EVOLVING MODELS OF ONLINE AUDIENCE MEASUREMENT: DEVELOPMENTS SINCE VANCOUVER

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

Background

At the Worldwide Readership Symposium in Berlin in 1995, I presented a paper that noted the rapid growth of the World Wide Web, especially among consumers in the United States, and speculated on the impact that exposures to magazines online might have upon print media audience estimates (McDonald, 1995). In 1997 at the Vancouver symposium, there were several more papers concerned, in one fashion or another, with the development of online media. Denman Maroney explained how log files from web servers were being used to profile web site activity, and he analyzed the considerable discrepancies that resulted from inconsistent definitions and tabulation protocols (Maroney, 1997). Nigel Jacklin and Peter Highland chronicled their experiences establishing a web site and extracting useful data from their own log files (Jacklin and Highland, 1997). Paul Donato presented evidence suggesting that readers read web pages in much the same way that they read magazine pages, but argued that print media and online media could probably not be measured with a common methdology (Donato, 1997). And I presented a paper that built upon my Berlin paper by making empirical estimates of the degree to which "source confusion" from Web-based magazine exposures contaminated audience estimates for regular magazines (McDonald, 1997). However with the partial exception of Denman Maroney's paper, there was little attention paid in Vancouver to the mechanisms by which web-based audiences were being measured. To some extent, this oversight was understandable. As of October of 1997, there was still considerable chaos in the field of online media audience measurement. Things have settled down a bit since then, though one could hardly claim that anything about the Internet – including its audience measurement methodologies – is yet stable. However the approaches have taken enough coherent shape to merit summary here.

Growth of the Internet

That the Internet has grown sharply is beyond doubt, though there is less than perfect consensus about its exact size and character. If we look at how many domain names have been registered, the growth is staggering. When we met in Berlin in 1995 there were 177,000 domains registered as .com, .net, or .org in the United States. When we met in Vancouver, that number had grown to 1.2 million. As we gather here in Florence in the fall of 1999, the number is just shy of 4 million (eStats, 1999). Of course not all of these domain name registrations represent actual web sites; many of these name registrations are the work of cybersquatters hoping to parlay their claim to a domain name into cash should a company want to make use of that name for an actual site. Certainly, many of the sites are corporate sites offering brochure-ware, or commerce sites selling products; only a minority are "media sites" offering news, information, or entertainment. Only a minority are ad supported, though there is little agreement regarding how many are in that number. The presence of large ad networks like DoubleClick (with 2300 client sites) and LinkExchange (with over 400,000 affiliated sites who make non-cash barter exchanges of ads) muddies the picture. The most conservative estimate was that provided by PricewaterhouseCoopers for the Internet Advertising Bureau; it claims that there were really only 1,263 entities competing for ad dollars on the Internet as of June of 1999, with the top 30 entities getting a vastly disproportionate share of the ad revenue. E-stats compiled a "consensus estimate" for 1999 that out of a total of 3,600,000 websites, 339,658 (or 9.4%) were active business websites, and 4,530 (or 0.13%) were ad-supported websites (eStats, 1999).

Universe estimates for the Internet also vary widely, though most of that variance can be explained by examining the details of how the estimates were derived. Since the Internet has been growing rapidly, one needs to know exactly when the survey was conducted on which the universe estimate was based. Was the survey asking about access or actual usage of the Internet? If it was asking about actual usage, what time frame is used – past week, past 30 days, past 6 months? Was it asking about actual Web usage, or could mere use of e-mail over the Internet constitute usage. Was it asking about home usage, work usage, school usage? Was it assuming that the Internet use was taking place over a PC, or could access via a telephone, a PDA or a TV also constitute Internet use? Did use of a proprietary online service like AOL qualify, or must the user have actually gone out onto the Internet itself? Did the survey cover only adults 18+, or did it also include teenagers or children? If the latter, were they self-reported or was their Web usage reported by their parents? Is the universe estimate bounded by one country's borders, or is it an amalgam of multiple countries?

