Recommendations for Benchmarking Web Site Usage among Academic Libraries

Christy Hightower, Julie Sih, and Adam Tilghman

The Web sites that academic libraries are developing for their research communities represent an important new aspect of information management. Comparative statistical analysis of Web site usage among similar institutions would improve librarians' ability to evaluate the effectiveness of their efforts. A centralized voluntary reporting structure for Web server usage statistics, coordinated by the Association of Research Libraries' (ARL's) Office of Statistics, would provide a significant service to academic librarians. Factors to consider in designing such a benchmarking program are discussed, based on a pilot study of Web site usage statistics from fourteen science and technology libraries.

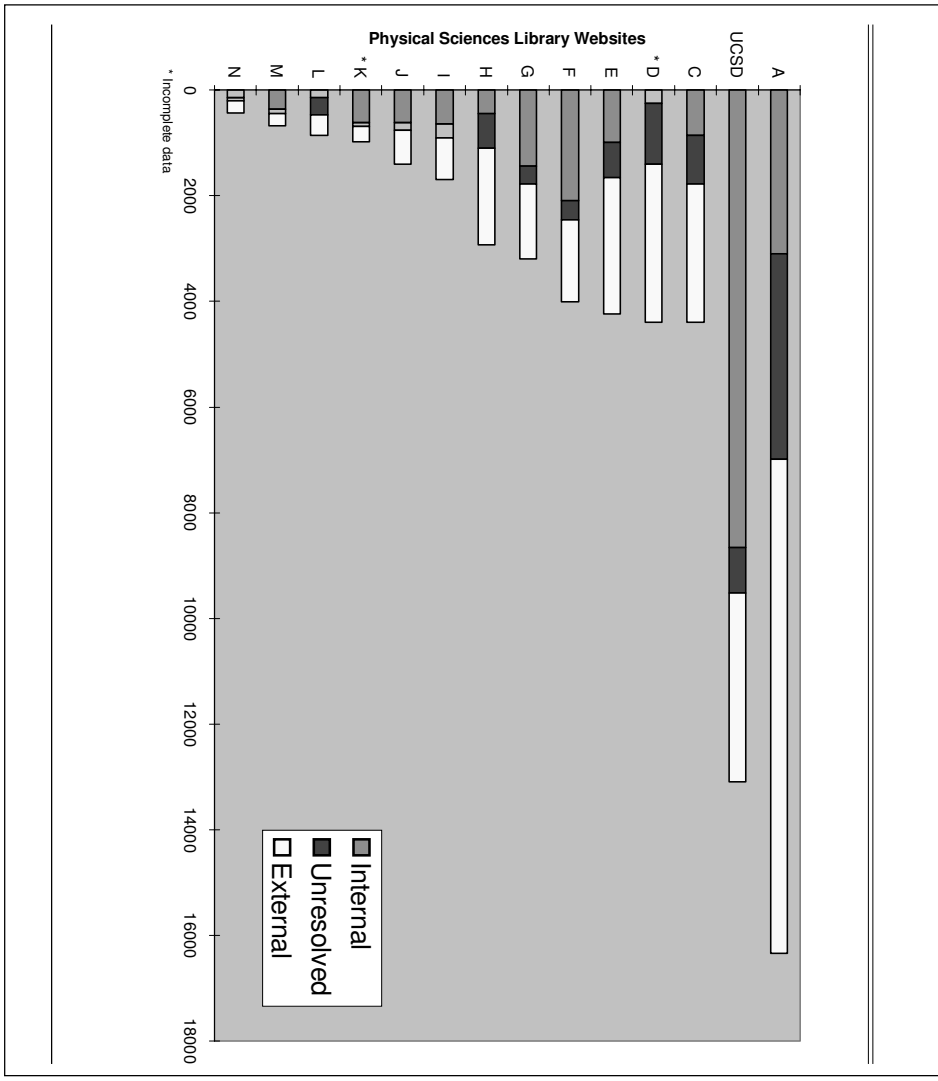
New measures of library activity attract the attention of practitioners and administrators alike because they promise answers to two eternal questions: (1) How effectively are librarians meeting the information needs of their primary clientele? and (2) Have their own approaches to budgetary and technological challenges been more or less successful than those of comparable institutions?

