

A Primer in Internet Audience Measurement

**By
Bruce Jeffries-Fox**

Introduction

There is a growing trend toward people using the Internet to get their news and to investigate particular issues and organizations. There is also growing evidence that what the public believes about you or your client's organization—in good times and bad—is substantially shaped by what they see, read and hear on the mass media, including the Internet. Therefore, it is important for Public Relations professionals to understand and assess cyber-media content, just as it is important to monitor and assess traditional media content.

But how do you begin? The Internet is huge and getting bigger every day. Some estimates place the number of web pages at 3 billion. There are at least 80,000 newsgroups with people chatting about products, issues, and organizations. There is no way you could, or should, attempt to monitor and analyze everything. In short, you need a way to determine which Internet sites are talking about you and, of these, which sites matter most.¹

Determining which sites are talking about you is relatively easy. There are a number of commercial services available that will scour a huge number of sites daily and identify when your organization is mentioned. Typically, these services pull down the stories and send them to you via email. Some services archive the stories on their servers. It's also possible to do-it-yourself, using search engines such as Google and Yahoo to find stories about you and load them into your computer. Unfortunately, these activities can produce so much information that it would be impossible to read and absorb everything. It is very easy to get swamped if you do not specify some limits.

The challenge, therefore, is to have some way to select websites that really matter, and then monitor and analyze them. This paper will help you make this selection by introducing audience metrics you can use to compare web sites and then prioritize them in terms of their importance to your organization.

Before jumping into Internet metrics however, let's review how we answer the question of which media matter for traditional media such as newspapers, magazines and television. Typically we make this assessment by comparing news outlets in the following dimensions:

¹ For the sake of simplicity we will use the terms "Internet site" and "website" interchangeably.

- Audience size (e.g., circulation, viewers, listeners)
- Audience composition (e.g., % of audience that is your target market)
- Status (e.g., the generally agreed upon credibility of the news outlet)

Can we use the same metrics to assess Internet news sites? For the most part, the answer is yes. Internet audience size and composition data are offered by several commercial services. Status data is not generally available but can be easily generated on a custom basis. So while the Internet certainly does have characteristics setting it apart from traditional media—most notably its interactivity, it is really no different from other media when it comes to the basic criteria we use to judge the importance of news outlets.

After reading this paper you will be able to evaluate the offerings of the various commercial services that have begun providing data on Internet audiences. You will understand the jargon and how Internet measurement terms relate to the traditional media measurement terms that you probably already know. Armed with this knowledge, you will be able to make a first cut at determining which sites to monitor and analyze.²

ISSUES IN MEASURING AUDIENCE SIZE AND COMPOSITION

Internet Metrics:

1. “Hits”

- Definition: “Hits” typically refers to the number of times a “file” is requested from an Internet site. A *file* may include an entire page (perhaps the site’s home page) or elements within a page (e.g., an image, a sound file, a text file). Typically, the number of *hits* is reported weekly or monthly.
- Pros: *Hits* are quite easy to calculate and, consequently, are frequently used to indicate how popular a site is. While there are problems with *hits* as a measure of audience size, it is a good measure for comparing the relative popularity (or “reach”) of different websites.

² Audience size and composition are only two ways of assessing which sites matter. Subsequent Commission papers will address other factors such as site credibility, interactivity and persuasion.

- Cons: *Hits* is a very crude measure of audience size because one *hit* is counted as one visitor. I.e., each and every visit to a site or request for a file within a site is counted as one *hit*.

In actuality, *hits* does not take into account return visits--the fact that one person may generate any number of *hits*. For example, 1,000 hits may be the result of 1,000 people visiting a site once each, or 100 people visiting 10 times each, or any combination in between. This inflates audience size figures.

The crudeness of this measure is exacerbated by the fact that sites differ widely in number of pages/files. Every visit to a site, page or file is counted as one *hit*, creating a bias in favor of larger sites with more pages and files. This leads to downgrading the relative importance of smaller sites that may have the same number of visitors.

- Bottom line: Use *hits* only if you can't afford or gain access to a better metric. *Hits* generated by individuals in your target group are a better metric than overall hits because it better reflects the people you care about.

2. "Unique visitors"

- Definition: This metric refers to the number of different people who access an Internet site. It is based upon one or more of the following types of information garnered from visitors: their IP address, their member name (if the particular site uses membership or if the visitor is part of a commercial service's panel), or their "cookies" (personal identification information).
- Pros: Because it measures the number of unique individuals accessing a site, *unique visitors* is a finer measure than *hits* of audience size. Today it is possible to obtain *unique visitor* information through several commercial services. Services using panels assign each panelist a unique membership number, and provide reasonably accurate measurements of the sites they visit.
- Cons: There are two main sources of imprecision in measuring *unique visitors*, both of which may lead to understating audience size: IP addresses are assigned dynamically, so the same address can be used by several different people over time, thereby understating the number of *unique visitors*. Cookies can be turned off or deleted at any time by individuals, which again leads to an understatement of *unique visitors*.

