

Cornell Hotel and Restaurant Administration Quarterly

<http://cqx.sagepub.com>

Website-generated Market-research Data: Tracing the Tracks Left Behind by Visitors

Jamie Murphy, Charles F. Hofacker and Michelle Bennett
Cornell Hotel and Restaurant Administration Quarterly 2001; 42; 82
DOI: 10.1177/0010880401421010

The online version of this article can be found at:
<http://cqx.sagepub.com>

Published by:



<http://www.sagepublications.com>

On behalf of:

[The Center for Hospitality Research of Cornell University](#)

Additional services and information for *Cornell Hotel and Restaurant Administration Quarterly* can be found at:

Email Alerts: <http://cqx.sagepub.com/cgi/alerts>

Subscriptions: <http://cqx.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Website-generated Market-research Data

Tracing the Tracks Left Behind by Visitors

by Jamie Murphy,
Charles F. Hofacker, and
Michelle Bennett

The computers that host websites automatically produce log files that record visitors' surfing and clicking behavior. An analysis of that data can reveal valuable marketing information.

While today's press is rife with internet news, the tenor is shifting from sanguine success stories to business owners' worst nightmares.¹ Stock markets slammed anything-dot-com stocks beginning in mid-April 2000, with many companies' share value dropping 40 or 50 percent in less than a month.² Similar to tulipmania, the South Sea bubble, and the 1841 Mississippi scheme—chronicled by Charles MacKay³—

¹ L. Kaufman, "After Taking a Beating, Dot-Coms Now Seek Financial Saviors," *New York Times on the Web*, April 18, 2000, as retrieved on February 18, 2001, from: www.nytimes.com/library/financial/041800market-ecommerce.html

² S. Lohr, "Stocks' Slide May Spark a Dot-Com Shakeout," *New York Times on the Web*, April 17, 2000, as retrieved on April 18, 2000, from: www.nytimes.com/library/tech/00/04/biztech/articles/17stox.html

³ C. MacKay, *Extraordinary Popular Delusions and the Madness of Crowds* (New York: Three Rivers Press, 1995).

investors woke up and asked, "Where's the profit?"⁴

So far most internet enterprises have generated little profit, according to a January 2000 survey of

⁴ B. Tedeschi, "Web Retailers Face Demands for Efficiency," *New York Times on the Web*, April 24, 2000, as retrieved on February 18, 2001, from: <http://www.nytimes.com/2000/04/24/technology/24commerce.html>

Jamie Murphy, Ph.D., is a visiting fellow in the department of information management and marketing, University of Western Australia (jmurphy@kroner.ecel.uwa.edu.au), where **Michelle Bennett**, B.Com., is an associate lecturer (mbennett@kroner.ecel.uwa.edu.au). **Charles F. Hofacker**, Ph.D., is a professor in the marketing department at Florida State University (chofack@garnet.acns.fsu.edu).

© 2001, Cornell University

Perth, Brisbane, Melbourne, and Sydney business owners.⁵ Market researcher Barry Urquhart found that only 2 percent of Australian companies on the internet profited from web-based transactions during 1999; that less than one half had a documented internet strategy; and only one in eight had a dedicated marketing budget for their website.⁶

The cloud over internet profitability has a silver lining, however, especially for small-budget entrepreneurs, and that is the marketing-research potential of the internet. Thanks to the web's digital delivery of marketing data, small and large businesses alike now can track users' browsing and purchasing behavior more effectively, more quickly, and less expensively than they could by using conventional media or syndicated research.⁷ Since entrepreneurs generally can react quicker than can large organizations, the internet seems to be custom made for nimble operators.

This paper opens with a brief discussion of traditional media research and its evolution to measuring consumer behavior through data mining and the worldwide web. An explanation of the web's client-server system and server log files follows. That technological information will be used to explain the promise—and limitations—of web-generated consumer data for marketing purposes.

Along the way we present a case study of a restaurateur's website to illustrate how one entrepreneur

successfully used the visitor information gleaned from his server log files.

Media Research

An advertising adage goes something like: "Half of my advertising is wasted, but I'm not sure which half." Marketers know they must advertise, but lament not being able to measure efficiently the effect of their advertising dollars. In general, marketers rely on imperfect measures such as surveys and behavioral measures—for example, scanner data, panel data, and observation—to guide their marketing strategies.

Surveys. Surveys were seen as superior in terms of time efficiency, cost, and effort relative to other data-collection methods. There is abundant research, however, that suggests that survey participants' reports of their behavior are subject to errors and biases.⁸ Technological advances have enabled and increased the efficiency of behavior measures via scanner data and, more recently, server log files. Those technological advances, combined with survey biases, have directed marketers increasingly toward actual-behavior measures. Some behavior is difficult to observe, however, and thus many syndicated research services combine survey and behavioral data.

