interface designer, provides valuable Web design guidance in his monthly "Alertbox" column. Two columns in particular [17, 18] describe mistakes commonly made by Web designers. These include (1) page elements that are in a constant state of animation, (2) long scrolling pages, (3) non-standard link colors, (4) long download times, (5) launching new browser windows, (6) lack of author information, (7) moving pages to new URLs, and (8) designs that look like advertisements.

Nielsen also offered "Ten Good Deeds in Web Design" [19]. These include (1) placing a name and logo on every page, (2) providing search on a Website with more than 100 pages, (3) using simple headlines and page titles, (4) using a page structure that makes scanning easy, (5) providing pages that are accessible for disabled users, (6) following what large Websites are doing.

Even though Nielsen [20] recommends that Web designers follow what large Websites are doing, he observes that the most commonly used design elements on Web pages may not be the most usable approach: While a design element may not be the best method for the situation, users expect to see design elements that they have learned through their experiences with larger Websites. Such conventions include using blue hypertext links (blue does not increase reading speed where other colors would), the use of horizontal tabs across the top of the screen to indicate main topics, or a colored stripe to indicate primary navigation elements on the left side of a Web page.

In contrast, a small number of empirical studies on Web use suggest design decisions, as well. For instance, Spool et al. [23] found that when users are looking for information they are focused, and that design approaches that are meant for "surfers" (e.g., advertisements) are distracting for information seekers. Catledge and Pitkow [7] determined that a hierarchy of information or database search might work for the goal-oriented user, but that these methods prove frustrating to a user whose desire is to happen across unexpected information. In their study of user Web behavior, they found that eighty percent of document requests were requests for a document (of type http) from a Web server. They also found that the navigation method preferred by users were hyperlinks, where hyperlinks made up 52% of all requests for documents. Therefore, Catledge and Pitkow determined that to a great degree, hyperlinks are the preferred method of Website navigation. Second to hyperlinks in popularity among users was the browser "Back" command, which accounted for 41% of all requests for documents. The researchers determined that the users in their study interacted in a small area of a Website and frequently backtracked. They also found that users typically navigate two layers within a site before they return to the point at which they entered it. They offered the following design advice from their study:

- Important information should be located within two to three clicks from the home page, given that users accessed an average of 10 pages per Web server.
- Too many links on a Web page may increase the time it takes users to find the information they desire.
- Groups of related information should be used, given that users interact with small areas of a Website.

Tauscher and Greenberg [25] reported on the patterns of user revisitation to Web pages. Users revisit Web pages at a rate of 58%. Interviews indicated that users revisit sites in order to (1) monitor changing information, (2) further explore a Web page, (3) use a "special purpose" page (e.g., a search engine page), (4) modify a page as its author, or (5) access another revisited page. In contrast, people access *new* Web pages in order to (1) satisfy changing information needs, (2) explore a Website, (3) visit a recommended Web page, or (4) explore a page while browsing for another item.

Byrne et al. [6] studied user tasks in the context of daily use of the World Wide Web. They built a "taskonomy" or a taxonomy of Web tasks. They established six general categories of Web tasks, as follows.

- Use information: Tasks include most Web browsing actions and are defined as one or more tasks in which a user attempted to make use of (read, listen to, view, watch, duplicate, download, display) information from the Web

- Locate on page: Tasks where a user must find a link on a Web page in order to use information or go to a URL.
- Go to page: Tasks include when a user types a URL, uses the Back or Forward button in a browser, etc.
- Provide information: Tasks might involve the completion of a Web-based form.
- Configure browser: Tasks where a user may resize a browser window.
- React to environment: Tasks where a user may respond do a dialogue box that is displayed by the browser.

Byrne et al. expected a hierarchy of tasks but instead found a "flat" structure (any one of the above general task categories can have any other of the tasks as a subgoal). They concluded that the category with the most numerous events was Configure but that Use Information tasks accounted for the most time spent by users, followed by Locate.

The rationale for the Byrne et al. [6] study was based on their observation that research into user navigation patterns on the Web, or "click-studies," includes very little information about user tasks and user context. Their study, therefore, focused on user context and user tasks. Their findings suggest that users spend more time reading Web pages, visually searching Web pages, and waiting for Web pages to load than they do interacting with graphical user interface buttons and browser history mechanisms. They also found that users are willing to scroll and read long texts, although there is a tradeoff between designing a Web page for reading (in the case of online documents that users find and are willing to read) versus scanning (in the case of users who are searching among online documents for a particular topic). Byrne et al. suggest that a task-oriented Web behavior study like their own could be combined with "click studies," where the click studies would provide more detailed information on users interactions with Web pages (e.g., most frequently visited links on a page).