As we progress toward the largely digital library of the future, the active role of librarians in designing user interfaces and expert systems further whets their already voracious appetite for usage data. The foundation for many institutions' digital library efforts is their development of highly customized Web sites. This is fortuitous because Web servers automatically log data about the demand for specific resources within these sites.

One expects to find copious professional literature on how academic libraries can capitalize on readily available data on the size and characteristics of their own Web site audiences. After all, the popular press has seen fit to devote lengthy articles to the scores of tools and services available for Web server log analysis, as have trade journals in the fields of business and computing. Therefore, it is astonishing to discover how little the library and information science journals have published about the potential of these statistics as a measure of library activity.

This dearth of literature has left librarians ill-informed as to the capabilities of Web server log analysis software. Many

Christy Hightower is the Web Coordinator in the Science & Engineering Library at the University of California-San Diego; e-mail: chightow@ucsd.edu. Julie Sih is formerly the Corporate Programs Librarian at the University of California-San Diego; e-mail: jsih@gort.ucsd.edu. Adam Tilghman is the Web Programmer Analyst in the Science Libraries at the University of California-San Diego; e-mail: agt@ucsd.edu.



academic libraries undervalue, or even ignore, their Web traffic data. Even libraries that regularly analyze their server logs have difficulty interpreting the results in the absence of benchmarks. Without external comparisons, judging the success of a site is difficult because the size of the potential audience is unknown. Fourteen thousand page requests per month may sound impressive, but how does one know whether that indicates stellar or abysmal demand for a midsize university's sci/tech library Web site?

To help library directors and Web developers make sense of their own Web site traffic measurements, the authors of this article examine how a benchmarking program might be developed to compare the statistics of one academic library Web site against those of others. In so doing, the authors identify several practical and philosophical issues concerning intercampus comparisons of library Web site traffic. Because such Web sites are unique resources developed for specific audiences, meaningful benchmarking can oc-

TABLE 1
Page Requests by Site versus Selected Pages (Feb. 1996)

Physical Sciences Library Web Sites	Total Site		Top Ten Pages		Engineering Web Resources Pages[1]	
	Requests	Rank	Requests	Rank	Requests	Rank
A	16,339	1	15,951	1	4,956	1
UCSD	13,086	2	8,698	2	512	2
C	4,394	3.5	4,340	3	n/a	n/a
D[2]	4,390	3.5	2,712	5	218	6
E	4,241	5	2,408	6	88	9
F	4,001	6	3,684	4	294	4
G	3,189	7	1,886	7	222	5
H	2,927	8	1,514	9	155	8
I	1,694	9	1,694	8	n/a	n/a
J	1,395	10	1,131	10	157	7
K[2]	1,081	11	610	13	46	10
L	851	12	785	11	16	12
M	679	13	679	12	360	3
N	429	14	329	14	27	11

1 Includes only those pages with external links
2 Incomplete data available

cur only among carefully selected peers whose Web sites share essential characteristics. Moreover, equitable comparisons require uniform definition of measurement units, as well as establishment of a standardized approach to collection, analysis, and reporting.

Based on their experience analyzing the Web server log files of fourteen universities' science and engineering libraries, as well as their evaluation of standards proposed by two Internet advertising bodies, the authors propose voluntary guidelines and a common set of metrics for estimation of library Web site audiences. They also assert that the ARL might provide a valuable service to its member institutions by facilitating this standardization process and providing a mechanism whereby academic libraries can choose appropriate peers and models.

Methodology

With no relevant research models in library and information science journals, the authors' main sources of technical information were the Web site of the wwwstat 2.0 analysis software,[1] the FAQ (frequently asked questions) document for the comp.infosystems.www USENET newsgroups,[2] and anecdotal advice shared in forums such as the Web4Lib electronic discussion list.[3] The authors also consulted several business and computing articles whose reports of known problems with Web site audience estimation helped in planning their approach.[4]

Because such Web sites are unique resources developed for specific audiences, meaningful benchmarking can occur only among carefully selected peers whose Web sites share essential characteristics.

