TOWARDS PANEL-CENTRIC HYBRID MEASUREMENT: ENHANCING ONLINE AUDIENCE MEASUREMENT

Dr Pasquale A. Pellegrini, comScore, Inc.

Introduction

The notion that the internet is the most measurable medium – self-evident perhaps – is confounding in that the internet also provides the *most* measures to deal with, understand, and translate into useful metrics because it turns out that not everything measureable is what is seems. Considerable experience in reconciling panel-based internet audience measurement with site-centric (or web analytics, server-side) measurement has produced the basis for comScore's new internet audience measurement methodology called *panel-centric hybrid*. This approach combines the strengths of both site and panel-centric measurement and so is a 'best of breed' methodology. Panel-centric hybrid measurement involves the 'beaconing' of a broad range of publisher web sites processed with comScore's existing filtering technology (for non-human and non-user initiated traffic) and then combined with the existing panel-based measurement structure designed to capture accurate demographics and usage intensity at the person level. The addition of site-centric beacon data to panel-based measurement is a significant enhancement to online audience measurement in several ways: it provides increased accuracy for estimating niche audiences, removes barriers to work place traffic measurement and out of home audiences, and effectively deals with the highly fragmented digital media environment with its extremely long tail.

The emergence of panel-centric hybrid measurement, the methodology of central focus in this paper - is of particular significance for news media in general, and newspaper sites, in particular. Ironically, a scan of any recent US newspaper reveals the dire straits faced by some of the largest and most prestigious newspapers in the country, and the issue is taking on global dimensions. Declining advertising revenue in traditional print newspapers coupled with the steep recession have hastened the seemingly unstoppable migration to the web; the number of people who began to rely on the internet as a regular or even their main news source keeps growing. If declining advertising revenue in legacy news platforms is to be counterbalanced to some degree by display advertising, the measurement of the size, composition, quality and behavior of this online audience must be measured effectively. Furthermore, evidence of the advertising effectiveness of display advertising beyond measurement of 'the click' must be provided to help shape a new revenue model for news media. The panel-centric hybrid methodology improves the measurement of work place based visitation, out-of-home surfing and provides a view of mobile news consumption – three areas where news sites have considerable audience but are also very difficult to measure in the past.

The Canadian Beacon Test: The Road to Panel-Centric Hybrid Measurement

The term 'hybrid' is used to describe several methodologies for internet audience measurement and so it is useful to recognize the roots of the distinctive panel-centric hybrid described in this paper and that forms the basis comScore's *Media Metrix 360 –* the next generation of digital audience measurement. The notion of using information from census based measurement alongside panel based measurement has been floated for some time, but the execution and full scale development of a panel-centric hybrid methodology was initiated with the formation of a group of willing Canadian publishers and agencies to participate in a large-scale test of this methodology in Canada (conducted from November 2008 to May 2009). This same group, along with representation from the Internet Advertising Bureau (IAB) Canada, forms the comScore Canadian Research Advisory Council (CRAC). As each phase of testing is completed, comScore has held monthly meetings with this committee to enable a thorough and transparent review of our new methodology.

The effort put into the Canadian testing is significant not only because Canada is an important and strategic market for comScore, but also because this new hybrid measurement methodology is to be rolled out worldwide. Thus, the support of the Canadian publisher, agency and IAB community in getting the methodology tested thoroughly is a significant cooperative achievement from all perspectives.   All markets measured by comScore will potentially benefit from this advanced methodology, with Canada, the US and UK moving to commercial hybrid measurement this fall. At the time of writing this paper, the Canadian market is well beyond the beacon testing and is now reviewing and reconciling their 'beta' panel-centric hybrid audience numbers.

