



Another contribution to *the science beneath the art*™ from the
Commission on Public Relations Measurement and Evaluation
at the Institute for Public Relations

**Using Web Analytics to Measure the Impact of Earned Online
Media on Business Outcomes:
A Methodological Approach**

By:
Seth Duncan
Context Analytics and Text100
San Francisco, California

Seth Duncan is Research Director at Context Analytics, the research affiliate of Text100. Seth specializes in social media research, statistical modeling and return-on-investment analysis, working with a broad group of Fortune 500 clients, ranging from healthcare companies to clients in financial services and consumer electronics. Prior to joining Context Analytics, Seth worked as a Statistician at the University of Pittsburgh Medical Center and University of California at San Francisco, conducting statistical analyses for Phase III clinical trials and acting as the lead statistician for numerous clinical trials. Seth has also taught statistics and research methodology at San Francisco State University. Seth received a B.A. in psychology from Reed College, an M.A. in psychological research at San Francisco State University, and has completed coursework in advanced statistical analysis at Boston College and the Massachusetts Institute of Technology.

Introduction

In recent years, there has been a large push for public relations measurement to focus more on “business outcomes” than “communications outputs.” However, many organizations continue to measure the success of their communications efforts using metrics, such as advertising value equivalencies, impressions, sentiment, message penetration, or simply volume of coverage. While each of these metrics can be valuable if used correctly, they say little about how coverage has influenced business goals. It’s good to know, for example, that corporate messages appeared in 20 percent of your coverage in top-tier publications (a PR output). This information can help a communications team understand what messages are gaining traction in the media, where messages are appearing or not appearing, etc. But, when the CEO asks, “what is message penetration doing for the business?” it would be much more helpful to show, for example, that the 20 percent of coverage containing messages produced two times more revenue as coverage without messages (a business outcome).

PR practitioners and measurement professionals widely agree that measuring outcomes is a desirable next frontier for the industry, but adoption has been slow because the tools needed to do so are out of reach for many communications teams. However, the relatively recent adoption of web analytics is likely to expedite PR’s shift towards this outcome-focused measurement. Online marketing has been successful at integrating a wide-range of web data, including search trends and web analytics, to optimize advertising and search strategy and spending. This has led corporations to shift sizable portions of their advertising budgets to the online vehicles with the best business outcome metrics. With the explosion of social media and mainstream media’s shift to an online format, it is now possible to apply the same measurement strategies to PR that have been effective for other forms of online marketing. Through the use of web analytics, PR professionals can directly measure the business value of carefully targeted earned media opportunities in the same way that marketers have been able to for over a decade. We believe that this data-driven approach leads to the best strategic PR decisions and greatest investment in PR campaigns.

This paper is intended to provide PR practitioners with an overview of web analytics and how they can be used in regular communications measurement programs. After a brief overview of web analytics technology, we provide an overview of different methodologies that PR practitioners can use to measure how PR and earned media are influencing customer behavior at their organization’s website. These methodologies are divided into two broad groups. We call the first basic analytics, since they require little or no expertise with advanced research

methodologies, such as integrating different data sources and using statistical modeling. These analytics answer broad questions about which types of sites are best at referring traffic, about conversions at the organization's website, and about comparing earned media with other forms of paid media. We call the second "advanced analytics" since they require a high degree of familiarity with advanced statistics. These analytics specifically address messages and story content that effectively drive traffic to a website and accomplish business goals.

Web Analytics: A Brief Primer

Web analytics tools are used to track web traffic patterns, particularly where visitors are coming from, what web browsing technology they are using, and how visitors are interacting with a website. Web analytics had a humble start as server error logs and were first used by IT professionals to evaluate the user interfaces of websites, and by and by web designers to find errors on web pages such as incorrect links. By 2000, enterprise web analytics providers, including WebTrends and CoreMetrics, were offering commercial web analytics tools designed for non-technical audiences, especially marketing departments. More recently, web analytics have been adopted as an optimization tool for marketers.¹ Web analytics are now one of the most important and widely used tools for internal marketing teams who sell or obtain a large portion of sales leads through their corporate website.

Here is a quick overview of how most contemporary enterprise web analytics data is collected. First, an individual types a URL into their web browser (or clicks on an advertisement, search result, or hyperlink). Second, the URL request is sent to and received by the company's website server, which then sends page data, along with embedded JavaScript code to collect web metrics, back to the browser. Finally, the JavaScript code is executed to collect various pieces of page browsing information from the user and then sends that information to a data collection server.

JavaScript tags will track the following types of visitor behaviors:

- Number of unique visitors
- Number of visits
- Number of pageviews
- Time on site
- Bounce rate (*e.g.*, percentage of visitors who leave after visiting one page)
- Number of goals reached (*e.g.*, downloads/registrations/orders)
- Conversion rate (*e.g.*, number of goals reached divided by number of unique visitors)

¹ For excellent in-depth guides into the history and many marketing uses of web analytics, see Avinash Kaushik, *Web Analytics: An Hour a Day* (Indianapolis, IN: Wiley Publishing, 2007); Avinash Kaushik, *Web Analytics 2.0* (Indianapolis, IN: Wiley Publishing, 2009); Alistair Croll and Sean Power, *Complete Web Monitoring: Watching Your Visitors, Performance, Communities, and Competitors* (Sebastopol, CA: O'Reilly, 2009); and Steve Jackson, *Cult of Analytics: Driving Online Marketing Strategies Using Web Analytics* (Burlington, MA: Butterworth-Heinemann, 2009).