Each of the above approaches provides an example of using some form of analysis to articulate or to validate design decisions. At the same time, the actual browsing behavior of Web users creates numerous discernable data, speaking to a number of different variables, the inferences from which offer a rich source of usability information. Such data on the actual experience of users on a particular Website are needed to validate Web design decisions and to enhance the design of a Website to increase ease of use and value. The following sections will explore the sources of Website user behavior, or use-based, data. Three primary sources of Web interaction user data are considered: Web server logs, client-side logs and usability data.

## 2. Sources of Web use-based data

### 2.1. Web server logs

A Web server is a software program that runs on a "networked machine." The server waits for connections from outside of it, and serves particular documents in response to a request from a Web browser [12]. Web server data are created from the relationship between a person(s) interacting with a Website and the Web server. A Web server log, containing Web server data, is created as a result of the httpd process that is run on Web servers [5].

All server activity–success, errors, and lack of response–is logged into a server log file [12]. Web servers produce and update dynamically four types of "usage" log files: access log, agent log, error log, and referrer log [4].

**Access Logs** provide the bulk of the Web server data, including the date, time, user's IP address, and user action (e.g., whether or not the user downloaded a document or image). The following information can be obtained from an access log:

a) Domain name (e.g., .jp, .org or Internet Protocol (IP) number)
b) User access date and time
c) Item (e.g., image, html file, etc. accessed)

**Agent Logs** supply data on the browser, browser version, and operating system of the accessing user.

**Error Logs** contain information on two specific events: (1) 404 (file not found) errors; the time, user domain name, and the page on which a user received the error is recorded, providing a server administrator with information on "problematic and erroneous" links on the server. (2) Stopped transmissions; information on "user-interrupted" transfers are recorded (e.g., a user might click the browser "Stop" button which would produce a "stopped transmission" error message).

**Referrer Logs** provide information on what other Websites link to a particular Web server. Information, such as the sites that most frequently refer to a particular server (e.g., users may often arrive at a particular Website through a search engine and the search engine would be logged as a site that frequently refers a user to a particular Website), can be obtained from the Referrer Log [4].

Web server logs are stored, in general, in Common Logfile Format or Extended Logfile Format. Common Logfile format includes date (date, time and timezone of a request), client IP (remote host IP and/or DNS entry), user name (remote log name of a user), bytes transferred, server name, request (URI query), and status (http status code returned). Extended Logfile Format includes bytes sent and received, server (name, IP address and port), request (URI query and stem), requested service name,

time taken for transaction to complete, version of transfer protocol used, user agent (service provider), cookie ID, and referrer (previous page) [5].

## 2.2. A Web server log data primer

The following discussion of server log data is divided into three groups: Navigation and activity, demographic, and performance.

**2.2.1. Navigation and activity server log data.** Navigation and activity server log data provide information on user interaction aspects of a Website i.e., navigation paths, number of accesses, time spent on a page, etc.

Various server logging tools provide an array of information about the activities of users on a Website. The information collected by these tools, and the terms used to describe data, are not always consistent between server logging tools. Following is a list of information that is collected by server logging software programs, accompanied by a discussion of the meaning of the data.

**Visitors and visits**. A *visitor* may be defined as a unique Internet Protocol (IP) address. Although an IP address may represent one person only, an IP address is in mose cases shared by more than one person [2].

A *visit* is a set of requests depicting all the pages and graphics seen by a unique visitor at one time. For example, a visitor to a Website may go to eight HTML pages and in the process request fifteen graphics. In total, these 24 requests equal one visit. Note that the total number of visits is typically greater than the total number of visitors as each visitor can visit a Website more than once.

Inferences about visits are imperfect, however, and visits are merely estimates: one cannot be certain that a series of requests is associated with one person, or the same person, within the same visit [2]. AccessWatch [1] defines a visit as "a unique host active during the period of an hour". Examples of unique hosts are www.aol.proxy.com or an IP address such as 133.80.61.2. According to AccessWatch, this type of data provides an indication of the degree to which users are interested in a particular Website.

