THE STATE OF DIGITAL MULTI-PLATFORM MEASUREMENT IN 2013 AND BEYOND: DEPLOYING BIG DATA SOLUTIONS TO OTHERWISE INTRACTABLE AUDIENCE MEASUREMENT CHALLENGES

Josh Chasin, comScore

Introduction

Two years ago, in the fall of 2011, Dr. Pascal Pellegrini and Cameron Meierhoefer of comScore presented a paper at the Print & Digital Research Forum (PDRF) in San Francisco, titled "Advances in Digital Measurement: Mobile Web Usage, Bias Correction, and Universe Coverage." In this paper, Dr. Pellegrini and Mr. Meierhoefer reported on "the very latest technology and methodology for measuring the digital media landscape," with particular emphasis placed on the emerging challenge of reporting on the holistic, or device-agnostic, digital audience across computers as well as mobile devices.

In digital measurement, two years is a long time.

In this paper, the author intends to delineate some of the core challenges in developing digital media audience measurement services, first for single-platform (i.e. "computer") Internet consumption; then, for multi-platform (computers, smart phones, tablets) Internet consumption. In building such measurement services, the media researcher is confronted with problems and challenges that would be intractable given the traditional toolkit and approaches to audience measurement.

Fortunately, the very digital technologies that render audience measurement so challenging, also provide new tools and opportunities for addressing these challenges, in the form of naturalistic, or Big Data, assets. As the author noted seven years ago (Harris, Chasin; "The Impact of Technological Innovation on Media Exposure Tracking: In Search of 'The New Traditional'", 2006), "The new complexity of audience behavior, and the burgeoning availability of media devices with which audiences can engage, suggested to us that the industry may well be more willing to jettison previously cherished 'gold standards' that had defined media research, in favor of more flexible, and wide-ranging, types of measurement and analytics."

Once the media researcher accepts the premise that the science of audience measurement is fundamentally changing, then these new tools—these data assets—may be successfully deployed in the service of creating robust and actionable estimates of the size and composition of media vehicle (and, indeed, media schedule) audiences, for individual platforms, and across multiple platforms.

The author will show how digital media measurement generates Big Data assets, and that these assets may be deployed in the solution of essential audience measurement challenges that the traditional media researcher toolkit cannot solve. Like all measurement tools, Big Data assets are not perfect. But, they are extremely powerful and eminently useful.

What is "Big Data"?

The term "Big Data" is one of the most common buzz phrases in the media landscape today. But what, exactly, IS "Big Data," and why is it interesting in a media measurement context?

Gartner offers an oft-cited definition of Big Data: "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization."[1] Similarly, the White House says that Big Data "are data sets so large, complex, or rapidly-generated that they can't be processed by traditional information and communication technologies."[2]

It is probably fair to say that a common defining characteristic of Big Data, certainly in the context of media audience measurement, is that the data is a byproduct of some other function or process. The best example is digital Set Top Box (STB) Data. In the course of normal operation, the cable operator or satellite provider generates a dataset containing, among other things, channels tuned by time, for every box in the system. This dataset—far larger than traditional TV ratings datasets—has tremendous potential to inform a TV audience measurement system. However, it is important to note that the dataset does not in and of itself *constitute* an audience measurement system; rather, it is a valuable, and imperfect, input into such a system.

Another core defining characteristic of Big Data is the potential for availability of new actionable insights emerging from combing through now-available vast stores of data, where such insights were not previously available. (Sadly, this is a potential that too often goes unrealized.)

In this paper, we will assume that-- with respect to audience measurement-- Big Data refers to naturally occurring, machine-

[1] Douglas, Laney. "The Importance of 'Big Data': A Definition". Gartner. Retrieved 21 June 2012.

[2] http://m.whitehouse.gov/blog/2013/04/18/unleashing-power-big-data

generated datasets of sufficient size and coverage that may be mined to reveal insights into patterns of media consumption at a level of granularity that traditional sample, survey, and panel-based data collection cannot provide. Of course, we further assume that these new and granular insights are essential in engineering the cross-platform measures of digital media consumption that the marketplace demands today.

In the Beginning

When broadcast media first emerged, the audience to broadcast media vehicles (specifically, at the time, radio stations and networks) were wholly unknown; unlike print media, wherein at least circulation could be known via counting (how many copies were printed and distributed), broadcast signals are cast out into the ether, and from there it is catch as catch can. Then, as now, advertisers demanded data on the size of the vehicle audience in order to wholeheartedly commit to placing dollars in the new medium of radio. In 1930, market researcher Archibald Crossley, with financial support from the Association of National Advertisers, launched the Cooperative Analysis of Broadcasting (CAB), a survey of US households to collect radio listening data on a recall basis, asking the respondent about stations heard over the previous 24 hours. Shortly thereafter advertising agencies, and finally broadcasters, became part of the revenue base of the CAB. (Beville; 1988).