Particularly useful for PR is that web analytics solutions segment referring traffic sources, usually into the following categories:

- Direct traffic
- Natural search (sometimes called organic search)
- Paid search (*e.g.*, Google Ad Words)
- Referring Search Keywords
- Email campaigns
- Digital Advertisements (*e.g.*, banner ads)
- Earned media

For the purposes of this paper, we will focus on the *earned media* category. Earned media includes referrals to an organization's website from traditional media, such as NYTimes.com, CNN.com or even Oprah.com (*i.e.*, any news site with a print or broadcast companion), or from strictly online media, such as Cnet.com, Marketwatch.com, as well as social media sites, such as blogs, forums, Facebook, LinkedIn, Twitter, etc. Now that so many organizations are using some form of web analytics, it's possible for PR professionals to measure the same sort of outcomes as digital advertisers. These metrics of engagement include the number of unique visitors coming to the organization's site from a story, the amount of time visitors referred from a story spend on the site, how many pages or which pages they read, and whether visitors from the story are helping the organization reach a goal, such as purchasing a product, registering for email alerts, or downloading white papers.

Why Web Analytics Are Important to Public Relations

PR practitioners are just beginning to utilize web analytics, and, over the next several years, this relatively new technology will help organizations optimize strategy and demonstrate PR value. At this early stage, there are three primary reasons that organizations should adopt this technology:

- 1) Web analytics allow PR efforts to be compared, side-by-side, with other forms of marketing. Traffic, conversion rates, and other web analytics metrics are collected from nearly all referral sources including advertisements, paid search, unpaid search, and earned media. This means that, for the first time, it is easy to track exactly how the outputs from both PR and advertising each affect business outcomes using the very same metrics.

Exhibit 1 illustrates such a comparison that Context Analytics found for a financial services company that sells a credit monitoring service online. The table compares sales conversion rates from customers who came to the branded website through traditional media (*e.g.*, NYTimes.com), from blogs, and from paid search over a 5 month period in 2009. While each of the conversion rates is high, earned media produced a higher conversion rate than paid search during that time period. It is a good example of how effective earned media can be in driving purchasing decisions relative to paid media, since paid search is widely considered one of the most targeted and effective forms of marketing.

Exhibit 1

Referring Source	Conversion Rate
Paid Search (N=361,013 unique visitors)	8.2%
Traditional Media (N=68,224 unique visitors)	10.5%
Blogs and Forums (N=204,662 unique visitors)	9.0%

- 2) Web analytics can inform strategies based upon empirical evidence, since web analytics can show which types of coverage (both in traditional media and in social media) help the organization reach its business goals. Rather than relying on intuition, data can be used to answer questions such as:
- Is a corporate Twitter account effective at getting potential customers to download marketing collateral such as white papers?
 - Is one key message better than another at getting customers to visit the corporate website and engage through registrations, downloads, or purchases?
 - Should an organization shift more resources and attention from traditional to social media?
 - Which online audiences are responding to the media by making purchases, and which are not?

Using web analytics, PR practitioners can arrive at answers to these questions. Instead of seeing online media, particularly social media, as a mysterious frontier, web analytics can help PR practitioners focus on media that are likely to be effective in driving business goals.

- 3) Web analytics provide an affordable alternative to marketing mix modeling.² Before the arrival of web analytics, the only method for measuring the effect of PR activity on business outcomes was to track media coverage over long periods of time and then use advanced statistical modeling to tie media coverage to metrics associated with organizational goals (*e.g.*, sales). Marketing mix models have been too costly and time-consuming for most organizations to adopt with any regularity, due to the large amount of historical data and statistical expertise involved. With web analytics,

² For a discussion of the application of marketing mix modeling in public relations, see Mark Weiner, Liney Arnorsdottir, Rainer Lang, and Brian Smith "Isolating the effects of public relations on sales: public relations optimization through marketing mix modeling," published by the Institute for Public Relations Commission on Public Relations Measurement and Evaluation, 2010 (forthcoming); www.instituteforpr.org.

it is possible to directly measure the behavior of customers who have been exposed to a range of traditional and social media over any period of time with minimal need for advanced regression, Bayesian, or time series-based statistics. Web analytics are certainly not a replacement for marketing mix modeling, which also takes offline marketing and customer behavior into consideration. Web analytics do, however, provide an accessible and cost-effective means of measuring the relative contribution of PR to business goals in a way that has been long been unavailable for most organizations.

Methods for Using Web Analytics for Communications

The remainder of this paper focuses on practical steps PR practitioners can take, using web analytics, to demonstrate the value of their efforts and optimize strategy. We will discuss two broad types of web analytics analysis. We'll refer to the first type of analysis as "basic," not because it lacks value, but because PR practitioners with little or no web analytics or statistical training can conduct these analyses. These analyses are similar to those that might be conducted using existing integrated solutions, such as Radian6 and WebTrends, or Omniture's SiteCatalyst, and can address questions such as:

- Which traditional media sites (*i.e.*, sites with a print or broadcast companion) and social media sites are driving traffic to an organization's website?
- Which types of sites (*e.g.*, traditional media, blogs, forums, etc.) are driving visitors and engagement at the organizations website?
- How do conversion rates for earned media compare to other forms of marketing?

We refer to the second set of analysts as "advanced," since they involve integrating different types of data, and, in many cases, conducting advanced statistical analyses. As a caution, PR practitioners will require help to conduct these sorts of analyses, either from a statistician or database expert from another internal marketing group, or from an external research partner. These analyses address more strategic questions such as:

- To which audiences should we be reaching out (*e.g.*, individuals of a specific age, gender, or income level)?
- Which messages are most effective at driving traffic and engagement?
- Which messages should be matched with specific audiences to optimize communications efforts?

