

Confusion Online: Faulty Metrics and the Future of Digital Journalism

Authors: Lucas Graves and John Kelly
Contributor: Marissa Gluck

Tow Center for Digital Journalism
Columbia University Graduate School of Journalism
September 2010

Dear Readers,

We are proud to present to you the first article of research from the Tow Center for Digital Journalism, which is launching this fall at the Columbia University Graduate School of Journalism.

In this report, our researchers dig deeply into the many traffic metrics that confront online news sites, and in the process, they provide insight into how that cacophony of numbers affects everything from advertising models to editorial decisions.

You'll notice that the authors also take an unusual additional step at the end of the report: They offer recommendations on how and why the Tow Center could play a role in explaining these metrics and helping the industry deal with them.

We think it is most appropriate, given the nature of the Internet, that we openly solicit your thoughts about how to shape some of the research goals of the Tow Center. And so we invite you to read this report, provide feedback on the findings, and give us your ideas about what research areas would be most valuable for the Center to pursue. You can leave comments on the accompanying article on the Columbia Journalism Review, http://www.cjr.org/reports/traffic_jam.php, or write us at our email addresses below.

We also would like to thank Mary Graham, a member of the Journalism School's Board of Visitors, whose generosity enabled the Tow Center to produce this report.

Thank you, and we look forward to hearing from you.

Sincerely,

Bill Grueskin
Dean of Academic Affairs
Columbia School of Journalism
bgrueskin@columbia.edu

Emily Bell
Director, Tow Center for Digital Journalism
Columbia School of Journalism
eb2740@columbia.edu

Executive Summary

The *New York Times* is one of the most popular news destinations on the Internet. Its online audience has been growing steadily for years, but over the first half of 2010 the number of monthly visitors to Times-owned sites surged — from 53 million to 72 million people, according to comScore, one of the leading firms tracking Web usage.

Why the sudden boost? The surge was a methodological anomaly: comScore decided to change the way it counts online users. Both comScore and its biggest rival, Nielsen NetRatings, have been revamping their secret formulas to bring their audience tallies closer to what online media outlets claim. But the two firms don't often agree with each other, either. In May comScore gave Washingtonpost.com an audience of 17 million "unique visitors," while Nielsen recorded fewer than 10 million. Their calculations of Yahoo's audience differed by 34 million people, roughly the population of Canada.

Media measurement has never been an exact science. But even by its imperfect standards, Internet audience estimates vary to an astonishing degree, depending on who does the counting and what methodology is applied. These swings are especially challenging for smaller, online-only properties trying to get on the radar of major advertisers.

Consider *The Daily Beast*, the news-and-opinion outlet edited by Tina Brown and owned by Interactive Corp. As reported in an *LA Times* profile early this year, Nielsen put the site's audience at 1 million for October of 2009; the same month comScore counted 2.2 million visitors, more than twice as many. Meanwhile, according to the site's own servers, close to 4 million different users were reading the *Beast* each month.

Comparisons like these are far from unusual. A striking paradox exists in the world of Internet measurement, whose math shapes the fortunes of news organizations both new and old: What is supposedly the most measurable medium in history is beset by a frightening tangle of incompatible standards and contradictory results.

This messy landscape poses a stark contrast to traditional media. Journalists working in newspapers, magazines, radio, and television can

rely on a dominant “currency” (though always an imperfect one) to measure audiences, close advertising deals, and assess the competition.

This report explores the industry of Internet measurement and its impact on news organizations working online. It investigates this landscape through a combination of documentary research and interviews with measurement companies, trade groups, advertising agencies, media scholars, and journalists from national newspapers, regional papers, and online-only news ventures. Principal findings include the following:

- * Major online news outlets routinely subscribe to multiple, incompatible, and quite expensive sources of audience measurement, picking and choosing data to tell a compelling story to advertisers. Smaller ventures rely mainly on Google Analytics.
- * Uncertainty about audience measurement hinders online ad spending, with buyers and sellers of media favoring incompatible metrics. (A 2009 study by McKinsey & Co., commissioned by the Internet Advertising Bureau, echoed this finding.)
- * Uncertainty about audience measurement impedes editorial decision-making, with editors unsure of which readers favor what coverage. Editors still choose among costly projects by instinct; as one reported, “You have more data, but it’s conflicting.”
- * The media-planning dynamic is inverted online: marketers allocate more resources to optimizing campaigns as they run, rather than to planning them beforehand. Even brand advertisers have adopted the thinking of the less-fashionable world of direct marketing.
- * Advertising technologies used online, such as behavioral targeting, tend to erode the value of a news outlet’s audience profile. Increasingly, the decisive information resides not with the publisher but in the databases of intermediaries such as ad networks or profile brokers.
- * As a result, despite widespread calls for a common currency, the online ad industry does not depend on having a single measurement standard like Nielsen’s TV ratings. In contrast to the world of television or magazines, space for advertising is not a scarce resource on the Internet,

and online marketers don't rely on ratings in the same way to purchase media or to evaluate their campaigns.

Thus the chaos of competing metrics online does not represent the failure of the Internet's promise as the most "accountable" medium, but, in some ways, its realization. The global network generates far more natural data about audiences than any prior mass medium; at the same time it diminishes the need to anoint a single, arbitrary standard for the sake of agreement. It turns out that accountability is a messy business.

This report identifies two routes which may bring a measure of consensus to this fractured landscape: A merger between the two top online ratings firms, comScore and Nielsen, or the emergence of Google Analytics as an accepted standard. Neither of these paths is assured, however, and neither would achieve the "clarity" of traditional media currencies — a clarity that results more from a lack of data than from good data.

In this environment, the report identifies three future avenues of inquiry in pursuit of the Tow Center's mission to foster viable and effective digital journalism:

First, educating journalists to navigate the chaos of data about online audiences, and in particular about journalism on the Internet.

Second, developing resources to help journalists understand the impact of their work, beyond counting "eyeballs" — thoughtful measures of how news travels and what effect it has in a networked information economy.

Third, producing much-needed research on emerging business models for professional journalism, in order to understand how high-quality reporting can thrive when old media economics no longer apply.

I. Fractured Media Metrics: The Lack of an Online “Currency”

A striking contradiction exists at the center of the confusing world of Internet metrics: What is by all accounts the most precisely measurable medium in history, in which every act of reading, watching or listening is a discrete, recorded event, is beset by a frightening tangle of incompatible standards for gauging traffic.

Every new medium has endured a period of statistical upheaval. Without exception, though, major ad-supported media platforms — newspapers, magazines, TV networks, radio stations — have settled on one dominant, third-party standard for counting audiences. In contrast, the online landscape today is, if anything, more fractured and confusing than in the Internet's earliest days as a popular medium, still characterized by basic disagreements over not just how but what to measure. This cacophony persists despite the clear maturation of the online advertising industry, which according to Forrester Research will claim \$29 billion in the United States in 2010, or 13 percent of total ad spending (though search-engine marketing accounts for more than half of the online share).

Among online news outlets of various stripes, the perception of a chaos of competing metrics seems to be universal. This is a troubling issue for these publishers, editors, and reporters. As they seek to perform powerful journalism with a wide impact, they are befuddled by contradictory data sets that fail to capture how their stories are being distributed or read, and what sort of impact they are having on their audience or on the institutions they cover. Furthermore, as they seek to build sustainable business models in the online economy, they have a hard time finding the reliable, consistent data that allow other industries to grow and thrive.

Several industry groups, representing publishers as well as advertisers and the measurement firms themselves, have launched initiatives that aim to bring clarity and consensus to the online measurement landscape. To understand what online news ventures can or even should hope for in such efforts, and what kind of contribution the Tow Center can make, requires first understanding *why* a measurement currency has not emerged online and *whether* one is likely to.

The key question, in other words, is whether the continuing disagreement over online measurement standards on the Internet is evidence of a young

medium, or of a fundamentally different one. Will an agreed-upon currency emerge for counting audiences on the Internet? Why hasn't one taken hold thus far? Answering these questions begins with a closer look at measurement currencies in traditional media.

Approaches to traditional media measurement

Every major platform for news, as for media more broadly, relies heavily on third-party measurement firms. This is especially true in the corners of journalism that count on advertising as a main revenue source. From the perspective of a publisher or broadcaster of news, media measurement — which typically means audience measurement, even when audiences aren't being directly polled — fulfills three distinct but overlapping needs:

- Understanding audiences, for editorial as well as commercial development.
- Evaluating competitors.
- Selling ad space. This includes marketing audiences to advertisers as well as setting ad rates and closing deals.

In each case, the most basic role of media measurement is to achieve a consensus on the number of people reading, watching, or listening to a particular news outlet — and to the ads it carries. Has readership increased since the redesign? Did last week's feature win viewers over? Can we command a premium with advertisers, based on our demographic profile? All of these questions revolve around measuring audiences.

Data Sources & Methods for Audience Measurement

Company	Medium	Data source	Methodology
Audit Bureau of Circulation	Print	Circulation audit	Census
BPA Worldwide	Print	Circulation audit	Census
Mediamark Research	Print	Survey	Panel
Simmons	Print	Survey	Panel
Scarborough Research	Print	Survey	Panel
Arbitron	Radio	Meters and Diaries	Panel
Nielsen Company	Television	Meters and Diaries	Panel
Nielsen NetRatings	Internet	Meter Software	Panel
comScore	Internet	Meter Software	Panel
Omniture	Internet	Server records	Census
Google Analytics	Internet	Server records	Census
Hitwise	Internet	ISP & panel	Hybrid
Quantcast	Internet	ISP & server	Hybrid

Source: authors' research

The \$19 billion U.S. market research industry¹ harbors many media tracking firms, offering a dizzying array of products based on various methodologies and data sources. However, among firms that measure media audiences, two broad approaches prevail. **Panel-based** measures such as Nielsen's TV ratings operate by tracking media usage within a small, carefully maintained panel of media users and extrapolating their habits to the broader population. **Census-based** measures, so-called because they purport to reflect the entire universe of media users rather than just a sample, are possible only where distribution offers some clue about that the size of universe — for instance, in records of the number of copies of a newspaper printed and sold each day. (A more accurate split might be between estimates derived directly from audiences and those that begin with media producers.)

