

The internet as a template for today's audience measurement

The internet is very measurable – but in almost too many ways, which aren't always compatible, explains **Josh Chasin**, comScore



THE INTERNET, WE have been promised from the very beginning, is the 'most measurable medium'. It therefore seems paradoxical that, almost 15 years on, so little consensus exists on the divergent measures available.

However, this is really not paradoxical at all. One consequence of being the most measurable medium is that the internet ends up as the medium with the most measures. We have an embarrassment of riches - panel-based-measurement, ISP data, site-centric data, third-party ad server data - all deployed in different configurations to provide different (sometimes overlapping) measurement solutions.

In the media business, where so many daily decisions depend on empirical quantification of the efficacy of the alternatives, multiple data sources can create chaos - especially when different sets of data point to different decisions,

But it may turn out that the sense of order instilled by a single set of numbers is an illusory artefact of the analogue age. In the digital age, the age of 'Audience Measurement 2.0', success may depend

on the extent to which we can derive order from an increasingly fragmented landscape.

Digital changes everything

For media audience measurement, digital technology introduces two profound changes.

The first is fragmentation.

Back in a simpler time (the 1990s), the advent of fibre-optic delivery of cable television promised ten times as many channels as before; in the States we called this the '500-channel environment'. How, media researchers asked, would we ever manage to measure and report on the TV audience when each household could receive 500 channels? Of course the answer was that we couldn't - because traditional audience measurement techniques proved inadequate for measuring the long tail implied by 500 channels.

The very nature of digital media profoundly changed the media landscape. A media operator no longer need possess a broadcast-spectrum licence, or own a printing press, to reach an audience. In

March 2008, the comScore Media Metrix service reported on fully 16000 web entities in the UK. Compare this to the number of vehicles reported in traditional media (see Figure i).

The implication for media researchers of this fragmentation is that traditional audience measurement techniques, on which we've relied for 80 years, simply don't scale. The smaller average audiences for digital media vehicles require dramatically larger sample sizes in order to report them with sufficient robustness. Thus the two most tune-honoured techniques for respondent acquisition - in-home and random digit dial - become inadequate for recruitment and maintenance of digital-age sample sizes; comScore recognised this problem - and its solution - early on, and uses online recruitment around the world (i).

The second profound measurement challenge is perhaps best characterised as the dichotomy between machines and people. This challenge bears careful consideration, because disentangling the measurement of people and of machines is central to understanding the

differences between the various internet audience metrics.

Simply put, in the digital age, it is easier than ever to know the behaviour of machines. Hence the 'most measurable medium' label. Digital technology enables us to know much of what happens online. We know, for example, how many times an ad is served; we know how many clicks this generated. We can track server calls and pages served. We can even know how many subsequent online sales accrued as a direct result of those clicks - at least as long as they occur on the same user session.

If a server has done it, chances are it has been tracked, counted and logged.

In the whole digital pipeline, there is only one thing we have trouble understanding: the last two feet of the funnel. These are the two feet that separate the user from the screen. Online metrics practitioners can track everything that goes on 'in here', but we find ourselves huddled against the screen from the inside, rapping on the glass, wondering 'Hey, what's that fellow out there doing?'

What he's doing is multi-tasking, engaging with content, seeing ads, making purchases. The medium is inherently measurable; but the people consuming the medium remain elusive - as elusive as newspaper readers, radio listeners or TV viewers. To truly understand the internet as an advertising medium, it is not enough to understand the behaviour of machines; we still need person-centric measurement systems to tell us who is engaged with content; how many there are; how much they consume; what else do they do; how often they return; how they interact with advertising; and, ultimately, what they buy. In media we call these things demographics, cumulative audience, duration/time spent; reach and frequency; ad exposure and advertising response.

Server analytics is not audience measurement

So, person-centric measurement is necessary to generate the data that buyers and sellers of advertising require. Indeed, the UK's Joint Industry Committee on Internet Measurement Systems (JICIMS) stipulated, in its recent call for proposals:

'JICIMS intends that this planning system will be based on a user-centric approach to measurement. Such an approach is seen as contrasting with a site-centric, census-based approach.'

This begs the question: how - if at all - should person-centric and site-centric data co-exist?

The internet metrics space comprise two different but overlapping disciplines: site analytics (*oka* web analytics) and audience measurement. Site analytics is a business science that emerged in the mid-1990s as a result of digital technology allowing site operators to accumulate and analyse server data. Third-party services like Omniture, WebTrends, Hitbox and Google Analytics provide systems that allow web operators to capture and manipulate server data, to generate detailed reports and key performance indicators (KPIs). Because the data encompass all activity accruing to a site, site-centric data are sometimes called 'census' data.

Site-analytic data are remarkably useful for website optimisation. Indeed, the Web Analytics Association defines site-analytic data as 'the measurement, collection, analysis and reporting of internet data for the purposes of understanding and optimizing Web usage' (2). Site analytics works much like measuring direct marketing or CRM; changes in website design are stimuli that elicit changes in response (measured by clicks, page views, transactions), and web operators use the

data to track how varied website stimuli can elicit optimal response.