What makes these analyses advanced is that they draw upon a wide range of data sources, including demographic data, for the sites that are referring visitors, as well as detailed information about the content of the stories and posts that are driving traffic to the organization's website.

Basic Web Analytics for PR Practitioners

Basic web analytics analysis for PR practitioners is fairly straightforward. It involves downloading the relevant referral data from an organization's web analytics software, cleaning the data, and organizing the results based on specific communications goals. The basic analyses outlined here can provide PR professionals with a clear overview of which types of earned media sites are driving traffic and engagement (*e.g.*, software developer forums, mainstream news, gaming blogs, etc.), compare the effectiveness of earned media with other marketing channels, as well as conduct a range of "return on investment"-type analyses. In this section, general guidelines for obtaining and cleaning the data are provided, along with recommendations for basic analysis that most PR professionals should be able to complete without extensive knowledge of research methodology or statistics.

Downloading Relevant Data

Most enterprise web analytics software systems allow downloading a list of sites that have referred traffic to the organization's website (and associated metrics) into a .csv data file or similar file format. Many web analytics providers, such as CoreMetrics and Omniture, organize referring sources by paid and unpaid types. Some web analytics software allows downloading only unpaid referring sites. Other web analytics solutions only allow the user to download reports of all referring sites, regardless of whether the sites contained ads or earned media, and the list will have to be manually screened to pull out content that is relevant to PR.

Some software solutions allow users to download referring site data at both the domain level (*e.g.*, Techcrunch.com) and URL level (*i.e.*, the specific post or story on Techcrunch.com). Reports at the URL level are particularly important, since that level of detail provides insight into which stories are effective at driving traffic and engagement.

For most web analytics software solutions, unpaid media reports can be customized. When downloaded, most reports will have individual domains (or specific URL) in each row and a series of user-selected web metrics in each column. When downloaded as a .csv file, it is easy to sort referring domains by the number of unique visitors they produced, total sales, conversion rates, etc. This structure is similar to the downloadable reports from social media providers, such as Radian6; consequently it is fairly easy to merge these two different sets of data in database applications.

The exact steps for downloading unpaid referring domains and URL reports vary significantly from provider to provider. If you are using a paid enterprise web analytics provider, contact them (or your internal web analytics manager) for details on how to download unpaid referral reports that do not include traffic from search results, emails, or advertisements. If you are using free web analytics solutions, such as Google Analytics, you might need to identify unpaid referrals manually.

Data Cleaning and Quality Control

Raw data from web analytics software providers can be very messy. That does not mean that the data is necessarily inaccurate, but unpaid referral reports often include significant numbers of email and search domains that are not directly relevant to PR or originated from paid marketing, such as an email campaign. Unpaid media reports that are downloaded directly from web analytics software will also contain referrals from other organizations' websites. Depending on your organization's goals, referrals from these sources might be irrelevant to your reporting and research.

Although it can take quite a bit of extra work to remove referrals from search and email in the unpaid media reports, it is very important if you want to collect valid web metrics for media coverage. In our experience at Context Analytics, conversion rates tend to be significantly lower for referrals from email and rarely-used search engines than from traditional and social media. Including data from these erroneous sources can artificially lower average conversion rates and make earned media appear less effective than it actually is.

If your organization has a web analytics team, you should be able to work with them to produce a report of earned media referrals that excludes referrals from email or search. If that is not possible, it is easy to identify email and search referrals. Very often, search pages will actually contain the word "search" in the URL, such as "http://search.imesh.com/" or "http://smartbizsearch.com/." Similarly, email referrals will almost always contain the word, "email" in the URL. Examples include URLs, such as "us.f302.mail.yahoo.com" or "106w.bay106.mail.live.com." Fortunately, since these erroneous referring URLs are readily identifiable using specific phrases, it is easy to find and remove them using any database or spreadsheet application, such as Excel.

Analysis of Media Types

One of the most straightforward yet useful things that a PR practitioner can do to understand what types of media coverage are driving unique visitors to and engagement with their organization's website is to categorize the referring domains. Knowing whether lifestyle blogs or financial forums are effective at sending traffic to your organization's website, for example, is the first step in understanding what aspects of a communications strategy are currently working or are likely to work in the future. If a communications team is spending a great deal of resources trying to pitch tech bloggers to write about a new service, but find that traffic from these blogs rarely leads to sales leads, the team should probably either pitch different types of stories to these bloggers, or abandon the blogger outreach altogether in favor of a different media strategy.

Fortunately, the analysis required to understand which types of websites are effective or ineffective at meeting PR goals are quite accessible to individuals with little or no experience with quantitative research or analytics (basic knowledge of Excel should be sufficient). It is as simple as developing a media categorization scheme, applying the categorization to a downloaded and cleaned unpaid referral web analytics report, and then comparing the media categories to understand which were the most effective at meeting your goal.

Categorizing Unpaid Referrals

There are many categorization systems that could apply to unpaid media sources. Stacks³ has provided three general rules for creating useful media categories that apply just as well here as in any other form of PR measurement:

- 1) Categories must be relevant to the purposes of the research. For example, if testing whether blogs with blue or red backgrounds are more likely to drive traffic, you would categorize blogs by background color instead of industry vertical.
- 2) Categories must be exhaustive, meaning that every type of unpaid media must fall into one of the categories.
- 3) Categories must be mutually exhaustive, meaning that a referring source cannot fall into more than one category within a single category system (*e.g.*, coding an article as being both mainstream and a blog).