¹ According to the 2010 Gale *Encyclopedia of American Industries*.

The role of media measurement “currencies”

Historically the advertising market has been the basis for powerful, decades-long monopolies in audience measurement: Firms such as the Nielsen Company, Arbitron, and the Audit Bureau of Circulations provide the “currency” that buyers and sellers of media use to set ad rates based on the size and quality of the audience being reached.² This is true despite the fact that media measurement has been plagued by controversy over methodology and business practices.

Consider Nielsen’s eponymous TV ratings, the undisputed currency of both broadcast and cable television. As early as 1963, doubts about the accuracy and fairness of broadcast ratings led to a series of Congressional hearings. One contemporary account summed up the findings this way:

“The hearings suggested that the illusion of exact accuracy was necessary to the ratings industry in order to heighten the confidence of their clients in the validity of the data they sell. This myth was sustained by the practice of reporting audience ratings down to the decimal point, even when the sampling tolerances ranged over several percentage points. It was reinforced by keeping as a closely guarded secret the elaborate weighting procedures which were used to translate interviews into published projections of audience size. It was manifested in the monolithic self-assurance with which the statistical uncertainties of survey data were transformed into beautiful, solid, clean-looking bar charts.”³

Still, Nielsen’s weekly ratings have enjoyed a half-century reign despite any number of critiques leveled against the system since then: that self-reported viewer “diaries” are unreliable; that the household panel has been too small; that the panel undercounts out-of-home viewers, for instance in college dorms; that the panel undercounts minorities; and most gravely, that the panel is not a truly random statistical sample of US

² This discussion of “currency” metrics draws on M.R. Sales & M. Gluck, *The Future of Television: Advertising, Technology and the Pursuit of Audiences*, The Norman Lear Center at the Annenberg School for Communication, September 2008.

³ L. Bogart, “Is it Time to Discard the Audience Concept?,” *Journal of Marketing* 30 (1), January 1966.

households.⁴ Nielsen's many methodological tweaks over the decades — such as increasing the size of its panel and deploying automated measurement technologies — suggest that these critiques have carried some weight.

It is a long-appreciated irony of media measurement that accuracy matters less than consensus. A media executive may have little faith in the formula used to infer the radio choices of millions of commuters, or in the “pass-along” multiple that transforms a small newspaper circulation into a much wider assumed readership; these doubts don't matter much as long as no competitor is seen to benefit. As an ABC executive put it in a 1992 PBS documentary about flaws in Nielsen's methodology, “Everybody's dealing off the same deck.”⁵

The history of radio, television, and print strongly supports the conclusion that buyers and sellers of media invariably anoint a single, third-party “currency” for counting audiences. This may be a messy process — for instance, Arbitron originally formed in 1949 (as the American Research Bureau) to track television viewing, and competed with Nielsen for decades before finally conceding defeat in 1993. Nevertheless, Nielsen led its rival throughout that period, and only became stronger as fewer and fewer clients were willing to subscribe to more than one ratings firm. Similarly, Arbitron has faced a number of challengers in ranking radio audiences (including Nielsen, which launched the first radio ratings service in 1942 and re-entered the business last year) but has dominated the industry for four decades. Each firm has also weathered profound change in its industry, for instance the growth of cable TV in the 1980s, and the dramatic consolidation of radio after 1996.

At first glance, print media seem to offer an exception, with two measurement firms maintaining competing currencies. However, each firm dominates a different segment of the print landscape: The Audit Bureau of Circulations, founded in 1914, is the audience currency among newspapers while BPA Worldwide, founded in 1931 as Controlled Circulation Audit, has much deeper support in the magazine industry and overseas.

⁴ For recent examples see M. Schneider, “Fox Wants Answers from Nielsen,” *Variety*, 18 May 2009; and A. Segal, “Nielsen Ratings: An Inaccurate Truth,” *Cornell Daily Sun*, 26 April 2007.

⁵ “Can You Believe TV Ratings,” *Nova*, PBS, 1992.

How measurement monopolies survive

Clearly, incumbent media measurement standards benefit from network effects. The more widely a currency such as Nielsen's ratings system is used to negotiate TV ad deals, the more necessary it is for any TV network or ad agency to subscribe, further cementing Nielsen's status as an industry standard. (This competitive advantage is enhanced by barriers to entry, such as the high fixed costs of establishing a viable panel without a base of clients already in place.) Given such a feedback loop, even a small lead by one measurement firm will – all other things being equal – eventually lead to outright dominance. This pattern suggests that a divided measurement markets are inherently unstable, and thus that a single measurement currency will emerge online as it has in other platforms.

The reality is messier than this. Two crucial features of existing measurement monopolies deserve close attention. First, and paradoxically, a single currency appears to be most dominant precisely in the broadcast platforms — television and radio — where natural data about audiences are absent. Conversely in print, where more complete, census-like information about the size (and to some extent the quality) of the reading universe has always been available from subscription and newsstand figures, the measurement landscape remains more fractured. Upon reflection, the paradox disappears. To compare separately audited (or even unaudited) newspaper circulation figures is less than ideal. But to compare reach estimates projected from separate audience panels makes almost no sense, since the panels may be biased in different ways. (This has become painfully clear in the disagreement between the two online panels, discussed in the next section.)

Second, the role of advertisers in anointing currency metrics cannot be ignored. Advertisers and ad agencies drove the formation of the circulation auditors ABC and BPA; funding for the two nonprofit auditors comes from dues paid by advertisers, agencies, and media companies, but agencies and advertisers dominate boards of both organizations. (Newspaper publishers are wary of ABC's online auditing proposals partly because advertisers have so much influence over the group.) Likewise, both broadcasters and advertising agencies pay to subscribe to Nielsen's television ratings. The firm's pricing is opaque, but full subscriptions for commercial clients run to tens of millions of dollars per

year. Media producers have a clear interest in understanding and comparing their own audiences. But it remains an open question whether a single measurement currency would emerge in the absence of pressure from advertisers and agencies.

Finally, established measurement firms work hard to reinforce network effects and protect their monopoly. A report from USC Annenberg's Lear Center points to Nielsen's "carefully staggered annual contracts," which make it extremely difficult for a challenger to win over a critical mass of clients.⁶ As Arbitron's CEO told the *New York Times* about its decision not to get back into TV ratings in the 1990s, "We looked at this and saw that there's a long history of people taking runs at the incumbent. ... But there's no halfway here. If we were to go after Nielsen, it would be war, and at the end of the day there would be one person standing. And believe me, there are skeletons littering the trail."⁷

Why no currency for online measurement?

As noted at the outset, the online measurement landscape remains extremely fractured. Two major firms, Nielsen NetRatings and comScore, are vying to become the industry standard in panel-based audience measurement; several smaller firms also compete in this arena. But as the next section explores in detail, publishers who subscribe to one (or both) of these typically also employ census-based audience measurement, which in the online world means analyzing server-side records of audio, video, and text pages served out over the Internet. (Many companies offer this kind of analysis; two leading services are Omniture and Google Analytics.) Meanwhile a third kind of measurement firm — Hitwise is an example — measures audiences using data aggregated from internet service providers (ISPs).⁸ The closest analogy in the offline world might be polling retailers about who is buying what magazines. Finally any number of startups are attempting to measure audiences and activity on new platforms such as mobile devices and social networks.

⁶ *The Future of Television*.

⁷ J. Gertner, "Our Ratings, Ourselves," *New York Times*, 10 April 1995; cited in *The Future of Television*.

⁸ For a good review of contenders see "Sizing Up Online Audience Measurement Services" from the Newspaper Association of America, at <http://www.naa.org/Resources/Articles/Digital-Edge-Sizing-Up-Online-Audience-Measurement-Services.aspx>.



Why hasn't a measurement currency emerged online? The preceding review of traditional media currencies begins to point to an answer — first, in the unprecedented abundance of data about online audiences and behaviors available from multiple sources, and second, in the limited role third-party ratings play in planning and paying for online ad campaigns. The following two sections will review each of these threads in turn, pointing to a basic decoupling of audience measurement and advertising. Ultimately this analysis suggests that though a single standard for estimating audience size may emerge, it won't play the pivotal role that measurement currencies have in the past.

II. Measuring Online Media: Disputes About Data

The term “banner ad” was coined by the site Hotwired, which standardized the novel advertising format and began selling it on a wide basis in late 1994. (By most accounts AT&T was first to try the format, with a come-on that read, “Have you ever clicked here? You will.”) Almost immediately Hotwired took the logical next step and began to report on “click-through rates” to its advertisers, giving them a new way to assess the success of their campaigns.

From its earliest days, the nascent online advertising industry was taken to herald a revolution, offering a precision and depth of information unmatched by any other advertising platform. The Internet was touted as the first truly “accountable” medium. A new class of consultants and agencies sprang into being to develop a vocabulary of techniques to exploit the medium’s capabilities.

