



## Audit Guidance

Version:	0.9
Date created:	30/05/2009
Policy official:	Adam Bailin
Date last updated:	02/03/2009
Date issued:	tba
Lead official:	David Pullinger
Author:	Andy Flint
Guidance number:	TG116a

### Purpose

This document is intended as a technical annex to TG116 'Measuring website usage'. This document provides a full description of the auditing process and includes technical specifications which should be given to web analytics suppliers to prepare for an ABCe audit.



## contents

<b>Audit Guidance .....</b>	<b>1</b>
<b>Table of contents .....</b>	<b>1</b>
<b>The ABCe Audit.....</b>	<b>3</b>
About this Guide.....	3
The Role and Purpose of ABCe .....	3
Overview of the Audit Process.....	3
Setting up your Web Analytics Tool for Audit.....	4
Self-hosted Web Analytics Tool.....	4
Third Party Analytics Tool or Service.....	5
<b>Using a 2-star Accredited ABCe Associate Web Analytics Product.....</b>	<b>7</b>
General Notes .....	7
<b>Data logging .....</b>	<b>10</b>
Log Format – Minimum Required Fields.....	10
Consistency of Log Format.....	11
<b>Data filtering and exclusion rules .....</b>	<b>12</b>
Invalid Filetypes.....	13
Subsited and other Pushed Traffic.....	13
Invalid User-Agents .....	14
Invalid IP Addresses.....	15
Invalid HTTP Transactions .....	16
Invalid URLs.....	16
Other Material Non-Human Activity .....	17
Additional Exclusions.....	17
<b>Technical Guidance – Mandatory Metrics .....</b>	<b>18</b>
Proxies and Caching .....	18
Page Tagging.....	18
Unique User/Browsers.....	18
User-Agents .....	19
Auditing Monthly UU/Bs.....	19
Cookies .....	20
UserID+Date File Requirements.....	24
Automated content (Web Feed and Podcast) UU/Bs.....	25
Compound UU/B Identifiers.....	25
Page Impressions.....	25
Identifying Automated Page Impressions.....	26
Identifying Interactive TV traffic .....	27
Identifying Page Impressions within Rich Media .....	27
Identifying Mobile Traffic.....	28
Identifying HTML Chat Traffic.....	28
Visits & Visit Duration .....	28
Calculating and Claiming Visit Duration.....	28
Measuring Visits and Visit Duration in Rich Media Content.....	29
<b>Technical Guidance – Optional Metrics.....</b>	<b>30</b>

Searches .....	30
AV Content.....	30
Referrals In.....	31
PDA Traffic.....	31
Web Feed Aggregators and other Automated Content Syndication Agents .....	31
Chat (IRC) traffic .....	32
Clicks and Clickouts .....	32
Podcasts and Downloads .....	32
Geographical IP Analysis .....	32
<b>Certification and Publicity .....</b>	<b>34</b>
Confidentiality.....	34
Certificate Layout .....	34
Certificate Issue and Release.....	34
Declaration of Inventory.....	35
Comparability in Metrics .....	35
Transparency .....	35
Syndicated Content .....	35
Breakouts of Inventory.....	36
Breakouts of Page Impression Types .....	36
Site Publisher's Statement.....	36
Site Publisher's Logo.....	36
Publicity and Use of ABCe Kitemarks and Logo .....	37
<b>Complaints and Appeals.....</b>	<b>38</b>
Publicity.....	38
Conduct.....	38
Appeals against an ABCe decision.....	38
<b>The JICWEBS Standards .....</b>	<b>39</b>
About JICWEBS .....	39
How are the Standards developed and changed .....	39
<b>JICWEBS web traffic metrics .....</b>	<b>40</b>
About the standard metrics.....	40
Unique User/Browser Metrics (Reach).....	40
Impression Metrics (Volume) .....	43
Other Inventory Metrics (Volume) .....	45
Visit and Duration Metrics (Frequency).....	48
Click Metrics (Interaction) .....	49
CHAT metrics (non-HTML).....	50
Geographical IP-based metrics .....	51
<b>Defined terms used in this document.....</b>	<b>52</b>

