

## THE CHALLENGE OF INTERACTIVE MEDIA TO CONVENTIONAL PRINT MEASUREMENT MODELS

Scott C. McDonald, Ph.D. , Time Warner Inc.

---

### The Growth Of Print Media On The Web

In the short time since our last meetings in San Francisco, interactive media have seen explosive growth -- led by publishers establishing sites on the Internet's World Wide Web. This "network of networks" allows dynamic updating of content that is distributed through larger computer file servers and downloaded on request to computers around the world. Though slow transmission speeds can make the process of reading magazines over "the Web" cumbersome, higher data rates are enhancing transmission speeds just as millions more individuals gain access to the Web through the on-line services (like America On Line, CompuServe, Prodigy and now the Microsoft Network). Technical trials for cable modems promise a quantum leap in transmission speeds over the very near term.

Meanwhile, the number of Websites has expanded very rapidly. Only a year ago, the Internet's registry showed only 8000 registered sites; now there are more than 50,000. Though estimates of the number of individual users are notoriously unreliable, there seems little doubt that the number has been increasing at a rapid clip. At the Pathfinder website established by Time Inc. only last December, traffic has grown phenomenally. The first six weeks of operation averaged only 400,000 visits per week; however by 29th week, weekly visits were numbering over 4 million. As of July, only 8 months after Pathfinder started, weekly visits to the site numbered over 6 million.

At least in theory, publishers who offer their magazine pages on the World Wide Web are offering their subscribers an alternative, electronic way of getting access to their editorial content -- and they are offering non-subscribers equal access to those pages without having to endure the formality of paying the publisher for that content. For now, no one really knows the extent to which this largesse adds incremental readers per copy -- just as no one knows the extent to which it cannibalizes paid circulation. In the headlong rush to get onto the Internet, publishers have deferred such questions for the time being; however in theory, at least, such free electronic distribution could imperil the consumer revenue stream for magazines, thereby making them ever more dependent upon advertising support.

Also in theory, advertising in electronic versions of magazines should afford advertisers much clearer measurement of audience exposure to their advertisements. After all, they should be able to know exactly how many people see an ad, how much it costs per person to deliver it, and how customers respond: what additional information they request and, in some cases, how many sales result (at what cost per sale). By this reckoning, interactive media should eliminate guesswork and make advertising completely accountable.

### Problems in Measuring Audiences on the Web

But in practice, interactive media are currently less accountable than conventional print media. In the United States at least 10 companies have announced their intention or hope of measuring media audiences on the Web; however they first must confront a five very basic methodological challenges:

1. **Auditability.** In most scenarios, the measurement of activity on the Web involves capturing the "clickstream" that, at the atomic level, is the basic incremental indication of user activity. However this clickstream flows into the file servers of the very media that are being measured, thus setting up a fundamental conflict of interest: the publishers who stand to profit from selling advertising are also the exclusive source of numbers to justify those ad expenditures. Ultimately, if audience estimates are derived from server-side data, some form of audit will be needed to establish credibility for the numbers. However this methodological or procedural issue is simple compared to the others.
2. **Unduplicated Counting.** Currently the most common measure of usage is the "hit". Websites routinely report the number of hits for different parts of their site, be that pages of editorial content or advertising-oriented links. These "hit reports" reflect the number of times files have been requested from the host Web server. However this is a very gross number. One cannot differentiate between 100 hits by different people and 100 hits by the same person. Indeed, one cannot even be sure that a hit is a person. Technically, a hit is simply a request for information from another server, and the communication protocols of the Internet keep the identities of various servers anonymous. Unless a Website establishes a registration protocol to require individual