Monitors. Survey instruments have evolved from questionnaires to monitoring devices installed on televisions and, now, on computers. AC Nielsen, known for media mea-

surement, today offers internet-use data through eRatings.com, a global panel that tracks 95,000 internet users' online behavior. Other firms include Media Metrix, with 55,000 panelists; and PC Data Online, which boasts 120,000 panelists.⁹

Simmons's Fall 1999 National Consumer Survey, for example, contains lifestyles, media use, and product-consumption habits of over 33,000 adult Americans.¹⁰ Subscribers to such services use the available data to target their advertising at selected audiences. Coca-Cola, for example, could search Roy Morgan's proprietary database, "Asteroid," for television shows watched by known Coca-Cola drinkers and then run advertisements on those shows.

Plastic. Transactional Data Solutions (TDS), a venture between Symmetrical Resources—parent company of Simmons—and MasterCard International, tracks 670,000 randomly selected MasterCard users' purchases. TDS clusters consumers by shopping behavior and then uses the Simmons data to further define clusters by media habits, brand preferences, and lifestyle.¹¹

Scanners. Similarly, AC Nielsen tracks scanner data in over 4,800 stores representing more than 800 retailers in 50 major markets. Led by grocery stores—with over 90 percent of all items scanned—"millions of products bleep their way across scanners in stores all over the country. Each bleep adds a new entry to the trillions of transactions already

⁵Sue Peacock, "E-commerce a Big Loser for Many: Survey," *The West Australian*, February 2, 2000, p. 18.

⁶B. Urquhart, "The Internet—Where's the Profit?," *Marketing Focus*, March 27, 2000, as retrieved on April 2, 2000, from: www.marketingfocus.net.au/pages/texts.htm

⁷For example, the services provided by such firms as Roy Morgan, in Australia (www.roymorgan.com.au); Simmons Market Research, in the United States (www.smr.com); and AC Nielsen internationally (acnielsen.com).

⁸See: E. Lee, M. Hu, and R. Toh, "Are Consumer Survey Results Distorted? Systematic Impact of Behavioral Frequency and Duration on Survey Response Errors," *Journal of Marketing Research*, Vol. XXXVII (February 2000), pp. 125–133; E. Blair and S. Burton, "Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions," *Journal of Consumer Research*, Vol. 14, 1987, pp. 280–288; and R. Nisbett and T. Wilson, "Telling More than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, Vol. 84, May 1977, pp. 231–259.

⁹Roy Morgan uses PC Data Online information for its Australian product, "Internet Monitor."

¹⁰"Simmons Announces Release of Fall 1999 National Consumer Survey Stand-Alone Version," press release, *Simmons Market Research*, April 13, 2000; as retrieved on February 18, 2001, from: www.smr.com/fall_stand.html

¹¹J. Lach, "It's in the (Credit) Cards," *American Demographics*, April 1999, as retrieved on February 19, 2001, from: www.americandemographics.com/publications/ad/99_ad/9904_ad/ad990409c.htm

Leading internet companies are increasing profitability by combining their knowledge of consumers' purchase behavior with data on website visitors' behavior.

stored in computer data banks in retail chains nationwide."¹²

A recent AC Nielsen study used scanner data to measure the impact that print campaigns have on product sales. Comparing two demographically matched groups of about 4,000 households, the data showed that "for nine of the ten packaged-good brands measured, households exposed to magazine ads were more likely to purchase the advertised product than those that had not seen the magazine ads."¹³

Mail and ads. The Direct Marketing Association (DMA), founded in 1917, suggests measuring behavioral responses to help target advertising expenditures. The DMA defines direct marketing as "...direct communication designed to generate a response in the form of an order (direct order), a request for further information (lead generation), or a visit to the place of business for purchase of a specific product(s) or service(s) (traffic generation)."¹⁴ Therefore, direct marketers test different advertisements and media to see which yields the most cost-effective results.

Direct marketing, according to a DMA-commissioned study, will generate \$1.7 trillion in U.S. sales in 2000 and \$2.7 trillion by 2005. Direct-response-advertising expenditures now account for more than half of all U.S. ad dollars, with \$176.5 billion spent on 1999 direct-response advertising. Direct marketers employ various media, including

catalogs, direct mail, teleservices (e.g., 30-minute infomercials; unsolicited phone calls to consumers), and, increasingly, the worldwide web. Direct marketers' web-driven sales were estimated at \$24.0 billion in 2000 and are projected to reach \$136 billion in five years.¹⁵

Direct-response concepts of "direct order," "lead generation," and "traffic generation" seem appropriate for the web's interactive environment.¹⁶ Furthermore, the web lends itself to empirical testing of direct-marketing effects as shown in other media.¹⁷

Data Mining, Electronic Loyalty, and Log Files

Some internet companies are increasing profitability by adding their knowledge of consumers' purchase behavior to data on website visitors' behavior. Analyzing this information—known as data mining—helps businesses plan effective electronic loyalty programs based on customer lifetime value, acquisition cost, and customer life-cycle economics.¹⁸ Bricks-and-mortar customers "leave