## The ABCe Audit

### About this Guide

1. This annex contains information relevant to all those involved in the audit, with, in particular, technical guidance for those responsible for the configuration of web analytics tools, the preparation of reported traffic claims and the delivery of web traffic data to ABCe.
2. Changes to the industry standards and hence this guidance can occur at any time. ABCe will inform COI of any such changes and ABCe will update this guide as and when required. ABCe therefore advises you to check with COI ([transformationalgovernment@coi.gsi.gov.uk](mailto:transformationalgovernment@coi.gsi.gov.uk)) regularly to ensure that you are using the latest version.

### The Role and Purpose of ABCe

3. ABCe is a division of the ABC (The Audit Bureau of Circulations UK & Ireland), the industry-owned, tri-partite, non-profit distributing organisation that works with and for site publishers, advertisers and media buyers to help them better understand and gain confidence in the data they use. ABCe works in confidence with all its subscribers to help them understand and implement JICWEBS (Joint Industry Committee for Web Standards) industry-agreed, compliant processes, to measure their electronic media.

### Overview of the Audit Process

4. An ABCe audit is an independent check that the traffic claims made by a website have been derived from the supporting data in compliance with the industry-agreed (JICWEBS) reporting standards.
5. The ABCe Audit is **not** a simple Pass/Fail process. The ABCe audit will seek to ensure that the website publisher is clear on any issues of non-compliance identified and that a process of re-iteration or adjustment is used to enable certifiable totals to be established wherever possible. Audits can fail but this is usually only when the data provided for audit does not meet the minimum industry standards and, consequently, is not auditable.
6. The audit includes a detailed analysis of the supporting website traffic data which incorporates tests to:
  - i. Apply Industry-agreed standard filtering rules to supplied data to test compliance.
  - ii. Apply site specific filtering rules to supplied data in accordance with the industry-agreed rules to test compliance.

- iii. Apply the current IAB/ABCe Robots List of known Robotic User-Agents to test the exclusion of material amounts of such invalid traffic.
- iv. Analyse the data to test the exclusion of material amounts of additional non-human activity.
- v. Analyse the data to test for the exclusion of material amounts of "internal traffic".
- vi. Confirm that the website inventory is consistent with the Domain listing provided.
- vii. Confirm the authenticity of the data provided.
- viii. Analyse the data and evaluate site structure to test for 'automatically refreshed' URLs, PDF URLs and Mobile URLs and to ensure these are broken out where they represent more than 5% of the total.
- ix. Count all valid Page Impressions by day and for the audit month.
- x. Count and de-duplicate Unique User/Browsers by day and month from [ix].
- xi. Evaluate the method of logging cookies to ensure compliance with the industry-agreed rules (if applicable).
- xii. Analyse the data to test for the presence of 'Subsiting'.
- xiii. Analyse the data and evaluate site structure to test the exclusion of 'concurrently served' and 'bounce through' URLs.

## Setting up your Web Analytics Tool for Audit

7. The steps you should follow depend on the Analytics tool or service you are using and are described below. In all cases please ensure that you have notified the auditor of all the domains to be included in the audit **prior** to the audit month.

## Self-hosted Web Analytics Tool

8. If using a self-hosted web analytics tool, ensure that the logs to be analysed contain the minimum fields required (see Log format section below) in a consistent format across all log files.
9. Ensure that the analytics tool is properly configured for compliance with the industry agreed rules (see Data Filtering and Exclusion Rules section below).
10. Ensure that you:
  - can capture all the logs to be analysed for **all** the domains in the website;
  - use IP+User-Agent as the Unique User/Browser (UU/B) identifier or, if using cookie, have a single, persistent cookie that is consistent across all the domains in the website and complies with the industry-agreed rules (see Cookie section below);
  - know which reports (from which account if applicable) will be sent to ABCe (for the website's Page Impressions, UU/Bs, Visits and Visit

Duration metrics) – totalled (and for UU/Bs, deduplicated) for the month and broken down by day.