Given the ever-growing range of web sites that contain earned media (Facebook, Twitter, blogs, social bookmarking sites, etc.), it is usually necessary to create multiple categorization systems to best capture the diversity of possible online media referrals. At a bare minimum, organizations should use the following two systems: site type and site content. Site type refers to the actual format of the website, while site content refers to the types of coverage or posts that the site contains. Together, these two categorization systems provide an acceptable degree of granularity into exactly what types of sites tend to drive unique visitors and engagement at an organization's website.

The examples we provide here should only serve as examples. Exact category systems should always be customized to be relevant to an organization's measurement goals.

- 1) Site Type Examples:
 - Mainstream media (*i.e.*, sites that have a print or broadcast companion)
 - Online media (*i.e.*, sites with paid journalists that do not have a print or broadcast companion)
 - Blogs
 - Forums
 - Video
 - Microblogs (*i.e.*, Twitter)
 - Bookmarking sites (*i.e.*, Digg, Delicious)
 - Social networking sites (*i.e.*, Facebook, LinkedIn)
- 2) Site Content Examples:
 - National news
 - Regional news
 - Business press
 - Investor sites
 - Lifestyle

³ Don W. Stacks, *Primer of Public Relations Research* (New York: Guilford Press, 2002).

- Personal/biographical
- Consumer review

Comparing Web Metrics across Site Categories

Once PR practitioners have created categorization systems and applied them to the referring sites from their web analytics software, the next logical step is to start to compare the different categories to see which are effective or ineffective at driving traffic and engagement.

There are numerous ways to compare web metrics across referring site categories. The two most basic approaches are to compare web metric totals and web metric averages across category systems. The first approach, total or sum-based comparisons, can, for example, indicate whether mainstream media are producing as many unique visitors as blogs. While total values are quite useful for seeing which broad types of media are effective at driving traffic and engagement, the second approach, comparing averages, can provide a greater deal of strategic insight.

If you are going to compare different types of sites, such as blog vs. forums, there will likely be different sample sizes. That means that 100 blogs posts, for example, might have sent 10,000 unique visitors to an organization's website while 10 forums sent 2,000 unique visitors. If you just look at overall unique visitor counts, the surface appearance is that blogs are significantly more effective at driving traffic (10,000 vs. 2000 unique visitors).

Looking at averages tells a different story, however. In this example, blogs sent an average of 100 unique visitors per post while forums sent an average of 200 visitors per thread. Looking at averages in addition to overall totals can provide further insight into the effectiveness of certain media channels. Looking at both can help PR professionals identify categories of sites that, while not writing very often about the organization or a product, are extremely effective when they do and should be considered a high priority target for PR outreach.

If a large number of different sites refer earned media to an organization's website, it is useful to take categorizations one step further by comparing web metrics in cross-tabulations of two or more category systems. A hypothetical example, shown in Exhibit 2, compares sales conversion rates across both site type and site content categories.

Exhibit 2

	Main-stream	Blogs	Forums	Video
General News	.01	.03	.01	.005
Business Press	.10	.11	.02	N/A
Lifestyle	.005	.04	.03	N/A
Technology	.02	.02	.03	.09

One final consideration when comparing web metrics across referring site categories is whether or not to use inferential statistics or measures of statistical significance. Since these analyses involve comparing totals and averages across categories, it is tempting to use t-tests, ANOVAs, chi-square tests of independence and similar statistical tests that help researchers make inferences about populations of data (*e.g.*, all URLs that referred web traffic) based on a smaller sample. If you have categorized all sites in an unpaid referrals web analytics report, though, these types of statistics are inappropriate, since you are not working with a sample, but rather a census of every single site that referred traffic to the organization's website. The only situation where inferential statistics would be necessary in these analyses is when you have selected a random sample of sites from a web analytics report for practical purposes. In these cases, you would need to use inferential statistics to test whether differences between means and percentages in the sample are representative of all sites in the report.

Recommended "Return on Investment" Analyses

For for-profit and some non-profit organizations that sell services or products, or collect donations on their website, web analytics provide a great opportunity to demonstrate the financial value of PR. Below are three relatively easy ways, in order of increasing complexity, to calculate the value of PR and similarity to a true return on investment analysis (ROI).⁴

⁴ See Fraser Likely, David Rockland, and Mark Weiner, "Perspectives on the ROI of Media Relations Publicity Efforts," published by the Institute for Public Relations Commission on Public Relations Measurement and Evaluation, 2006; www.instituteforpr.org.

1) *Comparing overall orders or conversions between earned media and paid media*

After an unpaid media report has been cleaned of any email-based or search-based referrals, their average conversion rate or total sales numbers can be compared with those from other marketing channels, such as paid search or content advertisements. Although this comparison will not necessarily represent a true return on investment, it will allow PR teams to demonstrate their relative value within the organization using metrics that are familiar to most chief marketing officers.

2) *Comparing earned media known to be garnered through outreach with earned media gained “spontaneously”*

Most PR teams keep a close record of media placements for which they are responsible or journalists/bloggers with whom they have kept in contact. As long as PR professionals keep close track of sites with which they have had some degree of outreach, it is possible to categorize earned media into “outreach” and “no outreach” sites or posts. This comparison can be used to simply address whether or not outreach appears to have had any effect on sales. In a more complex analysis, you can subtract the total sales from “no outreach” sites from “outreach sites” and use the remainder as an estimate of revenue earned from PR.

