COI

Measuring website usage

Version:	1.0
Date created:	14/01/2009
Policy official:	Adam Bailin
Date last updated:	10/03/2009
Date issued:	March 2009
Lead official:	David Pullinger
Guidance number:	TG116

Purpose

This guidance for central government describes a common approach to measuring website traffic. This enables departments to answer Parliamentary Questions and Freedom of Information Requests about website usage consistently and reliably. It also facilitates the development of performance measures for use in the planning, design and evaluation of websites.

Table of contents

Introduction

Purpose

This guidance for central government describes a common approach to measuring website traffic. This enables departments to answer Parliamentary Questions and Freedom of Information Requests about website usage consistently and reliably. It also facilitates the development of performance measures for use in the planning, design and evaluation of websites.

Background

The National Audit Office surveyed 153 central government organisations between October and December 2006 for the report Government on the internet[1]. Sixteen percent of organisations did not provide data about how their websites were being used.

The Government has agreed to the Public Accounts Committee (PAC) Sixteenth Report[2] recommendation that a single set of reporting metrics for website usage needs to be developed. This guidance describes those metrics and how to measure them accurately and consistently so that website usage can be audited against industry standards.

Additionally the PAC recommended that a method for measuring website costs and a single set of website quality measures be identified. These requirements are covered in two separate pieces of guidance. These three pieces of guidance provide central government with a common set of tools to evaluate the cost-effectiveness of websites.

Audience

This guidance is intended for Heads of e-Communications (or equivalent) within central government departments, executive agencies and NDPBs.

Definition of terms

The following terms are used throughout this document:

- **Page Impression** (or **Page View**)**:** A file, or combination of files, sent to a valid user as a result of that user's request being received by the server.
- **Unique User/Browser** (or **Unique User** or **Unique Visitor**)**:** A unique and valid identifier. Websites may use:
    i.) IP address + User-Agent and/or
    ii.) Cookie

[1] Government on the internet: progress in delivering information and services online
http://www.nao.org.uk/publications/0607/government_on_the_internet.aspx
[2] Public Accounts Committee – Sixteenth Report
http://www.publications.parliament.uk/pa/cm200708/cmselect/cmpubacc/143/14302.htm

- **Visit:** A series of one or more Page Impressions, served to one valid user, which ends when that user has not made a Page Impression for a 30-minute period.
- **Visit Duration:** The total time in seconds for all Visits of two or more Page Impressions, divided by the total number of Visits of two or more Page Impressions.
- **Website** (or **site**)**:** One or more domains grouped together by the Site Publisher as a single entity.
- **Web metrics:** the measures used to quantify the performance of a website (e.g. Page Impression, Unique User/Browser, Visit, Visit Duration etc.)
- **Web analytics:** the analysis, interpretation and combination of web metrics in order to understand and optimise website usage.

The terminology used in web analytics tools may differ from the standard industry terms. For example, many tools use the term Page Views rather than Page Impressions and Unique Visitors rather than Unique User/Browsers.

Reporting requirements

1. The following web metrics, as defined by the Joint Industry Committee for Web Standards[3] (JICWEBS), must be measured for each and every publicly accessible website operated by an organisation:
    - Unique User/Browsers
    - Page Impressions
    - Visits
    - Visit Duration

2. Central government departments must measure Unique User/Browsers, Page Impressions, Visits and Visit Duration starting from 1 April 2009 for every website open on 1 April 2010.

3. Executive agencies and non-departmental public bodies (NDPBs) must measure Unique User/Browsers, Page Impressions, Visits and Visit Duration starting from 1 April 2010 for every website open on 1 April 2011.

4. The following information must be provided to COI at the end of each quarter:
    - Number of monthly Unique User/Browsers
    - Number of monthly Page Impressions
    - Number of monthly Visits
    - Number of Visits of at least two Page Impressions
    - Total time in seconds for all Visits of at least two Page Impressions

5. Each report should contain figures for each of the previous three months. This information should be provided in the format shown in the reporting template in Appendix A.

6. All figures should exclude internal web development activity, performance monitoring, automated broken link detection and other types of non-human activity (e.g. robots and spiders). Further details on what to exclude are found in the Page Impressions section.