The authors began their study in November of 1995 by identifying possible peers for the Science & Engineering (S&E) Library at the University of California-San Diego (UCSD). They invited partici-

TABLE 2
Total Page Requests versus Edited[1] Byte Totals (Feb. 1996)

Physical Sciences Library Web Sites	Total Requests	Rank	Total Kilobytes	Rank
A	16,339	1	70,610	2
UCSD	13,086	2	74,346	1
C	4,394	3.5	60,932	3
D[2]	4,390	3.5	40,007	4
E	4,241	5	763	14
F	4,001	6	16,050	6
G	3,189	7	19,938	5
H	2,927	8	10,349	7
I	1,694	9	4,548	11
J	1,395	10	8,490	9
K[2]	1,081	11	4,380	12
L	851	12	2,038	13
M	679	13	8,724	8
N	429	14	5,201	10

1 Multimedia files excluded from byte count
2 Incomplete data available

pation from the sci/tech libraries of the following institutions: sister University of California campuses, the eight institutions used by the University of California for other benchmarking purposes (e.g., faculty salary comparisons), and some additional universities noted for their sci/tech programs. Of the twenty ARL and non-ARL institutions invited to participate, fourteen (including UCSD) participated fully by supplying their sci/tech libraries' raw Web server log files for the month of February 1996. The participating library institutions were: the University of California campuses at Berkeley, Davis, Irvine, Riverside, San Diego, Santa Barbara, and Santa Cruz; Cornell University; the Massachusetts Institute of Technology; Stanford University's Engineering Library; the State University of New York at Buffalo; the University of Illinois at Urbana-Champaign; the University of Michigan; and the University of Southern California.

It was not the authors' intent to assign relative performance outcomes to the participants based on the data collected. Rather, the study's aim was to gain valuable experience in the process of benchmarking in order to make informed recommendations on what data to measure and how best to collect and analyze those data for benchmarking in the library setting. The rankings presented in figure 1 and tables 1 and 2 are for purposes of comparing the metrics used, not the institutions. The authors felt that a small sample size was sufficient to accomplish these objectives and were not overly concerned when institutions with multiple Web servers were able to provide data for only a single server. Other aspects of the methodology used, which are identified below, also reflect the pilot nature of the authors' efforts.

Most participants made their raw February 1996 Web server access log files available by assigning them a uniform resource locator (URL) and allowing the authors to grab the data through their Web browsers, although two institutions sent their files via file transfer protocol (FTP). After obtaining the log files, the authors isolated the data for the specific Web pages they were interested in before running the files through analysis software (wwwstat 2.0).

Why Raw Log Files Were Requested

As they currently exist, Web server statistics are based on the data contained in a server's access log files. Each request for a document from a site is recorded as a line in that Web server's log (see figure 2). Most servers support the common log file format (CLF), which keeps very simple request information. It contains the visitor's host name (the machine

of today. Librarians' needs in terms of Web usage data, scarcely articulated even in our own professional literature, certainly have not caught the attention of analysis software programmers. Consequently, in whatever respects academic libraries' situations differ from those of companies who advertise on the Web, librarians must learn to adapt their measurement methodologies. This requires a technical understanding of not only the analysis tools themselves but also the motives of Web-based advertisers and the sites that cater to them.

One of the most interesting conclusions from the authors' study is that libraries cannot determine their Web peer group simply by comparing numerical usage statistics.

