

## ◆ Online Reader Usage across Disciplines

*Moderator:*

**Bill Silberg**  
Commonwealth Fund  
New York, New York

*Speakers:*

**Kent Anderson**  
New England Journal of Medicine  
Waltham, Massachusetts

**Stewart Wills**  
Science  
Washington, DC

*Reporter:*

**Alex Williamson**  
BMJ Journals  
London, England

Why do we spend so much time wallowing in the mounds of data that our Web sites generate? Because the data are so much more precise than those generated from circulation, syndicated studies, eye tracking, and commissioned qualitative and quantitative analyses of our print products. We can use Web data to set goals and strategy, improve editorial content, conduct reader and user surveys, and leverage technology to serve the needs of readers and authors better. Bill Silberg introduced the session with those opening comments and then stimulated us by asking what we should consider tracking: total visits, unique visitors and demographics, hits and page views, downloads, session time, return visitors, referring and trailing sites, browser geographic analysis, Clickpath, or open and click-through rates? A staggering array of data can be generated, but tools are available to help us. Silberg mentioned a few: traffic usage software including Web Trends, HitBox, DeepMetrics/LiveStats, and custom tools. Among the online survey tools, SurveyMonkey and Zoomerang

were recommended.

According to Kent Anderson, the *New England Journal of Medicine (NEJM)* recognized years ago that its “data sucked”: they were incomplete, hard to analyze, and deposited randomly in various databases and table structures. By leveraging the wealth of data generated by the Web site, *NEJM* has recently launched its author center. This site, exclusively for authors, allows each author to view the print circulation for the issue in which his or her

*Wills cautioned that  
“downloads don’t  
necessarily mean  
readers”.*

article was published, its online usage by month, any media coverage generated, its citations elsewhere with toll-free links where appropriate, and any syndicated use. The work involved parsing weekly print circulation statistics and passing them to the online host, passing usage and citation data into specific bins, creating mass-media citation-tracking databases, and passing them to the online host.

*Science* is a multidisciplinary journal, and usage differs among disciplines. Therefore, data should be compared only within subjects or fields. The tools for tracking usage consist of server logs, summary logfile compilations, custom tools, and Web trends reports. The weekly statistics from HighWire and a suite of additional tools provide an array of usage statistics plus user habits. Before we try to draw meaningful comparisons, we need to consider usage over time. In the first week or two

of publication, usage is high, but it quickly declines and then has a long tail that is significant. Early use is attributed to current awareness, news value, publicity, links, and many other factors. The later usage might be defined as “quality” usage when the article has become valuable and is being used by other researchers.

Is there a correlation between usage and impact factor? From the *Science* data, it appears that a correlation exists, but only after time, so no quick answers there! Featuring articles on the home page contributes to higher usage, but maybe that is only a short-term phenomenon. Stewart Wills proposed that a good comparative measure might be “interest factor”, which equals the ratio of article usage to average usage of articles in the same subject.

Another topic of controversy is what to count. Plenty of measures are available: abstracts, full-text HTML, PDF. Wills cautioned that “downloads don’t necessarily mean readers”. He suggested that the ratio of full-text views to abstract views might be used as a possible “quality factor”. Usage data can also be muddied by access-control problems and robots, automated downloads, and all sorts of other mischief. In addition, users take holidays: there are traditional downturns in usage around Christmas and over the summer.

Finally, Web sites are great places for conducting online surveys, but beware of oversurveying. Users tire quickly. A number of survey tools were recommended that could build a survey on the fly.

For the editorial team, usage data can be a gold mine and can guide editorial policy. Marketing teams will also find the data invaluable. However, data have their limitations; Web data are no exception, and this needs to be recognized in any interpretations. 