For example, MRI's Spring 1999 survey estimated that 50.13 million adults 18+ used the Internet from home within the 30 days prior the interview, and 33.68 million used the Internet from work during that same interval. However their format for asking the question makes it difficult to get an unduplicated count of adults who used the Internet from either location; my own derived estimate from their numbers would put that at about 55.98 million. However this would also include people who only used e-mail and do not venture out onto the web, as well as those who go to online services like AOL but do not visit the web itself.

Perhaps a more fruitful comparison can be made between the two leading rival Web audience measurement companies, MediaMetrix and Nielsen/NetRatings. Using very comparable definitions (persons aged 2+, in the United States, actively used the Internet or an online service within the month of July 1999), they come to very similar size estimates. MediaMetrix puts it at 62.89 million; Nielsen/NetRatings is a hair lower at 62.75 million. For those interested in some of the technical issues in reconciling different universe definitions, I would refer you to the MediaMetrix paper published in October 1998 on the occasion of their acquisition and integration of the rival RelevantKnowledge service (MediaMetrix, 1998).

Three Approaches to Measuring Online Audiences

In the "real world" of online advertising, the debates about universe size almost seem to be academic. Even if we knew the exact size of the Internet at any given point in time or the exact number of visitors to a particular web site, advertisers still would not have a satisfactory count of the number of people who saw their ads. The fact that I might go to Yahoo.com does not imply that I have seen an ad that is posted only to the travel section of Yahoo. Large sites serve different ads to different part of the site, sometimes in fairly complex rotations. Some sites have targeting mechanisms that track user behavior on the site, make inferences about the user's interests and match ads accordingly. Thus, in the online world, there is a weaker correspondence between the audience for the media vehicle and the audience for any given advertisement. To a large extent, the ad unit is detachable from the media vehicle. While advertisers still want to know the size and characteristics of the audience for ad-supported web sites (just as they want to know the audience and demographics of a TV program, a cable channel, or a magazine), this information is less satisfying to the online advertiser. The online advertiser wants to know how many ads were sent, how many were received by users, who those users were, and whether they did anything as a result of the exposure (like click on the ad). Even though 27% of online advertising is for general consumer products that are not sold through the online channel (IAB, 1999) there still is a tendency to crave click-through and to regard it as the criterion of success for online advertising. Click-through is a "hard" measure, irrefutable, not subject to quibbling as to whether or not it really happened. It can be manipulated by the advertiser and ultimately can work against branding and marketing goals if used improperly, but it remains the preferred measure of a campaign's success for many. As such, online ad effectiveness is being analyzed using metrics more typical of direct marketing than of conventional media advertising. Indeed, despite efforts by ad-supported web sites to defend a CPM-based pricing scheme, the IAB reports that more than 50% of the contracts include some hybrid elements combining impression guarantees with a click-through performance criterion. Typically, the performance goal is a "kicker" that offers financial upside for the web site rather than a punitive requirement for "make goods".

Against this backdrop, three approaches have emerged for the measurement of online audiences:

- Site-Centric – in which the website server log entries are the immediate subject of measurement. Site-centric measurement is built up from raw logfiles that indicate what files were requested and sent from the website. In theory, site-centric measurement should provide a comprehensive census of all activity at any given site; however in practice there are many potential sources of over- or under-counting. What's more, site-centric measurement provides little or no information on the characteristics of the visitors to the site who are exposed to the ads.

- Ad-Centric – in which the ad server log entries are the immediate subject of measurement. Ad-centric measurement asks neither the user-centric question of who is the audience for a website nor the site-centric question of which specific pages were sent by the site to the user, but rather the question of how many ads were sent to users at any given site. As with site-centric measurement, it is vulnerable to various forms of over- or under-counting, and it provides scant information on the characteristics of the people who are the targets of the advertising

- User-Centric – in which the person using the online media is the immediate subject of measurement. User-centric measurement is built on a panel model, akin to that used for TV measurement. Members of the panel have their computer use metered. In this case, however, the meter is not a physical device like TV's People Meter, but rather a software application that monitors the activity on the PC. Results are used to project ratings for websites in much the same way that ratings for TV programs are derived from metered panels of TV viewers.