In 1942, the AC Nielsen Company introduced the audiometer to commercial audience measurement—a device affixed to the panelist radio set that involved a stylus that moved across a slowly unrolling spool of coated paper as the tuning dial was moved, thus providing passive electronic measurement of vehicle consumption. At that point, the general model for modern media audience measurement was set—a sample or panel of persons (or households, and the persons comprising them), who provide, via some survey instrument or via passive electronic metering, an ongoing record of their media consumption-- the aggregation of which can then be processed, weighted, and projected to represent the relevant universe at large.

This model remained largely in place as radio gave way to TV, and as TV expanded to include cable and other forms of distribution, straight through to the digital age. When the (computer-accessed) Internet emerged as an advertising medium in the early '90s, companies like PC Meter, Relevant Knowledge, and Netratings emerged and ported the basic broadcast measurement construct online—passive metered measurement of a panel of persons, from whom demographic classification data and media vehicle exposure was collected, for the single medium in question. The major difference was that in the digital space, the "meter" was no longer a piece of hardware affixed to the set; it was a piece of software, downloaded onto the panelist's computer.

Early Digital Measurement Challenges

Digital media consumption introduced several profound challenges that tax the limitations of traditional audience measurement:

- **Fragmentation and granularity:** Even independent of the advent of the Internet, digital technology was wreaking havoc on traditional audience measurement. The buzz phrase in TV throughout the mid-90s was the "500 channel environment;" digital cable had led to an environment wherein media consumption had become far too fragmented for the old audience measurement model to support.

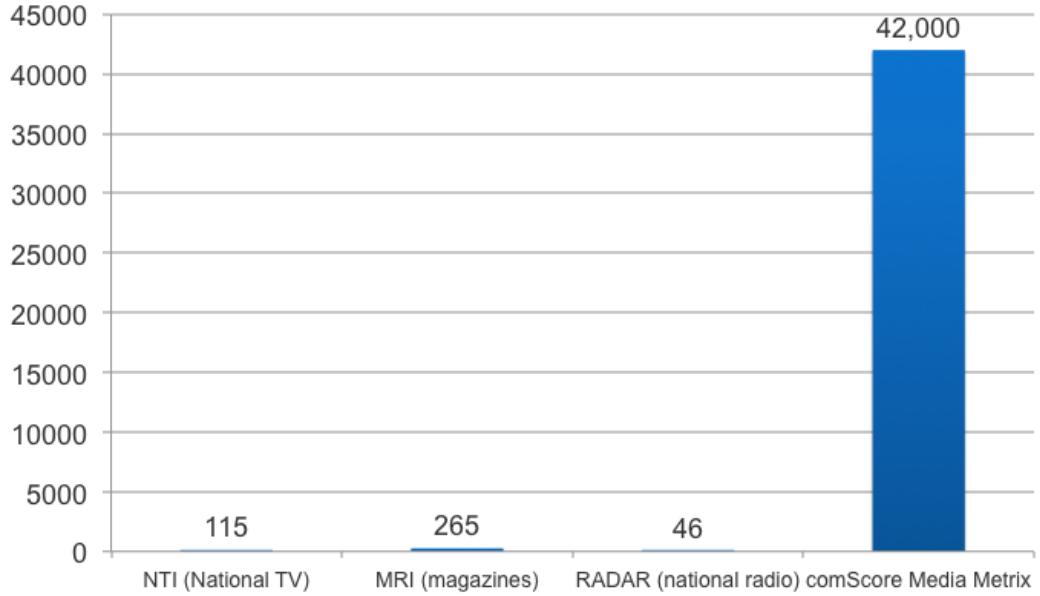
- **Universe definition:** In broadcast audience measurement, the household provided a means of entrée into both universe definition and sampling; for example, TV ratings were based on viewing in the home, to metered sets in the panel household, by persons living in the household. With respect to digital media, since many users access the Internet from both a home and a work computer, it became necessary to recruit against and integrate at least two separate universes: persons accessing the Internet from a home-owned computer; and, persons accessing the Internet from an employer-owned computer. (Indeed, in many countries, where computer ownership rates are low, it is essential to accommodate a third universe: "shared use," or Internet access from cafes, libraries, and other publicly owned computers.) For a period of time comScore also identified a third universe, Universities, which was essentially comprised of students in on campus group living quarters (i.e. dorms); but now these residences are treated as part of the home universe.

- **Workplace measurement:** Once we accept that it is necessary to measure the universe of employer-owned, or work, machines, the audience measurement practitioner is confronted with the reality that many such machines are effectively impossible to meter, owing to corporate restrictions on visitation to certain website types (generally including those from which audience measurement companies might recruit); to similar corporate restrictions on employee downloads (the meter is generally downloaded by the panelist); and, a requirement by some measurement companies that the person joining the panel has admin privileges on the machine in question (often not the case for employees with a work-owned machine.)