¹⁵ *Ibid.*

¹⁶ T.P. Novak and D.L. Hoffman, "New Metrics for New Media: Toward the Development of Web Measurement Standards," *World Wide Web Journal*, Vol. 2, No. 1 (Winter 1996), pp. 213–246, (as retrieved February 19, 2001, from: www2000.ogsm.vanderbilt.edu/novak/Web_standards/Webstand.html)

¹⁷ See: A.K. Basu, A. Basu, and R. Batra, "Modeling the Response Pattern to Direct Marketing Campaigns," *Journal of Marketing Research*, Vol. 32, May 1993, pp. 204–213; and G.E. Smith and P.D. Berger, "The Impact of Direct Marketing Appeals on Charitable Marketing Effectiveness," *Journal of the Academy of Marketing Sciences*, Vol. 24, Summer 1996, pp. 219–232.

¹⁸ See: F. Newell, *Loyalty.com: Customer Relationship Management in the New Era of Internet Marketing* (New York: McGraw Hill Professional Book Group, 2000); P. Carpenter, *eBrands: Building an Internet Business at Breakneck Speed* (Boston, MA: Harvard Business School Press, 2000); F.F. Reicheld and P. Schefter, "E-Loyalty: Your Secret Weapon on the Web," *Harvard Business Review*, July–August 2000, pp. 105–114; and Robert K. Griffith, "Data Warehousing: The Latest Strategic Weapon for the Lodging Industry," *Cornell Hotel and Restaurant Administration Quarterly*, Vol 39, No. 4 (August 1998), pp. 28–35.

¹² R.P. Heath, "Wake of the Flood," *American Demographics*, November–December 1996, as retrieved on February 19, 2001, from: www.americandemographics.com/publications/mt/96_mt/9611_mt/9611m59.htm

¹³ This study was sponsored by the Magazine Publishers of America; see: L. Calvacca, "Making a Case for the Glossies," *American Demographics*, July 1999, as retrieved on April 23, 2000, from: www.americandemographics.com/publications/ad/99_ad/9907_ad/ad990708c.htm

¹⁴ The Direct Marketing Association, as retrieved on February 18, 2001, from: www.the-dma.org/aboutdma/whatisthedma.shtml

no record of their behavior unless they buy something—and even then the data are often sketchy. But in virtual stores, their shopping and purchase patterns are transparent. Every move they make can be documented electronically, click by click.”¹⁹

Judith McGarry, vice president of strategic partnerships for drugstore.com, tests the success of customer-acquisition strategies with a direct-response model. “What gets us really jazzed here, or what gets us depressed,” she said, “is how we do with our customers. And they vote with their clicks.”²⁰ She measures those clicks through server-log files.

Log files. Log files help address questions about the behavior of the visitor, including:

- first page visited,
- last page visited,
- typical navigational sequences (i.e., movement among the website’s pages),
- referring site,
- average number of pages visited, and
- time on site.

Data-mining techniques can be applied to those and related questions to empirically investigate what visitors do on a site. In a 1995 article discussing worldwide-web research, communication researcher Sheizaf Rafaeli had this to say:

Not only does it [internet users’ behavior] occur on a computer, communication on the ‘net leaves tracks to an extent unmatched by that in any other context—the content is easily observable, recorded, and copied. Participant demography and behavior of consumption, choice, attention, reaction, learning, and so forth are widely captured and logged. Anyone who has an opportunity to watch logs of WWW

servers, and who is even a little bit of a social scientist, cannot help but marvel at the research opportunities these logs open.²¹

No cookies. Others, however, question the privacy concerns of analyzing server log files, especially when a website places a unique identifier—known as a cookie—on an individual computer’s hard drive.²² This paper discusses log-file analysis without the use of cookies.

The server that houses a website keeps a running record of web-browser requests. Each time a user clicks on a link or types in a web address (URL), the user’s browser requests a file from the server hosting the websites. The site’s server software records every request, be it a web page (.html or .htm file) or graphic (.jpg or .gif file) within that page.²³

At first glance, log files seem like gobbledygook. But free or commercial analysis tools transform these data into useful information.²⁴ Internet service providers (ISPs) often offer log-file analysis as well. Dissecting log files helps explain the available information, limitations, and possible uses of this information. Later on we introduce a case study of Norwood’s Seafood Restaurant. In the meantime, here are two entries from the server for Norwood’s.

²¹ J.E. Newhagen and S. Rafaeli, “Why Communication Researchers Should Study the Internet: A Dialogue,” *Journal of Computer-Mediated Communication*, Vol. 1, No. 4 (1995), as retrieved on April 25, 2000, from: cwis.usc.edu/dept/annenber/vol1/issue4/rafaeli.html

²² M.K. Reiter and A.D. Rubin, “Anonymous Web Transactions with Crowds,” *Communications of the ACM*, Vol. 42, No. 2 (February 1999), pp. 32–38.