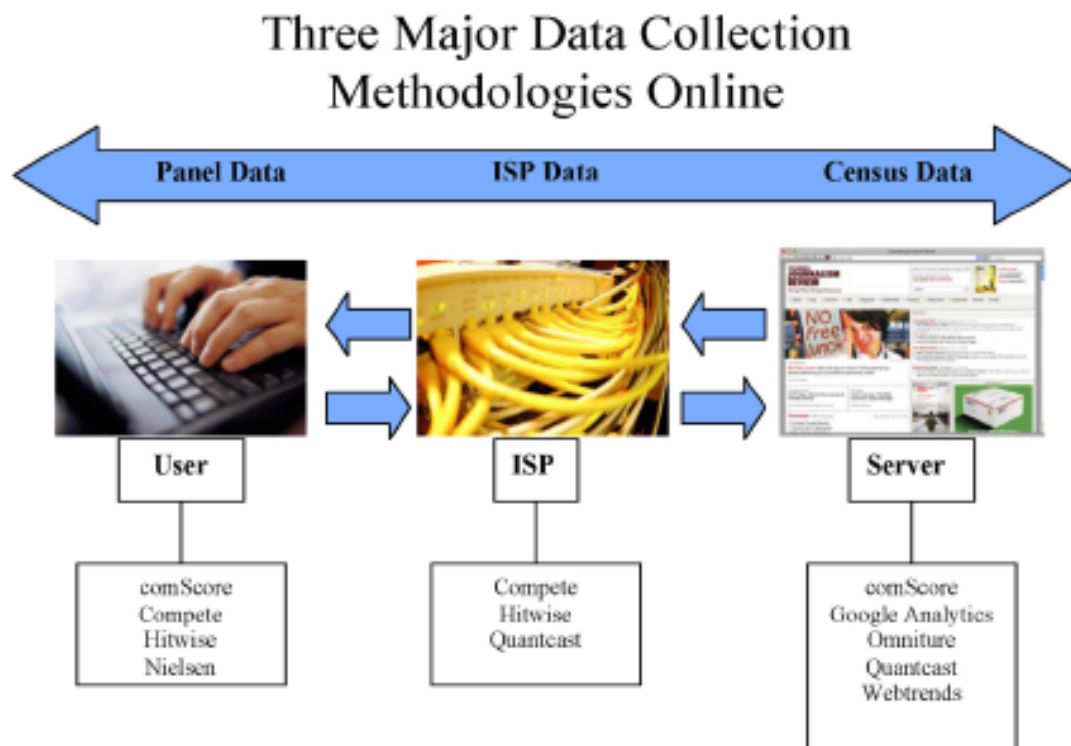
And yet, the reigning perception online is one of chaos and confusion. The industry cannot agree even on basic conceptual definitions, such as what constitutes a “unique visitor.” In the world of online news, individual publishers routinely negotiate a number of basic audience metrics which are not only mutually incompatible, but also vary wildly from month to month. Publishers seem to agree that much of the available data are unreliable, but disagree about precisely which.

Reconciling these two basic features of the online measurement world — abundant information and persistent confusion — is the key to understanding this world and how it is likely to develop. The chaos of audience information online does not represent the failure of the Internet’s promise as an advertising medium, but rather its realization. Information abundance is chaotic.

An embarrassment of data

In spite of the ceaseless business hype surrounding the Internet, it can be easy to understate the shift it marked in terms of the quantity and variety of data generated about audiences. As noted previously, traditional media operate with a relative paucity of “natural” information. Without conducting surveys, a TV or radio broadcaster has no direct indication of the number of people watching or listening.

Print publishers have more natural information at their disposal. Newsstand-driven publications record the number of copies of each issue distributed and returned, but have to estimate total readership. Subscription-based periodicals also have basic demographic information about their readers. (Controlled-circulation titles know much more, and use reader surveys and free subscriptions to actively shape an audience profile desirable to their advertisers.)



Source: authors' research

On the Internet every action by a reader generates a data trail in a chain of computers running, at a minimum, from the Web site being visited, to the ISP, to the user's own browser. Each of these three tiers has become the basis for a competing approach to audience measurement: Server-based (or census-based) analytics software runs on publisher servers; ISP-based estimates collect traffic data from major ISPs such as Verizon and Time Warner; and panel-based metrics use tracking software installed on the computers of a panel of Internet users.

Ad servers and advertising networks introduce additional layers of data collection: a single page request by a reader may result in calls to the publisher's in-house ad server, to individual third-party ad servers, and to an advertising network. Server activity at each of these layers can be aggregated and analyzed. More importantly, both content and advertising servers at each layer may introduce a "cookie" to identify the reader's computer in the future as he or she returns to the current site, or visits other sites in the same editorial or advertising network. As a result, a single user's actions may be simultaneously tracked by multiple categories of observers.

Over time, then, the number of parties who can produce meaningful information about online audiences and audience behavior has increased. The vocabulary of audience-related statistics has increased with it: To note only a few of the most basic metrics, publishers and advertisers today must be conversant in "page views," "click-throughs," "unique visitors," "usage intensity," "engagement time," and "interaction rates," in addition to the demographic and behavioral profiles of their audience.

The irony of expectations

This flood of data from different sources has resulted in a level of complexity that can be difficult to manage. The key selling point of digital media remains the ability to track consumer behavior: which pages were viewed, which banners were clicked, and when this viewing and clicking produced an "action" such as a request for information or an online purchase. In practice, though, making sense of the massive amounts of data collected is hard work.

Tom Heslin, senior vice president and executive editor of the *Providence Journal*, calls this the "irony of expectations": neither publishers nor advertisers have been able to keep up with the flood of data. "Our biggest challenge is to simplify solutions for our clients, even for national advertisers," he explains. "The development of metrics has far outstripped knowledge of ad buyers and sellers. There is a real disconnect between the technology and how it can be applied and used."

The main effect of the rising tide of information has been to increase uncertainty among advertisers as well as publishers, according to Rick Hirsch, multimedia editor of the *Miami Herald*. "Ironically it's still like

being a traditional editor making calls based on your gut instinct — you have more data, but it's conflicting,” he explains.

Hirsch says server data (analyzed via Omniture) gives him some idea of what share of his overall audience comes from Miami rather than from the Caribbean or Latino communities in other parts of the U.S. (This is based on visitors' IP addresses.) However, he has no way to match that data to particular stories. As a result, for instance, Hirsch can't confirm his suspicion that a core *Herald* audience consists of government employees working in Miami, which would argue for augmenting that beat. Likewise, he is unsure how much to invest in edgy video projects because he doesn't know whom they appeal to.

The Web-native news site *Talking Points Memo* offered a dramatic illustration of the abundance of information available today. TPM has been beta-testing a new server analytics package, Chartbeat, which offers a detailed real-time picture of the last 15 seconds of activity at the popular site. The software provides an instrument panel with a minute-by-minute picture of what articles people are reading, how far into each piece they read, which pieces they're commenting on, what readers are searching for on the site, who's linking to the site from elsewhere on the Web, and what people are saying about TPM on Twitter, among other data.

“I've been working on the Web for 15 years, but this blew my mind,” declares Kourosh Karimkhany, COO of TPM Media. “It was a real epiphany.” Karimkhany says that the real-time information is having a dramatic impact on editorial and design decisions, for instance by revealing exactly where readers drop off in each story (halfway down a page, there's almost no audience left) and by challenging expectations about which breaking stories deserve top billing. For instance, editors were surprised to see news of Al and Tipper Gore's divorce (even before a Portland masseuse came into the story) outperforming the political bombshell about General Stanley McChrystal's profile in *Rolling Stone*, and moved the divorce story into a more prominent spot.

Measurement companies themselves appreciate the systemic effect produced by the many kinds of audience data now available. Marketing copy from comScore concedes the point frankly:

“The frequent disparity between census-based site analytics data and panel-based audience measurement data has long been the Achille’s heel of digital media measurement. Because the two measurement techniques have different objectives, they employ different counting technologies, which often results in differing metrics that can cause confusion and uncertainty among publishers and advertisers.”⁹

Marc Johnson, CMO of Experian Hitwise, agrees that more information hasn’t always yielded greater clarity. “Digital media is much more complicated. There are many more things that are measurable. There are many more moving parts,” he says. “It’s not always agreed upon what is the most important thing to measure, and what those measurements mean or how they should be applied.”

This complexity does not appear likely to abate in the near future. If anything, the variety of audience measures available seems to be increasing as sites and advertisers try to accommodate mobile devices such as smartphones and e-books.

Publishers finesse multiple metrics

The result of this abundance has been that unlike their counterparts in traditional media, publishers working online routinely subscribe to both panel-based and census-based measurement services — that is, to multiple, incompatible estimates of audience size. Each of the newspapers interviewed for this report (though notably not the online-only news site, TPM) subscribed to either Nielsen or comScore, or to both, while also relying on a server-side Web analytics package, usually Omniture. Most also incorporate at least one additional source of audience data, such as Scarborough, Hitwise, Google Analytics, or Alexa, into their internal analysis and their pitches to advertisers.

However, impressions of the relative merits of these data sources vary widely. Some publishers find server-side data much more reliable. The *Providence Journal* subscribes to comScore, but sees hard-to-credit fluctuations in its online audience from quarter to quarter. As a result, the paper relies on audits of its server traffic, collected via Omniture, to come

⁹ http://comscore.com/About_comScore/Methodology/Unified_Digital_Measurement.

up with its official online readership. The *Journal* relies on comScore data mainly for “product development” — to gauge the success of niche sites among particular demographic targets.

Similarly, the *Miami Herald* uses comScore and also used to subscribe to Nielsen. But Hirsch reports that his paper’s position in either of the panel-based rankings varies for no apparent reason. “I don’t know when to believe them,” he says. Meanwhile traffic recorded by the *Herald*’s own servers, analyzed with Omniture, tends to match his own editorial sense of when certain stories, or entire editions, are commanding a great deal of attention in Miami.

As an example, Hirsch points to January of 2010, the month of the Haitian earthquake that claimed an estimated 230,000 lives. “We know our traffic went through the roof, because of our history of coverage in the region,” Hirsch says. The paper’s internal figures matched expectations: as recorded by Omniture, traffic spiked 36 percent over December, to 35 million pageviews, while unique visitors jumped 11 percent, to almost 6 million people. Meanwhile, though, comScore recorded less than half as much traffic for January, and fewer than an third as many unique visitors. Unaccountably, comScore had the *Herald*’s pageviews dropping by 40 percent the month of the earthquake — and falling again in February, despite the fact that Miami hosted the Super Bowl that month.

At larger national papers, the story is somewhat different. The *Washington Post* relies on both comScore and Nielsen data to understand how it fares against major competitors online, while also using server-side traffic figures for internal strategic analysis. Managing editor Raju Narisetti acknowledges that the cacophony of competing measurements has been a serious issue, with both the panels undercounting the *Post*’s audience.

“However, over time you can recognize the strengths and weaknesses of each and start to understand how one approach relates to the other,” he explains. “It is less of an issue now” — in part because of the hybrid measures in the works from comScore and Nielsen, which bring these audience estimates closer to internal data.