- if you are running a number of accounts (aka profiles) in a web analytics tool, for different parts of your website, to ensure that you can claim an overall figure for the whole website (those accounts added together) you must have a global or roll-up account set up that captures one dataset for the entity as a whole. This is to ensure you have a single compliant cookie across the whole website.
- archive the data;
- can supply the supporting data to ABCe upon request;
- can facilitate secure data transfer to ABCe via CD-ROM, DVD (not Blu-Ray), or SFTP (as agreed);
- can supply a claim (reported statistics) for the website traffic for each metric to be certified, broken down by day and in total for the audit month.

### **Third Party Analytics Tool or Service**

11. If using a third party analytics tool or service, notify them of the month chosen for the audit and your requirement that they facilitate the ABCe audit process.

12. Agree any applicable fees. Please note these are the sole responsibility of the Site Publisher.

13. Ensure that the analytics tool or service is properly configured for compliance with the industry agreed rules (see Data Filtering and Exclusion Rules section below).

14. Ensure that you or the supplier:

- can capture all the log file data for **all** the domains in the website (NB – some web analytic providers need data logging and/or have an ABCe module that must be activated to support an audit);
- use IP+User-Agent as the unique browser identifier or, if using cookie, have a single persistent cookie that is consistent across all the domains in the defined website and complies with the industry-agreed rules (see Cookie section below);
- know which reports, from which account, will be sent to ABCe (for the website's Page Impressions, UU/Bs, Visits and Visit Duration metrics) – totalled (and for UU/Bs, deduplicated) for the month and broken down by day.
  - if you are running a number of accounts (aka profiles) in a web analytics tool, for different parts of your website, to ensure that you can claim an overall figure for the whole website (those accounts added together) you must have a global or roll-up account set up that captures one dataset for the entity as a whole. This is to

ensure you have a single compliant cookie across the whole website.

- archive the data;
- can supply the supporting data to ABCe upon request (NB – some web analytic providers only make data available for a relatively short time, often over-writing data on the basis of a rolling number of days. In such circumstances the data must be downloaded from the supplier and archived by the Site Publisher;
- can facilitate secure data transfer to ABCe via CD-ROM, DVD (not Blu-Ray), SFTP (as agreed);
- can supply a claim (reported statistics) for the website traffic for each metric to be certified, broken down by day and in total for the audit month.

### **Larger Websites**

15. If your website is expected to have more than 20 million Page Impressions in the audit month, please ensure that the log file data provided to ABCe in support of the metric claims is delivered in two parts: a monthly UserID+Date file and separate sample days data (as requested). For more details see Auditing Monthly UU/Bs (below) or email: [transformationalgovernment@coi.gsi.gov.uk](mailto:transformationalgovernment@coi.gsi.gov.uk).

## Using a 2-star Accredited ABCe Associate Web Analytics Product

### General Notes

In alphabetical order (Company, Product, Website, Notes):

16. Foviance Limited, WebAbacus, [www.webabacus.com](http://www.webabacus.com) - A data analytics tool that can process either web server log files or page tag log files. It can be self hosted or delivered as a managed service. If using page tagging be sure to confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's web server log files or page tag log files, as applicable, should be provided to ABCe either by the Site Publisher or by Foviance (the latter is dependent upon a support contract being in place with Foviance). Contact: Sean Burton: [sean.burton@foviance.com](mailto:sean.burton@foviance.com)
17. Intellitracker Limited, Intellitracker Enterprise, [www.intellitracker.com](http://www.intellitracker.com) – An outsourced Application Service Provider (ASP) page tagging service. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A UserID+date file and 3 sample days data can be generated by Intellitracker for the audit month on request. Contact: Simon Roberts: [sroberts@intellitracker.com](mailto:sroberts@intellitracker.com)
18. ISSEL Limited, Pilot HitList, [www.issel.co.uk](http://www.issel.co.uk) – A self hosted software analytics tool that is usually used to process either web server log files or packet sniffed log data. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's filtered data (and sample raw data) should be provided by the Site Publisher. Contact: Bjorn Svensson: [bjorn@issel.co.uk](mailto:bjorn@issel.co.uk)
19. ISSEL Limited, ISSEL Log Companion, [www.issel.co.uk](http://www.issel.co.uk) - This is a simple filtering script that can be used by the Site Publisher to process web server log files. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's filtered web server logs and sample raw log files should be provided by the Site Publisher. Contact: Bjorn Svensson: [bjorn@issel.co.uk](mailto:bjorn@issel.co.uk)
20. Maxsi Limited, eVisit Analyst, [www.evisitanalyst.com](http://www.evisitanalyst.com) - A self hosted software tool that can be used to process either web server log files or page tag log data. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's web server log files or page tag log files (as applicable) should be provided by the Site Publisher. Contact: John Mawson: [jm@evisitanalyst.co.uk](mailto:jm@evisitanalyst.co.uk)