²³ For further information on log files, see: R. Stout, *Web Site STATS: Tracking Hits and Analyzing Traffic* (Berkeley, CA: Osborne/McGraw Hill, 1997); and C. Hofacker, *Internet Marketing*, third edition (New York: John Wiley and Sons, 2001), chapter 18; and R. Zeff and B. Aronson, *Advertising on the Internet*, second edition (New York: John Wiley and Sons, 1999), chapter 4.

²⁴ For sources of free and commercial log-file analysis tools, see: dir.yahoo.com/Computers_and_Internet/Software/Internet/World_Wide_Web/Servers/Log_Analysis_Tools/

Example 1:

```
crawler3.anzwers.ozemail.net—
[22/Sep/1997:02:03:29 -0400]
“GET /~norwoods/wine/selections/
Heitz.html HTTP/1.0” 200 1623 “ ”
“Slurp/2.0 (slurp@inktomi.com;
http://www.inktomi.com/slurp.html)”
```

Example 2:

```
n4-102-238.thegrid.net—[22/Sep/
1997:02:43:22 -0400] “GET /
~norwoods/wine/selections/
wineselection.html HTTP/1.0” 200
3062 “http://av.yahoo.com/bin/
query?p=groth+winery&hc=0&hs=0”
“Mozilla/4.03 [en] (Win95; I)”
```

Domain or IP address. The first entry, ending at the “—” symbol, is the domain or “internet protocol” (IP) address of the visitor. In the two examples above, both visitors originated from a domain. They could have originated from an internet-protocol address as well.²⁵

A running joke is that “on the internet, no one knows if you’re a dog.” Indeed, as the two files noted above show, not all website visitors are human. The first visited from *crawler3.anzwers.ozemail.net*, and the second from *n4-102-238.thegrid.net*. As we explain in a moment, one of these “hits” was not by an individual.

Time of day. The next entry, in brackets, records when the server received the request. The first visitor requested a file at 22/Sep/1997: 02:03:29-0400, while the second request was at 22/Sep/1997: 02:43:22-0400. The last set of numbers, -0400, refers to the difference between the server’s time and Greenwich Mean Time (GMT).

An analysis of time of day tells when visitors arrive. This is useful for taking the site offline for maintenance. Moreover, using these data

²⁵ J. Murphy, “Putting Boompapa on the World Wide Web: Registering a Domain Name, Step by Step,” *New York Times on the Web*, April 12, 1997, as retrieved on February 18, 2001, from: www.nytimes.com/library/cyber/week/041297boompapa.html

¹⁹ Reicheld and Scheffer, p. 111.

²⁰ S.H. Verhovek, “Dot-Com Leaders Await a Shakeout of Losers,” *New York Times on the Web*, April 23, 2000, as retrieved on April 23, 2000, from: www.nytimes.com/library/tech/00/04/biztech/articles/23dotcom-future.html

We suggest that website managers look at their own sites with commonly available browsers. This will let them see what their potential customers see.

combined with the domain data, one can infer whether visitors arrive from work or from home. If the domain is `www.ford.com` and the time of day is Monday afternoon in the United States, it is probably a worker at Ford Motor Company.

Log-file analysis tools also combine time with the first field—IP address or domain name—to approximate how long a visitor visited. Log files tell when a visitor arrived and the last page they requested. But they fail to reveal when a visitor stopped requesting pages.

“GET” request. The next entry, inside the first set of quotation marks, is the visitor’s requested file and the protocol used. File formats vary. Again, two popular file formats are `.html` or `.htm` (“hypertext markup” file). Other common formats include `.gif` and `.jpg`, two popular file formats for graphics. In the examples above, the first visitor requested `“/~norwoods/wine/selections/Heitz.html”` and the second requested `“/~norwoods/wine/selections/wineselection.html”`

As our Norwood’s case study will illustrate, visitors do not always start at the home page. As such, this field shows what interests visitors. (In the second example above, for instance, the visitor was initially looking for websites that included the terms “groth” and “winery.”)

Transfer code and file size.

The next two entries, both numbers, represent the server’s action with the requested file and the file’s size. “200,” the most common and most desired action, represents a successfully sent file. An undesirable code, “404,” means “file not found.” (Website managers should review code-404 entries to fix bad links on the websites or to resolve bad links from referring websites.)

The server successfully sent two files, code 200. The second numbers, “1623” and “3062,” represent the number of bytes in the sent files.

An important clarification is that servers record what was sent, not what arrived. Site-centric measures (e.g., server log files) measure files served. User-centric measures, such as Roy Morgan’s Internet Monitor, measure what the visitor received.

Previous page. The next entry, inside the second set of quotation marks, represents the visitor’s previous—or referring—web page. In other words, the visitor followed a link from this page when it requested the current page.

The first case, with no entry between the quotation marks, denotes one of several scenarios. The user typed the requested page’s address, or URL, rather than following a link; the requested page was “bookmarked” on the user’s browser and the user clicked on the bookmark; this page was the default opening page on the user’s browser; or, as the next field—type of browser—will show, the user is not human.