- Bottom line: If your primary concern is to assess a site's reach, *unique visitors*, while not perfect, is the best metric available today. If you can obtain a breakdown of *unique visitors* by demographics, you will be able to better assess the numbers of your target market that are visiting different web sites, vs. simply using total *unique visitors*.

3. "Page views/requests/impressions"

- Definition: The term "page" is a holdover from traditional print media. In this context page refers to a particular screen appearing on an Internet site. Page views, page requests and page impressions are all the same measurement, referred to here as *page views*.

As with *hits*, *page views* measures the number of times a particular page is accessed during the measurement period, without regard to the number of different individuals doing the accessing.³

- Pros: This metric deals with a much more specific portion of an Internet site—a single page, vs. a *hit* which deals with the site as a whole. This metric is highly suitable for PR-oriented web measurement because it is possible to get *page view* data for news stories appearing on particular web pages. It is a much better reflection of the number of people exposed to a particular story than either *hits* or *unique visitors* (both referring to the site as a whole).
- Cons: There is modest problem with interpreting the number of *page views* as the number of different people exposed to a particular page, because one person could access the same page multiple times during a measurement period (e.g., to re-read an article in order to obtain very specific information, such as a name or address).⁴
- The bottom line: While this metric is not perfectly precise, it gives us a reasonably accurate measure of the number of people exposed to a particular story. Until more precise measures become readily available, *page view* data is definitely one of the best measures you can use when you are concerned with how many people saw your story. And again, if you want to know how many of your target market were exposed to your story, ask for a demographic breakdown of the *page view* data.

³ Some commercial services break down page views by time period, by referrer (specific banner or link from which a visitor has come) and by visitors' host (e.g. AOL, Earthlink).

⁴ While it is reasonable to assume that most people will view a story on a particular website only once, there will definitely be exceptions to the rule.

4. “Unique page views”

- Definition: *Unique page views* is used interchangeably with *unique page requests* and *unique page impressions*. This metric goes beyond *hits* and *page views* by reflecting the number of *unique visitors* who have accessed a *particular web page* during the measurement period.

This metric is generally not available, even through commercial services. This will likely change in the near future, but today it is beyond the technical capability of most services.

- Pros: The *unique page views* metric is the Holy Grail of Internet audience size measurement. It provides a quite precise measure of a particular story’s reach.
- Cons: It does not take frequency into account, and therefore cannot reflect level of audience involvement.
- The bottom line: This is a metric to keep in mind and to keep requesting of commercial services. If enough clients request it, it will be implemented. It is worth pursuing because it is an excellent reflection of a story’s exposure.

Derived measures

When the above metrics are not available (or are not affordable) they can be estimated through the use of some modestly clever math.

Some commercial services have studied the quantitative relationships between three audience metrics--hits, unique visitors, and page views--over hundreds of websites. By creating “average” ratios, they are able to derive an estimate of one metric from another. For a fee, these ratios can be obtained from audience data providers. The following are examples of how they can be used:

1. Derived unique visitors: Let’s say that the typical ratio between *hits* and *unique visitors*⁵ is found to be 10-to-1. Using this ratio, if you know the number of *hits* a site gets in a month, you can divide this number by 10 to derive a crude estimate of the number of *unique visitors*.
2. Derived unique page views: If the typical ratio between number of *unique visitors* and unique page views is 20-to-1, then if you know the number of

⁵ From sites similar to those you care about.

unique visitors a site gets in a month, you can divide this number by 20 to derive a crude estimate of the number of *unique page views*.

3. **Derived frequency:** If you are concerned with frequency—i.e., the average number of times a person visits a website⁶--this can be estimated. The average frequency with which a person visits a site in a measurement period can be derived by dividing the total number of *hits* by the number of *unique visitors*.
 - Pros: If you are willing to live with an estimate rather than an actual measurement, a derived metric can save you the expense of buying multiple metrics. Further, you can derive an estimate for an important metric—*unique page views*—that is not yet available.
 - Cons: Deriving estimates in this way applies generalities to specific websites and stories. This may or may not do justice to your web coverage. It could yield numbers either larger or smaller than the story's actual exposure.
 - The bottom line: While actual measurements are always better than derived estimates, this is a reasonable approach if the numbers you want are not available or affordable. If you use a derived-measure approach, be consistent with the ratios you apply. Change ratios once a year or less. This will ensure that any changes you see in the data are really a result of audience behavior and not an artifact of your changing the measuring stick.