Business Week reports that ad-supported Web sites have used server log analysis to justify an average ad rate of $17 CPM (cost per thousand viewers) for 1997, whereas television's average CPM hovers between $5 and $6.[10] According to Matthew Kinsman, these Web sites' ability to document their growing audiences—and especially their ability to target specific niches within those audiences—allowed them to raise their ad rates more than 200 percent between the first quarters of 1996 and 1997.[11] Advertisers' enthusiasm for "the most measurable of all media by far"[12] explains why estimates for Internet ad expenditures range from $400 million[13] to $940 million[14] for 1997, and are expected to surpass $4.8 billion by the year 2000.[15]

Advertisers' well-publicized doubts about the reliability of Web server statistics have spawned scores of companies offering independent, third-party measurements and/or auditing.[16] These include ventures of such newsworthy companies as Nielsen Media Research; magazine trackers ABC (Audit Bureau of Circulation) and BPA (Business Publications Audit) International; and the so-called Big Three of financial accounting (Ernst & Young, Coopers & Lybrand, and Price Waterhouse).[17]

In 1997, the proliferation of analysis software and services, each of which had developed its own units of measurement, prompted two advertising trade associations to issue standards and guidelines for gauging Web audiences.[18] Both bodies attempt to define metrics and methodologies for cross-site comparisons. The first of these, the Coalition for Advertising-Supported Information and Entertainment (CASIE), is a joint project of the American Association of Advertising Agencies and the Association of National Advertisers. *The CASIE Guiding Principles of Interactive Media Audience Measurements*, released on April 3, 1997, are endorsed by the Advertising Research Foundation (ARF); indeed, they are based on ARF's long-standing principles for determining print, radio, and television audiences.[19]

On September 15, 1997, the Internet Advertising Bureau (IAB) released its own document, entitled *Metrics & Methodology*.[20] The thirty-eight-member Media Measurement Task Force that produced this document included representation from the owners of such immensely popular sites as Yahoo and Playboy, as well as from ad buyers such as Microsoft. Although both sets of guidelines are tailored to the advertising industry, librarians can use them as models for standards appropriate to the needs of their profession.

Recommendations for Developing a Benchmarking Program

A Web site benchmarking program for academic libraries could be designed and implemented either informally among a few institutions or formally among many libraries. Based on experience in collecting and comparing usage data from fourteen test institutions, this article proposes some voluntary guidelines and a common set of metrics for the estimation of library Web site audiences so as to make future cross-site comparisons a possibility.

Determination of Web Peers

Whether seeking to establish an informal benchmarking network of a few libraries or to identify which libraries within a formal reporting structure to benchmark against, identification of the library's Web peer group is an important first step toward obtaining and using Web server usage statistics profitably. One of the most interesting conclusions from the authors' study is that libraries cannot determine their Web peer group simply by comparing numerical usage statistics. There are important nonnumeric characteristics to consider as well, and these defining characteristics should be reported along with the numerical usage statistics in benchmarking programs. Finding a Web peer group is a three-step process:

Step 1. Identify the starting pool. The first step is to identify a starting pool of those institutions or individual branch libraries whose character or activities are of interest. The pool could contain the library's ARL peers, other institutions it usually benchmarks against because of similarities in student population or academic programs, institutions in the collection development consortia, or institutions with particularly noteworthy Web sites the library admires and wants to emulate.

Surprisingly, a library's ARL peer is not necessarily its Web peer. Admittedly, a larger sample might prove otherwise, but among the fourteen libraries in this study, an institution's overall ARL ranking bore no statistical relationship to the number of hits to its Web site or to the number of bytes transferred. In the study group of ARL institutions, Web site hits were not statistically correlated with circulation statistics, reference queries, number of full-time students, number of teaching faculty, or dollar amount of research grants received by the institution. In addition, age of the Web site showed no correlation to the number of hits: some young sites received more hits than older sites. Apparently, Web site character and quality are more influential in affecting usage than are the characteristics of the parent institution. By all means, ARL peers should be included in the starting pool because the library should benchmark itself against institutions it cares about; however, the starting pool should not be limited exclusively to the library's peers in the traditional sense.