**Hits, accesses and requests**. A *hit* is "any file from a Website that a user downloads," and *accesses* are "an entire page downloaded by a users regardless of the number of images, sounds, or movies" [4, p. 375]. A user accesses one Web page only of a Website even if that downloaded page has a number of images on it. Neither hits nor accesses represent unique users: many Internet Service Providers use proxy servers, which further complicates the situation because the Access Log will reflect the number of hits or accesses by a referring server, instead of by the number of users [4].

Definitions of hits and accesses vary somewhat among server log analysis tools. Some server log analysis tools track "requests." In the Web logfile analysis program Analog, a request is "when a Web server is asked to provide a page, graphic or other object". A request may be created by a "visitor going to a page or by the page itself requesting an object (usually a graphic)" [3]. The Analog [3] program distinguishes between requests (the number of transfers of any file type) and page requests (the number of transfers of HTML pages).

Analog tracks the "success" of requests. Success is defined in terms of HTTP status codes i.e., status codes in the 200 range (meaning a document was returned) or with a code of 304 (a user could use a cached copy of a document so the document was not required from the server). Analog treats logfile lines with no status code as a success. A redirected request has a status code in the 300 range (with the exception of 304) and indicates that a user was directed to a file other than the file originally requested. A common use of redirected requests is for click-through advertising banners. HTTP status codes also fall in the 400 range (indicating an error in the request) or in the 500 range (indicating a server error). The most common failure in this range occurs when a file is read-protected or not found. HTTP status codes in the 100 range are information status codes and are rare [3].

The following are a subset of hit and access data that are provided by server log analysis tools: number of hits, number of visits per hour, visitor view of pages (AccessWatch)[1], number of requests, average successful requests per day, average successful requests for pages per day, failed requests, redirected requests, distinct files requested (Analog) [3], most requested pages, least requested pages (WebTrends) [27], 7 most popular pages, average requests per day of the week (Accrue HitList) [2], top 40 pages that were requested at least once (SurfReport) [24], most common single page visits, and most popular directories, which could reveal the most requested information types (Accrue HitList) [2].

**Paths**. Paths may be defined as "the average length of a user's sessions, specific location duration (e.g., average time on a page), average download times, and how the user navigated through the site (e.g., entrance and exit points)" [4, p. 376]. Server logging software tools provide the following output related to paths: unique paths, average path length (reported in pages), longest path (reported in pages) (SurfReport) [24], previous pages viewed within a site--which may aid in determining how a user navigated to pages within the site--and jumps from the home page, which can indicate the most often used links (Accrue HitList) [2].

**Entry and exit pages**. Entry and exit page data provide information on where users enter and exit a Website. Server log analysis tool output include data such as top ten entry pages, top ten exit pages (SurfReport) [24], most popular entry pages, and most common exit pages (Accrue HitList) [2].

**Click-throughs**. A click-through occurs when a visitor to a different Website clicks on an advertisement that ultimately redirects the visitor to the logged site. Web servers track click throughs only when the HTML that contains the advertisement has been written so that a click does not go directly to another Website but instead goes to an application on the site where the advertisement was displayed, and then to the final destination. Applications that handle these transactions are called redirection programs [2].

The Accrue HitList [2] server log analysis product tracks "impressions," or the number of times an advertisement is requested from a Web server. The number of impressions may be less than the actual numbers of times an advertisment was viewed by a visitor due to Web browser caching. Accrue HitList calculates the click-through rate as the number of click-throughs divided by the number of impressions. According to Accrue HitList, the click-through rate is an "indirect estimate" of the effectiveness of an advertisement.

**Duration**. Duration may be defined in terms of average time per visit or average time per page [2]. Duration, however, may not describe interactive use of a Website. Users often leave a Web browser open and running for extended lengths of time without interacting with the browser [7]. Also, a Web user may not look at an entire page.

**Downloads**. Server log data can provide information on items that are downloaded from a Website. In the Accrue Hit List [2] product, downloads can include.zip files and applications (e.g., .exe files). Download information that is tracked by Web server analysis tools includes visitor downloads (AccessWatch) [1], and 7 most popular downloads (Accrue HitList) [2].