21. Nedstat Limited, Sitestat 5 or higher, [www.nedstat.com](http://www.nedstat.com) - An outsourced (ASP) page tagging service, a full month's page tag log can be provided by Nedstat on request. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. In addition, an ABCe Module is required to support the ABCe audit and must be implemented prior to the audit month. A full month's page tag log dataset or 3 sample days' data and a cookie file per day can be produced by Nedstat for the audit month on request. Contact: Bettina Hamm [support@nedstat.com](mailto:support@nedstat.com)
22. Nielsen Online, SiteCensus, [www.nielsen-online.com](http://www.nielsen-online.com) – An outsourced (ASP) page tagging service. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's page tag log dataset or 3 sample days' data and a cookie file per day can be produced by Nielsen for the audit month on request. Contact: [Raj.Dubasia@nielsen.com](mailto:Raj.Dubasia@nielsen.com)
23. Omniture Inc. (was Visual Sciences), SiteCatalyst HBX, [www.omniture.com](http://www.omniture.com) - An outsourced (ASP) page tagging service. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's page tag log dataset or 3 sample days' data and a UserID+Date file can be produced by Omniture for the audit month on request. Logging must be activated in advance of the audit period. Contact: Paul Kirk: [pkirk@omniture.com](mailto:pkirk@omniture.com)
24. Omniture Inc., SiteCatalyst 10, [www.omniture.com](http://www.omniture.com) - An outsourced (ASP) page tagging service. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's page tag log dataset or 3 sample days' data and a UserID+Date file can be produced by Omniture for the audit month on request. Contact: David Brown: [dbrown@omniture.com](mailto:dbrown@omniture.com)
25. SageMetrics SageAnalyst, [www.sagemetrics.com](http://www.sagemetrics.com) - An outsourced (ASP) page tagging service. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's page tag log dataset or 3 sample days' data and a UserID+Date file can be produced by Sagemetrics for the audit month on request. Contact: Benoit Droulez: [bdroulez@sagemetrics.com](mailto:bdroulez@sagemetrics.com)
26. Site Intelligence Limited, VBIS, [www.site-intelligence.co.uk](http://www.site-intelligence.co.uk) – A data analytics tool that can process either web server log files or page tag log data. It can be self hosted or delivered as a managed (ASP) service. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's web server log files or page tag log dataset can be provided to ABCe either by the Site Publisher or by Site

Intelligence (the latter is dependent upon a support contract being in place).  
Contact: Chris Brown [chris.brown@site-intelligence.co.uk](mailto:chris.brown@site-intelligence.co.uk)