3) *Estimated “Real” ROI*

A third, more involved, method for determining the financial value of PR is to actually carry out a real ROI calculation:

$$\text{ROI} = \frac{(\text{Sales from earned media} - \text{cost of obtaining earned media})}{(\text{Cost of obtaining earned media})}$$

Since there are no direct placement costs for PR professionals, just hours invested in outreach and content creation, it is possible to use either billable hours on the agency side or percentage of time spent on online media outreach on the internal side of PR and calculate the cost of obtaining coverage. This is certainly more of an accounting exercise than typical PR research project but, if done, correctly, it can provide a persuasive estimate of PR ROI for an organization.

Basic Analytics Summary

Although obtaining the metrics outlined in this section requires some time and patience, the resulting analyses will be invaluable to any PR professional working within an organization that conducts a great deal of its business activity through its website. These analyses will probably require some assistance from a web analytics software provider and/or internal web analytics team, but the analyses require only limited data management expertise, or other data from within the organization. Once these online data are being collected regularly, though, it will be possible to compare PR activities online with other marketing channels in a way that, until recently, has been very difficult if not impossible.

Advanced Web Analytics for PR Practitioners

The advanced analytics covered in this section integrate a wide range of data types into a single model that can help PR practitioners optimize their online media strategies. As a caution, readers should be warned that advanced web analytics analysis requires a moderate degree of expertise in data management and statistical analysis. Many PR professionals will need to seek assistance from outside of their organization's communications team to complete the types of analyses discussed in this section.

The methodology outlined here is analogous to the multivariate testing (also known as multi-variable testing) that is used in online marketing. Throughout the past decade, multivariate testing has been used to assess which combination of text, colors, images, and organizational layout of a webpage, landing page, or a digital advertisement are most likely to get customers to convert. In these types of analyses, marketers and web designers create many versions of an ad or webpage and systematically vary one piece of the content at a time, while holding all other attributes constant. After collecting this data, statistical modeling, such as multiple regression, ANOVAs, or Bayesian statistics, is used to identify which groups of attributes are most effective at driving an online goal.

Fortunately for most website and digital ad designers, numerous pieces of software (sometimes free or very inexpensive) exist that will design, implement, and conduct the statistical analyses (Google Website Optimizer is a popular example). Part of the reason that out-of-the-box software solutions are necessary to run complicated statistics for websites and ads is the high degree of experimental control over each variable (*e.g.*, text, color, etc.) – control that is not possible in offline PR. Instead, PR professionals have to rely on content that occurs “naturally” online, and assign attributes or variables (*e.g.*, messages, topics, sentiment, word count, etc.) to that content. This means that tying PR efforts to web metrics is a much more manual process than it is for other forms of digital marketing. It is still possible to apply the exact same multivariate testing logic to earned media, though. The only difference is that the media attributes must be discovered and coded in a systematic way by researchers instead of created and tested by web designers and web analytics professionals. Other than being more exploratory than for advertising or website creation, the methods and result of the analysis outlined here are exactly the same as in multivariate testing: it applies statistical analysis to determine which aspects of coverage are most important at driving business goals online.

The Goals of Using Advanced Statistical Applications to Integrate PR and Web Metrics

We outline three different ways to tie web analytics data to online media coverage in this section. Each draws upon the same set of integrated data, but the data are used to answer slightly different, but overlapping questions:

- 1) Which site and post attributes are most effective and least effective at driving traffic and engagement?
- 2) What are the cause-and-effect relationships between site attributes, post attributes, other forms of marketing, and web metrics, such as unique visitors and orders?

- 3) Which stories and types of sites should be paired to optimize the organization's website traffic and business goals?

Choosing which type of analysis to conduct depends on the degree of familiarity with different advanced statistical methods as well as the final audience for the results. Questions one and two differ only in terms of statistical complexity and tend to be more favored by PR professionals who like to use advanced statistical methods, while question three tends to be favored by PR professionals who like a clear direction or playbook from research results.

Data Sources

The methodologies outlined in this section rely on at least three of the data sources below:

- *Web Analytics*

Web analytics data should be collected at the URL level, rather than at the domain level. This allows researchers to tie web metrics to individual posts rather than to the entire domain of the referring site. Of course, researchers can look at any metric that is being collected by an organization's web analytics program. For the sake of simplicity, we will use three metrics as examples in this section: number of unique visitors from referring site, total orders, and total revenue. When applying statistical analyses to web analytics data, these metrics should usually be treated as either dependent variables or mediating variables in a path analysis.⁵

- *Site Attributes*

Site attributes refer to the overall qualities of the site that is referring visitors, rather than the specific post (*e.g.*, nytimes.com vs. the specific story appearing on the nytimes.com website). The only practical way to obtain important metrics about an earned media referral, such as traffic at the referring site or the demographic profile of the referring site's readers, is to do so at the overall site rather than post level. In the statistical methods described below, these types of attributes usually serve as independent variables.

There are several broad types of site attribute data we have found to be particularly useful when analyzing web analytics data:

- Site categorizations: These refer to the site categorization systems described earlier, including site type and site content.
- Demographic data: The demographic profiles of website audiences, including age, gender, income, education level, and family size, are now easily accessible

⁵ For a discussion of mediation in statistical analysis, see Andrew Gelman and Jennifer Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models* (London: Cambridge University Press, 2007) and David MacKinnon, *Introduction to Statistical Mediation Analysis* (New York: Lawrence Erlbaum, 2008).

through panel-based data that is either free or very inexpensive through sites including Google Adplanner, Microsoft Adcenter, Compete, and Quantcast. This is panel-based data so its reliability is greatly diminished for sites that receive relatively few visits each month, and thus should be interpreted cautiously.