In example 2, `“http://av.yahoo.com/bin/query?p=groth+winery&hc=0&hs=0,”` indicates that the web surfer was searching Yahoo’s database for a web page containing the key words “groth” and “winery.”

The referring field helps website owners to evaluate how users arrived, for example, via links or search engines. A key-word analysis, available with many log-analysis tools, such as Web Trends, tells what visitors seek on the site and provides insights for future modifications to the site.

Type of browser. The final entry signifies the type of browser that requested a specific file. The first is `“Slurp/2.0 (slurp@inktomi.com; http://www.inktomi.com/slurp.html)”`. Normally, this entry contains specifics on the browser (such as the platform). This case, however, includes an e-mail address and web address. As the name “Slurp” suggests, this browser digests

rather than browses. Visiting www.inktomi.com/slurp.html reveals that what's known as a spider indexed this page. In other words, a computer program, not a person, requested this page.

Robotic "spiders" crawl the web, indexing pages for search engines such as AltaVista and Google. These search-engine spiders follow links on a web page in an ever-increasing quest to add pages to their database. Along the way they generate "hits" on the pages that they visit.

The second log file (Example 2, on page X) represents the browser type "Mozilla/4.03 [en] (Win95; I)." More specifically, this is a Netscape 4.03-compatible browser, on a Windows 95 platform.

Studying browser types helps entrepreneurs plan what features to include—or not to include. If Java-based enhancements are a key element of the site, for example, then the majority of visiting browsers should support Java (a popular programming language).

We also suggest that website owners and managers look at their sites through the various browsers that are commonly available. This will let them see exactly what their potential customers see when they visit the site, depending on the browser used.

Limitations. A note on log-file-analysis limitations is necessary before moving into the Norwood's case study.²⁶ First, not all website visitors are human. Second, not all visits can be registered by log files.

To deliver web pages faster, ISPs and organizations, such as America

Online or the University of Western Australia, cache pages on their proxy server. Popular browsers, such as Netscape's Communicator and Microsoft's Internet Explorer, also cache pages. Storing pages locally, on a user's hard drive or on an organization's proxy server, yields faster download times for the cached pages should the user decide to return to those pages (for example, returning to a site's home page multiple times during a browsing session). This understates the number of pages that visitors requested, however.²⁷

Research applications of server log files are beginning to surface. A case study of the *South Pacific Journal of Psychology's* transition to being available in both an electronic format as well as its traditional printed version showed that server logs answered questions regarding popular content areas and countries where advertising had been successful. The data also suggested reasons for low-access rates in some targeted countries. Log-file analysis enabled the journal's editors to address readership questions, which would have been impossible using just the print version of the journal.²⁸

The Case of Norwood's

The building housing Norwood's Seafood Restaurant opened in 1929 as a Pan Am gas station and general store. Over the next few years the building housed a mosquito-control

center, a piggy-bank factory, another general store, and, finally, a restaurant. A family tradition for over half a century, Norwood's is a central-Florida institution known for fresh seafood and over 10,000 different wines.²⁹

In 1996 Norwood's owner—Don Simmons—worked with a team of Florida State University postgraduates to develop a website. He wanted to explore this new marketing medium, promote his restaurant, and possibly sell wine online. The team recommended, and Simmons chose, New South Network Services (www.nsns.com) as his ISP. A key factor in determining the team's choice was that New South Network Services' site-hosting package included log-file analysis.

Next, the web pages were designed and posted on the web. The first weekly report showed that Norwood's home page was the most popular among all the pages, with 125 requests, followed by the main wine page (34 requests). Visitors requested about 125 pages daily, following links from leading search engines such as AltaVista, Yahoo!, Lycos, Infoseek, Excite, Hotbot, and Northern Light Search.

As the referring domains demonstrated, the student team had indeed properly registered the site and search engines easily were finding it. This field also showed that the students had established valid links with other wine and restaurant sites.

Similar results continued over the next few weeks, except that the second-most-popular page became the red-wine list, not the main wine page. Further analysis showed that although the team had registered just the home page, search engines' robots automatically had indexed other Norwood's pages. For example, visitors queried AltaVista for

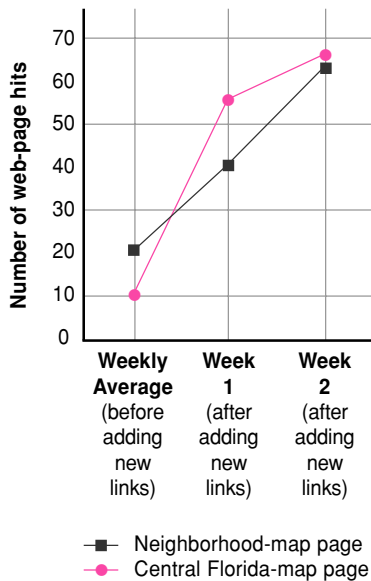
²⁶ See: J. Murphy and E. Forrest, "Hits, Views, Clicks, and Visits: Web Advertisers Face Data Jungle," *New York Times on the Web*, May 26, 1996, as retrieved on February 18, 2001, from: www.nytimes.com/library/cyber/week/0526measure.html; and P. Berthon, L. Pitt, and G. Pendergast, "Visits, Hits, Caching, and Counting on the World Wide Web: Old Wine in New Bottles?," *Internet Research: Electronic Networking Applications and Policy*, Vol. 7, No. 1 (1997), pp. 5–8.