- **Convergence**: Convergence typically refers to the phenomenon of disparate media types coming together. In digital, oddly enough, while the measurement archtype was based largely on a broadcast model, the original consumption archetype was based on print; media companies were publishers, and content was served in pages. Quickly though, the Internet became a media environment where consumers could read news paper or magazine articles, watch TV shows, or listen to the radio (not to mention, go into a store.) Additionally, native digital content types developed, including search apps, and social media. This variety of media experiences complicates the measurement landscape.

In order to understand the impact of fragmentation in digital media on the audience measurement practitioner, consider the number of media vehicles that respective national media measurement services report on, in the US.

Table 1
Approximate Number of Media Vehicles Measured and Reported
US National Syndicated Media Audience Measurement Systems



comScore pioneered the concept of the very large online panel, by developing ways to accumulate large numbers of panelists with dramatically lower respondent acquisition costs than had ever been possible in traditional panel measurement. The medium itself became the mode of both recruitment and installation; comScore computer metered panels are recruited exclusively online, including provision of the meter and collection of demographic information about the panelists. This was the second major breakthrough in digital audience measurement (after the development of the software meter). comScore's monthly Media Metrix in-tab in the US runs in the hundreds of thousands, supporting such granular reporting.

Naturalistic Data

While the migration through the digital looking glass creates myriad audience measurement problems that would be intractable given the traditional practitioner's toolkit, the digital landscape itself actually provides a new set of tools that may be deployed in the solution of these problems. The ability to use the Internet itself as a vehicle for recruitment is one such illustration. Another is the fact that consumer consumption of digital media can, in and of itself, result in a data stream about that consumption, as a byproduct of the process of the user requesting, and the media provider fulfilling the request for, media content. This data is generally not engineered to be audience measurement data—at least not in the traditional sense—but it can have great utility in the construction of a digital audience measurement service.

(It is ironic that the tiny text files placed on user computers by web entities are called "cookies." Perhaps a better name for these files would have been "bread crumbs."[3])

The first, and perhaps best, illustration of such naturally occurring, or "naturalistic" data sets, is TV Set Top Box (STB) Return Path Data. STB data requires adjustment and calibration in order to become useful in an audience measurement context (Shabbab, Taylor; 2005). Notable issues in deploying STB data as audience measurement data include (1) the fact that STB data reports on the status of the box itself, but not the TV (which might be turned off even as the box is left on); (2) the fact of not all TV sets in the household being connected to a Set Top Box; and, (3) the absence of data on who, or how many persons, are actually in view of the set.

Still, STB data provides a powerful "Big Data" asset—second-by-second tuning data for a census of those sets attached to an operator's set top boxes—and an argument may easily be made that anyone zero-basing a TV audience measurement system today would be foolish to do so without incorporating STB data as available. In the US, Rentrak has made inroads in midsized local TV markets by offering a STB-based TV measurement service that provides far more granular data than traditional sample- or panel-based services, thus reporting on smaller audiences with greater visibility, stability and precision. Incorporation of STB data into prevailing national Television Audience Measurement services has been slow, both in the US and internationally, but ultimately such an evolution is inevitable.

[3] Brothers Grimm, "Hansel & Gretel," traditional folk tale, published in Kinder- und Hausmarchen, 1812

Site Centric vs. Person-Centric Measurement

From almost the dawn of the commercial Internet age, there has been debate about the relative efficacy of two different approaches to audience measurement: person- or user-centric on the one hand; and, site-centric on the other. The former is based on traditional panel-based measurement; the latter, on naturalistic Big Data.

Person-centric measurement, based on a representative panel of users via passive metering of their computers in order to track media exposure, offers the advantages associated with person-based measurement: specifically, measurement of individuals as opposed to machines; user demographics; and vehicle duplication[4] (do visitors to website A also visit website B?). Duplication is essential to providing media operators with competitive context; and for disentangling reach from frequency in the creation of audience projections for advertising schedules that are comprised of multiple media vehicles. However, person-centric measurement is susceptible to both sampling and non-sampling error, owing to the fact that a priori, such measures are based on a sample.

Site-centric measures are sometimes referred to as "census measurement," because by placing measurement tags on the website, the publisher and their measurement provider may collect, track, and analyze data on a census of all activity taking place at the site. This data is therefore not prone to sampling error, representing instead an empirical tally of all site-based activity. The drawbacks of site-centric data are generally associated with measurement of reach, or the number of Unique Users; "Uniques" are tallied based on the number of cookies dropped or observed at the site during the measurement period. However, for a myriad of reasons—multiple browsers per machine, multiple machines per user, multiple users per machine, and, significantly, cookie deletion and rejection—cookies are in no way equivalent to persons.

Throughout the '90s and '00s, the debates about the relative efficacy of site-centric versus person-centric digital measurement raged at industry research and measurement forums. Site-centric measurement has become the domain of the practice of Web Analytics, or so-called "internal data," wherein publishers instrument their sites with tags from an analytics provider (e.g. Google Analytics, Omniture, Webtrends, comScore Digital Analytix) to track activity and provide Key Performance Indicators, or KPIs, made available exclusively to, and about, the publisher in question. Person-centric measurement remained the provenance of practitioners of audience measurement, such as comScore Media Metrix®, Nielsen Digital, TNS, and Gemius, who provide so-called "currency" or consensus-view data on exposure to all web entities meeting minimum reporting standards, and made available on a syndicated basis to multiple subscribers, for use in the planning, buying and selling of advertising.