Step 2. Narrow the list after site examination. The second step is to find Web peers from among those in the starting pool by examining each Web site. In the study, the authors were tempted to skip this step and merely rank their site against those with similar numbers of page requests because they had the luxury of having data from so many institutions in hand. When the authors did this, institution A (see table 1) appeared to be the only candidate to consider partnering with for future exchanges of benchmarking information. However, after examining the Web sites more carefully, it became clear that institution E was actually the closest peer because of its similarity in subject scope and design philosophy, site architecture, and target audience. These site characteristics are the important nonnumerical data elements that benchmarking programs should also collect. Figure 4 illustrates how the essential nonnumeric characteristics of each participant, as well as the numeric data, might be reported, whether the reporting form is paper or electronic. In locating one's Web peer, the ability to sort on these nonnumeric characteristics to narrow the starting pool becomes important.

Subject Scope and Philosophy. Obviously, a good subject match is important among peers. The information-seeking behavior of individuals, the availability of Web data sources, and the suitability of subject-specific data to presentation in a Web environment do differ somewhat by discipline. Equally important, however, is how the site is designed to present the various subjects to the visitor.

It bears repeating that in the study the authors attempted to ignore the structure

FIGURE 4
Sample Web Site Usage Data Report

Participant Profile

University Name: University of California, San Diego
Site Architecture: ■ traditional □ on the fly
Target Audience: ■ internal □ external □ both
Design Philosophy: □ single virtual library ■ virtual branches
Sample Period: Nov. 1997

Page Type:	Homepages		Reference Pages		Directional Pages		Unclassified		Totals	
Visitor Type:	Internal Requests	External Requests	Internal Requests	External Requests	Internal Requests	External Requests	Internal Requests	External Requests	Internal Requests	External Requests
□ Single Virtual Library Site										
■ Multiple Virtual Libraries										
InfoPath[1] library pages										
Art & Architecture										
Biomedical/Medical Center Libraries[2]										
Int'l Relations/Pacific Studies										
Science & Engineering										
Social Sciences & Humanities										
Scripps Institution of Oceanography										
Undergraduate Library										
Totals										

1. Umbrella library page
2. Two *physical* libraries; one single *virtual* library

of the Web sites sampled, with unsatisfactory results. Their attempts to compile statistics from multiple sci/tech library sites in order to construct at each institution a virtual counterpart to their own library's Web site (see figure 3) was time-consuming and awkward, and left lingering doubts about equality because the structure of any site influences the visitor's path (and thus page request counts) so greatly. Combining separate virtual branch library sites statistically after the fact is not the same as presenting a combined, unified site to users and then measuring the resulting traffic.

The authors contend that a visitor to a site designed as a single virtual library, containing links to all subject areas, has a different experience than one who comes to a site that presents itself as several virtual branches, each serving different disciplines. This virtual structure will affect hit counts in much the same manner as physical gate counts are affected by the existence of physical branch libraries. Single virtual "central" libraries and separate virtual branches are each valid design choices, but because the choice may affect page request counts, it is wise to select peers whose design is a match. In figure 4, each participant's design philosophy is categorized as being either a single virtual library site or a site with multiple virtual branch libraries.

In cases of differing numbers of virtual branch libraries, it may be useful to consider constructing a peer group of individual subject pages rather than whole Web sites. Thus, an engineering (or patents, or chemistry) subject page, for example, could be compared to other libraries' engineering (or patents, or chemistry) page rather than to their whole engineering library Web site. (See table 1 for engineering page request totals in the study group.)

Design factors other than the presence or absence of virtual branches also can affect page request counts. In this study, the authors considered using edited byte counts (site totals minus bytes for multimedia files) in relation to hit counts as a numeric indicator of the size and richness of libraries' HTML documents (see table 2). However, subjective assessments, based on visits to each Web site, were found to be more informative for peer choices. The presence or absence of link annotations, the number of useful links on each page, the presence or absence of a "home" button on every page, and the number of layers or "clicks" separating common starting and ending points also affect hit counts.[21] The number of additional design factors to consider in choosing a peer group is a judgment call.

