Why does Urchin report different values than some other systems?

1. Overview

Among commercial web analytics packages, you will normally find fairly similar numbers being reported for most common measurements. However, calculation methodologies differ sufficiently that it is unlikely that you will see identical numbers between any two systems, other than for very simple calculations such as hits. It is usually the case that definitions of terms such as Visitors, Sessions, Pages, etc. vary significantly from vendor to vendor, so before you can compare any values, you must first fully understand how each of the different reporting systems calculate them. The following information provides an overview on how Urchin arrives at several common reporting parameters. This information will allow you to do your own comparative analysis between the analytics software you are considering for your organization.

2. Visitors and Sessions (Urchin)

The Visitors and Sessions reports exist specifically to provide information about unique visitors and unique visitor sessions. Urchin's definition of a Visitor requires a web transaction that can be positively identified as unique. This means that Urchin can track that visitor if they return to the site even a month or more later. Urchin does not use the term "visitor" in any other context. If Urchin cannot identify a unique quality about the visitor, it uses the term "session." Here is an example of why that is the case: If a visitor uses a dial-up service to reach your website, that person will have a different IP address every time he/she visits your site. Many lesser web analytics solutions use IP addresses to identify visitors. So, if a visitor comes to the site 10 times in one week (with a different IP every time), those packages would report 10 different visitors, which is of course highly inaccurate. So, rather then identify such a transaction as a "visitor", Urchin identifies these visits as "sessions". It should also be noted that Urchin 5 employs special logic to handle the case of AOL visitors, where the proxying behavior of AOL may cause each click from the visitor to originate from a different IP address. For web analytics solutions that track strictly based on the IP address of the visitor, this will result in highly inflated session counts for AOL visitors. For sites that have a lot of AOL traffic, this effect can be significant. Urchin's additional logic helps to mitigate the effect of the AOL proxy and provide a more realistic view of AOL visitor sessions. The net result is that Urchin may show a more accurate, but lower overall session count than other analytics packages when AOL visitor traffic is factored in. In order for Urchin to identify the existence of a Unique Visitor, Urchin must recognize a particular cookie format, the Urchin Traffic Monitor (UTM), in the web server log file. This is a 3-part, first-party cookie system that allows Urchin to track first time and returning visitors with a high degree of accuracy. The UTM is a small quantity of javascript that is installed on each page of a website, usually via server-side includes, that calls a 1x1 pixel images, thus forcing a hit to the server. This hit includes information on which page was requested, the computer environment the visitor used, the visitor's unique identifier code, and other information. Please see the UTM articles under the Visitor Tracking section for more information.

3. Visitors & Sessions (ASP systems)

ASP systems calculate unique visitors by using a 3rd-party cookie system (the visitor being the 1st, the website being the 2nd, and the ASP being the 3rd). This requires that a bit of javascript code be embedded on each web page in a similar manner as the UTM (the chief difference being that the UTM uses a first-party cookie). When the page is requested, that javascript invokes the cookie handling from a remote location managed by the ASP provider. This approach is flawed for several reasons, including these:

Many visitors have cookies disabled

Many browsers, such as Internet Explorer 6, give security warnings when 3rd-party cookies are used

Non-HTML files, such as images and PDF downloads, cannot be tracked

Basic information such as hits and bytes transferred cannot be tracked

4. Pages

Urchin5

Urchin 5 uses both exclusive and inclusive "mime type" filters. One is applied to pages and the other to downloads. By default, Urchin will include any file with one of the following extensions in the Downloads report: .pdf, .zip, .exe, .sh, .tar, .gz, .pkg, .doc, .xls, .ppt. Any files that don't match one of those extensions will be included in the Pages report, unless they match one of the following image extensions: .gif, .jpg, .jpeg, .png, .js, .css, .cur, .ico, .ida. As with Urchin 4, pages that result in an error are reported on in the Status/Errors report.

New in Urchin 5

Urchin 5 has two different methods of categorizing valid pages. The method used depends on the Visitor Tracking method used for the report:

If Urchin is configured to use the Urchin Traffic Monitor (the most accurate reporting technique), the pages reports will only include pages that were viewed by visitors that were tracked with UTM cookies. However, there are usually many other pages viewed that are not requested by visitors. For example, search engine robots, spiders, and other monitoring agents may account for a significant number of the total pages viewed. Those pages are only displayed in the All Files report and Robots by Hits report.

Non-UTM

All other Visitor Tracking Methods used with Urchin 5 use the same parameters to quantify pages. In order for the page to display in the Pages reports, they must be requested from valid visitor sessions. This means that pages requested by known search engine robots will be moved to the All Files and Robots by Hit reports and not appear in the Pages reports. You may notice that the Pages reports contains less pages when the UTM is used. That is because Urchin is better able to track valid visitors and will move reporting on pages requested by non-valid visitors to the aforementioned All Files and Robots by Hits reports as a result, since the pages were not requested by human visitors

5. Referrals
Referral data is taken directly from the beginning of each session in the web server log file. Accuracy of referral reporting depends on the availability of accurate referral log data. It is common for many referral hits to be missing from the local web server log file. This happens when a remote cache server responds to a page request instead of the local web server. When that happens, the hit that contains the referral information

does not get recorded. As a result, the web analytics software cannot determine where the referral for a given session came from. It is important to note that when the UTM is used, all referrals (even those that are cached) are forced back to the local web server and accuracy in referrals is restored - another reason to always use the highly accurate UTM tracking method. Just as with the Pages report, Urchin 5 only lists Referrals that are generated by real human visitors when the UTM is being used.

https://support.lexiconn.com/kb/questions/21/