[3] Joint Industry Committee for Web Standards http://www.jicwebs.org/standards.php

Auditing requirements

7. Unique User/Browsers, Page Impressions, Visits and Visit Duration, must be audited in line with the industry-agreed standards defined by the Joint Industry Committee for Web Standards[4] (JICWEBS).

8. Website auditing will be carried out on behalf of COI by the industry-owned auditors ABCe.[5] Site owners should contact COI[6] to register for website auditing services. Details of website auditing costs are found in Appendix B.

9. In the Financial Year 2009-10, at least one month's data needs to be audited. This month may be nominated by the site owner but will be subject to availability. Site owners are encouraged to register for auditing as early as possible to ensure their preferred month is audited. COI will work with the site owner and the auditor to select an alternative month if the preferred month is not available.

10. In the Financial Year 2010-11 and thereafter, at least one month's data needs to be audited every six months with the month chosen at random by the auditor from each half year.

11. There is a three month lead time from the end of the audit period to the issuance of a certificate. Hence, central government departments must ensure that every website planned to be open on 1 April 2010 has an audit scheduled for a period ending no later than 31 December 2009.

12. Executive agencies and NDPBs must ensure that every website planned to be open on 1 April 2011 has an audit scheduled for a period ending no later than 30 June 2010.

The auditing process

13. The auditing process consists of the following key steps:
    i. Register for website auditing (contact COI)
    ii. Collect the data
    iii. Filter the data
    iv. Make a claim

    Auditing website usage is not a simple pass-or-fail process. The ABCe audit will seek to ensure that the website publisher is clear on any issues of non-compliance and that an iterative process is used to enable certifiable totals to

[4] Joint Industry Committee for Web Standards http://www.jicwebs.org/standards.php
[5] ABCe http://www.abce.org.uk/
[6] mailto:transformationalgovernment@coi.gsi.gov.uk

be established. However, it is essential that data is collected in the correct format for auditing. Further details are found in 'Audit guidance' (TG116a).

14. Data must be filtered according to the specific rules outlined in TG116a. An overview of what to exclude is shown in the Page Impressions section.

15. Claims (statistics for the selected month, broken down by day, per metric) should be entered on a standard ABCe Audit Claim Return Form. All claims and data should be submitted to ABCe within 5 working days of the end of the audit month. Please see TG116a for further details.

16. COI will work with the site owner to help with the audit process to ensure that the data supporting the claim is transferred securely to the auditor. This is normally done using the Secure File Transfer Protocol (sFTP).

17. Further details of the auditing process, including technical specifications, are found in TG116a. Site owners should inform their web analytics providers that they intend to carry out and ABCe audit and refer them to TG116a for technical guidance. Please email transformationalgovernment@coi.gsi.gov.uk for a copy of TG116a.

Why audit?

18. Auditing website usage data is essential to ensure that the rules for measuring the number of Unique User/Browsers, Page Impressions, Visits and Visit Duration have been implemented correctly. This guidance aims to help you avoid the common pitfalls in measuring website usage. However, the best way to ensure the figures are accurate is to have them audited by an independent third party.

19. Auditing website usage figures means greater confidence in their accuracy. This confidence enables departments to define web-based key performance indicators (KPIs). This in turn enables departments to set business goals and measure the progress of a website towards them, which is an essential part of strategic communications planning.

20. The value of auditing to government websites is:
    - The resulting certified numbers can be compared with confidence.
    - Audited data gives internal management transparency in the information presented.
    - This data can be trusted as consistent and therefore reliable both over time and between websites.
    - Being able to make such comparisons is critical to the benchmarking of individual websites.
    - Audited data offers proof to management of usage trends, avoiding mis-statement.

- Trend analysis that is based on data certified to industry standards ensures better investment decision-making.

21. In summary, auditing website usage delivers key business intelligence that an online service requires to aid its management in making the right decisions for its future success.

Minimum standard for web metrics

22. **Web metrics** are the measures used to quantify the performance of a website, for example Page Impressions, Unique User/Browsers, Visits, Visit Duration and so on. These metrics are defined according to industry standards set by the Joint Industry Committee for Web Standards[7] (JICWEBS). Web metrics should be referred to using the current JICWEBS standard terms.

23. Avoid use of non-standard terms such as 'hits' which do not provide reliable or comparable measures of website usage.