Both site-centric web analytics data, and person-centric audience measurement data, traffic in a metric called "Unique Visitors," and often truncated as "Uniques." Naturally, the reported Uniques from the different measurement types are widely divergent, owing to a number of factors, but largely to the fact that in web analytics, "Uniques" refers to Unique Cookies (or, more precisely, Unique Cookied Browsers), whereas in audience measurement, "Uniques" refers to Unique Persons.

To understand the extent to which site-centric measurement can overstate the number of "Uniques" relative to person-centric measurement, consider the number of discrete computer-based Scorecard Research cookies comScore observed in the US, the UK, and globally in July 2013, compared to the actual number of persons estimated to have accessed the Internet from a computer during the same time period in each geography (Scorecard Research is the comScore cookie domain):

	Cookies Observed	Estimated Internet Users*	Monthly Cookies per Person
United States	1,737,812,075	225,359,000	7.71
United Kingdom	214,403,705	44,603,000	4.81
Worldwide	6,687,198,846	1,582,029,000	4.23

**Persons 2+ in the US; 6+ in the UK; 15+ globally*

Hybrid Measurement

Not surprisingly, the idea began to emerge of creating a "hybrid" audience measurement system—integrating site-centric data with person-centric data. Site-centric, "census" data provides a powerful source for measurement calibration and validation, by introducing empirical, holistic measurement of volumetric consumption.

In a very real sense, digital hybrid measurement was presaged by Billboard (Out-of-Home) audience measurement; in the US, for example, the Traffic Audit Bureau's EYES ON ratings (renamed TAB Out of Home Ratings in 2012) involves adjusting census-level traffic counts (from government auditing; called Daily Effective Circulation) based on person-centric data to understand demographic composition and to disentangle reach from frequency (and—even more prescient—to account for Viewability.) Indeed, Philport, Matlin, and Walsh wrote about the Traffic Audit Bureau's EYES ON system, "A combination of site centric and survey centric approaches allows the outdoor industry in the United States to take advantage of the strengths of each one – to provide reliable metrics of gross volume for all outdoor sites and user data on the demographic composition of those site "audiences" and their frequency of exposure." (Philport, Matlin, Walsh, 2007).

[4] Note that Set Top Box data is naturalistic data that can also provide insight into vehicle duplication.

Several measurement organizations commercialized hybrid measurement systems in the digital space in isolated countries in the mid-00s; Nielsen, for example, was offering a methodology integrating site-centric measurement from tagging publishers in Spain, France and Italy by 2007.

In the Fall of 2008, in conjunction with a coalition of leading Canadian publishers, comScore undertook a test to determine the feasibility and efficacy of incorporating site-centric census measurement into the previously panel-based Media Metrix® audience measurement service. A year later, in August 2009, comScore launched what was branded as Media Metrix® 360, featuring Unified Digital Measurement™ (UDM™). UDM™ refers to the specific implementation of hybrid measurement deployed by comScore. The fundamental elements of UDM™ are as follows:

- Publishers agree to place comScore tags on their digital assets (websites, mobile websites, apps, etc.);
- These tags communicate with comScore's Scorecard Research server, dropping a comScore cookie (or reading the comScore cookie on computers where the cookie already exists);
- The tags fire a beacon, or pixel, associated with the specific cookie, to comScore's server upon the request of content, enabling empirical tracking of Page Views, video streams, and other content elements and types (images, ads, etc.)
- Page Views are tallied based on observed census beacon hits, filtered or sorted into the appropriately country, and filtered to exclude non-human traffic. As a result, for tagged publishers, reported Media Metrix® Page Views generally deviate from the publisher's "internal" counts by 10% or less;
- Unique Visitors are calculated and reported based on a census count of cookies (similarly sorted by country, and filtered to exclude non-human traffic), divided by a panel-observed, site-specific, country-specific measure of cookies per person (CPP). So for example, if for a given a website in the United Kingdom, comScore observes 2 million census cookies originating within the UK, and observations from the comScore UK panel indicate that there were 2 cookies per person among visitors to that site in the month, then reported Unique Visitors would be (2 million cookies / 2 CPP = 1 million persons);
- Duration (time spent) and visits are derived based on the application of a panel-derived measurement of time spent per Page View, and Page Views per Session, to census-derived Page Views.

Publishers have the option to use a redirect of the Omniture tag, as opposed to instrumenting the site with a new comScore tag, to provide comScore with census measurement. Publishers who opt not to participate in UDM™ tagging at all are measured via the panel only.