**Browsers**. Server log data can provide information on Web browsers used to access a Website, including browser software (Microsoft Internet Explorer, Netscape Navigator, Other) [1], and most popular browsers [2].

**Errors**. Server logging tools can provide error information: status code report (e.g., "Access Forbidden") (Analog) [3]; top 10 bad requests, and top 10 bad source pages (Accrue HitList) [2].

**Search engines**. Server log data can indicate the use of search engines in the context of a particular Website, indicating the following: percentage of traffic generated by search engine (based on a percentage of visitors and a percentage of visits), top 10 keywords used to find the site, top 9 search engines referring to the site (SurfReport) [24], and most common search engine crawls (Accrue HitList) [2].

**2.2.2. Demographic server log data**. Demographic data describe the "kinds of people" accessing a site. Examples of these data include: accesses by domain e.g., .org, .com, .jp [1], most active organizations (as determined by IP address or domain name e.g., aol.com), most active countries (as determined by the suffix of the domain name), new versus returning users (reported as a percent of the total number of visitor sessions) (WebTrends) [27], visits by distinctly authorized (requires a user id and password) users (Wusage) [29].

**2.2.3. Performance server log data**. Performance describes the load on a Web server and the responsiveness of the Web server. Performance information can include: megabytes of information served by the site, page demand, defined as the average number of pages traversed and average time to download a given amount of information for a specified number of visitors within a specified amount of time (AccessWatch) [1], or average data transferred per day (Analog) [3].

## 2.3. Web server log data validity issues

There are key issues associated with the completeness, accuracy and representativeness of server log data. These include caching and unique user identification. Due to these issues, it is suggested that Web server log data should be used for high level, general information [16]. Two relatively safe conclusions from server logs are (1) that hits received were *at least* as many as what the server log revealed, and (2) each different site/machine listed in a server log reflects at least one unique user access–it is impossible to determine if the site/machine represents one user, or more than one user using the same site/machine to access a Website.

**2.3.1. Caching and browsers**. A Web browser may make what is known as a "conditional request" to a Web server. In a conditional request, the browser only requests a document or inline object from the server if a page is not already stored in the browser's "disk cache". This method reduces network traffic. However, from a Web server logging perspective, pages that are served from the browser cache will not be recorded in the Web server log. Therefore, user data will not be captured in this situation [12].

**2.3.2. Caching and proxy servers.** Proxy servers are used by Internet service providers, and private and public

institutions with a large user base, in order to protect a network from unauthorized parties, and/or to reduce network traffic [12]. To reduce network traffic, pages that are requested and loaded into a browser, via a proxy server, are stored in the proxy server "disk cache." The idea is that documents that are often requested by users may be accessed from the proxy server cache, rather than from the Web server where the document originally resided. As in the case of browser cache, from a Web server logging perspective, pages that are served from the proxy server cache will not be recorded in the Web server log. Therefore, user data will not be captured in this situation.

Internet service providers and private and public institutions with users who are located in a limited geographic area might consider disabling proxy servers so that they could more accurately track server usage via Web server logging tools. However, the decision to disable a proxy server would need to be weighed carefully against any resulting degradation in performance – if users experience long wait times in page downloads, users may abandon the use of a Website.

**2.3.3. Unique user identification.** Each unique site in a server log may represent one or more unique user accesses [16]. For example, if 300 hits are recorded from Australia (.au), it is impossible to know if this is 300 hits from one person in Australia, or 300 people in Australia accessing a Website [11].

In addition, an IP address does not necessarily represent the same computer due to dynamic IP addressing via Dynamic Host Configuration Protocol (DHCP). Computers are manually assigned a permanent (fixed) IP address in traditional TCP/IP networks [26]. However, when DHCP is used, computers are assigned IP addresses dynamically without manual intervention. However, IP addresses, whether fixed or dynamic, are not reliable sources of user identification because it is the computer, not the user, who is assigned an IP address and multiple users may use a single computer to access Websites.

## 3. Client-side logs

Client-side serving logging tools have been used to address some of the shortcomings of server logging data and tools. Client-side logging tools can capture Website navigation through cached documents "overcoming some of the problems associated with analyzing standard Web server logs" [9, p. 573].