Site Architecture. In addition to these design variables, even more fundamental differences can exist in a site's architecture or technical implementation. Traditionally, Web sites are collections of HTML documents containing links and text. These HTML pages are "handcrafted," constructed individually by subject specialists, and sit ready and waiting for a visitor to browse to find the links and information he or she needs. However, an alternative model for constructing Web sites has emerged and is gaining support as the best way to scale up or "automate" the construction of Web pages. In this new model, the site is actually a database of individual links with their associated annotations, and the HTML page the visitor sees is created mechanically on the fly from this database of links and customized to match the visitor's typed-in query. In the new model, the vocabulary used greatly affects the content of the page viewed. The two models could produce either very similar or potentially very different page structures, but at the very least the keyword search entry point probably would inspire vastly different browsing and query behavior from visitors. Usage statistics based on page requests cannot honestly be compared between sites that offer visitors such different information-seeking experiences. (In the study, only institution C used this database or a machine-generated, on-the-fly model for

page generation.) In figure 4, the site architecture is noted in the participant's profile.

Target Audience. In narrowing the list of potential peers, the library also should look for a match in the site's target audience. For benchmarking purposes in academic libraries, page requests for internal and external users should be reported separately (see figure 1). Some library sites are designed and marketed to be a resource to the entire world. (In the study group, institutions A and C were of this type.) These sites have unique digitized holdings, offer unique programs, or in some other way significantly add to the value of their sites with the goal of serving the needs of users beyond their home institutions. The sites are consciously marketed to external users. Because the UCSD S&E Library's target audience is the faculty, staff, and students on campus, institutions A and C would fall outside its peer group.

Some library sites are designed and marketed to be a resource to the entire world.

Categorizing Web sites in the reporting structure according to design philosophy (single virtual library versus multiple virtual branches), site architecture (traditional handcrafted versus machine-generated, on the fly), and target audience (internal, external, or both) will give Web server page request counts more meaningful context and aid in peer identification. These three defining characteristics could easily be identified in a participant profile section of the Web site usage data report (see figure 4).

Finally, in examining the sites of each potential peer, the library also should look for whatever it values and admires most in a library Web site (e.g., sites with no stale pages, sites that consistently exhibit proactive and innovative uses of technology to improve service to their users). Its peer group should contain some exemplary sites that are inspiring.

Step 3. Evaluate following a trial run. Once the library has selected a peer group from the initial pool, the third and final step is to do a trial run and evaluate the usefulness of the match. Either internal, external, or total page accesses should be examined, as appropriate for the library's own site's goals. If the page access counts for some sites in the library's group are substantially lower, it is debatable how useful continued comparison with those sites will be. The library may want to consider dropping them from its pool if the time and effort necessary to obtain their data is high. (Of course, because the Web is always changing, the low-traffic sites may bear reexamination in the near future.) If a noteworthy site similar in architecture and purpose has page access counts that are substantially higher, it might be kept as a model to aim for. Also, it may be useful to consider constructing a peer group of selected subject pages rather than whole institutions if that would help the library's development goals.

Most Useful Statistics for Benchmarking Purposes

Statistical analysis of log files results in a great deal of data that are fascinating to a Web site's creators but of negligible interest to their peer institutions. This article has already discussed the negligible value of cross-site comparisons of byte transmissions in the context of peer selection; the authors further note that both sets of advertising industry guidelines examined omit mention of bytes entirely. What librarians—and advertisers—really want to know is this: How many unique individuals are using the resources on a particular Web site, and how does that number measure up to other, similar sites? Average and actual incidence of repeat visits also would be an indicator of sites' ability to maintain an audience.

Advertisers are hot on the pursuit of this "unique visitor" data, which Kirsner

calls "the holy grail of site measurement."[22] Unfortunately, most of the existing methods of visitor identification (e.g., sending cookie files or tracers,[23] or consulting IP address tables) identify individual computers—a futile strategy in campus computer lab and shared library workstation environments. Demanding self-identification for each session via surveys or passwords is intrusive upon the user and programming-intensive for the library.