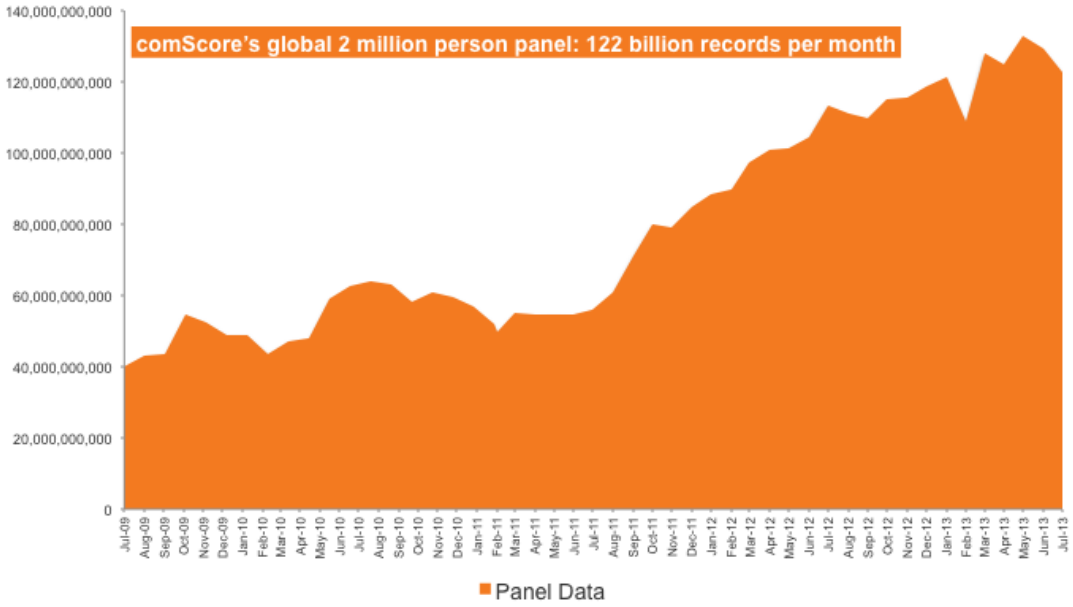
Today, comScore's UDM™ implementation of a hybrid measurement approach is available around the world, in 44 different countries, including the United Kingdom and Spain, where it is has been embraced by UKOM and MESA respectively as the marketplace currency. UDM™ is under audit in the US under the auspices of the Media Rating Council; in France through CESP, and in Spain through AIMC. Components of UDM™ have been audited in the UK by the Audit Bureau of Circulation, Ltd. ("ABC").

Is Site Centric Data, Big Data?
Does the integration of site-centric measurement constitute the deployment of a Big Data asset?

Consider that the size of comScore's global computer metered panel, at about 2 million persons, dwarfs traditional audience measurement panel sizes; and remember that metered measurement generates far more data than survey-based data.

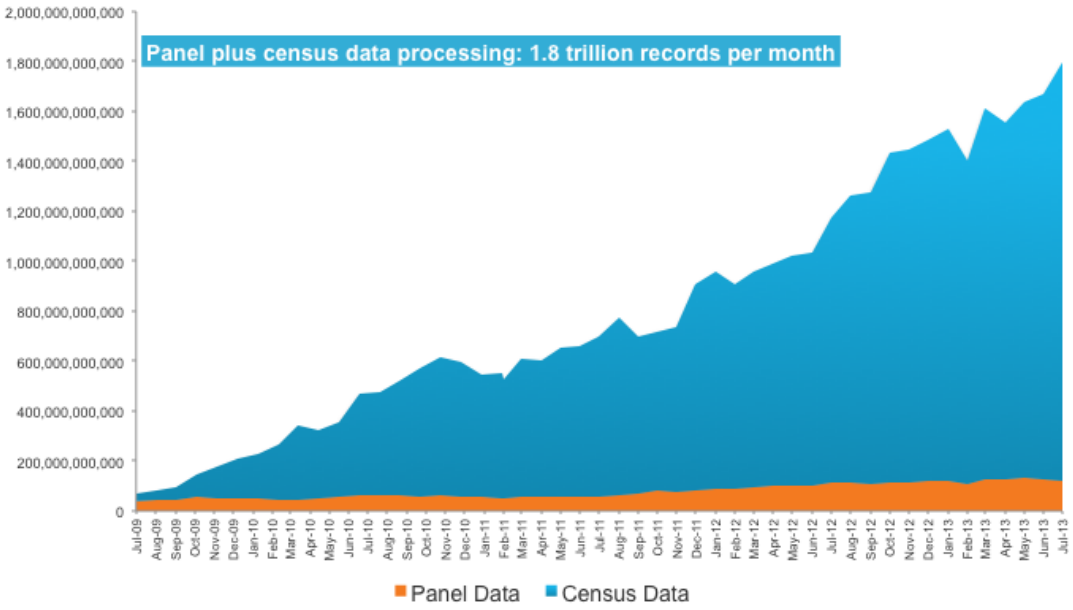
Table 2 below shows the volume of data generated by comScore's global computer panel by month, and processed by comScore, since 2009. As of July 2013, the panel was generating 122 billion records monthly:

Table 2
Data Records Processed Per Month:
comScore Global Panel Data
July 2009 – July 2013



In table 3 below, we layer in a second dataset—the volume of data generated and processed by comScore's site-centric measurement.