Each approach attempts to count or estimate the number of occasions when a person has the opportunity to see a unit of online media content, be it advertising or editorial. Each approach has its advantages and disadvantages, which I shall discuss now in greater detail.

Site-Centric Measurement

Site-centric measurement systems are presently the most commonly used among commercial websites. Site-centric measurement systems rely upon the site's server logs to count the total requests for pages or other content elements received and fulfilled by the site. Because they provide actual tabulations of those requests, they represent a census of server activity. Thus they are not vulnerable to sampling errors or statistical fluctuations that can pose problems for panel-based measurement systems. This feature also gives them the ability to provide far more fine-grained detail than could economically be possible for user-centric panels. It also means that site-centric systems count page requests regardless of whether they originated at workplaces, schools, or foreign countries; site-centric systems are equally indifferent to whether the page requests came from users on PC, PDA or TV-based platforms. This comprehensiveness represents probably the key advantage that site-centric systems have over the user-centric systems of today.

However site-centric systems also have some very conspicuous limitations. To begin with, the site logs on which they are based were not originally built with measurement in mind. As such, they are very messy and cumbersome to recode and distill into usable data. Different computers generate logs in different formats. As Denman Maroney demonstrated in his paper at the Vancouver Symposium, even small differences in definitional conventions can generate big discrepancies in the statistics reported for even a very simple web session. As more exotic elements have begun to populate the web – frames, daughter windows, java applets, interstitials, rich media elements – the difficulties of developing common counting conventions have only multiplied. Though this is not usually a problem for a Webmaster trying to understand traffic at his website, it makes site-centric data very difficult to use for cross-site comparisons or for cross-site media planning. Even if definitional differences could be eliminated, the conflict of interest issue would remain. Advertisers would still be justifiably reluctant to rely on ad sellers as the principal source of data on ad delivery.

Even though site-centric systems analyze all of the data entered into the site's log files, these raw data are themselves vulnerable to various distortions, omissions, blind spots and systematic biases. Most current systems can only count pages sent, not the pages displayed. Sometimes, the pages sent never get delivered because the user interrupts the page download – an occurrence that is more likely when the speed of transmission is too slow. Thus the server logs an event, but the end-user never experiences it. This is classic over-counting. On the other hand, often pages get displayed without any corresponding log entry occurring on the originating server. This situation arises because web pages are often cached on proxy servers (or in the individual user's browser) as a means of speeding up the delivery of that page. Studies of caching have indicated that it is the highly popular web pages that are the ones most likely to be undercounted -- in some instances, the undercount of particular pages exceeds 300%!

Keep in mind that the site-centric systems track computers, not people. A site log entry tells that a page was requested by a computer with a given IP address, but it does not know whether the user was Dad, Mom, or baby. Indeed, despite advances in recent years, it still has some difficulty knowing whether the page was requested by a non-human in the first place – the robots used to gather and assemble data for some personalized applications or the "spiders" used to index web pages for search engines. So not only do site-centric systems generally fail to provide demographic data on users, they have some trouble being certain that the user is a person at all. While many sites have implanted cookies (i.e. small pieces of identifying code) on the browsers of their users in order to recognize them upon their return, there still is some uncertainty as to whether the returning computer is being operated by same user as in the previous session. Moreover, many more advanced users refuse, disable or delete cookies. And because both browsers and computer hardware get updated with some frequency, cookies have a limited shelf life; thus they don't help site-centric systems develop a data framework for longitudinal analysis.