The *Wall Street Journal* subscribes to Nielsen, comScore, and Hitwise, in addition to using Omniture for server-side analysis. Kate Downey, the paper’s director of “audience analytics & insights,” observes that the

Nielsen and comScore ratings of wsj.com rarely agree with each other, or with the Journal's own records. However, she emphasizes that server data is also unreliable and prone to double-counting; to make their case to advertisers, salespeople rely mainly on demographic data from the panels and on the *Journal's* own registration records (all the more valuable since much of the site is behind a paywall). She appreciates having multiple sources of data at her disposal, each with its own strengths and weaknesses, and suspects that many of her peers at other papers agree.

"People use whatever numbers look good that month. It gives publishers some flexibility," Downey explains. "I think if everybody had the same numbers, we would hate that even more."

Talking Points Memo sees the same dramatic divergence in audience estimates. Google Analytics counted 1.8 million unique visitors for a recent 30-day span, while comScore typically gives it in the neighborhood of 300,000 visitors per month. But unlike its peers in the newspaper business, TPM's response is to ignore the panels outright — the site subscribes to neither comScore nor Nielsen, counting on advertisers and agencies to supply panel-based figures if they consider them necessary to the conversation. (For demographic data, TPM relies on its own, voluntary audience surveys; every six months or so founder Josh Marshall issues an appeal to readers, culling about 1,000 responses within 12 hours.)

"The panel-based numbers are atrocious," says Karimkhany flatly, pointing out that most of TPM's traffic comes from the workplace, which the panels don't capture well. "But as long as they're equally inaccurate for our competitors, it's okay. It's something we live with."

The controversy over "unique visitors"

Within the measurement industry, this overabundance of information works itself out in periodic disputes over data — disputes over what information is most important, and over best how to define or collect it. The fault line that surfaces most frequently is between panel-based and server-side measures. The current agitation for new standards (detailed below) springs in part from a very public disagreement in 2009 over what might be fairly called the atomic particle of online audience measurement, the "unique visitor."

For the first decade of online advertising, the total number of unique visitors to a site was usually defined as the count of unique “cookies,” de-duplicated over the period of analysis. This had become the *de facto* standard since most sites don’t require a log-in or authentication. But, it is a “technology-based” rather than a “people-based” standard. A single user visiting from multiple computers (or deleting cookies from his or her browser) will inflate the count; multiple users sharing a computer will produce undercounts.

In 2006 the Web Analytics Association, representing mainly server-side measurement firms, published a definition of unique visitors that added the option to use “authenticated users” when available. The precise meaning of this new standard was unclear; according to a recent article in *Mediapost*, “the goal of the standard was to educate the Web analyst to the most commonly used definition and to encourage vendors to openly document any variances from the standard, given that data collection and processing techniques may vary from vendor to vendor.”¹⁰

In 2009 another trade group, the Interactive Advertising Bureau, published a competing definition of “unique users” aimed mainly at panel-based measurement services such as Nielsen and comScore, but specifying that census-based tools such as Google Analytics should conform as well. The new guidelines require that the measurer “utilize in its identification and attribution processes underlying data that is, at least in a reasonable proportion, attributed directly to a person.”¹¹

The definition touched off a heated debate and drew heavy criticism for being overly vague. The IAB’s standard invited a new set of questions: Will sites be required to collect personally identifiable information? What would the privacy implications be for sites adopting this definition? Is this new guideline ultimately even applicable to Web analytics firms, or can it only be met by audience measurement companies?

As a result, what had been a fairly straightforward metric — if one whose relevance was sometimes questioned — now has multiple definitions,

¹⁰ J. McDermott, “Unique Visitors Come In Two Shapes And Sizes,” *Metrics Insider*, 20 March 2009, at http://www.mediapost.com/publications/?fa=Articles.showArticle&art_id=102512.

¹¹ Interactive Advertising Bureau, “Audience Reach Measurement Guidelines,” 23 February 2009.

used in multiple ways, by multiple firms. The episode suggests that online measurement is hamstrung not only by the abundance of data available, but also by the inevitable contentiousness of even well-intentioned efforts to define standards in a developed industry. The IAB and WAA have been working together to approve (though not to adopt) each other's definition of unique visitors, but they have yet to reach a consensus.

Well-known methodological weaknesses

Despite such disputes, the strength and weaknesses of various approaches to audience measurement are well known and widely agreed upon. This is especially true in the comparison between panel- and census-based metrics.

Assuming the panel is well-built, panel-based measurement offers two key advantages over server-side data. First, a panel permits demographic analysis, allowing a national news outlet to determine, for instance, its penetration among men in a certain age or income group. Second, panel-based research facilitates comparisons across competing sites and over time — knowing whether an outlet is improving its position vis-à-vis the competition.

Nielsen provided basic details about its methodology for this report. The company's Internet panel consists of about 200,000 people recruited online through various partners; panelists agree to install metering software that tracks their online activity on a click-by-click basis. To correct for biases among panelists recruited online, Nielsen uses a "calibration sample" culled via traditional offline techniques such as "random digit dialing" (some of these users are drawn from the 18,000 households in its TV panel). Nielsen is able to report on panelists Internet usage on a monthly, weekly and daily basis.

The validity of panel-based measures depends on how faithfully they reflect the larger public. Tracking software installed on member's computers has difficulty distinguishing between multiple users, and far more important, it misses what its members do on other computers — especially at school or in the workplace. As a result, sites that target professionals during business hours, such as the *Wall Street Journal* and *Talking Points Memo*, believe they are underreported since workplaces are reluctant to participate in the panels.

This basic flaw is widely recognized. “The over-reliance on panels whose members accept tracking software on computers has been seen as an acceptable way of measuring audience size,” explains the *Washington Post*’s Raju Narisetti. “However, this approach is non-random and violates the standards required to project to larger audiences with any degree of certainty. Further, there has been no effort made to determine alternative means of measuring usage for people who do not accept the software (such as government employees, companies with privacy policies, etc.).”

Meanwhile, having two major firms offering panel-based ratings exposes methodological inconsistencies and makes it much more difficult to use the ratings as a benchmark. Nielsen and comScore frequently disagree about even basic measurements — such as who is the No. 2 media company online, after Google. Per comScore, Yahoo is the runner-up, with 167 million unique visitors in May 2010. But that month Nielsen had Microsoft in second place and Yahoo in third. Nielsen counted just 133 million visitors to Yahoo properties — a difference of 34 million people, about the population of Canada.

Major Vendors Define Genres Differently: Will the Real Top 20 Please Stand Up?

Top 20 News & Info Sites (Nielsen)*

1	Wikipedia	58,617
2	Weather Channel Network	41,544
3	Yahoo! News	39,917
4	About.com	37,373
5	CNN Digital Network	36,343
6	MSNBC Digital Network	32,236
7	Yahoo! Answers	28,529
8	AOL News	24,811
9	eHow	18,778
10	WikiAnswers	18,460
11	Fox News Digital Network	16,573
12	SuperPages.com	15,850
13	Tribune Newspapers	15,844
14	WhitePages.com Network	15,646
15	NYTimes.com	14,849
16	SBC Yellow Pages	14,669
17	Google News	13,541
18	ABCNEWS Digital Network	12,084
19	Gannett Newspapers and Newspaper Division	11,842
20	McClatchy Newspaper Network	10,749

*Unique Visitors (000), Dec. 2009

Top 20 News/Info Sites (comScore)*

1	New York Times Digital	52,771
2	Weather Channel, The	48,084
3	Yahoo! News	46,287
4	CNN	40,092
5	MSNBC	30,021
6	AOL News	26,646
7	Gannett Sites	23,560
8	Weatherbug Property	22,799
9	WorldNow Sites	20,684
10	The Washington Post Company	18,996
11	Tribune Newspapers	13,608
12	ABCNEWS DIGITAL	11,019
13	FOXNEWS.COM	10,094
14	Advance Internet	9,976
15	Huffington Post	9,956
16	McClatchy Corporation	8,335
17	CBS NEWS Digital	8,142
18	Hearst Newspapers	8,141
19	Examiner.com Sites	7,657
20	BBC	7,376



Source: Interactive Advertising Bureau analysis

The Interactive Advertising Bureau has drawn attention to differences in the site rankings released by Nielsen and comScore, for instance in a slide on traffic in the “news and information” category, reproduced above. Some confusion results from way sites are grouped: comScore rolls up properties like Nytimes.com and About.com into “New York Times Digital,” while Nielsen counts them separately. But even apples-to-apples comparisons can differ widely. In May 2010, all Gannett-owned properties commanded 37.5 million unique visitors according to ComScore, but just 25.6 million according to Nielsen. The same month ComScore gave washingtonpost.com an audience of 17 million people, but Nielsen recorded fewer than 10 million.

The advantages of server-side measurement are similarly straightforward: Web analytics can claim to capture *everything* that happens at a given publisher’s site, in census-like fashion, with no need for dubious extrapolations. In addition, server-based measurement offers a level of behavioral detail panels cannot hope to match, following how individual

readers make their way through an online publication, how much time they spend on each article, and so on.

TPM's Karimkhany argues that statistical panels are an anachronism when concrete and comprehensive traffic data is available from Google Analytics and its ilk. "I have real concerns about Google's market power," he says. "But there is no reason not to trust Google Analytics. It's a rock-solid product, everybody uses it, and it's very difficult to game."

The core technical flaw in server-side measurement has been noted earlier: It tracks machines (or actually "cookies") rather than people, and so is highly vulnerable to miscounts when human either share computers, or use multiple computers and browsers. People who delete cookies may be counted more than once. Another major challenge is eliminating non-human visits to a site, especially from "bots" or "spiders" which search engines use to crawl the Web.

As important as these technical issues is the fact that advertisers and ad agencies tend to disregard server-side data. A study prepared by Bain & Co. for the Interactive Advertising Bureau in 2009 emphasized brand advertisers' dissatisfaction with the server-side metrics publishers generally make available, such as "page views," "time spent on page," and even "unique visitors."¹²

For this reason, the NAA's Randy Bennett suggests that publishers have made too much of the discrepancy between server-side and panel-based measurement. "They tend to focus on the discrepancy between metrics rather than on what advertisers want," says Bennett. "In the end it doesn't matter what publishers want. It only matters what advertisers want. And there's no standard around that."