Table 3:
Data Records Processed Per Month:
comScore Panel Data and Census Data
July 2009 – July 2013



Surely there can be little question that a database of 1.8 trillion records is Big Data!

Enter Mobile

When mobile began to emerge as a platform through which consumers consumed media content, notably with the advent of the iPhone in 2007, companies quickly began developing mobile measurement solutions. Over time though, it became clear that what was once seen as digital or Internet measurement, had in fact been "computer-accessed Internet measurement." In much the same fashion that the emergence of the VCR introduced the concept that "TV" and "Video" were not synonymous, so too did the emergence of mobile Internet access underscore the fact the Internet access and consumption no longer required a computer.

In order to properly measure the new digital audience, it became clear that a "computer access" silo and a "mobile access" silo simply would not do. Computer-accessed Internet and Mobile-accessed Internet are not different media to be measured and reported on separately— any more than AM and FM radio should be reported in different measurement systems, or broadcast and cable TV. From the perspective of the consumer of audience measurement (buyers and sellers of advertising), radio is radio, TV is TV—and the Internet is the Internet. The medium of the Internet must be measured and reported on in a holistic, device-agnostic fashion. Indeed, it has become essential that this device-agnosticism must also be forward-compatible. Systems today for measuring Internet usage from computers, smart phones and tablets, will certainly need to be flexible enough to accommodate the next mode of mass access-- which may turn out to be wearable technology, such as Google Glass. Or something else entirely, something that those of us actively engaged in the audience measurement practice cannot currently foresee.

The challenge thus becomes, how to measure Internet access and consumption, in a fashion that is platform-agnostic—that collects and allows for the reporting on data about media usage by persons in the universe, whether accessing from computers, smart phones, or tablets?

Single Source vs. Big Data

The fundamental premise of this paper is that the traditional media research and audience measurement is inadequate to solve contemporary challenges in audience measurement, but that Big Data provides a new set of tools, providing the practitioner is open to using them.

The traditional media research solution to measuring multi-platform Internet usage would be to recruit a panel of persons, each of whom installs a meter on all their computers, smart phones, and tablets—thus creating "single source" digital measurement.

In practice though, it appears that no one seems to think this is a good idea, because no one has yet endeavored to assemble such a panel.

This leaves the alternative of developing a computer Internet access measurement system, a smart phone Internet access measurement system, and a tablet Internet access measurement system, and then developing a strategy to integrate them. But, how best to do this integration? There appear to be two alternative approaches.

Fusion has been deployed to combine disparate databases in the audience measurement space for about 30 years. Indeed, comScore participates in several multi-media fusions, wherein Media Metrix® digital data has been fused with GfK/MRI data and Monroe Mendelsohn data in the US; PMB data in Canada; and NRS data in the UK. Fusion has efficacy when the goal is to combine two wholly different measurement studies from different companies, when the two have common variables and the desired goal is to hold the currency levels of each individual study constant. However, comScore doesn't believe fusion is the first, best solution to the problem of digital multi-platform measurement, because with fusion, estimates of cross-platform duplication are an outcome of the process, not an input into it.

comScore has come to see digital multi-platform as a challenge of understanding duplication.

In comScore parlance, there are two general types of measures: reach, and intensity. Intensity metrics are volumetric, and from the practitioner's perspective, have the benefit of being additive. In other words, a publisher generating 5 million Page Views from home computers, 2 million from work computers, and 1 million each from smart phones and tablets, has generated 9 million Page Views. The metrics may simply be summed together across platform. Reach, though (or Unique Visitors, or Cume) is subject to duplication across platform; a single person may access a website from computer, phone and tablet. Unless this duplication is accounted for, the reported audience of different Unique users will be overstated (and thus not "unique" at all). Therefore reach measures are the most challenging to get right in a multi-platform environment, and to get them right, the challenge is to solve for duplication.

