

Web Trends vs. Google Analytics for P2RIC.ORG

Web Trends (WT) gathers its data via server request logs. All HTTP requests and responses come into the server, and they are logged line by line. Web Trends extracts the data from each line and uses algorithms to calculate top visitors, top pages, and visits.

Google Analytics (GA), however, does not use server-side logs for data collection. Instead, Google Analytics only registers hits that can be processed by JavaScript. Once the JavaScript code has been loaded into the browser, extra client information can be gathered using GA's interface on browser types, OS, screen resolutions, even Flash or Java compatibility.

Server vs. Client Data

Server

On the server side, generally this information is collected at the raw data level:

- Request Date/Time
- Requestor's IP
- Request URL
- Response Code
- Number of Bytes Served
- Referring Page
- Query String Values
- User Agent String

From this, WT will calculate visits, most popular pages, HTTP errors, and referring pages. This data can be processed well enough that by combining IP and Request D/T, unique visitors are counted. By combining Referring Page and Query String values, search engines and search terms can be counted. The User Agent String can be used to detect OS and some browser types. Other combinations yield the data that we are used to seeing in the WT reports.

It is important to know that since this is at a *server* level, ALL HTTP requests are logged. This includes 404 Not Found, 401 UnAuth, 403 Forbidden, 302 Temporary, and 301 Permanent redirects, for ANY request. This is where the key difference between WT and GA comes in.

Client

By using JavaScript – a client-side scripting language – a service is able to gather much more user-based information. Client side data may include the following:

- Operating System
- Browser Name
- Screen Resolution and Color Depth
- Multimedia Capabilities (If Flash, Java, QT, etc plugins are loaded)

- IP Address
- Languages
- Connection Speeds
- Access to cookie values with some XSS may be possible

The data is not so different from server side requests, aside from the addition of specific browser capabilities which aren't recorded in server logs.

Problems with Relying on Google Analytics (JavaScript Data Collection) Alone

The JavaScript method of tracking data is only able to count successful HTTP requests that result in a viewable page. That viewable page must then load JavaScript into the browser. Most HTTP error pages are coded at the server level, and do not have JavaScript output. Let's consider the case of Google Analytics specifically.

In order for GA to process a page, the webmaster must add a JavaScript tag to the page. This can be done using a global header as it is with p2ric, so all pages that use the header will load the GA JavaScript tag. That's the rub. Any page that does not load the GA JavaScript tag will not be counted.

Times or requests where JavaScript is not usually included or cannot be processed:

- Direct file requests: readme.txt, AwesomeMovie.mov, FullPicture.jpg
- RSS or XML Feeds
- Server-side error pages (HTTP 401, 403, 404)
- Server-side redirects (HTTP 301, 302)
- Any page which does not include the JS header
- Any device or browser that does not support JavaScript (mobile, JS off)

As a result, many hits which would be counted using WT or other server-side tools would never be counted in GA.

Sample Data

Quickly, to compare one case.

For the 2008 Q4 Period, Web Trends reported 11,839 unique visitors. Google Analytics reported 2,541. However, looking into the data more, we see that the top page WT counted is the p2ricnews_rss file. Google would never be able to report this.

Sharing means Caring

The value that Google Analytics brings to the table is its ability to show information that is backed by Google's services. Bounce rate can be tied to AdWords. "Conversion Goals" can be set and watched. Google can automatically detect search engine and bot hits and count them separately than users. Since pages are loaded into JavaScript, there is a likelihood that the data collected in GA is a more-accurate representation of real-people activities. However, it does not count the other data feeds from the site.

WebTrends, on the other hand, is able to count all the missing requests that Google Analytics does not. Direct document requests, XML feeds, RSS feeds, and non-JavaScript page loads are still counted instead of being lost.

Each has a purpose and can be used to enhance the effectiveness of the site. We do not recommend using one over the other. A server side trending tool and a client side trending tool must be used together for the most accurate information.