

# USING GOOGLE ANALYTICS AND IMPLEMENTING THE MIRROR METHODOLOGY

**Government Website Traffic Audit Scheme 2009**

Version 1.0

## Introduction

Google Analytics is a free web analytics package provided by Google. It is not the same as Google Urchin, which is a separate software web analytics product supplied by Google. Neither Google Analytics nor Google Urchin are currently ABCe 2-star Associate products. Google Analytics supports the reporting of Page Impressions, Unique User/Browsers, Visits and Visit Durations. It is optimised to plug into Google's AdWords and AdSense paid search tools. Because it is zero-cost, it is used by a large number of publishers.

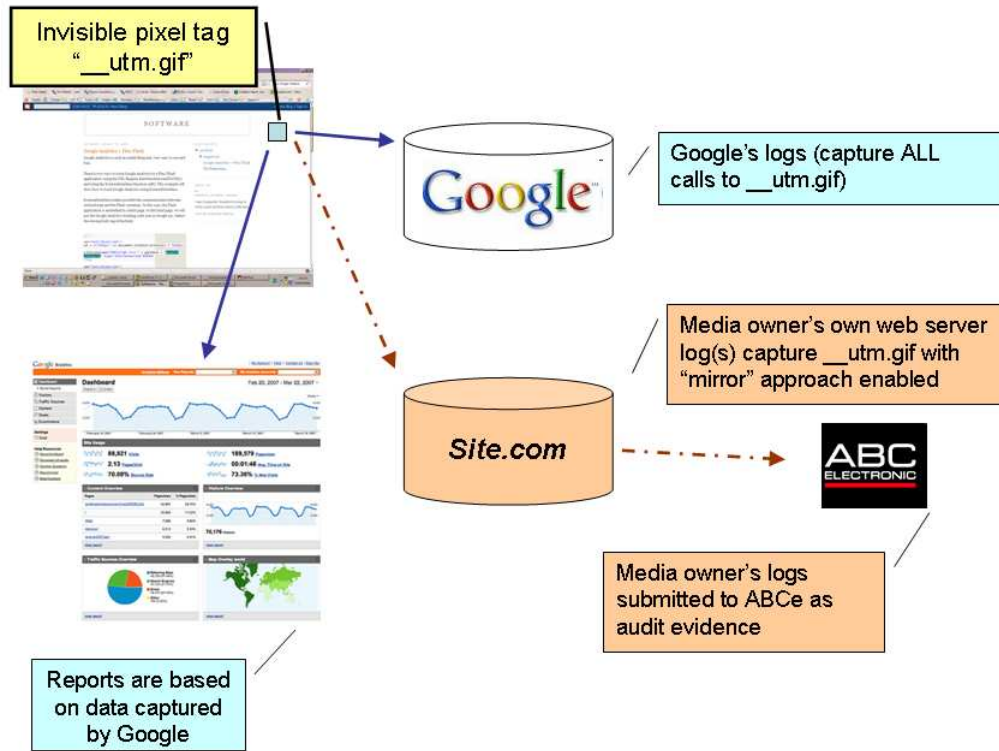
**Fig. 1 Sample Google Analytics Dashboard**



Google Analytics is a Javascript page tagging service based on Urchin, an analytics tool that was acquired by Google. Please note, however that, as stated above, Google Urchin is a separate software tool self-hosted by the site owner and is not covered by this paper; for details on ABCe auditing using Google Urchin please follow method [c] or contact COI at [transformationalgovernemnt@coi.gsi.gov.uk](mailto:transformationalgovernemnt@coi.gsi.gov.uk). Site owners wishing to run the Google Analytics code must incorporate some HTML into each page they wish to tag. This HTML makes a call to Google's servers, which host the Javascript code. The site owner does **not** host the Javascript code on their own web-servers.

The ABCe website audit requires the submission of supporting (logged) data. For legal and privacy reasons, Google have hitherto not released their data for any purpose, and so ABCe has suggested that site owners “mirror” the log file data. In other words, the site owner should configure their Google Analytics tag (or write their own supplementary code) to ensure that every request sent to Google has a matching request logged by a server under the site owner’s control.