Data collection

We will now examine where these data typically come from. There two main sources: Commercial services and the Internet sites themselves. The commercial services generally use consumer panels to collect the data whereas the Internet sites generally measure their own visitors. While it would seem that the latter approach would be more direct and therefore yield a more reliable measurement, this is very often not the case. The reasons for this are discussed below. The bottom line: We strongly recommend that you obtain these data from commercial services.

1. Panels. Commercial services typically use “panels” to gather their data on Internet audience size.

- **Definition.** A panel is a group of people who are paid (or otherwise compensated) by a research firm for providing data about themselves -- such as which products they are buying, their opinions on certain issues, which Internet sites they've visited, and so on.

⁶ Or exposure to a particular story on your site.

- *Pros.* The best part about using panels is that the research companies have already collected a lot of data on their panelists. Typically, they will have on file their panelists' demographics, psychographics, recent purchases. This allows them to tell you a lot about the people visiting your website without you having to pay for it on a custom basis.
- *Cons.* The greatest concern with using panels is their representativeness. That is, to what extent are these people reflective of the U.S. population or some sub-group, such as your target market? Question the firms. Make sure the demographics of their panel quantitatively mirrors the proportions in census data.⁷

The other main concern with panels is more subtle. Are people who will take the time to participate in market research somehow different from others in the demographic groups they are supposed to represent? While this is an important concern it is almost impossible to completely address. Data firms know this is an issue and most have taken substantial steps to minimize the possibility of bias in their panels. Ask them how they have tackled the issue and go with the firm you feel has done the best job in this area.

Research companies track their Internet panelists' surfing by either of two methods:

1. They ask panelists to write in a "diary" which sites they've visited, and
2. They put software on panelists' computers that automatically tracks the sites they've visited.

There is no *right* way to collect these data, and neither of these approaches is perfect. Requiring people to write things down runs the risk that they will be too busy some days and simply not catch everything. The software approach means that panelists know that everywhere they go will be recorded—so they may alter the places to avoid embarrassment. Not surprisingly, these different approaches produce different numbers.

Because there is little experience with these two data collection methods, it is premature to make a recommendation. In the meantime you should make your own judgment about which of these two approaches will produce data most suitable for your purposes.

Other factors that can help you select a panel are:

⁷ Most firms will have such data readily available for your review.

- *Panel size.* Some panels are much larger than others. In general, larger samples are better than smaller samples.
- *Panel composition:* The system with the most panelists meeting the demographics of your target market would probably be best.

2. Getting audience figures from Internet sites. Many Internet sites provide their own audience figures. It is not a good idea to rely on these numbers. The reason is that websites have a vested interest in inflating their numbers. You want as accurate a calculation as possible, and this means the company providing the data should have accuracy as its main goal, not a hidden agenda of boosting advertising rates based on the number of purported site visitors. Rather, we recommend obtaining data from a commercial service.

Standardization of metrics for Internet audience size

Several organizations are developing Internet audience metrics and are working toward standardizing these measurements. Among the leading organizations doing so are:

Surveillance Data, Inc. (SDI)
Comscore
Nielsen/Net Ratings
Standard Rate and Data Service (SRDS)
The Business Publishers Association
The Advertising Research Foundation
The Internet Advertising Bureau

As of this writing, these companies and others offer several different metrics that reflect, in various ways, the size and composition of Internet audiences. We suggest that you speak with them about the issues brought up in this paper if you are considering purchasing data from them.

Summary

While increasing numbers of people are obtaining their news from the Internet, the Net has grown so large that it is impossible to monitor and analyze every site carrying stories (or discussions) about your organization. With the help of monitoring services, you can determine which sites are carrying the most stories about you. Still, the number of sites and number of stories can be overwhelming.

The solution is to prioritize which news and discussion sites are most important. With traditional media we make this determination by assessing the size and demographic composition of a news outlet's audience and the outlet's "status". These same three criteria can be applied to the Internet as well, although at this time it is much more affordable to obtain audience size and demographic data than status.

The field of Internet audience measurement is quite new. It is spawning metrics with novel names such as *hits*, *unique visitors* and *page views*. However, upon closer examination we find that all of these new terms refer one way or another to the established concepts of reach and frequency—the cornerstones of traditional media analysis. Consequently, while it might be necessary to learn a few new measurement terms, it is likely that you are very familiar with the underlying concepts. (See Appendix A for a comparison of Internet terms to traditional media measurement terms.)

The Internet audience metrics most frequently used today and covered in this paper...) vary considerably in the precision with which they measure the number of people exposed to a particular story. The cost of obtaining these data varies with their precision.

In order for PR people to gain credibility among other communication disciplines and senior management, we need to demonstrate our worth through high-quality research. Such research requires precise measurements. Consequently we recommend that you use the most precise Internet audience metrics you can afford, whether your interest is assessing how many people came to your own site, or how many people were exposed to a story about you on someone else's site.