- Referring site traffic: This is, perhaps, the most important single piece of site-level data that researchers need when measuring the impact of earned media coverage on web metrics. Referring site traffic should be thought of as “circulation” data for online media, and it is consistently one of the best predictors of which specific sites are driving visitors to an organization’s website. Of course, sites with a very high volume of visitors, such as Techcrunch.com or Yahoo! News, are going to drive a significantly greater number of visits to an organization’s website when they mention the organization or a product than smaller sites that only receive a few thousand visitors each month. This means that, if there are systematic differences between the content of sites with higher traffic and lower traffic (*e.g.*, sites with higher traffic tend to have more negative stories than sites with lower traffic), referring site traffic will be acting as a confounding third variable that must be statistically controlled for.

- *Post Attributes*

The last types of data required for these advanced analyses come from the actual stories and posts that mention the organization or product. The metrics are similar to those that an organization might already be collecting for both print and online media, and often include the following metrics for each story/post:

- Sentiment
- The presence (or absence) of specific key messages
- Specific product mentions
- The presence (or absence) of specific corporate or industry topics
- Competitor mentions
- Word count for story/post
- Prominence of mention or hyperlink in story/post

These are just a sampling of post-level attributes to identify; others can be used depending on research needs. Some organizations might be interested to know, for example, whether the presence of pictures or images tends to drive conversions. Communications researchers should take the time to operationally define and quantify any characteristic of coverage content that is likely to be relevant to business goals so that these characteristics can be included as independent variables in statistical models.

Although there is no wrong way to organize data from the above sources, a database containing overall site level attributes and nested post-level attributes is most appropriate for the analyses we discuss in this section. This means that data or attributes for each referring story or post should be contained in a single row (post-level data), while overall site attributes (site-level data) can appear across as many rows as there are referring stories/posts from that site.

A hypothetical database structure for post-level and site-level data is shown in Exhibit 3. In this example, Techcrunch had three posts that referred traffic to the target website. Data such as post word count will vary between each of the Techcrunch posts, but site-level data, such as the traffic as unique visitors at referring site, will be the same for the three posts. All post attributes, site attributes, and web metrics should be entered or loaded in this (or similar format) into a database, such as SPSS, SAS, or R, to perform appropriate statistical analyses on both post and site-level data. Once data is integrated and organized into a database, then researchers can conduct multiple regression analysis, hierarchical linear modeling, path-analysis, or any other statistical analysis that provides measures of effect-size, that is, how much impact one variable is having on another.

Exhibit 3

Referring Site	Post ID	Post Word Count (Post-level data)	Unique Visitors At Referring Site (Site-level data)
techcrunch.com	1	532	1,871,199
techcrunch.com	2	1,855	1,871,199
techcrunch.com	3	1,546	1,871,199
cnn.com	4	965	29,160,255
cnn.com	5	2,187	29,160,255
nytimes.com	6	10,981	17,941,769
nytimes.com	7	4,226	17,941,769

Ranking Site and Post Attributes

Most multivariate statistical analyses can address the question, “Which site and post attributes are most effective and least effective at driving traffic and engagement?” By treating site-level and post-level attributes as independent variables within a single model and estimating their effect on a dependent variable web metric, such as unique visitors, orders, or conversion rate, it is possible understand what type of online media coverage is having the most significant impact on online business outcomes.

Ranking each site and post attribute by some measure of effect size is the most straightforward way to address which types of stories are having the biggest business impact. Researchers can use either the R^2 or beta coefficients from a multiple regression model to estimate the relative importance of each site and post attribute to a business outcome.⁶

⁶ Unless there is a specific hypothesis or model to test, we recommend using stepwise multiple regression, which is commonly used for data mining and available out-of-the-box for most statistical software tools such as SPSS. The relative contribution of each attribute can be estimated by looking at

Although multiple regression should be sufficient for ranking the contribution of site-level attributes, it can be problematic when regressing post-level attributes onto web metrics because posts within a site are likely to be similar to each other (*i.e.*, it violates the assumption of “homogeneous variance” across observations). This means that standard errors within the regression model could be artificially low and “statistical significance” could be artificially high. To properly analyze post-level data, researchers should use hierarchical linear modeling,⁷ an advanced form of multiple regression that allows web metrics to be analyzed at both the site and post-level within a single statistical model. This allows a proper regression analysis of site and post level data on web analytics, as well as the ability to test for interactions between both levels of data. HLM does require specialized statistical software.⁸

Modeling Cause-and-Effect Relationships between Site and Post Attributes and Web Metrics

After site and post attributes that contribute to web metrics of interest are identified using multiple regression or HLM, path analysis can be used to present the results in an intuitive cause-and-effect model. Statistical path analysis is a form of structural equation modeling that applies linear regression in a systematic way to test cause-and-effect relationships between independent variables and dependent variables. Path analysis can be used in a variety of PR measurement applications, but its application is particularly useful for measuring the effects of coverage on web analytics data since we know for certain, by using web analytics data, that customers viewed the earned media before visiting the organization’s website. This means that making inferences about cause-and-effect relationships between coverage and sales, for example, does not require the same “leap of faith” that is often required when analyzing correlational data.

While it might be worthwhile to understand the intricacies of cause-and-effect relationships between site types and post content, perhaps the most useful part of using path analysis in PR research is to create an easy-to-understand visualization of regression analysis for both PR professionals and the senior executives of an organization.