24. As a minimum, central government organisations must be able to measure the following web metrics:
    - Unique User/Browsers
    - Page Impressions
    - Visits
    - Visit Duration

25. The number of Unique User/Browsers, Page Impressions and Visits; the number of Visits of at least two Page Impressions; and the total time in seconds for all Visits of at least two Page Impressions must be reported on a quarterly basis as specified in the reporting requirements.

Unique User/Browsers

26. The number of Unique User/Browsers is the total number of unique devices (e.g. computers or mobile phones) that have made requests to the site in the period being measured.

27. The number of Unique User/Browsers gives an indication of the size of audience of a website over a given time period. However, it is impossible to tell how many *people* are using a site from looking at the data. We can find out the number of unique *devices* (e.g. PC browser, mobile phone, games console etc.) accessing the site by capturing and analysing a unique 'valid identifier' for each device. Valid identifiers include:
    - IP address + User-Agent (browser and operating system)
    - Cookie (a persistent not a session cookie)

28. There are several scenarios which may alter the number of Unique User/Browsers counted, for example:
    i. Several devices accessing the site with the same IP address (e.g. behind a corporate firewall)
    ii. A single device accessing the site with different IP addresses (dynamic IP address allocation)

[7] Joint Industry Committee for Web Standards http://www.jicwebs.org/standards.php

iii. Several people sharing the same device (i.e. having the same IP address + User-Agent or cookie)
iv. The same person using different devices to access the site (i.e. changing their IP address and/or User-Agent or cookie)

These effects are accepted by the industry and the number of Unique User/Browsers need not be altered as they are seen as having a 'balancing out' effect. The number of Unique User/Browsers is still seen as a useful comparable measure of audience size.

Page Impressions

29. The number of Page Impressions is the total number of requests (e.g. mouse clicks) made for a site's content by users of the site (i.e. unique devices) in the period being measured.

30. The number of Page Impressions gives an indication of the **volume** of content being delivered by the website over a given time period. In other words, it provides a measure of how busy a website is.

31. The number of Page Impressions is not the number of files returned by the server. A single request by a user (e.g. a mouse click) may result in a whole collection of files being returned, for example an HTML file plus a style sheet and several associated images. The key principle is that one page request is counted as one Page Impression.

32. To count the number Page Impressions accurately, web analytics tools must be configured to exclude certain types of page request, for example:
    - Non-human activity (i.e. robots and spiders)
    - Error messages (e.g. 404 errors - "file not found")
    - Component files such as style sheets, images etc.
    - HTTP request types[8] other than 'GET' and 'POST' (e.g. 'HEAD')
    - Redirects and concurrently served pop-ups

    A complete list is found in TG116a 'Audit guidance'.

33. Figures should also exclude internal web development activity, performance monitoring, and automated broken link detection. If IP addresses cannot be used to filter out these activities, for example where the GSI firewall displays a single IP address for all users from the host department, then an exclusion list should be created to filter out PCs used for site development work. Implementation will depend on local circumstances, but it may be possible to use a combination of cookies and User-Agent descriptions to maintain the required filter of site development and maintenance PCs.

[8] Hypertext Transfer Protocol – HTTP 1.1http://www.ietf.org/rfc/rfc2616.txt

34. If a website uses PDFs to deliver content, Page Impression figures should exclude partial file downloads (i.e. when the browser downloads a file in separate parts to manage the download efficiently). Such partial downloads are recorded in the log file with a 206 status code.

Visits

35. A Visit is a single period of activity by a Unique User/Browser which ends when that Unique User/Browser has not made a valid request for content for a 30 minute period. The idea is to capture a single user's almost continuous burst of interaction with the website. The 30 minute period may seem arbitrary but it is the industry standard definition as defined by JICWEBS. Tools must be configured to set the timeout period between Page Impressions to 30 minutes.

36. The number of Visits doesn't mean very much on its own. However, the number of Visits per Unique User/Browser provides a measure of **frequency**, or how often a user is returning to the site. This is important from a marketing perspective as it suggests a level of customer loyalty. A high number of Visits per Unique User/Browser indicates that a site is one to which users tend to return.

37. The number of Unique User/Browsers is dependent on the period of time over which the data is collected. Therefore the frequency should be calculated over a length of time that it would be expected to observe return visits, for example quarterly.