Finally, advertisers as well as media providers routinely bemoan the inability to measure audiences in a consistent way across multiple platforms — television, the internet, mobile phones, etc. In 2005, a report by the Advertising Research Foundation found that "many respondents included comments about the lack of multi-media comparability and difficulties that they experience in integrating data from the measurement

¹² Bain & Co., *Building Brands Online: An Interactive Advertising Action Plan*, November 2009.

of various media for which they provide integrated planning support.”¹³ And in 2009, a new “Coalition for Innovative Media Measurement” — led by giants such as CBS, NBC, and Disney — was formed to establish a new standard to gauge total media usage across broadcast and the Internet.

The promise of hybrids

The online media measurement landscape evolves quickly. While comScore and Nielsen remain the two top panel-based services, others are trying to encroach. Methodological differences have eroded between the two panels and competitors who derive their data partly from ISPs, such as Hitwise, Compete and Quantcast. According to Hitwise’s Johnson, “Everyone is trying to grow their business by offering a full suite of data to marketers. More and more we overlap in each other’s areas.”

Meanwhile, both Nielsen and comScore are adopting a “hybrid model” which combines their panel research with server-side data collected from clients. comScore’s Media Metrix 360 is the first such offering, a “panel-centric hybrid” that combines the company’s two million person global panel with server-side analysis. The goal is to deliver a unified count that reconciles discrepancies between panel and server data, as well as to provide more granular detail on Web-site usage.

Nielsen’s version has not been officially unveiled, but interviews with the firms and their clients indicate that both hybrids work similarly: Client sites embed “beacons” on their content servers that allow Nielsen or comScore to track visits from users who aren’t members of their panels. (comScore’s beacon has been integrated directly with Omniture’s popular Web analytics software.)

How are these conflicting data sources reconciled? Per comScore’s site, the firm “has developed a proprietary methodology to combine panel and server-side metrics in order to calculate audience reach in a manner that is not affected by variables such as cookie deletion and cookie blocking/rejection.” Or as a Nielsen analyst explained, “The essence of what we’re doing is creating ‘person-centric’ audience measurement data using the strengths of panel-based measurement (quality demographics) and server-side data (census-level tracking of content).”

¹³ Cited in *The Future of TV*.

Whatever its technical merits, the immediate effect of the hybrid approach has been to increase audience figures for many sites, pushing panel-based figures closer to publishers' own internal estimates. According to one comparison¹⁴ based on comScore's December 2009 data, unique visitor counts went up an average of 30 percent under the hybrid approach; some sites — *The Onion* is one — saw traffic nearly triple. (Not every site has been so lucky, however; according to a recent *New York Times* article, a methodological tweak at comScore slashed Hulu's traffic by 45 percent in June.¹⁵)

The boost has been especially dramatic for newspapers, according to comScore's Josh Chasin. (A discrepancy of 75 percent had not been uncommon for news outlets, due most likely to their high at-work traffic.) Both the *New York Times* and the *Providence Journal* report that the hybrid figures better reflect their own audience estimates. The increase has been substantial: comScore's audience estimate for *Times* properties jumped to 72 million in May 2010, up from 53 million in December 2009, before the new methodology was implemented (and up from 47 million in May 2009).

While hybrid measurement promises more reliable audience estimates, though, it is not clear that it will result in a single audience standard online. The new methodology adds another layer of complexity, and its implementation has been piecemeal: Sites that do not download and install beacon software on their servers cannot be measured with the new hybrid formula, and therefore should not be compared directly to sites that do participate — even though the measurement firms purport to rank *all* sites in the markets they cover.

This landscape is further complicated by the higher level of access afforded to paying clients. After initially limiting the service to its customers, comScore now allows any site to install a beacon, for free; it is not clear whether Nielsen will follow suit. According to Jon Gibs, vice president of analytics at Nielsen, "We can't do things just for the good of the industry. If there's no one paying for the service, it doesn't make sense to do it."

¹⁴ Analysis by campaign management firm XA.net; see <http://www.cpmadvisors.com/2010/02/06/comscore-hybrid-methodology-leads-to-large-traffic-increases/>

¹⁵ A.R. Sorkin & M.J. de la Merced, "Hulu Is Said to Be Ready for an I.P.O.," *New York Times*, 8 August 2010.

Reform in the air

The controversy over defining unique visitors — and in general over the very different pictures of the online world painted by various measurement firms — has provoked calls for an organized, industry-wide reform. The state of affairs was captured well in a column by Steve Smith, digital media editor at the *Media Industry Newsletter*:

“We are a decade and a half into the life of the ‘most accountable medium,’ the Web, and just this week we see some of the major online measurement firms still tweaking their models and arguing over methodology. I have major media companies reporting their monthly numbers to me, and I see staggering differences between the stats they claim from their internal logs via Google Analytics or Omniture and the third-party numbers. It’s not just mobile, either. It’s still a mess all over.”¹⁶

In a sign of the times, the Internet Advertising Bureau — a trade group “dedicated to the growth of the interactive advertising marketplace” — made Smith’s condemnation the opening slide of a presentation in early February to the Association of National Advertisers. (The presentation went on to highlight the wide differences in the monthly rankings released by Nielsen and comScore, reproduced above.)

IAB appears to be leading the reform charge, at least in the world of panel-based measurement. A confidential McKinsey & Co. study commissioned by the IAB and the American Association of Advertising Agencies concluded in 2009 that confusion over metrics stood in the way of greater online ad spending. Based on this report, the IAB has proposed a “cross-industry task force of senior marketers, agency executives and media leaders to reach consensus on standardizing and simplifying basic audience measurement.” The Newspaper Association of America has committed to support this task force and to recruit participants from the newspaper industry.

A parallel standardization effort is underway from the Media Ratings Council, a group comprising media companies, advertisers, and agencies which dates from the 1960s and whose mission is to accredit and audit

¹⁶ S. Smith, “Taking Mobile’s Measure, Again... And Again,” *Mobile Insider*, 26 January 2010.

ratings firms. The MRC has worked with the IAB to coordinate definitions (for instance of IAB's controversial 2009 guidelines for "unique visitors"). However the MRC is also in the process of accrediting comScore and Nielsen, an effort likely to continue into next year. It is conceivable that these ongoing audits will make the two panel-based measures more compatible, and perhaps more transparent to publishers and advertisers using them.

In the world of server-side measurement, any number of organizations offer third-party auditing of traffic data from Omniture and other Web analytics tools. Both leading circulation auditors, ABC and BPA, operate interactive arms that produce verified online readership figures. ABC has been particularly active here: Its "Audience-FAX" service, developed in partnership with the NAA and Scarborough, purports to measure a newspaper's net readership across print and the Internet. However, this service depends on newspapers to submit readership data in various categories; the competitive picture may be skewed if newspapers calculate these differently.

The emergence of new platforms and devices also poses a growing problem. For instance, as yet no consensus exists around how to measure streaming media and online video. How is an impression defined when content is continuously streamed? When it is short-form versus long form? When it plays in the background, as most online radio does today? How these terms are defined is no small matter – streaming music service Pandora today stops playback if a listener doesn't engage with their page in a certain amount of time. Each time a user has to click back to that page to hit "play," a new session is initiated, thereby boosting Pandora's traffic numbers.

Will an audience measurement currency take hold online?

A number of countervailing forces appear to be at work in the online media measurement today: on one hand, explicit standard-setting efforts and the emergence of "hybrid" audience measures, and on the other, a growing diversity of measurement companies, media types, and technology platforms.

Two points bear consideration. The first is that the most successful media measurement currencies have emerged not from industry task forces, but

through market power, which is to say research monopolies. The single development which would do the most to clarify audience measurement standards would be a union of comScore and Nielsen NetRatings — an event which, given the high costs of maintaining competing panels and the obvious benefit of eliminating embarrassing discrepancies, is not out of the question.

Combined with the assimilation of server-side data via “hybrid” approaches, such a union would establish a single, industry-wide standard for comparing online audiences. Alternatively, a server-side measurement may take hold as an audience standard if (as TPM’s Karimkhany suggests) the industry is gradually coming to see panel measures as obsolete. In this scenario, the most likely candidate for a standard is Google Analytics, which is much cheaper and far more widespread than alternatives like Omniture, especially among blogs and smaller, independent Web sites. (Google also has the technical advantage of being able to calibrate online measurements using its own vast audience and huge advertising network.)

The second point to consider is that even if a consensus emerges by either of these routes, the resulting standard will not automatically be a media “currency” in the way of Arbitron’s radio rankings or Nielsen’s TV ratings. Agreeing on a single estimate (however imperfect) of the number of people who read the *Times* online last month, or on whether the *Times* or the *Post* did better among high-income women, will not stanch the flow of information about what Web users are doing and what they care about. It will not prevent either of those papers from touting other statistics which strengthen their case. And most important, it will not necessarily satisfy the needs of advertisers, who no longer have to plan their campaigns — nor pay for them — on the basis of static readership profiles. The following section will investigate how media metrics are used in planning and executing advertising campaigns, and how this affects publishers.

III. Measuring Online Media: A New Planning Paradigm

Third-party research plays a critical role in the offline media ecosystem. As a measurement currency, information from monopoly ratings firms such as Nielsen and Arbitron guides media planning, governs media pricing, and is even used to assess the success of ad campaigns. However, audience measurement plays a substantially diminished role in advertising on the Internet. Two shifts help to account for the evolving role of media measurement online, and are explored below: the emergence of performance-based pricing models, and the increasing reliance on new kinds of audience targeting.

An online Gestalt switch

The role of media measurement in online advertising reflects the basic shift in the way media space is bought and sold on the Internet. One way to appreciate this Gestalt switch is to consider what “inventory” means online and off. In broadcast or print advertising, inventory is a scarce resource — no matter its circulation, a newspaper has a finite number of ad pages in each edition at a reasonable ad-edit ratio. This scarcity is even more pronounced in broadcast; hence the practice of reach “guarantees” promised to advertisers on the basis of past Nielsen ratings, and adjusted after the fact (via “makegoods” or “overdelivery”) once Nielsen results are in for a given campaign.