We hope this paper has provided you with an understanding of the various metrics available today, along with their advantages and disadvantages so that you may become a good consumer and a better user of these data.

Appendix A

COMPARISON OF MEASUREMENT TERMS FOR TRADITIONAL VS. INTERNET MEDIA

The basic ideas involved in measuring Internet audiences are very similar to those of traditional media measurement, despite differences in jargon. No doubt you are aware of the traditional measurement terms as they are commonly used in PR, advertising, and other forms of marketing communications. To facilitate an understanding of how the Internet measurement terms relate to traditional media measurement terms, we offer the following:⁸

1. Audience size metrics

- Websites and “media vehicles”

While in the Internet world, we speak of Websites or Internet sites. In traditional advertising language a website is a “media vehicle” akin to a TV program.

- Hits

There is no equivalent of Internet *hits* in traditional media measurement.

This is because *hits* does not separate reach and frequency (see below), which typically are individually addressed in traditional media measurement.

- Visitors and audiences

Visitors in the Internet world are referred to by advertisers as “audience members.”

“*Target audiences*” are those people who match the client’s target demographics, psychographics, etc. While advertisers routinely use this term, it is only starting to be used in Internet audience measurement. As more commercial services begin offering these data this information will

⁸ Our comparisons are based on standard advertising industry measurements, as defined in the American Association of Advertising Agencies’ Guide to Media Research (1995).

become more affordable and we can expect to see *target audience* become part of Internet audience analyses.

- Page views and exposure

When a person views a particular Internet page, it is the equivalent of “exposure” in the traditional media. There can be exposure to an Internet ad or Internet story, just as there is exposure to a traditional media ad or story.

Since some pages include stories, information and ads, it is often impossible to determine what people are attending to when they view the page. This is analogous to the situation in traditional media where there are metrics for the number of people viewing a particular media vehicle (e.g., program, magazine issue), but it is not known whether they have attended to a particular part of the program or the advertisement within a program.

- Reach

Reach is a key term in traditional media measurement. Conceptually, reach is a measure of the number of people exposed to, or reached by, a message. In the advertising world *reach* is typically expressed in numbers of people or as % of the total population (or total target market population). The reach concept applies equally well to the Internet and is increasingly used in Internet media analyses.

In both traditional and Internet media analyses there are different ways of measuring reach. Some are much more reflective of the underlying concept than others.

For example, in a lot of advertising research the number of viewers of a particular show is taken to be the number of people exposed to an advertisement within that show. In reality, this may not be the case for a number of rather obvious reasons. There are now more precise measures available, but more precision costs more money, and for cost reasons viewership figures are still often used.

Media analysts supporting PR managers often face a similar problem in measuring reach. Analysts often use a newspaper’s or magazine’s circulation figures as a count of the number of people exposed to a particular story appearing in a particular issue. In reality, of course, this is not likely to be the case but the cost for greater precision is often too high, so the practice continues.

In the early days of Internet media analysis (from which we are just emerging) we have seen a similar tendency. Most Internet *reach* measurements to date have utilized *hits* or *unique visitors*, using the assumption that anyone visiting a website was exposed to all stories within that website. The unaffordable cost for greater precision keeps this obviously flawed practice going.

As more firms offer the data, competition among them drives up the precision and drives down the cost. Consequently we have seen a progression in Internet measurement from hits to page views, and looking forward, to unique page views as measures of reach.

2. Audience frequency metrics

- *Frequency* refers to the number of times a particular person is exposed to a story, an advertisement, a picture, etc. It applies equally well to the Internet and traditional media. Frequency has been researched for years by the advertising industry because, in general, more information is communicated and more persuasion occurs when people are exposed more frequently to the same message. The concept of frequency is only occasionally discussed in Public Relations traditional media analysis.
- *Effective frequency* is a term used by advertisers to mean the number of advertising exposures judged to be required to produce a positive change in awareness, attitude or purchasing action. Discussion of *effective frequency* is rare in PR-oriented media analyses, although the concept is often discussed among PR managers and is gaining currency among research firms.
- The Internet's ability to measure frequency offers the opportunity for PR practitioners to begin using this metric to gain insight into the relationship between frequency and impact in an Internet context.

3. GRPs

- In the world of advertising, gross rating points (GRPs) are the established measure of an ad campaign's weight (the likely impressions generated by the campaign). It is calculated by multiplying the reach of a campaign by the audience's average frequency of exposure.
- In theory, the concept of GRPs could apply to Internet audience measurement, but at this time measurement of Internet reach and frequency is at a very early stage. When more commercial systems make such data available at a reasonable price, it is likely that the GRP concept will be introduced into Internet measurement.