Client-side server logging tools are predominantly used as a means of collecting data in a controlled study environment, rather than in commercial applications. Etgen and Cantor [10] developed the Web Event-logging Tool (WET) as an alternative to Web server log data, due to the following limitations of server log data: Web server logs do not collect data on client-side user interfaces, including Java applets and form element interactions; proxy server and browser caching impacts the validity of server log data. WET was designed to provide usability data on Website use. It is currently considered as a complement to other usability testing data collection techniques, including usability tester notes that are collected manually. Another client-side tool, Listener, was also developed to capture client-side Website usability data. Listener is designed to capture a user's navigation through a Website through navigation elements such as links [9]. The benefits of Listener are described as follows:

- Provides access to Website user interaction behavior in the case when a usability tester does not have access to server logs
- Records Website user interactions that are not captured by Web server logs i.e., records actions on Web pages that would not be recorded in a server log due to caching
- Listener will operate without an HTTP server connection, which could be an advantage for usability testers who do not have access to a Web server and/or network

Client-side logging tools have been used in an advantageous way to capture Web user interactions in a usability testing situation. Client-side logging tools provide more detail about user interactions with Websites and also address the problem of Web server caching. This detail could include a user load of a page, a user click in a checkbox in a form and submission of the form [9]. Data gathered in client-side logs includes event date or time, type of event (e.g., load, click, submit), elements of an online form including source type (e.g., checkbox), source name ("Submit Now") and source value, and event/source location (e.g., /orderstuff/order_form.html) [9]. Therefore, as stated by Catledge and Pitkow [7], "actual user behavior, as determined from client-side log file analysis, can supplement the understanding of Web users with more concrete data" (p. 1065). In other words, client-side log data could be used with server log data and other forms of usability data collection to provide a more complete description of user interaction on a Website. However, a method to capture client-side interactions in a large scale, commercial setting has not been developed.

## 4. Validating Web design

A combination of traditional usability testing techniques, client-side server logging and Web server logging may

provide the best opportunity to understand how users interact with a Website, what tasks users are trying to accomplish, and what improvements should be made to a Website to increase ease of use. Given the validity issues noted above with regard to interpreting logs at face value, triangulation of such data would wisely be combined with traditional approaches to usability testing. These approaches might include (a) observation of user interaction on Websites in a usability laboratory or in a field setting, using video and audio taping of free-form and/or predetermined tasks, recording navigation patterns and user comments for observers, who would note user interactions and analyze their notes and the recorded data for insights; and (b) remote evaluation of Websites, using online questionnaires and telephone interviews. As Kanerva et al. [13] suggest, "The ultimate success of (software) is difficult to measure in tangible, reliable behaviors like task time or number of errors. In addition to traditional measures, researchers have to make strong use of natural observation and subjective questionnaires." Whether or not to use one or more of these sources of use-based Website data depends upon the particular research question(s) at hand. The following section offers examples, for the purpose of illustration, of how Website design principles can or cannot be validated using each of three types of use-based data: Web server logs, client-side logs and/or usability testing.

### 4.1. Example 1: Provide easy access to information that is most frequently used.

**4.1.1. Server log data.** Analyze what pages users access most often. Analyze paths to determine if the most efficient access to important information is available. Look at common entry pages – does this indicate frequency of use and/or importance? Compare paths between most frequently and least frequently used to determine if a particular design impedes access to less frequently accessed pages. Also note that in the server logging product Accrue HitList, the use of requests to determine the popularity of a site is not appropriate because a page with many graphics will generate more requests than a page with fewer graphics. Therefore, the number of visits or accesses (the number of HTML page requests) will provide a more precise representation of activity [2].

**4.1.2. Client-side log data.** Analyze what page elements users access most often. Analyze paths, links and elements to determine if the most efficient access to important information is available.

**4.1.3. Usability testing data.** Observe what users use most often. Ask users to rank tasks and information needs by importance and frequency of use.

### 4.2. Example 2: Do not provide large files on a Website that require long download times [17] nor make a Website graphical at the expense of performance.

**4.2.1. Server log data.** According to Nielsen [17], human factors guidelines specify a 10 second maximum response time to mitigate the risk that users will lose interest in a Website. Nielsen further states that 15 seconds may be acceptable given that users are accustomed to long download times on the Internet. Bertot et al. [4] argue that stopped transmission data from the server Error Log can indicate that there is a pattern to users stopping the download of large files.

**4.2.1. Client-side log data.** Analyze the types of links that users are typically using i.e., are they text, graphic, a combination?