27. speed-trap.com Limited, e-intelligence suite, [www.speed-trap.com](http://www.speed-trap.com) – An outsourced (ASP) page tagging service. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's page tag log dataset can be made available or 3 sample days data and a UserID+Date file can be produced by speed-trap for the audit month on request. Contact: Dvae Vodden [david.vodden@speed-trap.com](mailto:david.vodden@speed-trap.com)
28. WebTrends Inc, WebTrends 7 & 8 On Demand, [www.webtrends.com](http://www.webtrends.com) - An outsourced (ASP) page tagging service. Always confirm **prior** to the audit month that the log data is being captured. Also note that this data is typically only retained for only 2 weeks, hence this data must be downloaded and retained by the Site Publisher. A full month's page tag log dataset can be produced by WebTrends for the audit month on request. Contact: Conrad Bennett [conrad.bennett@webtrends.com](mailto:conrad.bennett@webtrends.com)
29. WebTrends Inc, 7 & 8 Software (with SDC), [www.webtrends.com](http://www.webtrends.com) - A self hosted software tool that must be used with SDC (Server Data Capture, aka page tagging). Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's page tag dataset can be provided by the Site Publisher. Contact: Conrad Bennett [conrad.bennett@webtrends.com](mailto:conrad.bennett@webtrends.com)
30. WebtraffIQ Limited, WebtraffIQ, [www.webtraffiq.com](http://www.webtraffiq.com) - An outsourced (ASP) page tagging service. Always confirm **prior** to the audit month that the log data is being captured and will be retained for supply to ABCe. A full month's page tag log dataset can be made available or 3 sample days data and a UserID+Date file can be produced by WebTraffIQ for the audit month on request.

**Please Note:**

- 1) The above information is correct as of the date of publication of this guide but is subject to change at any time. ABCe therefore advises you to check with COI regularly ([transformationalgovernment@coi.gsi.gov.uk](mailto:transformationalgovernment@coi.gsi.gov.uk)) to ensure that you are using the latest version.
- 2) Support from an Associate is wholly dependent upon the contractual relationship between the Associate and the Site Publisher.
- 3) Fees may be levied by the Associate for the support of ABCe auditing. These are wholly a matter between the Site Publisher and the Associate and are **not** the responsibility of ABCe.

## Data logging

31. All claims must be supported by logged records which represent the **valid traffic** in the audit month. These logged records may be generated by
- The sites own Web servers
  - Dedicated Page Tag servers (typically collecting graphic requests generated by browser-side measurement tools)
  - Packet sniffers (not often used)

Whatever their nature, data collection servers must be date and time synchronised, preferably to GMT, so that their log files' date and time stamping is aligned. Appropriate archiving procedures must be in place to ensure you can substantiate your claims for the audit month. The log files used to support your claims must be retained.

### Log Format – Minimum Required Fields

32. Each logged record should contain at least the following data fields:
- Date and time stamp of the request, including any time adjustment
  - IP Address of the originating user (NB – Dotted-quad format e.g. aaa.bbb.ccc.ddd, **not** the resolved name)
  - Full request-URI, including:
    - Domain (Host)
    - Requested URL
    - Any applicable query parameters
  - Full unmodified User-Agent string
  - Referrer URL
  - User Identifier (e.g. Cookie) if not logged in other fields

Additionally, for web server logs (not required for Page Tag logged data):

- HTTP Status code (200, 302, 404 etc)
- HTTP method of the request (GET, POST etc.)
- IP Address (or name) of the **server**
- Bytes transferred

Other fields, such as Site-ID, Protocol or Screen Resolution, may be logged if desired. Fields may be logged in any order as long as the minimum standards above are met.

The W3C CLF (common log format) does **not** include fields essential to the accurate counting and auditing of website activity, such as User-Agent, and as

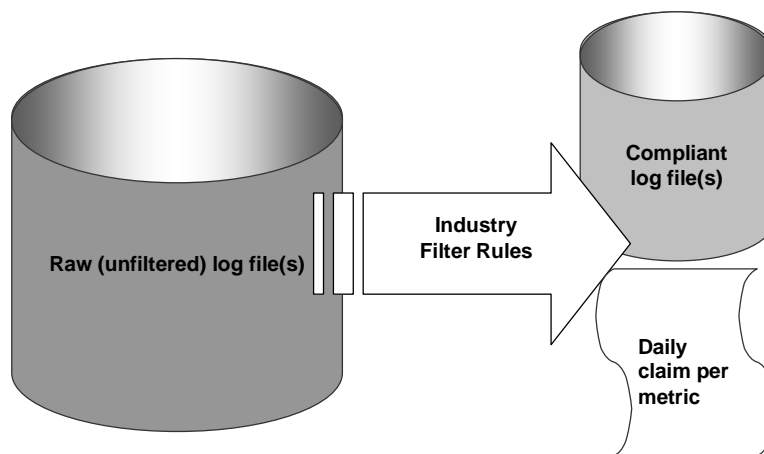
such cannot be used as audit evidence. Also, some proxy log types (e.g. Squid) do **not** contain the minimum required fields.