The Canadian experience will serve as 'proof of concept' for the methodology worldwide. The immediate reaction to the announcement of the Canadian test was overwhelmingly positive, and major publishers began the site beaconing process instantly. In the first 3 months of testing alone, some 300 domains and over 25 companies were fully beaconed. This positive response, echoed in the US, speaks to the support that such a methodology has garnered, and has permitted comScore to quickly test, refine and validate the beacon data against the Canadian Media Metrix panel data in addition to a series of other internal validation techniques [see Pellegrini (2009a) for details on hybrid validation, or Chasin & Pellegrini (2009) for an overview of the Canadian test].

How does comScore measure online audiences?

The basis for internet audience measurement by comScore is a large panel of online persons. Panels have a long history in both academic and applied business and market research, across a variety of disciplines from engineering, biostatistics, demography and sociology, employing a multitude of panel design methodologies. The reason for the continued use of panels, especially for market and media research, are the rich 'event histories' of individuals and their behavior that are collected; a longitudinal, continuous and repeated record used to provide social and behavioral dynamics. The key advantage of panel data over, say, cross-sectional or snapshot type data, is that the former permits an analysis of the exact timing and incidence of events (visits, purchases, etc) and relate the incidence of events to individual variability (Pellegrini 2009b. Panels and their associated designs, data collection and specialized analytical frameworks have provided insights into recidivism, unemployment dynamics, inter-regional mobility and consumer behavior.

comScore employs an online panel methodology for internet measurement which provides a direct means for associating each individual panelist, and their associated demographic composition, with their internet behavior. This means that the panel permits measurement of the delivery of digital content linked with demographic composition, along with metrics like how long individuals browsed a site (duration or engagement), the overall reach of a site, and the frequency of visitation. Panels permit an understanding of the context of what people were doing before or after being exposed to an advertising message so questions like 'what is the likelihood of a trademarked search after exposure to an advertisement?' can be considered. comScore is able to gather such information in a passive manner; panelists are never asked to "log in" when they begin an internet session. Thus, the behavior gathered is free of any bias associated with survey or electronic data collection where respondents remain aware that they are being observed and measured, and may alter their media consumption accordingly.

The panel-centric hybrid methodology is an extension and a significant step forward in terms of the art and science of internet measurement. To fully understand this enhancement, it is important to understand the basis of the methodology which uses a massive panel at its core. With data collected from over 171 countries, and reported on 41 countries, comScore panelists represent the global internet audience. A consistent panel methodology across all 41 reported markets means that regional and global views of internet behavior are possible, as well as market level analyses. All markets require high-quality enumeration surveys for weighting purposes so an estimate of the total internet audience for that market at home or work is required. comScore's panelists are recruited exclusively online which nicely overcomes issues such as how to reach or recruit cell-only households using random dialing approach, for example. A series of third-party application providers means that we are able to cast a wide net and recruit from a large range of sites; panelists are offered free downloads of games, utilities, data storage, etc, in return for their participation in the panel. A full discussion of the comScore methodology, from enumeration to the development of page views and unique visitors, is clearly beyond the scope of this paper, but the salient components are covered here.

Data collection from recruited panelists occurs via a software meter that is downloaded onto all machines used by the panelist at home and work (if possible) – the two reported universes. The software "meter" captures all web activity in a passive fashion, and so therefore does not require the panelist (or other panelists in the home who share the metered PC or laptop) to identify themselves at each web session. A proprietary 'session attribution technology' uses all possible information captured to "learn" how many users use the same machine, and then it differentiates their usage (i.e., sessions) via biometrics (e.g., specific attributes of keystrokes and mouse usage) and other session information. A final stage in the methodology involves weighting and minimizing any bias that may come from our recruitment philosophy. For example, there is some risk that heavier internet users will be reached more regularly by our recruitment offers and partners, so corrective weights are applied. The myriad digital data captured by our software meter from panelists requires a powerful taxonomy in order for this granular data to be digestible in a meaningful way. The so-called Client Focused Dictionary (or CFD) is a proprietary URL dictionary used to define sites' hierarchies and is created based on a combination of URL patterns that make up various sections of a site. Thus, a well-maintained and accurate CFD is paramount to internet measurement so that sites are reported in a purposeful and meaningful manner, while simultaneously filtering out non-essential domains such as back-end server calls or non-user-requested pages. The panel methodology is the only way to derive true unique visitor counts – real people – visiting pages on the web, not cookies[i], ISPs, or any other combination of non-human traffic, and so has played a major role in the movement of advertising from traditional media to digital media. Advertisers want to know who is seeing their display ad or searching as opposed to raw counts of visits, say, from server-based census data, which contains varying levels of noise which should not be counted as unique visits to a site or page.