38. The monthly, quarterly or yearly frequency is an example of how metrics can be combined to form a business KPI. For example, the current quarterly frequency might be 2.2 and the business goal might be to achieve 3 Visits per Unique User/Browser every quarter. By measuring the quarterly frequency, an organisation can monitor progress towards this goal and therefore gauge how well its website is performing.

39. Alternatively, the purpose of a site or page may be to deliver a key piece of information that, as a result, means a repeat Visit is not the desired outcome. In such cases a single Visit may be the optimum outcome and hence the goal may be to show decreasing, rather than increasing frequency.

Visit Duration

40. Visit Duration is the average length of time between the first and last Page Impression for Visits of at least two pages.

41. Visit Duration can provide a measure of the level of **engagement** of website. It can be argued that if users spend longer on a website, they are more engaged. This conclusion should be qualified with further evidence, such as

that obtained through user testing. Otherwise, it might be indicative of a site that is simply difficult to navigate.

42. The use of average Visit Duration as a KPI should be aligned with a site's business goals. For example, if the purpose of a website is to encourage exploration then performance might be measured as progression towards a high average Visit Duration. However, if the purpose of a website is to answer very specific and detailed questions then a low average Visit Duration might be desirable. In each case, it is important to supplement quantitative data with qualitative data such as that obtained through questionnaires or user testing.

Using web metrics for planning and design

43. The number of Unique User/Browsers, Page Impressions and Visits are useful metrics to gauge the size of a website's audience, its volume of traffic and the level of interaction, respectively. However, there are several other industry standard metrics that are useful in website planning and design, some of which are presented in this section. The analysis, interpretation and combination of web metrics in order to understand and optimise website usage is known as **web analytics**.

    Note: A complete list of industry standard metrics is found in TG116a 'Audit guidance'.

Internal search

44. Internal search on a website is tracked through the number of Searches – the number of completed searches carried out on the site. A Search is counted even if the search fails to yield any results for the end user. It represents the number of searches requested by end users of the site.

45. Even if a search produces results, it may be that the results are irrelevant. Therefore, it is important to monitor where users are navigating to after a search is requested. If they are leaving the site then it is unlikely that the search was useful to them. If they go to a particular page, then the search is likely to have been a success, and is said to generate a Search Click. A high number of Search Clicks per Search means that an internal search engine is performing well.

46. Monitoring the words and phrases used in internal searches is helpful to gain customer insight and steer users towards the information they are looking for. For example, if users are searching for a particular term, it is important to feed this into the site's editorial policy so that the relevant pages are more likely to match with the user's query in search results. It is also useful to follow this practice for searches originating in external search engines – a commonly used technique in search engine optimisation (SEO).

Country of origin

47. It is possible to determine country of origin from a user's IP address and therefore to carry out a geographical analysis of a website's audience. This may be of particular interest to websites with an international audience (e.g. Directgov, Number 10, Foreign & Commonwealth Office, Department for International Development).

Whole site analysis vs. granular analysis

48. Figures for the whole site may indicate that the website is known to potential users, search engines and other websites. A more detailed analysis is needed to answer specific questions about whether the services or information provided are meeting user needs. Analytical tools can be configured to show:
    - the most popular pages
    - the most popular areas (groups of pages)
    - usage of a group of web pages and documents relating to a specific subject or campaign
    - how demand for specific pages changes over a selected timeframe (e.g. showing daily activity over a week)

    This type of information can then be used to promote popular sections of the site (e.g. on the home page).

User experience

49. Web analytics can also be used to track user journeys, for example:
    - the search engine or website that led a user to the site
    - entry and exit pages
    - the route taken by the user through the site

50. A Referral In is where a user has requested content from a site where their previous activity has not been with that site. Analysis of Referrals In provides insight into where users are before they visit a website, for example if a user has been referred by a search engine or a link from another website. This is not always the case as a user may have typed a URL directly into their browser or followed a link from an email. Site owners should be in a position to report on Referrals In.

51. Web analytics can be used to inform design improvements to the website. Understanding how individual users move within the website can help web teams to design improvements to the structure of the site. For example, the home page was traditionally seen as the starting point for the majority of user journeys. However, web analytics is showing that this is not necessarily the case because search engines, bookmarks and links from other websites lead users directly to the most relevant page.