²⁷ Two Australian companies, Red Sheriff (www.redsheriff.com) and Top100 (www.top100.com.au), have products that address the bias caused by caching. Red Sheriff circumvents caching by adding a proprietary code to websites' pages, while Top100 gathers information from large organizations' proxy servers to report on popular sites among these organizations. See: L. Weihart, "Measure for Measure," *Australian Industry Standard*, April 19, 2000, as retrieved on February 19, 2001, from: www.thestandard.com.au/articles/display/0,1449,9125,00.html

²⁸ P.A. Watters, M.F. Watters, and S.C. Carr, "Evaluating Internet Information Services in the Asia-Pacific Region," *Internet Research: Electronic Networking Applications and Policy*, Vol. 8, No. 3 (1998), pp. 266–271.

²⁹ "Norwood's Seafood Restaurant—A History," as retrieved on February 19, 2001, from: www.norwoods.com/welcome/Norwood_history.html

Exhibit 1
Monitoring Norwood's page hits



“wine and cabernet and Mondavi” and then clicked straight to Norwood’s red-wine list. The lesson learned through the log files is that visitors may come in through any door, not just the website’s front door (i.e., home page).

This knowledge became important a month later. The home page was still the most popular page, averaging under 200 weekly requests. But a map of Florida showing directions to Norwood’s, buried four levels deep in the site, leapt from under 20 weekly requests to become the second-most-popular page, with 58 requests. That seemed like an aberration. Someone speculated that perhaps an individual was using this page as their browser’s default opening page. For the next five weeks this page averaged over 150 weekly requests, while requests for the home page stayed about the same.

The log files held the answer to the mystery. We could see that those search-engine visitors had searched not for “Norwood’s” or “directions to Norwood’s”; they had searched for “map of Florida.” Since Norwood’s map page was titled “Map of Florida,” it had generated those unintentional hits. The Norwood’s team regarded the visitors as potential customers in any event, and updated the “Map of Florida” page by adding links to a central Florida map and neighborhood pages.

Thereafter requests for those newly linked pages vaulted. For example, until the update, the neighborhood page averaged under 20 requests per week, and the central Florida page averaged half that. With new links to those pages, central Florida had 54 requests the following week, while the neighborhood had 40. More of the same happened the next week, with central Florida at 66, and the neighborhood at 63 (see Exhibit 1). The map pages, however, dropped in popular-

ity as other websites with Florida maps came online.

Thanks to an informal marketing arrangement (that is, occasional free meals), Casey’s Florida Map page (www.luddist.com/map.html) links to Norwood’s Florida and central Florida map pages. To track Casey’s referrals, Norwood’s created Map2a.html, which replicated Norwood’s existing map of Florida, and Map1a.html, which replicated its central Florida map. Links to those duplicate map pages (Map2a and Map1a) existed exclusively on Casey’s web site, to make tracking those referrals easy.

Using log files, we observed that during the first week of July 2000 Norwood’s had 1,118 requests for Map2a compared to 62 requests for Map2 (the original Florida map) and 746 requests for Map1a versus 41 requests for Map1 (the original central Florida map). Given the page designs, we know that the 1a and 2a requests originated predominantly from visitors who first found the link on Casey’s website; typographical errors and cached pages in search engines generated the balance of the requests. (As a result of the increased traffic, Casey now gets all the complementary fresh seafood and homemade onion rings she can eat!)

Archives. Archiving old pages is another lesson that was learned by looking at log files. When the wine list was first updated, the old wine list was kept for historical purposes. Although the site linked to the new list, and did not link to the archived list, visitors—robotic and human—continued requesting the old list. Search engines kept both lists, old and new, in their database. Based on this information, the old list now explains to the visitors that prices have gone up and the current wine list is just a click away.

Log Files and Experimental Research

It would help to know what motivates or influences visitors to navi-

Exhibit 2**Testing Norwood's web-page design**

Visitors to Norwood's home page received one of four randomly generated pages. One condition (i.e., one of the four different page designs) was the original home page (see graphic, above), an image with six text-based links underneath. Two new conditions placed a smaller image

to the left or right of the links, and the fourth used a smaller image with the links below. The new designs' compact layouts required less scrolling for visitors to see all six links. As a result, the revised designs yielded more clicks than the original design.

gate a website. That is, what features and design elements prompt surfers to click, scroll, or otherwise stick with ("click through") a particular site? Hoffman and Novak were among the first researchers to address clicking behavior.³⁰ In general, they explained, clicking behavior

ranges from goal-directed (searching) to experiential (surfing).