**4.2.3. Usability testing data.** Spool et al. [23] found no evidence that graphics helped users to retrieve information on a Website. They also found that "most users examined text links before considering image links" (p. 8). If there is a question of whether or not users prefer and are more successful with text versus graphic elements on a Website, usability testing could include tasks and questions pertaining to text and graphic elements. If it is found that users prefer and are more successful with graphics elements, the size of graphic elements must be considered if they are to be included on a Website.

### 4.3. Example 3: Do not use frames [17].

**4.3.1. Server log data**. Analyze data to determine if users abandon a frames-based homepage frequently. Since users can access a site from any number of locations within the site, path data should also be used to determine if the homepage is most frequently found as the beginning of a user path through a Website.

**4.3.2. Client-side log data.** Analyze path, link and element data to determine if a frames based approach produces longer or more indirect paths than a non-frames based approach.

**4.3.3. Usability testing data.** Observe users using a frames versus non-frames based approach to determine which approach is more appropriate and easier to use.

### 4.4. Example 4: Do not open new browser windows [17]

New browser windows could be opened to display online help and/or to display additional information pertaining to a Website.

**4.4.1. Server log data.** If the URL of the new browser window is different from the URL from which the new window was launched, the newly logged URL could be useful in determining if, for example, users select online help. However, if the URL of the new page is not different from the URL from which it was launched, this data could be misleading in a Web log because the data would appear as two accesses of the same Web page.

**4.4.2. Client-side log data.** The browser's "Back" command accounts for 41% of all user interaction requests for documents [7]. Nielsen [17] reasoned that since the Back button is "the second most used feature" on the Web, users would be able to navigate among information without the use of new browser windows. Nielsen did not offer empirical data to support this claim.

**4.4.3. Usability testing data.** Nielsen [17] found that users do not often notice when a new browser window is opened. This finding could be tested in a usability evaluation of a Website. Users could be asked questions about content contained in launched new browser windows to both determine whether or not they noticed new browser windows and to determine the perceived value of the new content if users noticed new browser window launches.

## 5. Experimental research and Website evaluation

The discussion so far has focused primarily on descriptive data. Interpretations of such data, alone, can only be used with extreme caution regarding users' actual responses to any particular element of a Web design. However, alternative approaches to the use of such data can unburden the need for triangulation. Collecting and analyzing data within pre-constructed experimental research designs overcomes most of the validity issues that plague the use of server logs and analysis programs for normative interpretations. This approach has received scarce attention in the literature, yet may be of great use in Website development or refinement.

### 5.1 Experiments through the Web

While experimental research using the Web to study a variety of non-Web topics is gaining popularity [e.g. 8, 21, 22], experimentation on Websites themselves (i.e. Website characteristics as independent variables) has been slow to follow. Web-based questionnaires using CGI forms are obtrusive, that is, subjects know they are being queried, which may skew accurate reporting, in some cases even moreso than traditional survey methods [28]. Yet the most available, unobtrusive data that describe the reactions of Web users to Web pages--behavior recorded by server logs--offer direct measures of behavior that can be compared to detect empirical differences between versions of Websites.

### 5.2 Experiments about the Web: Server logs as data in unobtrusive experimental designs

We propose that during protoype development and refinement, Web designers may create parallel Websites reflecting variations in design about which the designers want to know. For instance, if it is true that download speed, or intensive graphics, discourage users, yet if Web designers need a direct indication of whether such effects are affecting users' click-throughs or stops, parallel sites could be constructed which vary these attributes. The log data can be compared in order to test for significant differences between prototypes, rendering useful inferences about the direct effects of these design choices on specific user behaviors of interest.

One way to facilitate the efficient deployment of such an experiment is to use a javascript-based random redirect program on the home page of a site. A redirect script automatically sends the user to a different page. Thus, by including a redirect script on a home page, users can be sent to experimental pages transparently. By including a randomization routine in the redirect script, the home page can accomplish the random assignment of subjects to test conditions, which is the foundation of many experimental research designs. Users need not be given different URLs in order to be exposed to different prototypes and thus, users need not know that their reactions are solicited, preserving the unobtrusive nature of the research and accentuating data accuracy.