Panel versus Server Measurement
Advertisers are wary of moving their expenditure to the internet due to the seemingly persistent measurement conflicts between panel versus census (site-centric) measurement. The long standing disparities between panel-based online audience measurement data and site-centric data have been troublesome for practitioners of online marketing and advertising for over 10 years. More recently, it has become clear that the best way to develop an audience measurement system that gains consensus and maintains scientific audience measurement rigor, is to develop a system that bridges the gap between server data and panel data and that provides answers to questions like "why don't the panel numbers match my server numbers?" Considerable experience in resolving these questions, and the Canadian beacon test, has provided solid answers and the science behind panel-centric hybrid internet measurement. Differences in geographical definitions of markets (international versus domestic traffic),

site definitions (taxonomy) , server-side traffic counting versus audience measurement approaches (e.g., unique visitors versus total visits), and the type of traffic being counted (spiders, non-human traffic) can all account for panel versus census differences. Cookie deletion is perhaps the most prominent reason for differences most importantly because of their ubiquitous use. A number of studies have shown that cookie deletion results in an overstatement of visitor counts because the resulting data is known to misclassify some returning visitors as new visitors (Fulgoni and Mörn, 2008).

But there are two sides to every story, and the dynamics of the internet sometimes pose challenges for panel-based measurement which may account for differences. For example, niche sites with small audiences will be more challenging for a panel to measure accurately even with extremely large panels like those deployed by comScore. It is important to remember that in the internet world, audiences may be reported for thousands of sites each month, as opposed to traditional media where the number of radio stations or newspapers reported for a market tends to be small. Panels are also challenged in the measurement of out-of-home traffic or internet café usage since these 'public' computers will have multiple users and not be metered. A final challenge involves the measurement of work-based internet use. Here, a unique challenge for internet measurement at work stems from restrictions on employee software downloads which restricts panelists from metering their work computers which decreases work place coverage.

Panel-Centric Hybrid Measurement

By combining the power of panel measurement - for unique visitors, unduplicated reach, and user demographics – and the best of site-centric measurement – for measuring niche sites, at work visits – this hybrid methodology provides a best of breed approach. The key to this approach is keeping the consumer at the center of the measurement system, while using site-centric data to enrich the panel measurement. The panel centric hybrid methodology involves the beaconing (or tagging) of publisher web sites across a wide variety and representative set of entities. The beaconed sites provide site centric information on visitation that is used to augment the traditional strengths of panel measurement – persons based page views and unduplicated unique visitors. The audience levels are set by the beaconed site centric measurement, while the panel is used to bring the key aspects of demographics and unique visitors to bear on these measurements, thus providing a 'best of breed' solution to online audience measurement.

Beacons allow a third party to provide server logging to a web site, in this case, proprietary beacons for comScore. The type and configuration of browsers used by visitors to a site are reported by beacons, in addition to the number of times a particular web page has been viewed, along with the exact time of each visit. comScore is able to track the web pages a visitor views within a web site, and across many different web sites. In addition, beacons permit a count of banner advertisement appearances and can match an online purchase back to the web site that displayed the banner advertisement.