To sum up, site-centric systems give us loads of detail about site activity -- far more than can be obtained from other sources. Site-centric systems capture activity regardless of the platform that is supporting the user, regardless of the country of origin, regardless of whether the activity was initiated at home, work or school. However site-centric systems suffer from acute limitations. They are very vulnerable to forms of systematic over-and under-counting that can seriously distort their reports in ways for which, thus far, we have no means of detecting or correcting. They can tell you that pages were sent, but not that they were received. They can tell you that a computer requested the page, but they can't tell you who was operating the computer. They remain highly idiosyncratic in their operational definitions and in their counting systems. Many try to allay advertiser concerns about these problems by commissioning occasional audits; but the audits only tell us that the site is counting truthfully according to its own definitions and algorithms – not that those are valid, reliable, or comparable to other sites.

Some of these problems would be ameliorated by the steps proposed by FAST – adoption of a universal set of definitions and counting standards for web log data, comprehensive audit policies, and so forth. However some of the problems remain constitutive, a function of the raw data and the point at which the web activity itself is being observed.

Ad-Centric Measurement

Ad-centric measurement is a new development since Vancouver and one that has spread quickly. It has grown up from the practice, unique to the Web among all media, of having ads served not by the media, but by "remote" third parties. In some cases, these ads are simply being called from third-party ad servers hired by the advertiser (e.g. AdForce, MatchLogic). In other cases, the third-party server is actually selling the ads on behalf of an affiliated network of sites (e.g. DoubleClick, LinkExchange). In either instance, the net effect is to detach the responsibility for ad serving from the media. Ad-centric measurement then does not report on how many pages were delivered, but rather on how many ads units were delivered. By consolidating the reporting across multiple sites, ad-centric systems offer considerable convenience to advertisers. Rather than trying to sort out arcane definitional differences that might distort reports from different web sites, advertisers gain some simplification and uniformity from ad-centric measurement reports. (Even if there are definitional and counting differences among ad-serving networks, this is less of a problem because any given advertiser usually only uses one ad-serving network).

Ad-centric measurement offers many of the same benefits and limitations of site-centric systems. Their big advantage is that they count ad exposures consistently across web sites. And the log files from their servers represent a census of all ad units sent. Some ad-centric systems have technologies designed to defeat caching by returning a "ping" to the originating server confirming receipt of the ad unit sent regardless of whether it was called from the originating server, a proxy server, or local cache; however these technologies have not yet been verified and audited. Should they succeed, substantial progress can be made against this source of undercounting, to the general jubilation of web sites and ad servers everywhere. On the other hand, ad units are as vulnerable as web pages to being aborted before they load on a user's computer screen. A recent study by Ogilvy One and Thinking Media showed as many as 42% of the ads logged as sent by the originating server never arrived at the end-user's computer (Ogilvy One and Thinking Media, 1999). Thus an advertiser relying on ad-centric measurement reports could be overpaying by quite a lot for undelivered impressions.

Like site-centric systems, ad-centric systems usually cannot describe the characteristics of the audience receiving the ads, nor can they describe their media behavior over some period of time (as is done, say, in reach and frequency analysis). Like site-centric systems, the ad-centric systems face a conflict of interest problem since they are being asked to verify their performance on the job they were hired to do by the advertiser; thus, ad-centric systems also require audits.

As noted before, the Internet presently is the only medium where one finds the ad-centric approach to measuring opportunities to see advertising. However this is not likely to be the case in the future. The rapid developments in digital television could easily lead to the development of similar systems for TV. Digital cable boxes could include slots for remote or local ad serving of TV commercials. Personal video recorders, such as those on the market now from TIVO and Replay Networks, offer a similar opportunity to detach the ad unit from the program or content that surrounds it. For the time being, I don't see a similar scenario developing for magazines.

User-Centric Measurement

The user-centric approach to online measurement draws largely on the model established for television audience measurement. Large panels of consumers are recruited, presumably using the principles of random sampling. Software is installed on their computers that continuously monitors their behavior – what software is used, what web pages are loaded into the browser, what specific application has the screen's focus of attention at any given moment. To measure audiences for web sites, the software reads the URL – usually parsing it to at least three levels: domain, sub-domain, and section. For very large sites, the sample size will allow reasonable resolution to even finer levels of detail. However in most cases, it is only possible statistically to sustain reporting at the domain level. Thus, user-centric measurement does not provide much information of value for internal site analysis. Nor does it offer much for advertisers who are advertising only on specific parts of sites.