A further advantage that this approach offers over normative data interpretations, is that the various interpretive frailties to which normative interpretations are prone, do not apply in this context. For instance, it does not matter if 100 hits to experimental Site A come from 49 different users or 78 different users, because an equivalent hundred hits to experimental Site B has the same probability of profile distribution. Whatever frailties that occur in the measurement of one test site are assumed to occur in another; random error is cancelled out

in equivalent designs when random assignment of subjects is employed. One complication might be in assessing return visits, since a returning user may not actually be redirected to the same site as s/he saw previously. However, once again, the effect that this problem presents in assessing Site A is equal to that for Site B, and balances out. When tests for the effects of differences between Web pages are conducted, it matters less *who* is connecting to the site, rather than, among those who do connect, what navigations or other browsing behavior they perform.

Such approaches could be developed to employ inferential statistics to compare path lengths, duration, aborts, and almost any other feature of experimental Websites that may be of interest to designers. To do so requires that more than one prototype is employed, and that the various versions differ only on key, identified variables. For example, a *z*-test might determine, among Web design A, B, C, or D, which has the longest visitor duration. Or a *t*-test might reveal, among prototypes with and without consistent graphics on each page in the site, whether a longer path, with more steps through a site, is significantly more or less frequently taken due to graphics. Such experimental techniques, while previously employed in usability labs with small numbers of subjects and scored by human observers, can give way to large-scale research and development efforts that would be appropriate to test on large numbers of selected subjects (e.g. university students, corporate employees), or on the Web user population at large.

## 6. Conclusion

As we have argued above, analysis of Web design may be facilitated and improved using organic, spontaneous, unobtrusive, and inexpensive data that speak directly to how people actually utilize Websites, by analyzing server logs and client-side logs. While these data are notably imperfect, their weaknesses may be overcome in at least two distinct ways: by triangulating them with traditional usability testing, and/or by collecting them within the framework of experimental designs intended to test directly the differential effects of specific design options.

In most cases, no one source of user interaction information is as strong as when more than one source is combined. In some cases, however, only one method of collecting user interaction data is possible. Server administrators need to understand the meaning of server log analysis tool output (e.g., if the tool produces server "accesses" or server "hits") [4]. Server administrators also need to understand the extent to which server log output is useful in determining what improvements should be made to a Website to increase its value and ease of use. Web designers and usability professionals need to understand how to use the output of server log analysis tool output in their design and usability testing and analysis activities. In particular, designers and usability professionals need to understand what they can and cannot learn from server log data and further, how those data could be complemented by data from client-side analysis data, usability testing data, and other techniques. Website designers must evaluate whether or not to use one or more sources of use-based Web data based on available resources, maturity of a Website, the particular questions and issues that need to be addressed, and the existence of design options that may experimentally be tested.

Web designs that are based on user tasks, validated design principles, and use-based data collected over time will provide the most value to Website users and ultimately, to the owners of Websites. As noted by Nielsen [20], more research is needed in the area of Web usability. Use-based techniques of gathering data on Website designs will contribute to the understanding of how users use Websites and the analysis of use-based Web data will generate additional principles of Web design.

## 7. References

[1] AccessWatch, 1999. Retrieved May 15, 2000 from the World Wide Web: http://accesswatch.com/

[2] Accrue HitList, 1999. Retrieved May 15, 2000 from the World Wide Web: http://www.accrue.com/

[3] Analog, 1999. Retrieved May 15, 2000 from the World Wide Web: http://www.statslab.cam.ac.uk/~sret1/analog/

[4] Bertot, J. C., McClure, C. R., Moen, W. E. and Rubin, J., "Web usage statistics: measurement issues and analytical techniques," *Government Information Quarterly,* vol. 14, no. 4, pp. 373-395, 1997.

[5] Buchner, A. G., and Mulvenna, M. D., "Discovering Internet marketing intelligence through online analytical web usage mining," *SIGMOD Record,* vol. 27, no. 4 , pp. 54-61, 1998.

[6] Byrne, M. D., John, B. E., Wehrle, N.S., and Crow, D. C., "The tangled web we wove: A taskonomy of WWW use," *Proceedings of CHI99,* pp. 544-551, 1999.

[7] Catledge, L. D., and Pitkow, J. E., "Characterizing browsing strategies in the world-wide web," *Computer Networks and ISDN Systems,* vol. 27, pp. 1065-1073, 1995.