52. Deeper analysis of time spent within the website and numbers of pages visited can be used to understand whether goals are being achieved. Visit Duration is a measure of the time between the first and last Page Impression for a Visit of at least two pages. Whilst the average Visit Duration for a site may indicate the degree of user engagement, analysis should be carried out to explore specific user journeys in more detail. For example, a user might

take longer to complete a task if they take a sub-optimal route through the website.

53. Although error pages should be excluded from the audited figures, web teams can use information about ‘file not found’ 404 errors to reduce the number of broken links and improve the user experience.

54. The accurate collection and analysis of usage data is a powerful method for evaluating a website. However, web analytics alone does not paint a full picture of the way a website is used and who it is used by. Web metrics should be used alongside other data in order to gain a deeper level of insight, including:
    - qualitative data (e.g. from online satisfaction surveys, interviews, focus groups etc.)
    - observational data obtained from user testing
    - expert reviews (e.g. heuristic evaluation)

    The most effective website evaluation uses a variety of carefully chosen methods employed at appropriate times in the development lifecycle.

Web analytics tools

55. The two main methods of measuring website usage are:
    - Server-side, where data from the web server logs is analysed;
    - Client-side, where data from the user's browser is analysed (often referred to as page tagging).

    Both methods can support audits, but some tools/services are more audit friendly than others. The differences between the two are described briefly below.

Server-side (log file analysis)

56. Server-side methods
    - record every file requested by a user's browser (one page could consist of ten or more files including images and style sheets which need to be filtered out of reports);
    - contain a great deal of non-human traffic (e.g. search engine spiders and robots);
    - do not count pages fully loaded from caches and mirror sites.

Client-side (page tag file analysis)

57. Client-side methods
    - are anti-robot and hence require less filtering because they tend to be .cgi. or .java scripts and robots tend to ignore scripts;
    - can avoid the problems of under-counting caused by the use of the back button and page caching;
    - can record a great deal more about the way users interact with a site than log file analysis.

What sort of web analytics tool is best for your website?

58. Key considerations when choosing a web analytics tool are:
    i. Ease of use and depth of detail
    ii. Vendor viability and data ownership/access
    iii. Cost
    iv. Support for auditing
    v. What is the primary purpose of the tool?
    vi. What are the site's goals?

    These are discussed in turn.

Ease of use and depth of detail

59. Client-side analytics may involve time-consuming tagging of all elements within a web page. Server logs, on the other hand, tend to be collected by default by the site developer or host. However, work will be required to filter these logs to make the information contained therein accessible.

60. There are major advantages to using client-side web analytics tools: they eliminate extraneous traffic caused by search engine robots; they report on cached pages; and they can also show how users interact with a web site, often in real-time. For example, they can show which fields in a form are more likely to cause a user to abandon a transaction. Client-side analytics can also track other on-page events such as a user's mouse movement across a page. Moving from server-side to client-side analytics can deliver a much richer understanding of the customer journey.

Vendor viability and data ownership/access

61. A potential drawback in using client side technology is "vendor dependency". The user is reliant upon the vendor – both technologically (software and hardware as deployed by the web analytics supplier) and in terms of data access/ownership – he who collects the data owns the data! Server logs, on the other hand, are portable and can be analysed by any number of web analytics tools.

Cost

62. Log file (server-side) analytics tends to be cheaper (though not necessarily in terms of human resource) than solutions that use client-side technology, although the arrival of Google Analytics may have changed this dynamic. At the time of writing, the raw data captured by Google Analytics is not made available to the client for auditing purposes. However, Google offers a workaround which enables a website publisher to collect a mirror copy of the data locally and this can be used by ABCe for auditing purposes. For more information email transformationalgovernment@coi.gsi.gov.uk

Support for auditing

63. A key consideration when choosing an analytics tool is whether or not the tool can support a successful audit. Many web analytics tools and services are available but only some of these have a proven ability to facilitate auditing. Any web analytics package with ABCe Associate Subscriber[9] 2-star rating has successfully delivered auditable data completing an ABCe audit to industry agreed standards, publishing an ABCe Audit Certificate.