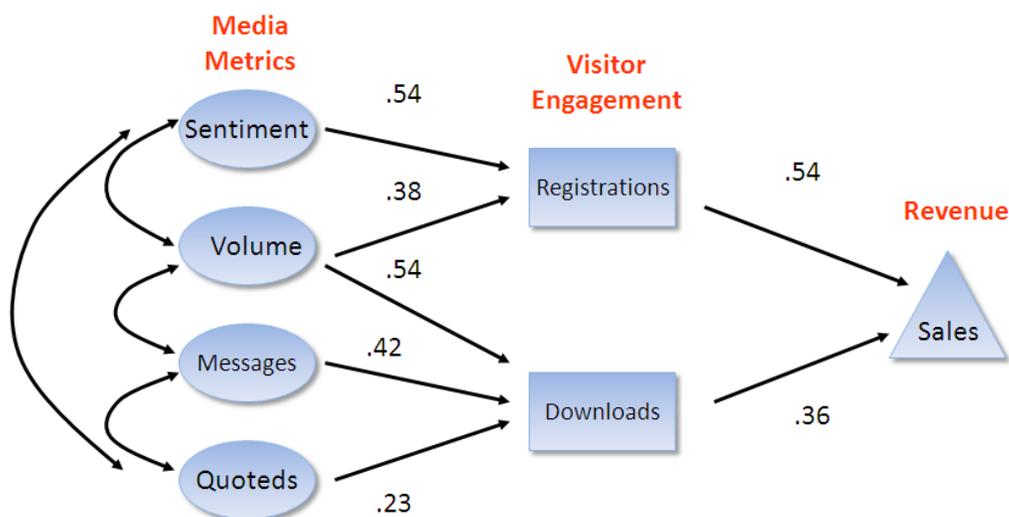
Exhibit 4 shows a hypothetical example of a path analysis assessing how post and site attributes affect registrations, downloads and online sales:

Exhibit 4

the R^2 change in the regression model when each independent variable is added or by the size of the standardized beta coefficients. Many of the metrics we have described in this paper are continuous interval variables, which work particularly well for linear regression models, especially web metrics such as unique visitors and sales. Other categorical variables will need to be converted into dummy variables (*i.e.*, values converted to 0s and 1s).

⁷ Anthony S. Bryk and Stephen W. Raudenbush, *Hierarchical Linear Models for Social and Behavioural Research: Applications and Data Analysis Methods* (Thousand Oaks, CA: Sage Publications, 1992).

⁸ The most popular HLM software is, appropriately, called “HLM” and can be found at www.ssicentral.net.



The path coefficients in path analysis are taken directly from the standardized beta coefficient from multiple regression analysis. The path coefficient for each dependent variable (registrations, downloads, and sales, in this example) is calculated by including each preceding variable in that path.⁹

To test the validity of a path analysis, researchers will need to apply structural equation modeling to evaluate the “goodness of fit” of the model. Structural equation modeling requires specialized statistical software, such as Lisrel, Amos (an SPSS program), or EQS. Structural equation modeling can also be used to compare which of multiple competing path models best fit observed data.

Creating a “Playbook” for PR Professionals

Once advanced statistical analyses have been applied to all of the integrated data, communicating the results clearly to a broad audience, most lacking statistical training, can be a significant challenge. One user-friendly format involves placing each site and post attribute into a quadrant with two axes (see Exhibit 5 below): the degree to which the attribute seems to drive

⁹ With the above figure, there would be three regression separate regression models used to determine the path coefficients (the notation b_{11} vs. b_{12} is used to indicate that these would be different beta coefficients):

- 1) Registrations = $beta_{11}$ Sentiment + $beta_{12}$ Volume + $beta_{13}$ Messages + $beta_{14}$ Quoteds
- 2) Downloads = $beta_{21}$ Sentiment + $beta_{22}$ Volume + $beta_{23}$ Messages + $beta_{24}$ Quoteds
- 3) Sales = $beta_{31}$ Sentiment + $beta_{32}$ Volume + $beta_{33}$ Messages + $beta_{34}$ Quoteds
+ $beta_{35}$ Registrations + $beta_{36}$ Downloads

unique visits, and the degree to which the attribute drives a business goal (*e.g.*, registrations, sales, etc.).

This quadrant graph categorizes each site and post attribute into one of four types:

- 1) Very effective. In the upper right quadrant are the types of sites and stories that tend to drive a lot of visitors to the organization's website and get the visitors to reach a business goal when they are there.
- 2) Opportunities. In the upper left quadrant are the types of sites and stories that do not necessarily produce a great deal of visitors yet result in sales and engagement when referrals do occur. This might, for example, be a very effective type of story that only rarely appears in high trafficked sites or not enough sites are writing the type of story. Alternatively, it could be a type of site that does drive a lot traffic and engagement (a lifestyle site, for example), but that type of site only rarely writes about the brand or product. These types of sites and posts represent an opportunity for communications teams, since the story or post type represents an effective, but under-used source of traffic.
- 3) Missed opportunities. In the lower right quadrant are types of sites and stories that produce a large number of visitors, but not very much engagement once those visitors appear at the site.
- 4) Ineffective. In the bottom left quadrant are the types of sites and stories that drive neither visitors nor conversations. This type of coverage might be important at some future time, perhaps when an organization's media or competitive environment changes, but, for the time being, outreach to these types of sites should receive a low priority.