Online inventory cannot be a scarce resource in the same way, since it is generated on the fly by each decision to view a page. In theory, it should be unnecessary to speak of audience guarantees at all on the Internet — an ad banner or pop-up can simply be shown until the purchased number of impressions has been reached. (In practice the most desirable online property is often sold as sponsorships, not impressions; and even impression-based campaigns will prefer outlets with a large enough audience to deliver the desired audience within a certain time frame.)

This shift is well-understood, of course, but what it means for media measurement has not always been appreciated. Advertisers purchasing space or time in traditional media are paying, in a very immediate way, for a set of audience numbers delivered by a trusted third party. Based on Nielsen or Arbitron figures, applying the standard formula of “reach and frequency,” a company like Proctor & Gamble can calculate (if with

disputed accuracy) what it will cost, say, to make sure 40 percent of TV viewers or radio listeners in a certain market hear a Duracell jingle an average of three times each.

Online such calculations are both less possible and, to many advertisers, less relevant. Neither user panels nor server-side analytics can realistically claim to gauge the “reach and frequency” of an ad campaign across multiple sites and ad networks. Ad impressions have been decoupled from media audiences. People reading the same article online will not necessarily see the same ad banner, making the link between a media property’s reach and that of its advertisers much more tenuous. More to the point, advertisers no longer *need* “reach and frequency” to plan campaigns or purchase media; they no longer need a measurement currency in the same way.

Marc Frons at the *New York Times* makes this point succinctly, discussing the disagreement between different rankings of top online news outlets. “I think it’s less important online because advertisers can see how well the ad is performing on their end,” Frons says. “So the Nielsen number and the comScore number are just bragging rights for publishers. They matter less.”

Pricing online ads: impressions versus performance

Online as in traditional media, the basic unit for pricing advertisements remains the CPM, or cost per thousand. Its persistence has defied predictions that performance-based pricing schemes would sweep aside the old-media habit of selling exposure.

However, the online CPM differs from its offline cousin. In broadcast, CPM is based on households or viewers, and in print on audited circulation; thus an advertiser can roughly compare the cost of reaching 1,000 people via a TV spot and a magazine spread. Online, though, CPM refers to impressions rather than viewers or readers, making cross-media comparisons difficult. One session at a news outlet online may generate a dozen impressions as the reader clicks around from story to story.

The most sought-after inventory online is usually sold on a CPM basis. However, several other pricing models exist, including CPC, or cost per click; CPL, or cost per lead (usually determined by a user registering

for a newsletter, account, etc.); or CPA, meaning cost per action or cost per acquisition (for users who convert to customers). Less desirable inventory is often sold on the basis of performance, so the advertiser only pays for the desired result.

These distinctions may not be as defining as they are made out to be. An advertiser who buys media on a performance basis knows the total number of impressions delivered and can easily calculate CPM, for instance in order to compare two sites on an apples-to-apples basis. The reverse is true as well — an advertiser who buys on a CPM basis also has records of click-throughs and purchases and so can derive the various performance measures. And again, all of these approaches differ fundamentally from offline cost-per-thousand deals in that on the Internet, the thousand impressions (or clicks or actions) are recorded one by one, not based on audience estimates.

Two tiers of online inventory

From a publisher's perspective, online pricing schemes creates a caste system based on the distinction, inherited from offline media, between premium and remnant inventory. Online the two categories are not formally defined, but still fairly clear:

- **Premium inventory** sells for a relatively high rate; it is usually sold by the media owner, rather than a third party like an ad network; it is more often the province of “brand advertisers”; and it typically sells on a CPM basis or as a sponsorship package.
- **Remnant inventory** sells at a low rate; deals are usually transacted through an ad network or aggregator, often on a performance basis (CPC or CPA); and buyers, who hail from the direct-response end of the spectrum, may have little idea where their ads end up running.

Talking Points Memo offers a good illustration. The site sells roughly one-third of its inventory direct, one-third as remnant, and the final third through Google's AdSense, though the ratios fluctuate from month to month. TPM's Karimkhany suggests that premium inventory might command a CPM of roughly \$10. Unsold or less desirable inventory is offloaded through remnant optimizers for a much lower price, ranging

from perhaps 40 cents to \$2. Meanwhile inventory sold through Google's AdSense network varies even more unpredictably depending on advertiser demand for a given keyword, from an effective CPM of about \$2 to as much as \$20 in extreme cases.

"All this is very dynamic," Karimkhany explains. "Sometimes we sell a lot of direct ads, such as when BP needed to get its message out and environmental groups wanted to give the counter-message. This crowds out AdSense and remnant. Sometimes AdSense goes crazy, like during elections when campaigns buy keywords. And sometimes, like January and February and summer vacation months, remnant predominates because there's very little active purchasing."

It is entirely possible for an advertiser to buy inventory on a single site that is both premium and remnant, if the advertiser is negotiating directly with the site and also working through an ad network. However, top-tier sites often have a policy against selling any inventory as remnant.

This is the case at both the *New York Times* and the *Wall Street Journal*, which sell most online inventory on a CPM or sponsorship basis and do not participate in ad networks (other than Google's AdSense, which the *Times* uses). "We sell brand, not click-through," declares the *Journal's* Kate Downey flatly. "We're selling our audience, not page counts."

Marc Frons echoes the sentiment, pointing out that the *Times* can afford to take the high road. "For us as the *New York Times*, brand is important," he says. "You really want to make the Internet a brand medium. To the extent CPC wins, that's a bad thing."

Other newspapers take a hybrid approach. The *Providence Journal* negotiates cross-platform, multimedia packages whenever possible — for instance, combining a quarter-page print ad with a certain number of banner impressions, a search term, and (via a partnership with Yahoo!) a behavioral targeting profile. These bundles arguably help to resist any erosion of "projo.com" into a second-class, performance-based ghetto. However because traffic is hard to predict and can vary greatly from month to month, the paper unloads "oversupply" as remnant inventory through Yahoo's ad network.

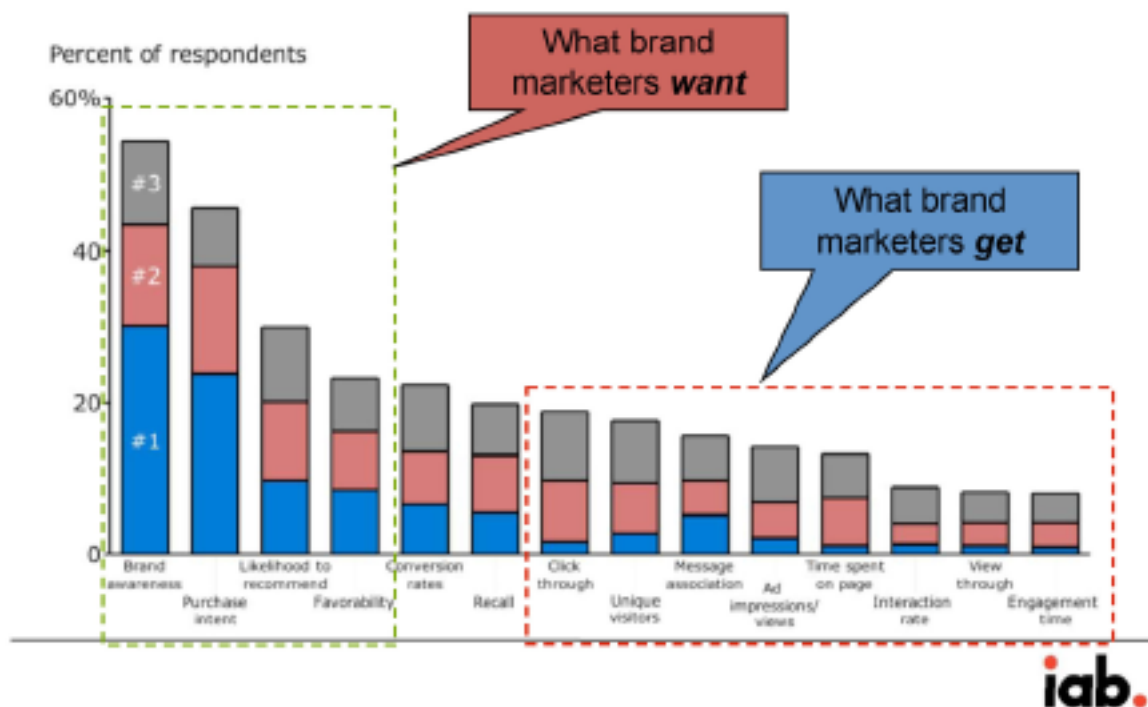
A new dynamic in media planning

A basic reason syndicated research plays such a critical role in broadcast and print media is that most advertisers have no good way to judge the effectiveness of their campaigns. Media planners thus frontload most of their analytical time and resources, using demographic data — imperfect as it is — to plot and plan campaigns before they run. To a great extent in traditional brand advertising, once a campaign launched the media planner's work was done, until the post-mortem.

That media-planning dynamic is inverted online. Advertising agencies as well as in-house marketers tend to allocate more resources to optimizing campaigns *as they run* than to planning them beforehand. Much has been made of the cultural and philosophical shift this required at major agencies, which have had to import the thinking of the much less fashionable world of direct marketing.

Brand marketers want actionable, traditional brand-impact metrics online...

Which metrics are most valuable for brand-building campaigns?



Source: IAB and Bain & Co., *Building Brands Online: An Interactive Advertising Action Plan*, Nov. 2009

It is still true that direct-response advertisers are more likely to use performance-based pricing online, while brand advertisers gravitate towards CPM. But there are no hard and fast rules, and even brand advertisers have an irresistible stream of data about how their campaigns are doing. Taken to a theoretical extreme, an advertiser concerned only with results would have no need for up-front demographic data at all, or indeed for media planning as it has traditionally been understood. When the only metric that matters is sales, who sees the ad or how often becomes irrelevant.

Ratings define the “consideration set”

Because of this inverted planning dynamic for online advertising, the role of syndicated research services like comScore and Nielsen is greatly diminished. Rather than defining a currency, these panel measurement companies act as a sieve, helping planners to come up with a list of sites and ad networks to potentially include in a campaign — the “consideration set.”