Beacon site centric data will not match site server data for important reasons. As mentioned above, since we measure people, not machines, we project to the national online population only and so filter out international traffic to web sites in any market. In addition, non-human traffic from web crawling 'bots' and 'spiders' is also excluded from our measurement. Furthermore, as mentioned earlier, site centric, cookie-based unique visitor counts will be overstated largely due to cookie deletion. Here, cookie-based site centric data is known to misclassify some returning visitors as new visitors. Finally, it is important to note that not every page view logged by a server counts as a page view in the comScore panel environment. In fact, our *exclusionary* edit rules are designed to ensure that people are viewing pages and viewing the pages they request, thus so-called non-essential domains, redirects and non-user-initiated traffic is omitted from the panel estimates. All these same filters are applied to the beacon data including looking at so called outlier sites (where the visits are extraordinarily high, for example, and so require investigation) (for a full discussion of our hybrid reconciliation and validation, see Pellegrini, 2009a).

The key to utilizing the rich and wide census data lies in the panel; panels are important for keeping people at the center of measurement but are equally important in transforming census internet data into accurate, actionable, people-based measurement. Panels are used by comScore for data quality and the development of verification processes to ensure that server-side data is capturing true user experiences, as intended. The panel is also used for differentiating between different users on the same computer, or the same user on multiple machines – and so elegantly avoids the cookie deletion problem, and provides the basis for an industry standard unique visitor. Page views are also reconciled via the panel; the panel allows comScore to ensure that server-side events conform to the stringent requirements of a page view as consumed by a panelist. Thus, panels are paramount to the development of a panel-centric hybrid system not only for the fact that they provide the demographic input, but also from the panel's ability to help filter the sheer volume of census data and develop metrics that align with persons-centered measurement. The accuracy of panel-centric hybrid data is ensured via three key modes of validation:

A. comScore Direct – a self-servicing interface for clients to check against internal metrics
B. URL Reviews – an exhaustive internal URL level review of beacons and metrics
C. Client Beacon Scorecards – a comprehensive set of metrics reviewed by comScore's quality assurance team

Online Newspaper and News; Hybrid versus Panel Measurement

The panel-centric hybrid methodology is particularly well-suited to enhancing the measurement of newspaper and news category sites in general which typically have a significant portion of their traffic from work-place computers, and increasingly, from mobile and hand held devices. It is difficult to avoid hearing or reading about the challenges facing traditional print newspapers today, with several moving to digital only publishing while others have filed for bankruptcy protection. Newspaper advertising revenues continue to fall in the US, Canada and other markets, while online advertising spend remains stable or has grown even during recessionary times (Sass, 2009). In the three year period from 2006, newspaper advertising revenue share slipped from 29.9% to 21%, while internet advertising rose from 9.6% to 19.2% (Sass, 2009). The fall of advertising revenues in the past few years have had a stark impact: one out of every five journalists working for a newspaper in 2001 is no longer there (PEJ, 2009). The increasing pace of audience migration to the internet coupled with the collapsing economy has hastened the call to reinvent the news advertising model or business structure.

It is certainly not clear that readers to traditional print are simply moving away to the internet and spending time visiting pages from their preferred print brand; no clear one to one relationship exists between declines in advertising revenue and the growth of page views and unique visitors at the corresponding digital publication (e.g., NY Times readers who now read news online do not necessarily go to NYTimes.com).   One reason for this is the sheer volume of online news destinations, from branded print sites to news aggregation sites, giving people a huge selection of online alternatives forming a highly fragmented and niche rich news category that disperses the online news audience. Evidence from Canada suggests that newspaper sites are seeing some success in moving readers to their branded sites versus the more general news sites, and this trend may continue as newspaper sites adopt news aggregation models as they redefine themselves in the digital marketplace. Charts 1 and 2 show total unique visitors (13% versus 4%) and percentage reach (11% versus 2%) growing faster year over year for newspaper versus news/information sites (source: comScore Media Metrix).