[8] Coomber, R., "Using the Internet for survey research," *Sociological Research Online,* vol. 2, no. 2, 1997. Retrieved May 9, 2000 from the World Wide Web: http://www.socresonline.org.uk/2/2/2.html

[9] Ellis, R. D., Jankowski, T. B., Jasper, J. E., and Tharuvai, B. S., "Listener: A tool for client-side investigation of hypermedia navigation behavior," *Behavior Research Methods, Instruments, & Computers,* vol. 30, no. 4, pp. 573-582, 1998.

[10] Etgen, M., and Cantor, J., "What does getting WET (web event-logging tool) mean for web usability?" *Proceedings of the 5th Conference on Human Factors and the Web,* 1999. [9]

[11] Goldberg, J., "Why web usage statistics are (worse than) meaningless," 1999. Retrieved May 15, 2000 from the World Wide Web: http://www.cranfield.ac.uk/docs/stats/
Retrieved May 15, 2000 from the World Wide Web: http://www.itl.nist.gov/iaui/vvrg/hfweb/proceedings/etgen-cantor/index.html
[12] HTTP-ANALYZE, 1999. Retrieved May 15, 2000 from the World Wide Web:
http://www.netstore.de/Supply/http-analyze/
[13] Kanerva, A., Keeker, K., Risden, K., Schuh, E., and Czerwinski, M., "Web usability research at microsoft corporation," In J. Ratner, E. Grosse, and C. Forsythe (Eds.), *Human Factors for World Wide Web Development.* Mahwah, NJ: Lawrence Erlbaum, 1997. Retrieved May 9, 2000 from the World Wide Web: http://www.research.microsoft.com/ui/papers/webchapter.htm
[14] Korgaonkar, P. K., and Wolin L. D., "A multivariate analysis of web usage," *Journal of Advertising Research*, vol. 39, no. 2, pp. 53-68, 1999.
[15] Larson, E., "Free money," *The New Yorker,* pp. 76-85, October 11 1999.
[16] Linder, D., "Interpreting WWW statistics," 1999. Retrieved May 15, 2000 from the World Wide Web:
http://gopher.nara.gov:70/0h/what/stats/webanal.html
[17] Nielsen, J., "Top ten mistakes in web design," May, 1996. Retrieved May 15, 2000 from the World Wide Web:
http://www.useit.com/alertbox/9605.html
[18] Nielsen, J., "The top ten new mistakes of web design," May, 1999. Retrieved May 15, 2000 from the World Wide Web: http://www.useit.com/alertbox/990530.html
[19] Nielsen, J., "Ten good deeds in web design," October, 1999. Retrieved May 15, 2000 from the World Wide Web: http://www.useit.com/alertbox/991003.html
[20] Nielsen, J., "When bad design elements become the standard," November, 1999. Retrieved May 15, 2000 from the World Wide Web:
http://www.useit.com/alertbox/991114.html
[21] Pettit, F. A.., "Exploring the use of the World Wide Web as a psychology data collection tool," *Computers in Human Behavior,* vol. 15, pp. 67-71, 1999.
[22] Reips, U-D., *"Why this?"* 1995. Retrieved May 9, 2000 from the World Wide Web:
http://www.psych.unizh.ch/genpsy/Ulf/Lab/WWWExpMethod.html
[23] Spool, J.M., Scanlon, T., Schroeder, W., Snyder, C. and DeAngelo, T., *Web site usability: a designer's guide.* North Andover, MA: User Interface Engineering, 1997.
[24] SurfReport, 1999. Retrieved May 15, 2000 from the World Wide Web: http://netrics.com/SurfReport/surfhome.html
[25] Tauscher, L., and Greenberg, S., "Revisitation patterns in world wide web navigation," *Proceedings of CHI97,* pp. 399-406, 1997.
[26] Vicomsoft, March 2000. Retrieved May 15, 2000 from the World Wide Web: http://www.vicomsoft.com/glossary/fixipaddress.html
[27] WebTrends, 1999. Retrieved May 15, 2000 from the World Wide Web: http://www.webtrends.com/default.htm
[28] Weisband, S., and Kiesler, S., "Self disclosure on computer forms: Meta-analysis and implications," *Proceedings of CHI 96*, pp. 3-10, 1996.
[29] Wusage, 1999. Retrieved May 15, 2000 from the World Wide Web: http://www.boutell.com/wusage/