### **Consistency of Log Format**

33. You must not change the format of your logs during the audit period. You must ensure that any third parties managing your logs for you are also aware of these requirements. If you operate a mixed logging format, you may fail to exclude robots, and miscount Users and hence Visits, unless you ensure that the User-Agent strings are normalised to the same format throughout the dataset before the data is analysed and the claimed metrics are calculated.

## Data filtering and exclusion rules

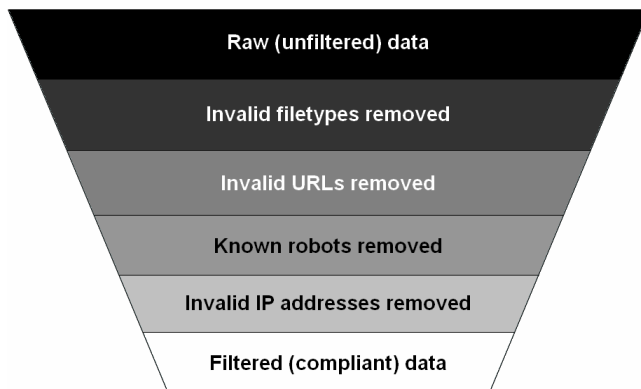
34. Why is filtering of the logged data necessary? In order to calculate metric counts correctly the Site Publisher must ensure that only **valid traffic** is counted. This is rarely the full dataset collected but is more often a subset of this that can be derived by filtering the original dataset to exclude certain records that do not comply with the industry-agreed rules. Site Publishers can deliver to ABCe either compliant (post processed) data (containing only records deemed to be of valid traffic) created by filtering the raw log files for the audit month or, as a minimum, the Site Publisher must deliver the raw (original) unfiltered data to ABCe. In the latter case additional charges apply.



All the required fields listed in the Log Format section (above) must be included in the data supplied to be audited. As a sanity check, the number of records in a day's compliant log file data should equal the number of Page Impressions claimed by the website for that day.

## ABCe Filtering Guidance

35. The following graphic sums up the basic filtering methodology:



Note that additional steps such as removal of invalid HTTP Method and/or Status codes may be needed depending on the type of data captured.

36. You must configure your web analytics tool to filter your traffic to exclude:

1. Invalid filetypes (e.g. graphics)
2. Substituted (pushed) traffic (e.g. substituted URLs)
3. Invalid User-Agents (e.g. robots)
4. Invalid IP addresses (e.g. internal addresses)
5. Invalid HTTP transactions (e.g. PUT, OPTIONS)
6. Invalid URLs (e.g. framesets)
7. Other invalid material (e.g. non HTTP transactions)

### Invalid Filetypes

37. Any filetype that can never represent, or is always served in conjunction with, a valid Page Impression (e.g. graphics, stylesheets) must be excluded from counts.
38. The following file types, usually bearing textual content, are likely to be served in response to **valid** requests from valid users.
- Any URL terminating in a slash or with no file suffix (e.g. www.foo.com/ , www.foo.com/index )
    - .htm .html .stm .shtml .shtm .h4
    - .asp .aspx .cfm .cgi .jsp .jspx .php .php3 .php4 .do
  - Conversely, ABCe consider files of the following types to be **invalid**, and require their exclusion from claimed figures (unless the site can prove, in an auditable manner, that these are the only files served in response to the user's request):
    - Graphics:
      - .art .bin .bmp .cdf .cgm .cls .con .emf .fpx .gif .hqx .ief .iff .image .jif .jpe .jpeg .jpg .lwf .pbm .pgm .pct .pic .pict .png .rgb .tga .tif .tiff .wbmp .xbm .xpm .xwd
    - Formatting and Page Information:
      - .css .ico .inc .xsl
    - On-Page Scripting:
      - .beans .class .dll .java .js
  - Note also that Flash Player/Shockwave Flash sequences (.swf) may represent the entire content of a particular page, and in such cases represent valid traffic. If certifying Page Impressions, the client must indicate to ABCe any pages holding graphical content only in order to include traffic to such pages in the audit.
  - Furthermore, ABCe recommends that URLs intended to be visited by RSS aggregators only – typically with the filetypes .rss, .rdf or .xml – be excluded from counts as Web Feed (RSS) traffic is not valid.