Chart 1: Year over year total unique visitors to newspaper and news/information categories, all locations, Canada, 2+.
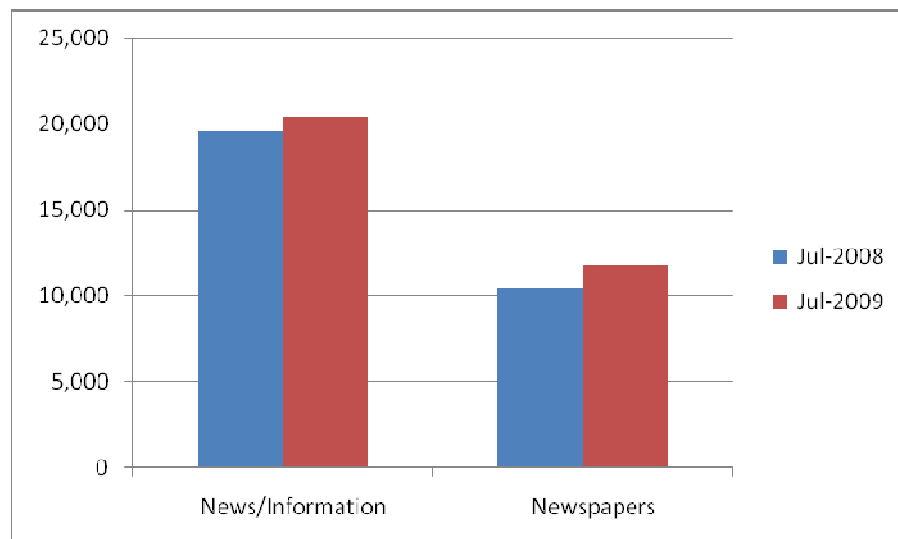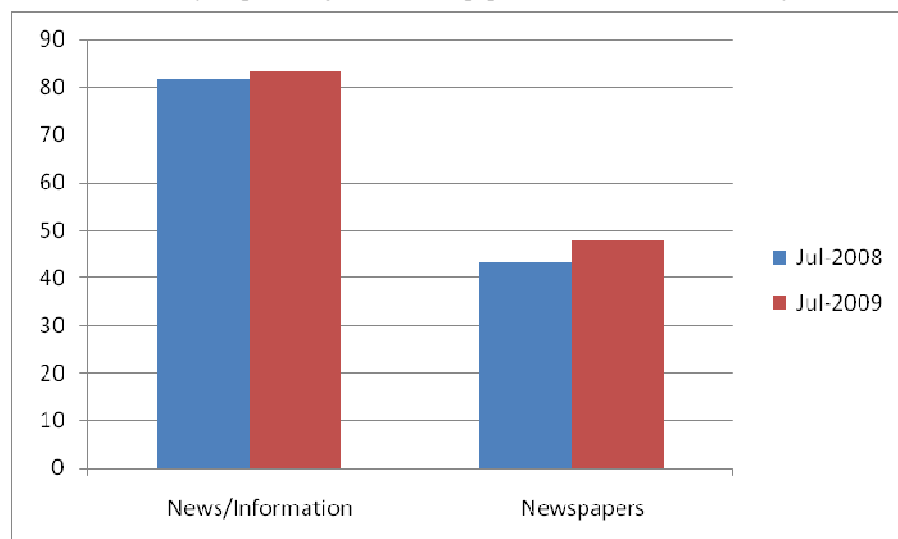

Chart 2: Year over year percentage reach, newspaper and news/information categories, all locations, Canada, 2+.

No matter how the move to online news is viewed, however, the stakes are high. PEJ (2009) reported that the ten richest media companies account for 28% of the most popular news sites, and two of the five most popular websites in the US – CNN and AOL News – are both owned by Time Warner, the largest media company in the world with revenues of $47 billion in 2008. These big players are fighting to reassemble online around an ad-supported model which has been accelerated by the economic downturn. Further adding to the need for accurate and consistent online measurement is the burgeoning mobile news category. In the US, our estimates show that more than 63 million people accessed online news and information from their mobile devices at least once in January 2009, up 71 percent from January of last year. During the same period, the number of consumers who used their mobile device to access online content daily doubled to more than 22 million (source: comScore Media Metrix). Canadian comScore data illustrate the growth of work based newspaper site visitors (20%) to be seven times faster than the home-based growth (3%) over the 15 month period from March 2008 until July 2009 (see Chart 3, source: comScore Media Metrix). The engagement at work was also higher versus home based at 21.6 versus 17.6 minutes, respectively. The need to measure home, work and, eventually, mobile newspaper audiences is clear.