[9] ABCe Associate Subscribers http://www.abce.org.uk/cgi-bin/gen5?runprog=abce/abce&type=page&p=current_associates.html&menuid=associates|current_associates#a

Hence, these tools have a proven track record for facilitating auditing. Both server-side and client-side 2-star Associate Subscribers exist. It should be noted that some levy additional charges for facilitating an audit.

64. To prepare for an audit, email transformationalgovernment@coi.gsi.gov.uk. COI will work with you to ensure that analytics data and reports for the month in question are available and can be transferred to ABCe. A full description of the auditing process and associated technical specifications are found in TG116a 'Audit guidance'.

What is the primary purpose of the tool?

65. Web managers have a need to ensure that the servers can cope with demand. In such cases the key metrics to focus on are Page Impressions per hour, most popular pages, and even the amount of data transferred (some pages can be "heavier" than others, depending upon the number of images and other files associated with it). If the primary purpose of the tool is to monitor traffic volumes, then server side analytics may be suitable.

66. If the primary purpose of the tool is to gain insight into user behaviour, client-side analytics may be an appropriate choice. Client-side analytics can provide detailed information about the way customers use a site. Some tools are even able to recreate an individual customer's journey by breaking a Visit down to a series of discrete "events".

What are the site's goals?

67. To choose the right package you will need to know what your management team's requirements are; what are the site's goals? What is the point of the site? These questions tend to be easier to answer on commercial sites because it is possible to measure the Return on Investment (ROI) by seeing whether the site makes money. For instance, if a marketing campaign costs £x, good web analytics will show whether it has sent more traffic to the site and whether this traffic resulted in an increase in sales.

68. With government websites this question is not so easy to answer because many tend to be information-based. The goal could merely be to get users to download a PDF, in which case simple log analysis may suffice. But if the goal were (say) to encourage channel-shift from off-line to online transactions, then the metrics required need to be wider than web analytics and include data about the other channels.

Alternative methods of measuring website usage

69. Alternative methods of measuring website usage include research based on activity by user panels. These panels are selected to be representative of a particular population (which can be a website's target audience) at the time they are recruited. Website usage is measured via a device or software attached to or installed on one of the user's PCs. These statistics are then extrapolated to provide estimates of usage for the total user population. Examples of this type of measurement are provided by Nielsen Online and Comscore.

70. Another approach is to monitor website usage by analysing data collected by one or more ISPs (Internet Service Providers). Again, the figures accumulated can be extrapolated to provide estimates for global usage. This is the approach taken by Hitwise and offers a different insight as they deliver percentage rather than absolute rankings. Independent validation of what is included and excluded is not currently in place for this offering.

71. The above methods provide useful insight into user activity but are outside of the control of the website owner and are not normally audited or verified by an independent industry owned third party. These research services may be used in addition to the auditable methods described above. However, they should not be the only method used for measuring website usage.

Appendix A: Reporting template

Site details	URL			
	Department			
Site usage	Month and year (MMM-YY)			
	Number of Unique User/Browsers			
	Number of Page Impressions			
	Number of Visits			
	Number of Visits of at least two pages			
	Total time of Visits of at least two pages (in seconds)			
Contact details	Name			
	Email address			
	Telephone			

Appendix B: Website auditing costs

The minimum cost per audit is £1,740 provided that:

- the web analytics provider is a 2-star rated ABCe Associate Subscriber[10];
- data is provided in the correct format;
- data is delivered on time.

COI will work with site owners to ensure that when the above conditions are not met, the cost per audit is as close to the minimum as possible.

Full details of the requirements for website auditing, including technical specifications, are found in the accompanying guidance (TG116a). This is available on request from transformationalgovernment@coi.gsi.gov.uk.

There is an additional fee of £200 for audits where the analytics provider is not a 2-star rated ABCe Associate Subscriber.

The above figure is inclusive of a £200 discount for the first 100 registrations. Please contact COI[11] as early as possible to qualify for the discounted rate.

Please note that the above figures are accurate as at March 2009 and non-inclusive of VAT.

[10] ABCe Associate Subscribers http://www.abce.org.uk/cgi-bin/gen5?runprog=abce/abce&type=page&p=current_associates.html&menuid=associates|current_associates#a

[11] mailto:transformationalgovernment@coi.gsi.gov.uk