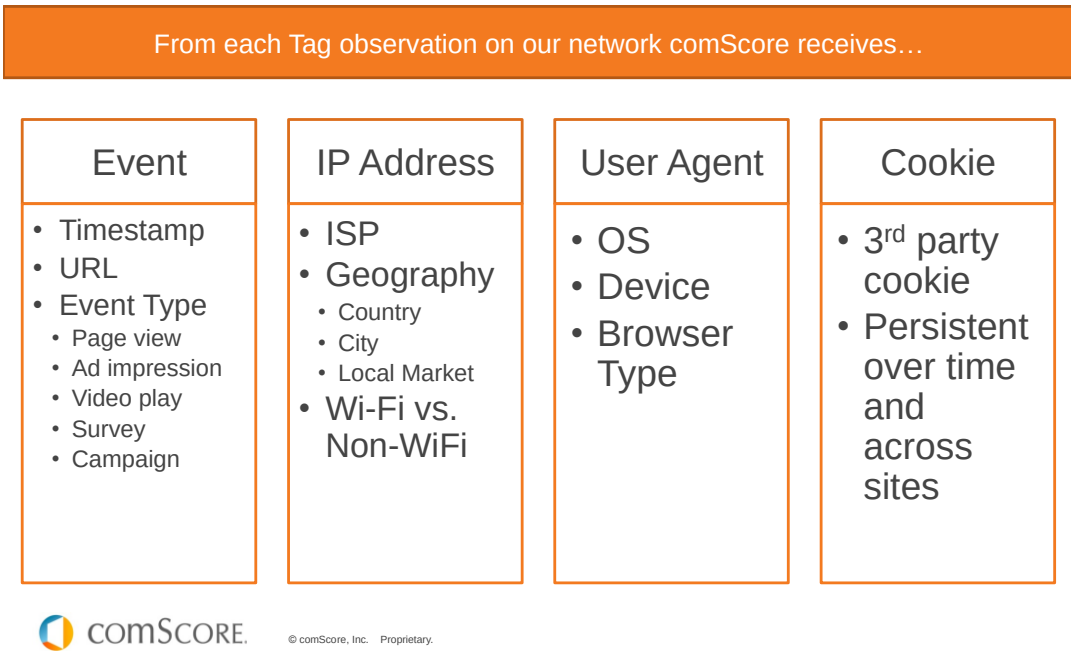
In providing "currency" audience estimates, comScore believes it is essential to observe duplication, not to generate it.

The question is, how?

Media Metrix® Multi-Platform: Introducing the Dynamic Panel
comScore launched Media Metrix® Multi-Platform (MMX MP) commercially in the US in February 2013. The service is also available in the UK, but the US implementation will be used here for illustrative purposes.

MMX MP reports on digital audiences across PCs (which is to say, computers, including both Windows and Mac machines) and mobile devices. Mobile is comprised of smart phones and tablets. In addition to the PC panel, comScore also maintains US panels of smart phone users and tablet users. Computer accessed Internet usage has traditionally been tracked through the service known as Media Metrix; mobile-accessed Internet has been tracked through Mobile Metrix® (which, similarly to Media Metrix®, deploys comScore's UDM™ implementation of a hybrid methodology.) MMX MP is the tool that combines and, for reach, de-duplicates measures from Media Metrix® and Mobile Metrix®.

As noted above, the 1.8 trillion monthly records comScore collects through publisher tagging comprises a powerful Big Data asset. Specifically, individual tag-based observations across the comScore network include the following data elements:



It is possible to sample from within the comScore census network to establish a panel-like asset to measure behavior and, significantly, overlap, or duplication. The comScore Census network deploys cookies or unique identifiers on upward of 90% of the computers and mobile devices accessing the Internet in the US, with visibility into vehicle usage for all tagging entities across these devices. From this dataset, comScore has been able to assemble a virtual, or "dynamic" household panel based on the census data, and comprised of Internet-accessing devices from within these households. Essentially, because the census data contains IP address and device type, and because of the prevalence of in-home Wi-Fi networks, it becomes possible to knit devices back together into household units, based on the assumption that devices accessing the Internet from a common IP address, are all devices associated with the household having that IP address (i.e., on the same household's Wi-Fi router).

Because we are concerned with cross-platform duplication at a person, as opposed to a household, level, and because digital census data is stubbornly agnostic with respect to the individual user associated with machine usage, the goal in developing a Dynamic Panel to measure duplication across devices becomes to isolate single person households within the census data. To do this, comScore makes the assumption that any household with one and only one computer, and one and only one smart phone, is a single person household. (While of course there are single person households with either a smartphone or a computer but not both, since the objective here is to assess cross-platform duplication, in practice these households are not helpful, and are thus excluded.)

Other governing constraints—or, in traditional parlance, Minimum Reporting Standards—include the requirement that the IP address be static throughout the month (since the measurement period is a month, if the IP address is not static across the month, we cannot make observations about monthly duplication); and, that comScore cookie on the computer in the household is static through the month (for the same reason.) It is not necessary that the mobile device cookies are static, because these devices have other unique identifiers that enable measurement throughout the month.

The application of these criteria results in a monthly Dynamic Panel of approximately one million persons:

	Approximate counts
Total observed monthly US household IPs	120 million
Total observed monthly static household IPs	40 million
Observed static IPs with static PC cookie	8 million
US monthly dynamic panel in-tab (Observed static IPs with static PC cookie and 1 smart phone)	1 million

Within these million single-person households, the device distribution is as follows:

Smart Phones	1,000,000
Computers	1,000,000
Tablets	440,000
Other Internet-enabled Devices (gaming consoles, smart TVs, etc.)	110,000

Note that there is a quite sufficiently robust sample to allow for vehicle-level observation of duplication across smart phones and tablets, as well as across smart phones and computers.