As a result, a site like the *Wall Street Journal* pays close attention to its ranking and especially to its demographic profile in the syndicated research services. “We live and die by the demographics,” says Downey. Once a potential advertiser has the *Journal* on its list, however, salespeople can make their case using other research sources, including the newspaper’s own records. This often means “educating” potential advertisers about the under-representation of the *Journal*’s valuable at-work audience on the major research panels.

While a good comScore or Nielsen profile may make a site attractive to a brand advertiser, the audience data are not used to set the terms of an ad buy. How a deal is structured will depend more than anything (and more than in offline media) on the specific objectives of the advertiser. But even in a straightforward CPM deal, media performance matters. If a site does not deliver desired impressions (or clicks, or conversions) quickly enough, it runs the risk of being cut from the media schedule.

In fact, cutting venues out has become routine campaign strategy. “You want to cast a fairly wide net initially, then refine it as the campaign runs,” explains Rudy Grahn, director of analytics at media agency Zenith Optimedia. Advertisers have other tactics at their disposal, including

renegotiating the price or the volume of impressions, but a site that consistently under-delivers will find itself eliminated from a campaign.

The threat from targeting

If one promised “revolution” in advertising online has been the ability to pay only for ads that perform, the other has been the ability to aim marketing messages at certain individuals with a precision not possible in the offline world. In fact three dominant models of targeting exist on the Internet, two of which have direct analogues in traditional media:

- **Demographic targeting** uses characteristics such as age, gender, income, and education to define a desirable audience. Most traditional advertising is targeted by demographic (and occasionally “psychographic”) profile at least in the selection of venues, and the same logic applies in well-known destinations online. If BMW advertises on the *Wall Street Journal*’s site, at least one reason is its audience of high-income, educated professionals.
- **Contextual targeting** based on the editorial focus of the venue also routinely figures in media planning both online and off. If BMW buys impressions on Kelley Blue Book or Edmunds.com, it hopes to capture potential car buyers already in the decision-making process.
- **Behavioral targeting** has no traditional precursor; it relies on cookies to sort Internet users based on what sites they’ve visited, what search terms they’ve used and what actions they’ve taken. If BMW wants to target people already in the purchasing “funnel” it can buy profiles of car-seeking consumers from a “profile exchange” such as BlueKai. Such a profile might be based on visits to an automotive comparison site, on a search for “Mercedes dealers in New York,” or on pre-applying for a car loan, for instance.

Some high-traffic publishers are able to deploy sophisticated targeting technologies in-house. Dow Jones uses all three approaches described above to serve ads across its network of business-and-finance sites. Two *Wall Street Journal* subscribers visiting wsj.com — say, a Cleveland-based

small business owner and a Manhattan financier — may see different ads based on anonymous profiles derived from their registration data. (The profiles are encoded in cookies attached to their browsers.) Or, someone who frequently visits the “technology” pages of *wsj.com* may be assigned to a profile that draws tech-related ads, even as the same user travels to *marketwatch.com* or *barrons.com*.

However, most behavioral targeting occurs through an ad network like AudienceScience or a profile exchange such as BlueKai. For behavioral targeting to be predictive, it generally requires wider inputs than a single site can provide; the goal is to assign users to particular profiles (i.e., “golf fanatic”) based on their behavior across a wide swath of the Internet, and then to be able to reach those users wherever they go. BlueKai calls this “audience portability,” and makes no bones about the disintermediating effect on publishers, or “inventory”:

“Today, media buying is constrained to only buying data that is tied to a particular inventory. ... At BlueKai, we understand that's not the way things should be. We know that online data should be separated from the media — and made customizable and accessible to everyone, at scale, anywhere on the Internet.”¹⁷

As a result, the rise of behavioral targeting represents a distinct threat to publishers: By discriminating among users individually, behavioral targeting diminishes the importance of a site’s overall brand and audience profile. Suddenly the decisive information resides not with the publisher but in the databases of intermediaries such as advertising networks or profile brokers. A similar threat may be emerging in the domain of demographic targeting. As it becomes more possible to attach permanent demographic profiles to individual users as they travel the Web, the selection of outlets will matter less in running a campaign.

This is why online media outlets tend not to participate in third-party ad networks if they can avoid it. “We don't want to be in a situation where someone can say, ‘I can get you somebody who reads the *Wall Street Journal* while they're on another site that costs half as much,’” explains Kate Downey.

¹⁷ See <http://www.bluekai.com/free-the-data.html>.

In fact, Dow Jones is looking at ways to not just resist the trend, but reverse it, by pulling outside behavioral or demographic profiles into its own ad servers. This would let the *Journal* and its sister sites target ads based on what their visitors — even unregistered ones — have done elsewhere on the Internet. “The goal is to be able to give a better ad experience and user experience to even the anonymous people coming to our site,” Downer says. “That would be fantastic.”

Publishers may also try to control contextual targeting by packaging their inventory as “content channels” designed to respond to advertiser priorities. For instance, when automotive inventory was at a premium, many sites rushed to create new sections to draw advertising from car manufacturers, in some cases pushing beyond their normal editorial range; a good example is Forbes’ Luxury Car channel online.

Publishers need competitive intelligence, local data

The advent both of performance models and of individual targeting makes third-party audience metrics (at least as traditionally understood) somewhat less decisive for advertisers, and argues against the emergence of a single online currency on the order of Nielsen’s TV ratings.

However, third-party metrics remain relevant for both buyers and sellers of media online — especially for the publishers, who express a clear need for research to clarify their competitive position, make editorial decisions, and present a case to advertisers. “Data on the competitive landscape is missing, and it’s important for both editorial development and for ad sales,” declares the *Miami Herald*’s Hirsch.

The two syndicated research services, Nielsen and comScore, appear to be the best option for this sort of competitive intelligence, despite publishers’ many complaints about the panel-based measures. This is in part because publishers are reluctant to advertise their server-side data, or to trust claims made by their competition. As the *New York Times* Marc Frons puts it, “Sharing our data with competitors is not something we’re eager to do.”

What newspaper publishers say they want, almost without exception, is more demographic granularity. The hybrid measures in development from Nielsen and comScore hold some promise in this regard. Publishers

who have experimented with the hybrids report that thus far they have a limited (if useful) objective, to provide more accurate audience counts. However it seems clear that the measurement firms aim to marry demographic data more closely to specific newspaper sections. Hirsch argues that this sort of information would be enormously useful:

“There are a lot of optional, high-cost, high-effort editorial projects a newspaper can choose to pursue. I wish I had the data to guide these editorial choices. Ironically it's still like being a traditional editor making calls based on your gut instinct — you have more data, but it's conflicting. Better data would make it easier not to resort to cheap tricks to spike traffic.”

However, especially for regional newspapers such as the *Herald* and the *Providence Journal*, this section-level demographic data has to be local to be really useful. The panel-based measurement tools typically don't have a large enough local sample outside of major metropolitan regions to be able to adequately report on traffic for local newspapers' web site. It is partly for this reason that both newspapers and sites like Talking Points Memo rely heavily on registration data and reader surveys to understand their audience.

The lack of local insight affects editorial decisions as well as a publisher's appeal to local advertisers. “Our philosophy is we want to sell some national stuff,” Hirsch continues. “But the advertisers who are traditionally important to us, like department stores, upscale restaurants — if I can tell them in a compelling way what our audience looks like, that's good. We need to show upscale professionals are our readers. We need to distinguish our audience.”

Finally, today's metrics landscape almost completely neglects the smaller, independent online news outlets so often credited with representing a crucial trajectory in the future of journalism. Even a major site like *Talking Points Memo* barely shows up in the comScore or Nielsen rankings; smaller, local journalism projects are off the radar completely.

Based on a series of summits with independent media producers in seven cities, Tracy Van Slyke of the Chicago-based Media Consortium asserts that these independent online journalists are very poorly served by available data sources. Most cannot afford comScore or Nielsen and are

too small to be captured by panel methodology. As a report produced on the basis of the “Impact Summits” noted,

Many summit attendees expressed frustration with the inconsistency of current social media analysis schemes. “Dashboards” — which combine and analyze a range of data points on one screen — are in wide use across the online media environment. For example, web traffic analysis tools such as Google Analytics or Mint allow webmasters to track the numbers of site visitors over time, page views, and time spent on a site. ... However, many public interest media projects are not only too small to show up on these larger comparative services, but are increasingly using social media sites like Facebook, Twitter and YouTube to distribute content, rather than centralizing their work on a single site.¹⁸

Like TPM, these independent outlets almost all rely on Google Analytics to gauge their own traffic from month to month. However, even more than audience estimates these news outlets want to understand and assess their impact in the wider journalistic sphere, and use any available tools to track inbound links or references by more established news outlets.

¹⁸ J. Clark & T. van Slyke, *Investing in Impact: Media Summits Reveal Pressing Needs, Tools for Evaluating Public Interest Media*, 12 May 2010.

IV. Conclusion & Recommendations: A Role for the Tow Center

The foregoing analysis of the world of Internet metrics points to several immediate conclusions:

First, the measurement landscape online is far more confused and uncertain than in traditional media, with no established “currency” metric for comparing audiences. This is generally understood to be a problem by both publishers and advertisers, though within each group some negotiate the confusion better than others.

Second, the lack of a single measurement standard is due mainly to two factors: a natural abundance of data about online audiences, and a diminished need for a currency metric to plan campaigns and close advertising deals. Put another way, the prevalence of a clear standard in television and radio has been a function of a relative dearth of information.

Third, the prospects for standardization remain uncertain despite various ongoing efforts to reconcile competing metrics. However, the adoption of “hybrid” methodology and the prospect of a merger between comScore and Nielsen suggest how a dominant audience measurement standard might emerge.

Fourth, such an audience-measurement standard, if it emerges, will not play the decisive role of traditional media currencies such as Nielsen’s TV ratings. The abundance of information and the diversity of advertising models strongly suggest that measuring and comparing online media will remain a complex endeavor involving many data sources.