Chart 3: Online newspapers total unique visitors (000s), home and work, Canada, persons 2+.
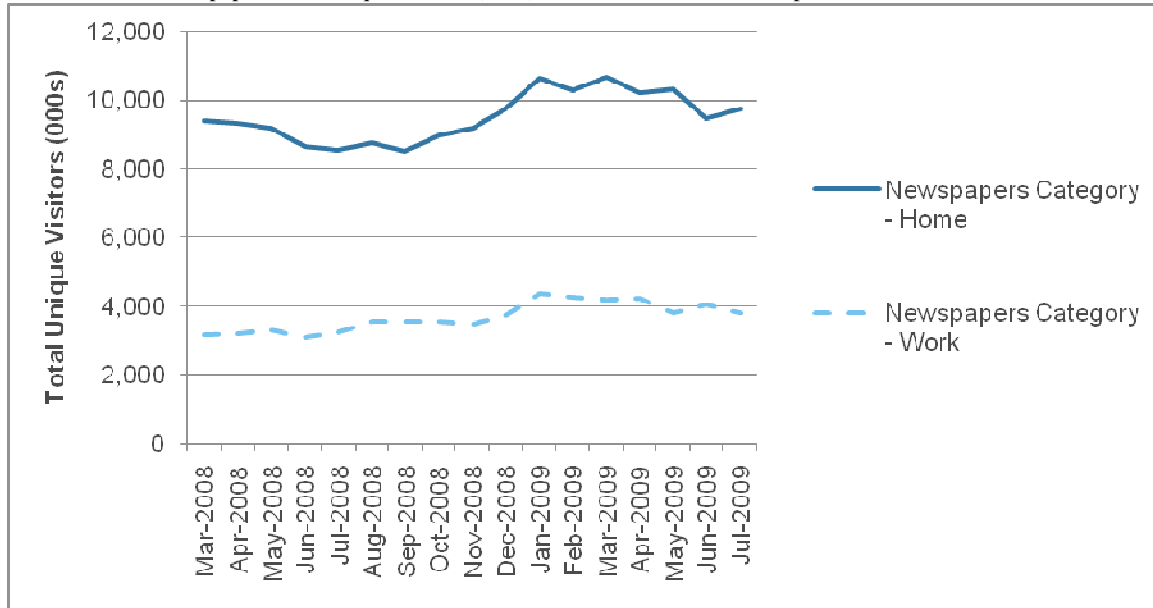
The validation and reconciliation processes between panel and beacon census data led to the development of page view and unique visitor metrics – two key metrics in online audience measurement. Page views in a hybrid world are essentially beacon counts properly filtered and validated for non-human traffic, international traffic, the application of Media Metrix page view and editing methodology (non-essential domains, multiple beacons per page, etc.), and non-eligible URLs. Despite sounding complex, this is rather straight-forward relative to the science behind unique visitors in a hybrid world. Here, unique visitors may be conceptualized as:

$$UV\left(hybrid\right) = \frac{Total\ Census\ Usage}{f\left(\alpha, \beta, \mu\right)}.$$

where the denominator is the usage rate per person *adjustment factor* and α is total usage, β is visitation frequency and μ is usage intensity. The elements in this adjustment factor take on functional forms of increasing complexity, based again on our experience reconciling census beacon data and panel measurement. This equation accounts for the fact that there are multiple machines per person households, multiple machines per person across home and work locations, multiple persons per machine, and that there are browser and operating system dependencies for cookie deletion factors, amongst other considerations. We are able to derive estimates for the various terms of this equation directly via our panels.