It's easy to see how it would be useful within site analytics to understand the behaviour of people, as opposed to machines. Suppose a site redesign elicits an increase of 3.5% in monthly visits - a statistic wholly knowable from site-centric data. The operator might want to know whether these incremental visits are coming from new visitors (increased reach), or from more frequent visits among existing visitors (increased frequency).

Cookies

To track user-level visitation, site analytics uses cookies - small pieces of code dropped onto the computers of site visitors in order to identify them upon return. When you visit a website for the first time, you get a cookie; when you visit again, the site knows you are a return visitor because it finds your cookie.

Cookie-based tracking may be sufficient for the needs of site analytics. But it is grossly insufficient for audience measurement. It is largely this issue - the use of cookies as a surrogate for individuals - that differentiates site-analytic from audience measurement data, which are engineered to track individual persons.

Some challenges for cookies as surrogates for individuals are as follows:

^ Multiple users on one machine: if three people use Internet Explorer to visit a given website from the same machine, that website has no way to parse out the individual users (required registration is a partial remedy).

^ Multiple browsers on one machine: conversely, if one user on one machine visits the same website using multiple browsers (say, Firefox and Explorer), the site will drop a cookie for each browser.

> Multiple locations: if a user visits a website from both home and work, that user will have a cookie on each machine and be counted as two visitors.

^ International visitors: audience measurement systems typically report on traffic from specific geographies (generally countries). Cookie-based data may or may not be able adequately to identify and filter international traffic. >

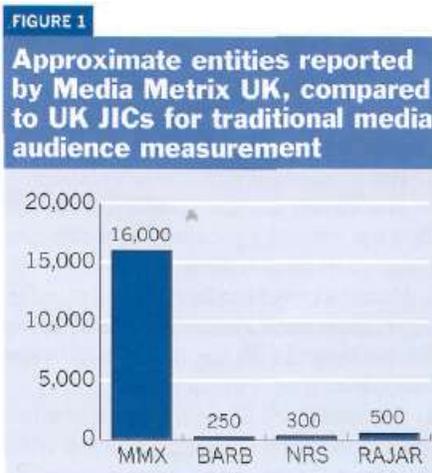


TABLE 1

Monthly unique visitors vis-à-vis cookie deletion

	Users per 100	Cookies per month	Site-centric visitors counted if frequency of visitation is:		
			monthly	weekly	daily
Delete cookies	31	4.7	31	124	146
Do not delete cookies	69	1.5*	104	104	104
Total	100	NA	135	228	250

*Note: because of access to websites via multiple browsers per machine, in the comScore cookie deletion study we observed 1.5 cookies per machine for non-deleters.

^ Non-human traffic: it has been estimated that 'hots' and 'spiders' crawling the web account for up to one-quarter of all traffic. This automated traffic will trigger server calls and be logged by site-centric data. Audience measurement systems, of course, have only humans in their panels. Web operators and site analytic providers endeavour to track and eliminate non-human traffic, but accurate identification of non-human traffic is a constant battle.

> Cookie deletion.

This last point is generally considered to be particularly problematic, and merits some discussion.

In 2007, comScore released a study of cookie deletion (3) in which we found that 31% of internet users deleted cookies; and that, of those who did, the average user exhibited 4.7 observed cookies a month (or one every 6.6 days) for the studied sites. Other studies (Belden, 2004; Jupiter, 2005; Nielsen/NetRatings, 2005) also found that at least 30% of internet users delete cookies in a month.

Website unique visitors are typically reported monthly. So let's consider the impact of cookie deletion in calculating monthly unique visitors (assuming one new cookie every 6.6 days for cookie deleters) (see Table 1).

For websites with relatively infrequent visitation (say, once a month) the impact of cookie deletion is minimal; however, the phenomenon of multiple browsers can still drive inflation in reported uniques of 35%. If a website's average audience returns once weekly, cookie deletion can result in site-centric overstatement of reach of around 128%.

In the ongoing dialogue between buyers and sellers about audience size, it is important to remember that site-centric measures will overestimate reach. The more frequently that site is visited, the greater the overstatement. This is why panel-based data have become the planning standard for evaluating online advertising opportunities. Certainly, there will always be criticisms of person-based measurement - this is the lot in life of the system that serves as the 'referee' in the competition for audiences. But, with panel-based metrics, buyers and sellers of advertising can agree on a framework for dialogue.

Internet metrics: what's next?

Like all media audience measurement, the two foundation metrics of internet measurement answer the questions, 'How many?' and 'How much?' These are necessary to develop advertising schedule audiences; 'How many?' is cumulative unduplicated audience, or reach. 'How much?' provides a measure of usage, that supports the development of frequency measures (the newspaper reader of five issues a week will be exposed to a print campaign with greater frequency than the reader of one issue a week).