Therefore, Web audiences must be estimated based on the number of "hits" (requests for individual files) made upon the server. Selecting which of these hits to count is essential for objective comparisons. For example, in the case of a library Web page that makes liberal use of decorative graphics (e.g., library logo, backgrounds, bullets), a single visit registers several hits on the server—one for the HTML file and one for each of the nontext elements, which are treated as individual files. This phenomenon can inflate a site's aggregate hit count significantly. The authors recommend reporting "page requests" rather than "hits" as defined by the IAB. For nonframed pages, a *page request* is defined as "An opportunity for an HTML document to be displayed within a browser window, which may contain text, images, media objects (i.e., Java, Shockwave, Real Audio) or other online elements."[24] Thus, multimedia files are eliminated from the log before analysis. Table 1 reports page requests rather than hits under this definition.

Participants in benchmarking programs should preprogram their analysis software to ignore hits resulting from unsuccessful or rerouted requests. In addition, agreement should be reached on the treatment of hits from content-rich multimedia files and OPAC interfaces.

"Multimedia files" is a class that spans from 120-byte GIF images of bullets to 100Mb+ audiovideo extravaganzas. At the high end are data-rich files such as those presented by art, music, architecture, and map libraries. Because multimedia hit count comparisons would be worthwhile only between items of the same data type, these resources would be better benchmarked in a separate program designed for that purpose.

Although Web versions of OPACs do contribute to institutions' Web presence, the navigational patterns that characterize their usage are significantly different from other library Web pages. For that reason, Web OPAC usage data also should be benchmarked as part of a separate program, if at all.

Level of Detail at Which Page Requests Should Be Reported

Frequency. Analysis of the authors' own site's usage over time reveals that traffic varies greatly depending on whether an academic term is in session. To facilitate comparisons between quarter-system and semester institutions, institutions should report their statistics on a monthly basis.

Another argument for reporting Web server traffic on a monthly basis, rather than by academic term or year, is the great speed at which Web functionality develops. Libraries that seek to be dynamic reorganize their Web sites frequently as they create new resources and employ new functions. Monthly reports are more likely to distinguish the effects of these newly added (or newly deleted) files on overall usage patterns.

As noted previously, the CLF format contains a wealth of information, and analysis software can be configured to present this information at varying levels of detail. This gives participants in statistical exchanges a great deal of flexibility. However, some options provide more useful benchmarking indicators than others. In the study, the authors evaluated these various measures to determine the ideal format for statistical reports among peers.

Page requests for selected pages. The authors investigated whether evaluation of page request counts for the ten most

3, 1997): 24I–26I.

17. Damon Darlin, "Ratings Game," *Forbes* 158 (Dec. 2, 1996): 226; Cris Beam, "ABC and BPA Audit the Body Electric," *Folio: The Magazine for Magazine Management* 25 (Oct. 15, 1996): 30–32; Jane Hodges, "BPA Questions Int'l Audit Standards," *Advertising Age* 68 (Feb. 3, 1997): 56; Himelstein, Neuborne, and Eng, "Web Ads Start to Click," 134.

18. Consult http://www.yahoo.com/Computers_and_Internet/Software/Internet/World_Wide_Web/Servers/Log_Analysis_Tools/Titles for an extensive list of Web server log analysis software and services.

19. Coalition for Advertising-Supported Information and Entertainment, *CASIE Guiding Principles of Interactive Media Audience Measurements*, available at http://www.commercepark.com/AAAA/casie/gp/guiding_principles.html (Apr. 3, 1997).

20. Internet Advertising Bureau Media Measurement Task Force, *Metrics and Methodology*, available at http://www.iab.net/advertise/metricsource.html (Sept. 15, 1997).

21. Kirsner, "Web of Confusion," 36–39.

22. Ibid., 36.

23. Neil Randall, "The New Cookie Monster," *PC Magazine* 16 (Apr. 22, 1997): 211–14; Stephen H. Wildstrom, "Privacy and the 'Cookie' Monster," *Business Week* no. 3506 (Dec. 16, 1996): 22.

24. IAB, *Metrics and Methodology*.

25. Association of Research Libraries, *ARL Statistics and Information*, available at http://www.arl.org/stats/statistics/stat.html (Apr. 1997).

26. See note 6 above for reviews of such software and services.