**Fig. 2 Google Analytics Logging**



## Technical Assistance

The code exists in two versions. The older version (Version 1, “Old Style”) is:

```
<script type="text/javascript" src=" http://www.google-
analytics.com/urchin.js "></script>

<script type="text/javascript">
_uacct = "UA-xxxxxxx-x"
urchinTracker();
</script>
```

And the newer version (Version 2, “New Style”) is:

```
<script type="text/javascript">
var gaJsHost = (("https:" == document.location.protocol) ?
"https://ssl." : "http://www.");
document.write(unescape("%3Cscript src='" + gaJsHost + "google-
analytics.com/ga.js' type='text/javascript'%3E%3C/script%3E"));
</script>
<script type="text/javascript">
var pageTracker = _gat._getTracker("UA-xxxxxxx-x");
pageTracker._trackPageview();
</script>
```

Guidance on how to set up the mirroring approach appropriately to the version of the Google Analytics code used on the site owner's website is available from:

<http://www.google.com/support/googleanalytics/bin/answer.py?answer=76305#local> .

The mirroring approach will direct requests for \_\_utm.gif to a different set of server logs, which should be accessible to the site owner. The site owner can then filter the logs to show only the \_\_utm.gif calls and submit this data to ABCe in support of the Page Impression, Unique User/Browser, Visit and Visit Duration claims reported by Google Analytics.

Site owners who are expecting a high level of traffic should consider directing all \_\_utm.gif requests to a dedicated server handling only these requests (for example, reportingserver.site.com). This may require additional technical configuration such as using a load-balancer to route all \_\_utm.gif requests to such a server. ABCe would then only need the log files from this server (or group of servers). This approach enables a site owner who uses Google Analytics to supply ABCe with a copy of just the data on which the Google Analytics reporting is based for auditing purposes.

ABCe recommends that site owners who have issues configuring Google Analytics, and who require a full expert opinion, follow Google's advice and request the services of a Google Analytics Consultant, usually a third party consultancy accredited by Google, via the following URL: [http://www.google.com/analytics/support\\_partner\\_provided.html](http://www.google.com/analytics/support_partner_provided.html).

## Notes

1. ABCe cannot provide consultancy services beyond suggesting the above approach. Furthermore, the cookie identifier used by Google Analytics ("utma") for the counting of Unique User/Browsers contains the domain information. Therefore, any network comprising more than one top-level domain will not be counted in compliance with the industry-agreed standard, since this requires all cookies to be persistent across all certified domains. Our understanding is that Google Analytics can be configured to count across a "tree" of sub-domains, but not across more than one top-level domain. Google provide instructions on configuring across sub-domains to report Unique User/Browsers compliantly – see: <http://www.google.com/support/googleanalytics/bin/answer.py?answer=74981> for Version 1 of the code, or: <http://www.google.com/support/googleanalytics/bin/answer.py?answer=55524> for Version 2 of the code.
2. In the light of this, some site owners have used the Google Analytics tag but chosen to count Unique User/Browsers via a different cookie implemented on their own servers collecting the mirror data, which will be logged because they are using their own server logs. Alternatively, IP + User-Agent combination can be used to count Unique User/Browsers and will **not** be subject to the domains issue outlined above. However, in **both** of these cases, the Unique User/Browser numbers reported by Google Analytics will probably **not** be substantiated by the data provided for audit – even though the Page Impressions may be.
3. The site owner's "mirror" log files will need to comply with the industry-agreed reporting standards for the minimum fields required which are:
  - Date
  - Time
  - IP Address (of the user) - not the resolved name

- URL and Query (Domain will be found within the query string of the \_\_utm.gif, as will Referrer of the page)
- User-Agent
- HTTP Method
- HTTP Status
- Cookie (if used)
- Referrer (which will show the URL of the page, since it is the Referrer of utm.gif)

The data submitted for auditing will also need to be in a single uniform format for ALL web servers. Furthermore, ABCe recommends that if the site owner expects to count more than 20 million Page Impressions in the month, the Google Analytics .gif calls should be captured in a separate dedicated log file. This will decrease the time and cost required to isolate these .gif calls. Note also that Google state the following in <http://www.google.com/support/analytics/bin/answer.py?answer=55476&ctx=sibling>:

“Google Analytics is a free service that offers users up to 5 million pageviews a month. If your site generates more than 5 million pageviews per month, you will need to link it to an active AdWords account in order to avoid interruption of your Google Analytics service.”

4. Google Analytics supports Unique User/Browsers (reported in Google Analytics as “Absolute Unique Visitors”), Page Impressions, Visits and Visit Duration. To report Visit Duration, in accordance with the Industry agreed standard, it is necessary to select all non-bounce Visits as the basis of the calculation i.e. to exclude from the calculation all single page Visits.