The building-block metrics for online measurement are the unique visitor and the page view; the product of the two generates gross impressions.

However, over the last 12 to 18 months, there have been questions raised about the continued efficacy of the page view. These arise from two factors.

1. Technologies like AJAX, which allow content to be refreshed dynamically without the need to serve a new page: for

example, if you go to Google Maps and navigate round a map, scrolling and zooming, each action creates a refreshed map, but the page viewed does not change.

2. Rich media (i.e. video) and new content types (for example, widgets) do not lend themselves to a page-view model.

Both developments point to an increased efficacy of duration-based metrics. The short answer to the question, 'Is the page view dead?' is: 'No-but it has company.' At comScore, we expect duration-based measures to increase in prominence and relevance; the page view will not go away, but it will become less vital.

One happy benefit of a migration from page views to duration-based metrics is that advertisers and agencies tend to see an implicit relationship between time spent and engagement; all things being equal, we assume that the media vehicles with which consumers spend most time are probably those generating the greatest engagement. There is room for debate around that assumption; however, an emphasis on online duration-based metrics gives us an opportunity to build new metrics that provide greater insights into the quality of engagement.

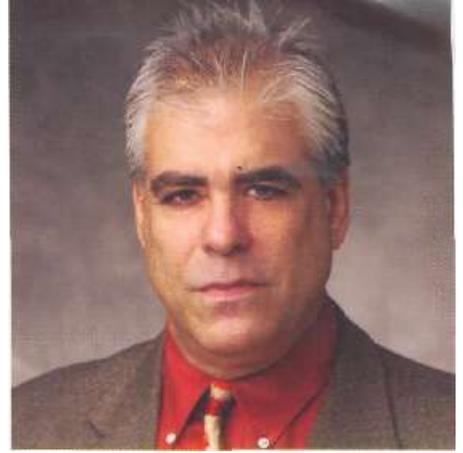
Conclusion: all media become digital

Since the earliest days of internet advertising, marketers have called for metrics that would 'let us evaluate the internet on the same terms as TV'. If we could put this new digital thing into the traditional analogue box; the thinking went, we will have a context for proper deployment of internet advertising.

But a funny thing happened on the way to the 21st century. The internet didn't grow up and resemble traditional media. Rather, traditional media are evolving into digital media, and increasingly resemble the internet. This seemingly terrifying prospect has come to pass: the internet has become the template for traditional media.

As television programme distribution disengages from the broadcast model, an increasing portion of TV consumption will take place on demand - 'served', if you

Josh Chasin joined comScore as chief research officer in May 2007. Prior to that, he owned and operated a media research consultancy, Warp Speed Marketing, whose clients included Arbitron, Scarborough, comScore, Simmons, the Online Publishers Association, the Traffic Audit Bureau, and Project Apollo. jchasin@comscore.com



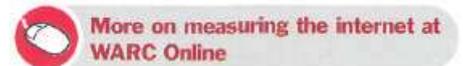
will, to the visitor who has navigated to the content and sent a request upstream. The linear construct of TV - a programme available at a certain time - is an artefact of broadcast technology: digital distribution (e.g. set-top boxes) imposes no requirement of schedule on content availability. Expect the concept of 'air date' to become more like a magazine's newsstand date; the velocity with which TV content accumulates audience will begin to resemble print audience accumulation. Expect, too, that as TV options fragment, the viewer will need a 'search' application to sort through them, suggesting a new market for search advertising in TV.

Emerging media are going to be digital. The next great media frontier is mobile/cellular. As media pundits talk about the 'three screens' - TV, computer and phone - content consumption grows increasingly device-impedent.

Consumers will read the paper online, search on TV, watch videos on the phone. A consensus is emerging that the industry will require cross-platform measures, tracking media exposure regardless of platform or device. The media measurement landscape is growing increasingly complex, despite our collective desire for a single, simple set of answers.

Understanding the complexities of internet audience measurement today can help prepare media practitioners for the coming, continued digital revolution and convergence in all media. If there is one thing we can say with certainty about digital media, collectively, it is this: that they will continue to be the most measured media. It would be nice if this abundance of metrics leads to a renaissance in the media research profession as a career path within media organisations.

- . *Another challenge to random digit dial, or ROD, sampling, is the widespread penetration of mobile phones and their exclusion from sampleframes. Persons who have mobile phones only are systematically excluded from RDD, and persons who are 'mobile-primary' will be systematically underrepresented. A December 2007 comScore study demonstrated that web usage varies by phone status - that these two sub-groups use the internet differently from the general population, so their exclusion or underrepresentation produces significant bias.*
- . *The reader interested in learning more about site analytics is directed to the Web Analytics Association, at: www.webanalyticsassociation.org/.*
- . *www.comscore.com/reauest/cookie_deletion.asp.*



Anúncio