User-centric measurement has many advantages over the other approaches. It measures pages that are received by the end user, not merely sent by the originating server. It counts those pages regardless of whether they were sent by the originating server, by a proxy server, or by local cache in the user's browser. User-centric systems require that individual users identify themselves, thereby allowing the reports to characterize demographic features of a web site's audience. To the extent that user-centric systems rely upon true random sampling techniques, they are projectable to such universes as "home users", "work users", etc. (The adequacy of these samples is one of the key points of contention about user-centric systems, as will be discussed below). Because user-centric measurement is conducted by independent, objective third-parties, there is no inherent conflict of interest.

However also has significant shortfalls. As already noted, due to the finite resolution of any sample-based technique, it has difficulty providing detail about activity below the domain level of web sites. This makes the measurement of exposure to specific ads difficult at best. Because of the fragmentation of the World Wide Web, many meaningful sites cannot be measured at all.

By far, the biggest problem with the user-centric approach centers around the problem of developing and maintaining an adequate sample. For previous media, samples were always drawn from a universe of households or individuals in a single country. For magazines, respondents were asked to recollect their prior reading. For radio, they were asked to keep diaries. For TV, some were asked to keep diaries and others had their TV sets metered. As TVs have migrated into ever more diverse locations, media companies, agencies, and advertisers have become increasingly uncomfortable about the exposures that being missed in airports, workplaces, dormitories, vacation homes, sports bars, and other venues; but the household-based panel model remains intact for the time being. However the Internet began, to a large extent, through work and school-based activity. Even though virtually all surveys tell us that now home use of online media surpasses work-based use, there still is a lot of online activity taking place in locations out of the home. Here we encounter new terrain in sampling. We don't have any census of PC's in the workplace, much less PC's with Internet access. Many PC stations at schools or libraries are truly "public", with many, many users. How can these be randomly sampled when, by definition, this presumes that we can calculate the probability of selection into the sample and use that probability as the basis for audience projection?

Of the two leading user-centric measurement systems in the U.S., both have wrestled with this problem. MediaMetrix has built up a workplace sample by asking panelists to install the software on their workplace computers as well as their home computers. They tell us that they have managed to get surprisingly good compliance, but they don't reveal exactly how many panelists are complying at work, nor do they specify how they handle situations where some household members install at work while others don't. They use MRI's survey as an enumeration survey to weight their work and home data, but this still begs the question of the validity of their work-based sample. Despite these issues, MediaMetrix has not equivocated, nor has it been paralyzed by Hamlet-like angst about how to proceed. Rather, it has responded to the industry's vociferous demand for a workplace sample (ANY workplace sample), and the MediaMetrix monthly reports contain estimates of home, work, and unduplicated reach for web sites. Nielsen/NetRatings has been more cautious to date, partly because its service only came to market earlier this year (three years behind MediaMetrix), and partly Nielsen regards itself as a standard bearer for orthodox sampling methodology. Nielsen/NetRatings is testing approaches, floating ideas, and deciding what to do. So in the meantime, its service only reports on home usage, putting it at a competitive disadvantage in the marketplace.

Another contentious area surrounding user-centric systems is the use of "roll ups" in the reporting. Each service has struggled to find consistent rules for reporting multi-domain aggregations (e.g. all Time Warner sites, all Turner sites, all CNN sites), but the complexity and flexibility of the Internet make it nearly impossible to impose these rules. Some sites sell their ads across a wide array of domains. For example, each of the movie studios creates individual sites for each of their movies – but sell ads across them in very flexible ways. Thus Warner Online wants to get "credit" for all 400 of its active web sites, though the pattern of ad serving across those sites is tailored to any individual advertiser's preferences. Uniform roll up rules can hinder the utility of the data provided by the user-centric systems. Analysts might want to compare the Wall Street Journal's site to Yahoo! Finance, but WSJ.com is a top-level domain while Yahoo Finance is not. Though the user-centric services do us a service by trying to create barriers to sites making false claims with user-centric data, their inflexibility in the roll up rules can ultimately do a disservice to their clients.

