

## Health System Web Statistics: Increasing Data Accuracy

To provide Departmental Web Managers (DWM) and the groups they represent with effective data on how their Health System websites being used, the Web Development Center (WDC) is in the process of changing the current method for analyzing web use. This change will provide a more accurate accounting of pages being viewed and user behavior while visiting those sites.

The Health System web environment uses a caching system to improve overall server performance. The caching system stores copies of frequently requested pages on our websites. When users request these pages which have recently been served, the server sends them the cached copy of the web page instead of retrieving it again from disk. The current implementation of the web statistics package, Google's Urchin, uses a method called IP+Agent. Because this method is susceptible to caching, not all web pages sent to users from the server cache are counted. In some instances, pageviews are undercounted using the IP+Agent method.

The new tracking method being implemented is Google's Urchin Traffic Monitor (UTM). This method is designed to negate the effects of caching and more accurately track pageviews and visitor activity on Health System websites.

The UTM method is more accurate because it relies on tracking codes added to all Health System web pages located in the Web Developer's Toolkit. The addition of tracking codes allows every page visit to be counted. In some instances, individual website pageview data will actually increase because UTM counts all visits to a web page, even if the page was loaded from the server cache.

There are some scenarios where overall website numbers will actually decrease with the UTM method. The existing IP+Agent counts certain file types as a pageview, including: PDF files, Word documents, PowerPoint presentations, Excel spreadsheets, Adobe Flash files, etc. Websites that include many of these types of files will see an overall decrease in website traffic reported because UTM classifies these files as file downloads and web pages. Only web page formats are counted as pageviews (HTML, CFM, etc.).

### Implementation of UTM

The WDC tested the UTM method in late 2007. At the conclusion of testing, UTM was implemented on the production servers in December to begin collecting live data for the Health System website. The Marketing department was enlisted to pilot this method prior to house-wide release.

During testing, it was discovered that several websites did not include the new tracking codes. This caused an 18-day gap in data collected for several Health System websites (no data exists for January 1-18). If departments require this data, please contact the WDC at: [wdc-support@virginia.edu](mailto:wdc-support@virginia.edu). The WDC can work with you to provide missing data from IP+Agent generated reports.

To ensure no data is lost between reporting methods, the WDC will use both tracking methods throughout 2008 (IP+Agent and UTM). However, UTM-generated reports is the default.

**Please note:** Since IP+Agent and UTM are different web traffic tracking methods, they generate reports from different sets of log files. The WDC is unable to generate new reports using the UTM method for data from previous years. The new reports will only be for 2008 going forward. Reports for all previous years will continue to use the IP+Agent method. Those reports are still available.