The panel-centric hybrid audience estimates for newspaper sites reveal, in almost all cases, marked increases in both page views and unique visitors across all locations (see Table 1). It is important to note that the numbers presented here are a total of work and home location visits, and at this point do not include mobile visits. We know from more detailed examination of these results that much of the increases shown here, and in many other examples from Canada, the US and the UK, that newspaper page view and unique visitor increases are the result of increased work-based audience – a traditionally difficult area of measurement for panels. Clearly, panel-centric hybrid measurement improves online newspaper measurement, and has the potential to further improve this measurement as mobile audience reports go online and as more niche newspaper and news sites are introduced as the legacy newspapers continue to reinvent themselves. Both small and large market newspaper sites, as well as national newspaper sites show consistent patterns of over 100% page view increases and double-digit unique visitor estimates (see Table 1). Clearly, the newspaper category is recognizing substantial improvement in measurement from the development of panel-centric hybrid measurement. It is also noteworthy that news aggregation sites see page view and unique visitor increases in a hybrid environment, although the magnitude is much smaller; about 40-50% in page views and unique visitors.

139

Table 1: Example panel-centric hybrid page view and unique visitor percentage change versus panel-based Media Metrix estimates for newspaper sites (site names withheld).

Newspaper Site/Type	Page View % change	Unique Visitor % change
1 large	107	54
2 small	156	73
3 small	165	47
4 small	126	22
5 large	183	65
6 national	161	74
7 national	134	85

Discussion and Summary

Online measurement experts have debated the pros and cons of panel-centric versus site-centric measures for some time, and the debate continues. The by-product of these healthy debates includes creative thinking towards ways to overcome the limitations of both, and naturally, ways to integrate the best of panel and site-centric data. The measurement of people is the cornerstone of publisher and advertiser requirements - to understand how people interact with content, view ads and buy products, and hence, the need for panel-centric measurement that puts the person at the core of the methodology. Nevertheless, panel-centric measurement can be enhanced by the integration of site-centric census data.

This paper reviewed panel measurement of online audiences in methodological and conceptual terms, and how this methodology can be enhanced with beacon census data (aka server-side data). The filtering and validation of both panel data, and now panel-centric hybrid data was presented, showing how the various filters and exclusions are important in getting at person level measurement. Newspaper sites were shown to have significantly higher hybrid numbers due mainly from work locations, so their all location page view and unique visitor numbers were considerably higher than their panel based audiences. As panel-centric hybrid methodology rolls out, newspaper sites will benefit from this more accurate and improved measurement in terms of what will surely be increases in advertising revenue and in terms of accurate audience metrics as feedback for design and service improvements. Panel-centric hybrid measurement represents a significant leap forward in art and science of online audience measurement in general, and an improvement of particular and timely importance to the newspaper industry as it struggles to change and adapt.

References:

Chasin, J. and P.A. Pellegrini (2009) Building the Panel-Centric Hybrid: A Canadian Perpective, **ARF Audience Measurement 4.0,** NY, NY, USA

Fulgoni, G.M and Mörn, M.P. (2008) How Online Advertising Works: Whither the Click? **Empirical Generalizations in Advertising Conference**, The Wharton School, Philadelphia, PA, USA.

Pellegrini, P.A. (2009a) Panel-Centric Hybrid Measurement: Successfully integrating traditional web analytics approaches to enrich panel-centric measurement. **ESOMAR WM**[3] Proceedings, Stockholm, Sweden.

Pellegrini, P.A. (2009b) Digital Audience Measurement: comScore's Panel-Centric Hybrid Approach Revealed, **eMetrics** conference, Toronto, Canada.

Pew **Project for Excellence in Journalism** (PEJ) (2009) The State of the News Media: An Annual Report on American Journalism, available at http://www.stateofthemedia.org/2009/index.htm.

Sass, Eric (2009) Internet Grows 37.5%, Traditional Media Declines 30%, 2006-2009. **MediaPost,** September 8.

---

[i] A cookie is a small piece of code inserted into the computer of the user whose behavior is being examined; cookies are used to try to uniquely identify the computer and thereby track its activity over time.