### Substituted and other Pushed Traffic

39. There are two main types of pushed traffic – substituting and contextual linking. In both cases, Page Impressions, and other metrics such as UU/Bs and

Visits, can be generated from content that has been "pushed" into the user's browser.

- Substituted traffic occurs when, upon a user requesting a page, a new browser window opens automatically on the user's PC (most often as a pop-under) which carries a different page, usually from another website. This second browser window therefore generates a Page Impression for a page (and usually a website) different to that which was intended by the user's action.
- Contextual linking occurs when the activity of a user in a non-browser application (such as an Instant Messenger client) is analysed and a new browser window is opened (usually as a pop-under) containing content from a website considered relevant to the subject of the user's conversation.

Note: Pushed traffic is different in nature from Automated Page Impressions, which are valid under the current industry agreed standards and, when material, declared separately on the certificate. In all cases, Automated Page Impressions result from an intentional user request (at least for the first Page Impression). Pushed traffic by comparison is not the result of an intentional user request. Hence, under the industry-agreed standards, pushed traffic is **not** valid and hence must be excluded from all certified figures.

## Invalid User-Agents

40. Websites must exclude material robotic activity from their claimed statistics. By material is meant activity accounting for more than 5% of Page Impressions in the audit month. The following types of robotic User-Agents are included in the standard exclusion process:

- Personal spiders and offline browsers can have significant and material effects on website traffic. Their activity levels are highly unpredictable over time and across websites. Hence, their User-Agents are not included in the standard robots list. Sites may need to justify the inclusion of such spiders or other proxy and caching activity to an auditor if the total impact on the website's traffic exceeds 5% of the Page Impressions audited in the audit month.
- PDA devices, Web Feed (RSS) aggregators and other automated syndication agents are included in the robots list. The Page Impressions certified for your site should **not** include any PDA or Web Feed (RSS) aggregator traffic.
- Records with unidentifiable User-Agents (usually nulls, "-") are also deemed invalid, since there is a risk that the activity was made by a robot. Therefore, any record with a null User-Agent must be excluded along with all other known (Listed) robots, unless the site can provide adequate justification for their inclusion.
- Records with the contype User-Agent (generated when old versions of Internet Explorer make surplus HTTP requests for a file) are also deemed invalid.

- To aid in the exclusion process, the IAB/ABCe Spider and Robot Exclusion List contains a list of those robots found to be material in audits. It is updated monthly, and is available from the Subscribers' Area on the ABCe website ([www.abce.org.uk](http://www.abce.org.uk)). This list is subject to alteration without notice, and the date of alteration will be shown. The list to be used in a given month is normally published by the 25th of the previous month. ABCe always makes the latest list available within the Subscribers' Area as robots\_current.txt.
- The site should apply the list by matching case insensitive strings anywhere in the User-Agent field in the web server log. A simple match of the start of the string is not adequate. Site Publishers may wish to ensure rigour in robot exclusion by adopting the dual-pass approach recommended by the IAB. The first stage includes only records matching the User-Agents on the Include list only; the second then excludes records from the remaining data based on matching the robotic User-Agents in the separate Exclude List. A one pass approach, applying the second Exclusion list to all records is, in ABCe's view adequate.

## Invalid IP Addresses

41. All traffic generated by internal activity, such as website development, performance monitoring, or automated broken link detection must be excluded (usually by excluding particular IP addresses or URLs).