- identification, one cannot be certain who exactly has requested the information. Registration imposes a barrier to use and thus reduces net traffic; and thus far, users have shown significant resistance to efforts to require registration. Moreover, the communication protocols used on the Internet are designed to try to automatically circumvent such barriers -- to automatically find an alternative way around the barrier and into the same site. Even if a Website were to be successful in establishing a registration procedure, the identity of the file requester can be disguised through the use of alias "screennames" -- alternative ID names often used on the Net in accord with its culture of anonymity. Thus, a Website might know that a file request comes from the server at the University of Michigan, but little more. Obviously, these problems make it difficult to develop computer programs that automatically unduplicate counts of exposures.
3. **Caching.** Because users pay for their use by the minute, there are economic incentives to download files to the hard disk of a personal computer, then read them later. This is called "caching" -- saving the file for later review when one is "off-line". Obviously, this opens the opportunity for repeated exposures to the same magazine pages, or for sharing them with other people -- all beyond the electronic gaze of the network itself.
  4. **Probability Sampling.** At one level, the Internet (and its graphical World Wide Web) is a direct marketing mechanism. If one is selling products directly, this can be terrific since one can engage the user's attention, answer the user's questions, and close the sale all in the same session. However if one is an advertiser who does not sell product directly to consumers, the Internet's similarity to direct marketing media constitutes a problem. While one can put up messages and get counts of the number of responses (within the limitations already outlined), one never knows what universe those respondents are drawn from, nor what the probability is of finding them in the population at large. Thus any audience measurement drawn from server-side data is going to find it nearly impossible to use the powerful tools of probability statistics to project exposures among a population at large. While direct marketers might be able to calculate the "CPT" (cost per transaction) or "CPL" (cost per lead) for a given ad on the Web, the more conventional "CPM" (cost per thousand exposures) measure will prove to be more elusive.
  5. **Global Reach.** The World Wide Web is just what it claims to be -- world wide. Although it is probably fair to say that the United States has been the most aggressive growth market for personal computers, computer networks, and for publisher and advertiser activity on those networks, inquiries come from far and wide. In a recent month on Pathfinder, the Time Inc. website, there were 6,731,100 hits (or requests for files) in total; 3,102,952 of those hits -- 46% -- came from servers outside of the United States. This is a special case of the audience inference problem already cited. Not only do we not know the probability that any given "hit" can be found in the population universe of a given country, but we don't even know whether that hit comes from the right universe.

This brings me to a broader problem that ought to be of concern to those of us charged with measuring audiences for conventional print media. Currently, audience measurement for both electronic and print media take place on a national basis -- yet here we have the domestic edition of TIME magazine being distributed electronically to audiences abroad. That version contains neither the ads found in the domestic edition of TIME nor the ads found in the international editions. Are audience estimates for a given international edition advertiser being accidentally inflated by claimed readership that has its true origins in an ad-free exposure to the domestic edition of the same magazine posted on the World Wide Web? Even within the confines of the United States, we don't really know the extent to which this rapidly growing vehicle for media distribution is influencing claimed audiences for existing media. While the magazine editors might be happy with the wider diffusion of their copy, advertisers might rightly worry that a large fraction of the audience for the magazine never even had an opportunity to be exposed to their ad -- since the exposure occurred in the electrosphere where a different set of ads ran.

How can we meet the challenges posed by the explosive growth of these new media? We certainly have some experience with asking "source of copy" questions, so one obvious and immediate step would be to include the possibility of exposure to issues via the World Wide Web in our battery of "source of copy" questions and to make audience adjustments accordingly when reporting the results. However recall on these matters is notoriously unreliable, and it is unlikely that the numerous issues I have raised can be solved with the mere introduction of a supplemental question or two in our existing surveys. Rather, we need more fundamental work to develop reliable measurement of exposures to the new media. We need to understand clearly the extent to which exposures in the electrosphere influence audience estimates for conventional print media. And we need a better evaluation of the way in which the World Wide Web causes a blurring of media across national boundaries.

Moreover, we need to develop new models for how advertising on the Web works and find appropriate metaphors to relate them back to traditional advertising metrics like reach and frequency. The new media ARE different in that they can provide a better idea for advertisers about how the audience is interacting with your material. But how can this information be captured, distilled and integrated back into more conventional measures of advertising effectiveness. For example, if a user sees an advertiser's button on a Web page, is that equivalent to seeing a whole ad? Probably not, because it is only a button with a logo. Is it equivalent to seeing a direct mail piece in the mail? Perhaps. If the user clicks the icon and goes into the ad, is that equivalent to opening a piece of direct mail? Or is it equivalent to calling an 800-number to request more information? Or is it equivalent to calling the 800 number AND reading the literature that is sent in response? And how does any of this relate to ultimate sales and to brand perception? Since interactive media offer the possibility of greater depth of content, can we arrive at any common measure of the "depth of exposure" to an ad?

It has been only two and half years since we last met in San Francisco. At that time, none of us remotely imagined that we would shortly be confronting issues like these. Certainly, our world is changing rapidly in ways that are hard to predict. If we don't make it our business to study these phenomena aggressively and energetically, we may find our entire enterprise on the way to extinction by the time we convene again three years from now.