Fifth, despite the clarion call for clear standards, what news outlets working online mostly wish for is more and better data: in particular, detailed information about their own local audience and about their direct competitors. What the best-placed online publishers possess is not clarity, but control — control of the detailed demographic and behavioral information which advertisers can seek from an ever-widening array of sources.

This initial analysis suggests three broad avenues for the Tow Center: First, educating journalists and journalism students to understand and

monitor the Internet measurement industry; second, cultivating data to track and assess business models in online news; and third, developing approaches to think about and measure *journalistic* success in the networked public sphere.

Broadening the scope

The challenge of Internet audience measurement is just one part of a larger shift in the media landscape, brought about by a rapid evolution of communications technology. These changes are ongoing — innovations in mobile devices, tablet computing, and other new platforms and content models may challenge the dominance of content delivered by Web browsers on personal computers. Furthermore, the growing importance of social media as the navigational frame that brings eyeballs to content will alter dynamics of exposure and impact in unforeseeable ways. Media industries will face an evolving challenge to their business models, and better solutions to today's problems may not have a long shelf life.

The news media face an indefinite period of uncertainty, with few anchors to hold old business models in place. Furthermore, advertising faces a challenge larger than just changes in the way people consume professionally produced media content. The Internet collapses the social and communicative distance between all parties: individuals, businesses, government, civil society, celebrities, subcultures, and knowledge elites, etc. For marketers, camping their signal onto professional media of any kind is no longer the obvious primary strategy. As Clay Shirky put it in his much-cited essay on the Internet's challenge to newspapers, "That the relationship between advertisers, publishers, and journalists has been ratified by a century of cultural practice doesn't make it any less accidental."

To paraphrase Doc Searls, the fact that all parties are now one degree away from each other changes everything. In fact, as this report has documented, improvements in online measurement may harm rather than help advertising as traditionally understood. As Searls says, "Advertising amounts to buying guesswork, and technology will reduce the guesswork in ways that will not be friendly to the business. The frontier of the media industry is not advertisement."

Whether Searls's claim will hold is far from certain, but the scope of the uncertainty itself is the point. Confused audience measurement is just part of a larger problem facing professional media, and journalism in particular. In contemplating a role for Columbia's new Tow Center, it is important to bring a wider landscape into the picture. How might an academic center make a meaningful contribution to the future of journalism, one that embraces advanced Internet data and metrics without being constrained by them?

A good way to start is by rethinking the question: How can academic research contribute to preserving and improving the practice of journalism as the institutional foundations of the profession shift unpredictably in the face of rapidly evolving technologies and consumer behavior? Analysis of Internet data can be enormously valuable, used in the service of examining the right questions.

A focus on accurate audience estimates and better behavior metrics draws the eye away from more basic questions of impact which capture the social value of journalism: Did this story spread? How has it made a difference? Does the work of reporters lead to more public accountability? Just what, actually, is the emerging role professional journalism in what Yochai Benkler calls the *networked public sphere*?¹⁹

The news profession and the society it serves need to know how journalistic outputs affect public affairs and political processes in a time when Internet data show so powerfully what survey research has always suggested: that the distribution of knowledge and attention to public affairs is extremely unequal. The current crisis in commercial media is an opportunity for the profession to reexamine old assumptions about the relationship between the press, the public, and democracy. It is not enough for journalists and editors to fall back on abstract moral claims about the profession's public role, while telling the business side of their organizations to get savvy about metrics for the sake of saving old business models.

Jay Rosen, professor of Journalism at New York University, says this about the collision of established forms of journalism with new kinds of

¹⁹ Y. Benkler, *The Wealth of Networks*, Yale University Press, 2006: chapter 7. (Accessible at http://www.benkler.org/Benkler_Wealth_Of_Networks_Chapter_7.pdf).

data: “The difficulty with Internet data is that the reaction to it is one of two extremes: journalists either ignore it completely, or become slaves to it. Neither is a rational position for 21st Century journalists.”

An opportunity exists for the Tow Center to help the profession move past this first-generation reaction of denial or submission, toward a more mature position of knowledge. Sophisticated empirical analysis can help understand the role of journalism in the evolving media ecosystem, in ways that better equip journalists to make a difference, get credit for the difference they are making, and find ways to get paid, whether through old or new models. Let us consider three approaches in turn.

1. Understanding measurement data

A clear role for the Tow Center echoed by academic experts as well as publishers lies in educating journalists to navigate the chaos of data about online audiences, and in particular about journalism on the Internet. The NAA’s Randy Bennett argues that the most critical contribution the Tow Center can make is to enhance journalists’ generally thin understanding of the measurement and ratings world.

Poorly imagined, of course, this reflex risks becoming a non-response. One obvious and immediate goal might be to support journalism which exposes the contentiousness of Internet metrics and monitors ongoing reforms — for instance, the methodologically opaque “hybrid” measures now being unveiled. The broader objective, however, must be to help journalists to use data to think critically about the evolving journalistic landscape itself.

“The question is how journalists can use the numbers to improve journalism by seeing how people use it,” Rosen explains, adding that the goal should be “to become data-savvy enough to find where short-term user interests and longer term (professional/societal) interests overlap.” When journalists use Internet data today, it tends to be in unsophisticated ways aimed at improving fairly simple metrics. The opportunity exists to marry data on traffic, citations, users, and text to generate sophisticated insight into how stories spread, agendas are set, and information on key topics is discovered and used by key audiences.

2. Measuring journalistic influence

In sharp contrast to the deluge of data about the size and shape of online audiences, there is an almost complete lack of thoughtful measures of how news stories travel and what effect they have. Journalists themselves care deeply about how their work fares in the so-called “link economy” — who is citing, linking to, and talking about the stories they produce. This requires closer attention to the journalistic ecosystem itself, to the relationships between and among different kinds of news outlets and journalists.

The Tow Center can play an important role in developing tools and approaches to measure journalistic influence in a rapidly changing news environment. Rosen suggests a key mission for research is to show how journalists’ work matters, and to help editorial culture maintain authority in the face of audience metrics aimed at replicating mass-media business logic. “When you don’t have great data, you use whatever you’ve got,” he says. “The need is for metrics that arise out of the mission and values of editorial workers, and can support alternative logics for better decisions.”

3. Assessing new business models for news

The academic experts we interviewed each suggested that research on emerging business models for professional journalism would make a valuable contribution. Academic research should, in Searls’ words, “study the economics of whatever future media might be.” The Tow Center should be positioned to make a contribution to journalism both where ad-supported media thrive, and where they do not.

Of course, access to intimate business data from publishers and news outlets is essential to develop this role in identifying, understanding, and teaching new business models for journalism. Based on initial queries, publishers do not reject out of hand the idea of sharing information on their investments — human, financial, and otherwise — in the Internet, and on the success of those investments. The history of similar “benchmarking” efforts in journalism (for instance, from the NAA) suggests that obtaining data is difficult and low response rates are the norm. Nevertheless, it is clear that news ventures are deeply interested in such comparative analysis and that an independent, university-based research center might be in a good position to aggregate data to provide it.

Conclusion: the opportunity

There is an opportunity for Columbia's Tow Center to lead by, first, asking challenging and relevant questions; second, initiating and participating in collaborative, multi-method studies that respond to those questions; and third, disseminating this knowledge within the community. Yochai Benkler emphasizes the importance of applying multiple research methods, in combinations appropriate for particular questions, and in dialogue with similar research organizations and efforts around the world.

The possibility of collecting rich datasets from publishers and other industry sources should be considered in the context of building this research capability. A danger is that sitting on a stockpile of data would, over time, bias the questions asked to fit the data available. But there is an enormous opportunity to use a high value archive of otherwise unavailable data as the centerpiece of a strategy to attract collaborators and research funding. Early work that successfully mines insights from novel data would attract follow-on work and spark new research ideas in the broader academic and commercial communities. The Center could thus participate in studies conceived externally as well as internally. Additionally, rich data could be a useful resource for the joint journalism/computer science degree program. Computer science students could contribute a lot to the mining of Internet data, as well as provide a good deal of the expertise that would be required to store, access and analyze it.

It is feasible to build a strong research component into the Tow Center's activities, and empirical work on Internet data can certainly build knowledge that will help journalists understand how their work finds traction in the networked public sphere. Rosen suggests that the culture of journalism suffers from an anti-technology, anti-data, anti-business bias that has "infantilized" the profession in face of the Internet's challenge. A valuable mission for the Tow Center is to overcome this bias with empirical research which arms journalists and editors with the knowledge to do their own jobs better, letting them take responsibility for using new media tools and Internet data to further the profession's objectives. This review of the chaotic landscape of online measurement suggests one good, informal mission statement: To ensure that the responsibility for



understanding and using Internet metrics in journalism does not rest solely with people selling ads.

Funding for this research was provided by Mary Graham, a member of the Board of Visitors of the Columbia University Graduate School of Journalism.

Author Biographies:

Lucas Graves is a PhD candidate in Communications at Columbia University. His dissertation research uses network analysis to study structural changes in the current-affairs ecosystem. Lucas has worked as a technology and media analyst with Digital Technology Consulting and before that with Jupiter Research, where he covered Web technology as well as online advertising and commerce. He is also is a longtime magazine journalist, now on the masthead of *Wired* magazine. Lucas holds an MA in Communications and an MS in Journalism from Columbia's J-School.

John Kelly is a graduate of the PhD program in Communications at Columbia University and has studied communications at Stanford and at Oxford's Internet Institute. His research blends Social Network Analysis, content analysis, and statistics to solve the problem of making complex online networks visible and understandable. John is the founder and lead scientist of Morningside Analytics, and previously served as Principal Investigator of Columbia's Interactive Design Lab. He is an Affiliate at the Berkman Center for Internet and Society at Harvard Law School.

Marissa Gluck is a writer, speaker and consultant covering the marketing and media industries. Marissa is founder and managing partner of Radar Research; formerly she worked as a senior analyst at Jupiter Research and as emerging technologies specialist at i-traffic. She has advised clients including Interpublic Group, WPP, DoubleClick, Nielsen NetRatings, Google, CBS, AOL Time Warner, and the Norman Lear Center. Marissa holds masters degrees in Communications from the London School of Economics and from USC's Annenberg School.