Implications for Magazine Measurement

For the moment, the battles about how best to measure opportunities to see ads on online media follow their own course and don't have much direct bearing on print media measurement. Though interactive TV systems of the future may well have their own version of site-centric and ad-centric systems to vie with the incumbent user-centric systems, no such scenario seems likely for print media. In the background, the broader problems of media fragmentation, declining response rates to surveys, increased respondent burden, and the problems of getting truly representative samples of heterogeneous populations certainly will put stress on audience measurement systems for all media.

There are perhaps two ways in which the debate on the Internet side of the house might spill over into these print media precincts.

Some of the more thoughtful participants in the online measurement debates have argued that our ability to develop credible random samples has now fallen into such disrepair as to render false the fundamental Gaussian assumptions on which our entire enterprise is premised. Though they have not presented the industry yet with a convincing Bayesian solution to the online measurement problems, there are efforts in that direction. Indeed, the "Reconciliation Study" sponsored by FAST and the IAB and presented at last month's ARF workshops, is one noteworthy attempt to develop an empirical basis for adjusting audience estimates based on prior assumptions about inherent biases in both the samples and in the point of observation. Should such a solution be embraced by the online media, it could spur interest in fusion techniques that, to date, have failed to raise much support in the United States.

A second possible spillover from the online world to the print world could come from the interactive nature of the medium itself. Virtually everyone involved in the development of online media measurement agrees that our assessment of the medium should not stop at mere assessment of OTS, at the raw tonnage of CPM. (Some partisans argue, against wisdom I think, that online media measurement should not be concerned with OTS at all, that it promote units of analysis that have no relationship to the other media). My discussion here of the three main methods for estimating OTS has focussed primarily on their strengths and weaknesses in performing that fundamental task. However, as I've already noted, advertisers still are happiest when they have click-through too, even when click-through for the curious 1% should matter less than the brand impact upon the non-clicking 99% who also saw the ads. As online media measurement moves beyond OTS to create better measures of responsiveness and interactivity, it could whet the appetite of advertisers for proof of impact from other media. TV soon will be able to deliver some form of that feedback to advertisers. What will print be offering to show that it, too, has impact?

References

Donato, Paul. (1997). "Website and the Readership Reflex – A Page by Any Other Name". *Worldwide Readership Research Symposium VIII, Vancouver. Session Papers.*

eStats. (1999) *The eAdvertising Report.* Published by eMarketer. Volume 1. April 1999.

IAB (1999). "First Quarter Advertising Report". Prepared by PricewaterhouseCoopers for the Internet Advertising Bureau. August 1999.

Jacklin, Nigel; Highland, Peter. (1997) "Not Drowning but Waving: Researching the Internet". *Worldwide Readership Research Symposium VIII, Vancouver. Session Papers.*

McDonald, Scott. (1995) "The Challenge of Interactive Media to Conventional Print Measurement Models. *Worldwide Readership Research Symposium VII, Berlin. Session Papers*.

McDonald, Scott (1997) "Are Web-Based Exposures Contaminating Print Media Audience Estimates?". *Worldwide Readership Research Symposium VIII, Vancouver. Session Papers*Maroney, Denman. (1997) "Measuring the World Wide Web". *Worldwide Readership Research Symposium VIII, Vancouver. Session Papers.*

Media Metrix. (1998) "A Comparison of World Wide Web Audience Estimates Utilizing Two Different Approaches". Methodology paper to clients, October 13, 1998.

Ogilvy One and Thinking Media. (1999) "Are Your Online Ads Getting Through?". *The Advertiser.* August-September 1999.