Since then researchers have shown that ad-banner copy, the number of ad banners, and previous ad-banner exposure all affect clicking on those ads.³¹ Mechanical fea-

tures such as a page's design, image size, background, sound-file displays, celebrity endorsements, Java scripts (i.e., software code for certain actions), frames (i.e., web pages within web pages on the same screen), and animation also influence website

³⁰D.L. Hoffman and T.P. Novak, "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations," *Journal of Marketing*, Vol. 60, No. 3 (July 1996), pp. 50-68.

³¹See: C.F. Hofacker and J. Murphy, "World Wide Web Banner Ad Copy Testing," *European Journal of Marketing*, Vol. 32, No. 7/8 (1998), pp. 703-712; C.F. Hofacker and J. Murphy, "Clickable World Wide Web Banner Ads and Content Sites," *Journal of Interactive Marketing*, Vol. 14, No. 1 (Winter 2000), pp. 49-59; and P. Chatterjee, D.L. Hoffman, and T.P. Novak, "Modeling the Clickstream: Implications for Web-Based Advertising Efforts," *eLab Manuscript*, 1998, as retrieved on February 19, 2001, from: www2000.ogsm.vanderbilt.edu/papers/clickstream/clickstream.html

navigation.³² As a wired lifestyle and limited discretionary time (for shopping, for example) are better predictors of online purchasing than demographic profiles, web-page design takes on additional importance.³³ Finally, a user's previous online behavior may also predict future online behavior.³⁴

Back to Norwood's. After reviewing the literature and examining the log files, we wondered whether Norwood's website would attract more clicks from visitors if the site had a different home-page design.

Using a previously tested methodology, we set up an experiment.³⁵ Visitors to Norwood's home page received one of four randomly generated pages. One condition (i.e., one of the four different page designs) was the original home page, which comprised a photographic image of the restaurant with six text-based links underneath the

photo. Two new conditions placed a smaller image to the left or right of the links, and the fourth used a smaller image with the links below. The new designs' compact layouts required less scrolling for visitors to see all six links. We predicted, therefore, that the revised designs would yield more clicks than the original design.

This was a true field experiment, as Norwood's visitors had no idea that their clicking behavior was being monitored. After cleaning the log-file data to eliminate search-engine spiders and access by researchers (versus potential consumers), the experiment yielded over 600 valid visits during the period February to April 1997. As Exhibit 3 illustrates, all three of the new designs had a higher percentage of clicks—requested pages—than the original home page.

A frequency distribution of the data in Exhibit 3 revealed a Poisson pattern, so that Poisson regression was used to analyze the data.³⁶ Comparing the number of clicks on the original home page with those at the three alternative designs was significant, with $\chi^2(1) = 5.5222$, $p = 0.0188$. Comparing the original design with the best design—which Norwood's subsequently used—provided these results: $\chi^2(1) = 5.7262$, $p = 0.0167$. Were these findings to hold for the future, Norwood's would see an almost 50-percent increase in clicks on its home page, simply by changing the page's layout.

Marketing Implications

For entrepreneurs, a website's marketing goal influences the site design as well as which of the log-file statistics will prove useful. If the marketing goal is to communicate product information (e.g., benefits,

prices, add-ons), the site should convey this information using multiple and detailed pages. The most useful log-file data are total pages visited per visitor.

By comparison, if the marketing goal is to convey an image, elicit an emotion, or build a relationship, then the site should be fun and affective. Effectively done, this would create flow (that is, a pleasant and engaged state stimulated by the balance between a task's challenges and the person's ability³⁷), brand involvement, and a positive attitude change among visitors. An appropriate measure of "surfing" behavior in this case would be the time that visitors spent at the site.

If the marketing goal is to promote an established product, accelerate repurchases, or increase market share, the site could tie in to contests, offer coupons, or relate to retail and offline media campaigns. In this case, an appropriate measure of success might be hits on contest and sign-up pages or the use of online coupons.

If the marketing goal is online sales, the site should funnel visitors to the page where they type in their credit-card number. Here, a good measure would be "conversion efficiency" at each page in the chain, from the home page to the page where the customer keys in her credit-card number.

Finally, if the marketing goal is to sell ad space (as would be the case in a content or portal site), the site should maximize the number of pages the visitor is willing to visit. An appropriate measure, therefore, would be total pages viewed (thus influencing ad revenue).