Current traditional mobile measurement metered panels (including those maintained by comScore as well as others) run as large as 10,000-20,000, but no larger. The ability to create a panel-like asset from Big Data that provides data, not just on mobile usage, but on computer usage as well—and that is discrete enough to support observations of site-specific duplication at a very granular level across platforms-- represents a profound breakthrough. By developing a Dynamic Panel with an in-tab of one million persons, comScore has been able to empirically observe cross-platform duplication—thus solving the primary challenge to cross-platform measurement.

Post Script: Sure, But What About Sites That Haven't Tagged?

One of the prerequisites for a currency measurement system is that all entities with a sufficiently large audience to pass minimum reporting thresholds be included in reporting. In Media Metrix, and indeed in Mobile Metrix, entities that choose not to deploy tags for comScore measurement are still included, via a default to a panel-only reportage. The astute reader will have no doubt realized that, if tagging is responsible for generating the data used to de-duplicate across platforms, and if de-duplication is essential to cross-platform reportage, then there is a problem as pertains to reporting on cross-platform usage for non-tagging entities.

In practice, what comScore has done is to develop an algorithm based on observation of real duplication across platforms, for all tagged entities. This algorithm is then applied to all entities in the reporting, in order that non-tagging entities may be included. The development of this algorithm is an interesting topic, but is beyond the scope of this paper. The reader should be aware, though, that development of such an algorithm is both possible, and—in order to report on non-tagged entities—necessary.

Caveats and Future Developments

Some users have questioned the efficacy of restricting the duplication observation to single person households. Ultimately the relevant question becomes, to what extent is the cross-platform duplication of persons in one-person households different from that of persons in multi-person households, and is this material with respect to the calculation of multi-platform estimates?

The answer here is somewhat uncomfortable for audience measurement traditionalists. The fact is that digital multi-platform audience measurement remains an emerging science. Technological adoption is far more rapid today than ever before, and the science of audience measurement is hard-pressed to keep pace with demand (and doing so would be wholly impossible with the traditional tool kit.) Consider that the consumer tablet marketplace effectively didn't exist until April 2010, when Apple introduced the iPad. According to comScore's TabLens report, in April 2013—just three years later—there were 66.4 million tablet users aged 13 or older in the US (By July that figure had increased to 70.6 million.)[5] By comparison, the first TV broadcasts in the US took place in 1928, and the first TV advertising ran in 1941; but by 1948 there were only about 35,000 sets in use.[6]

While observed duplication based on persons in single person households may not be optimal, it is state-of-the-art.

A second relevant topic to address is one of demographics; since the age and gender of the person in the single-person Dynamic Panel household is unknown, how can the demographic composition of the cross-platform audience be quantified? Here, the answer is simple. If multi-platform Unique audience may be expressed as (Computer audience) + (Mobile audience) – (Computer/Mobile overlap), then we simply perform this calculation at the discrete demographic level. The Unique computer audience, and the Unique mobile audience, for given entities within given target audiences are already known from Media Metrix and Mobile Metrix, respectively. The Computer/Mobile Overlap is then modeled based on the Dynamic Panel-observed algorithm.

[5] comScore TabLens, TabLens Owner Trend, July 2013, persons 13+.
[6] "Television." *The Encyclopedia Americana*. New York: Scholastic Library Publishing, 1984: 433

Conclusion

By developing a virtual, household-based "Dynamic Panel" from the Big Data asset of the comScore Census Network, comScore has managed to develop a tool for tracking cross-platform digital vehicle consumption duplication at an extremely precise and granular level. Development of such an asset is essential to the reporting of Internet usage across platforms, and would have been impossible without a Big Data solution.

Such Big Data deployments in the service of media vehicle audience measurement are of course imperfect. Then again, so are traditional measurement tools. But Big Data assets lie the Dynamic Panel provide audience measurement practitioners with the ability to measure and report on an increasingly complex and fragmented media landscape, with greater and greater precision.

Sources

Beville; Audience Ratings: Radio, Television, Cable; Lawrence Baum Associates; 1988

Bulgrin, Enoch; *Project Blueprint: A Path to a Multi-Platform Solution*; presented at the ARF Audience measurement Symposium, June, 2013

Carvalho; *Multi-Screen Media Planning: Audience Measurement*; AdMap, July/August 2012

Chasin; *The Internet as a Template for Today's Audience Measurement*; AdMap, July/August 2008 (issue 496)

Chasin, Harris; *The Impact of Technological Innovation on Media Exposure Tracking: In Search of "The New Traditional"*; presented at the ARF Audience Measurement Symposium, June 2006

Pellegrini, Meierhoefer; Advances in Digital Measurement: Mobile Web Usage, Bias Correction, and Universe Coverage; Presented at the Print Digital Readership Forum; 2011

Philport, Matlin, Walsh, *More Than Just a Survey: Considerations for Selecting an Integrated Approach to Audience Measurement*; ESOMAR Worldwide Multi Media Measurement (WM3), Dublin, June 2007

Shababb, Taylor; *From Set Top Meters to Set Top Boxes: Audience Measurement for the Digital World*; EOMAR TV Conference, June 2005