Exhibit 5



Once each of the site and post attributes are categorized into one of these quadrants, the graph provides an easy-to-follow playbook of what types of stories should be appearing more often in specific types of sites by matching stories or sites from the upper left quadrant

with sites or stories in the lower right quadrant. If done correctly, such pairings provide insights into which messages and audiences can be matched to more effectively drive business goals at the organization's website. The quadrant analysis also provides a clear visualization of what types of coverage are currently working or not working, so PR teams can plan for future outreach accordingly. If an online news site for a particular vertical consistently appears in the "not effective" quadrant, communications teams might shift resources and outreach towards sites that are currently very effective.

Advanced Analytics Summary

When integrated with other, "standard" PR measurement data, web analytics can provide PR professionals with very detailed statistics about exactly what sort of coverage is getting target audiences to visit an organization's website and complete business goals. This information is invaluable for creating new online PR strategies. Equipped with this sort of research and measurement, communications teams can make much more targeted decisions about which sorts of sites to reach, what type of stories should be "pitched" as well as what types of PR-created collateral is likely to be effective. Since PR professionals have much less control over earned media than other marketers do over other forms of digital marketing, there are currently no "out-of-the-box" software solutions to measure earned media referrals as efficiently as content advertisements or paid search. The methodology outlined in this section requires some degree of human-coding in earned media, as well as advanced statistical modeling, so PR professionals will most likely need to seek help from their PR agency, measurement providers, or other internal marketing teams to take the fullest advantage of advanced web analytics.

Problems with Integration

While web analytics hold the promise of providing valuable insights to PR practitioners, they have not been widely adopted in PR measurement. The largest hurdle for PR professionals is to integrate web analytics with other forms of traditional and social media measurement.¹⁰ For organizations that measure their traditional and social media coverage and PR efforts, understanding how to connect regular media reports with a web analytics report might seem like a daunting task. Discovering whether the same media universe is covered in each report, data quality control, and correlating the metrics from media reports (*e.g.*, sentiment, message penetration, etc.) with web analytics metrics (*e.g.*, unique visitors, sales conversions) requires time and quantitative or statistical expertise that many communications teams do not currently have.

Some social media monitoring tools and web analytics providers have started to integrate basic social media measurement with web analytics. Radian6, a popular social media

¹⁰ Integration is not just a problem for PR; a survey of North American marketers by eMedia in September, 2009 found 46 percent of respondents reported that integration was a challenge in using web analytics.

monitoring software tool has partnered with WebTrends, one of the earliest enterprise web analytics providers, to integrate their web analytics into Radian6's user interface. This integration of Radian6 and WebTrends allows users to understand how many visitors, sales, etc., were referred by a specific post or site. Omniture, a popular enterprise web analytics software provider, recently released a user-friendly interface for tracking web analytics metrics for social media through its Omniture Site Catalyst software. Much like the Radian6 and WebTrends integration, Omniture Site Catalyst allows users to see web analytics metrics for social networking sites, video, social bookmarking sites, blogs, and microblogs, such as Twitter.

Technical Limitations of Web Analytics

While very useful for tracking customer behavior online, there are currently several key limitations in web analytics software that are particularly relevant to PR. First, cookies must be enabled on individual computers in order for JavaScript to detect unique visitors and track their behavior. If an individual uses multiple computers, works at a company with strict IT regulations that do not allow cookies, or is using a mobile device that does not support JavaScript (most newer devices, such as the iPhone, do support JavaScript), their behavior will not be tracked reliably. Also, if customers tend to delete their cookies frequently, it might not be possible to link their behavior on your website with their initial referral source.

Most of the data issues associated with using JavaScript and cookies will likely amount to little more than "random error" in an organization's PR measurement program. Unless you believe that a particular customer segment is more likely to be visiting your website through a device that does not support JavaScript than other segments, there is little reason to believe there will be systematic errors in the type of data that matters most to communications professionals. In general, web analytics data should be reliable enough to create a good *model* of how earned online media are affecting online business outcomes. It is important to remember, though, that web analytics are imperfect, and will never provide a 100 percent accurate picture of where each customer and website visitor came from, or how they behaved once they arrived at your organization's website.

One of the largest challenges that PR faces in using web analytics is that most web analytics software is designed to track search and click-throughs (*i.e.*, an individual clicking on a hyperlink that directs them back to your website or a landing page). Although some brands are likely to consistently have hyperlinks to their website in earned media, most organizations will not. If a link to the organization's website is not available in a news story or blog post, a customer would have to type the organization's URL directly into their web browser, which would consequently be tagged by the web analytics software as "direct traffic." Luckily, there are some new web analytics technologies that track "view-throughs" as well as "click-throughs." Tealium (Tealium.com) is one provider with this capability using the same cache-checking technology as Google to determine which sites a search user has visited (the sites that appear as purple instead of blue in Google search results). Tealium can track which visitors came from up to 300 earned media sites, providing PR professionals with insight into a much wider range of earned media referrals than click-throughs.

Conclusion

For companies that use PR to drive activity at one or more web properties, web analytics are a necessary part of measurement. Since web analytics technology has some limitations and most organizations sell products and generate sales leads through offline channels, web analytics might not be the “holy grail” ROI measurement system that the PR industry has been waiting for. That being said, it might be the closest thing yet. Web analytics can give PR professionals insight into the value of PR and the types of business outcomes it helps drive in a way that has not been possible without extensive (and sometimes cost-prohibitive) primary research. In much the same way that online advertising has revolutionized how advertisers can measure and optimize their efforts, PR can leverage exactly the same tools and techniques to measure actual user behavior and optimize campaigns to get the best outcomes.