³² See: X. Drèze, and F. Zufryden, "Testing Web Site Design and Promotional Content," *Journal of Advertising Research*, March–April 1997, pp. 77–91; P. Zhang, "The Effects of Animation on Information Seeking Performance on the World Wide Web: Securing Attention or Interfering with Primary Tasks?," *Journal of the Association for Information Systems*, Vol. 1, Article 1 (March 2000), an online-only article found at: jais.aisnet.org/articles/default.asp?vol=1&art=1 (password required); and N. Mandel and E.J. Johnson, "Constructing Preferences Online: Can Web Pages Change What You Want?," *Wharton Forum for Electronic Commerce*, February 1999, as retrieved on February 19, 2001, from: ecom.wharton.upenn.edu/public/Constructing_Preferences.PDF

³³ S. Bellman, G.H. Lohse, and E.J. Johnson, "Predictors of Online Consumer Behavior," *Communications of the ACM*, Vol. 42, No. 12 (December 1999), pp. 32–38, as retrieved on February 18, 2001, from: www.acm.org/pubs/articles/journals/cacm/1999-42-12/p32-bellman/p32-bellman.pdf

³⁴ See: W.W. Moe and P.S. Fader, "Capturing Evolving Visit Behavior in Clickstream Data," working paper, Wharton School (department of marketing), March 2000, as retrieved on February 19, 2001, from: fourps.wharton.upenn.edu/ideas/pdf/fader-evolvingvisits.pdf; and Chatterjee, Hoffman, and Novak.

³⁵ See: Hofacker and Murphy, 1998; and Hofacker and Murphy, 2000.

³⁶ Drèze and Zufryden.

³⁷ For further information on flow, see: Hoffman and Novak (1996); and Mihaly Csikszentmihalyi, *Finding Flow: The Psychology of Engagement with Everyday Life* (New York, Basic Books, 1997).

Limitations and Future Research

The value of the aggregate information that log-file data can provide should not be underestimated. Not only do log files provide an effective means of monitoring consumer behavior, they are also a relatively low-cost market-research tool. There are, of course, limitations.

The page-layout results described here apply to Norwood's and may not apply to other websites. Moreover, visitors may change their behavior over time.³⁸ Further, log files can't identify individual visitors or repeat visitors; instead, they identify computers connected to the internet. Perhaps most important for the hospitality industry, there's no evidence that more clicks mean more revenue.

Gurley extols conversion rate—"the number of visitors who come to a particular website within a particular period divided into the number of people who take action on that site (purchase, register, and so on)"—as the single best measure of a website's success. Conversion rates measure site aspects such as convenience (e.g., Amazon's one-click purchasing), consumer interface (e.g., ease of navigation), and performance (e.g., fast and error free).³⁹ Log files provide the means to calculate conversion rates on websites.

Internet technology and "surfing" is still new to most consumers. As such, their behavior is likely to change as they become more famil-

Exhibit 3
Page design and average-click percentages

Condition	N	Clicks (requests for pages)	Average Click-through percentage
Original layout (image at top)*	167	67	40%
Smaller image (image on left)	164	96	59%
Smaller image (image on right)	134	76	57%
Smaller image (image at top)	161	80	50%

*This is the version shown in Exhibit 2.

iar with it. This highlights the need to monitor activity on websites on an ongoing basis. To do so, cookie technology and visitor registration may be employed to partially overcome limitations of tracking repeat visits of the same computer and same person.⁴⁰

Our study explains how to measure individuals' actual internet behavior. By comparison, whether consumers' reports of their behavior is reliable is debatable (as previously mentioned).⁴¹ It remains for future researchers to combine our methodology with self-reported measures to test the consistency between them.

While our experiment with Norwood's found that a certain page layout was more effective than the other three page designs, it remains to be seen whether other hospitality websites observe similar results. Further research may reveal that a certain layout is more conducive to certain consumers, or that what works for one type of business is different than what works for another.⁴²

Generating answers to such questions may require combining the log-file methodology discussed here with the use of cookies or site registration to attribute on-site behavior to individuals. Similarly, laboratory experiments using a methodology like the one we developed for the Norwood's field experiment—complemented by qualitative studies that probe why people click—could identify and classify individual users.

Finally, in addition to testing websites' popularity and effectiveness, this methodology could be used to test the effectiveness of opt-in e-mail,⁴³ print, and televised advertising campaigns. Various campaigns would use a unique URL to track responses to the corresponding campaign. Marketers could track clicks on specific URLs and thereby deduce which offers, copy, or layout generated the greatest response. **CQ**

³⁸ Moe and Fader.

³⁹ J. Gurley, "The Most Powerful Internet Metric of All," *CNET News.Com*, February 21, 2000, as retrieved on February 18, 2001, from: news.cnet.com/news/0-1270-210-3287257-1.html

⁴⁰ D.L. Hoffman and T.P. Novak, "How to Acquire Customers on the Web," *Harvard Business Review*, Vol. 78, No. 3 (May-June 2000), pp. 179-188.

⁴¹ Lee, Hu, and Toh.

⁴² J. Murphy, "Surfers and Searchers: An Examination of Web-site Visitors' Clicking Behavior," *Cornell Hotel and Restaurant Administration Quarterly*, Vol. 40, No. 2 (April 1999), pp. 84-95.

⁴³ For further information on e-mail campaigns such as opt-in, permission, and request, see: S. Godin, *Permission Marketing: Turning Strangers into Friends and Friends into Customers* (New York: Simon and Schuster, 1999); and J. Nielsen, "Request Marketing," *Alertbox*, October 15, 2000, as retrieved on February 19, 2001, from: <http://www.useit.com/alertbox/20001015.html>