- If any of this activity is performed by outside agencies on the website's behalf, this traffic is also deemed to be internal and must be excluded.
- 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 website development and maintenance PCs.
- Traffic generated by non-technical and non-development staff may be included if the website can demonstrate in an auditable manner that it can differentiate such activity from disallowed internal traffic.
- The standard internal network IPs (127.\*, 10.\*, 172.16.\* to 172.31.\*, and 192.168.\*) listed in RFC 1918 (<http://www.ietf.org/rfc/rfc1918.txt>), as well as the IPs used by known automated website monitoring tools, must be excluded.

These are:

See table on next page:

64.71.140.50, 64.71.140.52, 217.154.116.210	AlertSite
64.41.133.0-63, 64.41.145.*, 64.41.253.*	Attributor
213.166.22.44, 194.176.73.163, 193.201.201.75	Axzona
217.204.41.130-190	NewsNow
213.152.230.132-213.152.230.159, 213.161.84.0-213.161.84.31, 217.204.37.18-217.204.37.31, 213.161.80.128-213.161.80.143, 80.168.62.192-80.168.62.255, 82.211.104.128-82.211.104.159, 82.211.106.32-82.211.106.63	SiteConfidence
83.138.191.96, 89.234.59.74, 89.234.59.75, 212.100.226.110	SiteMorse

## Invalid HTTP Transactions

42. Only log records with the following HTTP Status Codes may be counted: 200, 201, 202, 203, 204, 205 and 304.

- HTTP response errors, i.e. all records that do not have a good status code, must be excluded.
- Status code 206 indicates a partial fulfilment of a request and will always be preceded by a 200; hence, it is not valid for counting of Page Impressions.

43. The HTTP commands "GET" and "POST" are the only two under which information is sent to a user as a result of a valid page request. Therefore, all other HTTP method requests are deemed invalid. NB: The above requirements will apply to tag (browser-side) logs by default.

## Invalid URLs

44. When Page Impressions (which must be counted first, all other metrics are derived from this dataset) are counted, the following additional exclusions must be applied:

- Concurrently Served or User-Invisible Content such as framesets, pop-ups and bounce-throughs are not valid Page Impressions, and therefore must be excluded from certified activity.
- However, if a pop-up is the only result of a user request or the user requests that a panel within a frameset, or a pop-up, is refreshed, then serving the refreshed panel may be counted as a Page Impression since it has been requested.
- Frameset exclusion can be complicated by the difficulty of distinguishing between wanted and unwanted records. Often the URLs associated with frames will all have the ".html" file extension. So, unlike the straightforward elimination of unwanted graphics (".gif" or ".jpeg") records, it will not be possible to filter records for inclusion or exclusion simply on the basis of their file extension.
- Pop-ups are deemed invalid wherever they are served concurrently with other content requested by a valid user.
- Departure pages (sometimes known as bounce-through pages or 'goto' pages) are pages to which the user is redirected without their

knowledge before leaving the website. These are deemed invalid unless they contain visible content that is delivered to the user. However, they are often used to enable the counting of Referrals In or Clickouts and so the website may wish to process them to count these metrics.

- Automatically Refreshed Content is valid (if otherwise compliant), since the user is deemed to have requested the refresh by staying on that URL. Such automated traffic must be broken out if greater than 5% of the total Page Impressions in the audit month.
- Splash pages do not need to be excluded from audited figures. The industry has agreed that they offer users a genuine opportunity to see. However, any page automatically refreshed from a splash page becomes refresh traffic, as above, and therefore must be broken out if greater than 5% of the total Page Impressions. Websites may wish to distinguish the homepage URL from the URL to which the splash page automatically sends the user, so as to avoid all homepage Page Impressions being recorded as automated.
- 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. Initial requests for PDF files (those with an HTTP status code of 200) may contribute to the Page Impression total, but must be broken out if greater than 5% of the total Page Impressions in the audit month.

### **Other Material Non-Human Activity**

45. This is non-human activity (e.g. robots and spiders) not excluded by the User-Agent filtering based on the published IAB/ABCe Exclusion list. This requires behavioral pattern analysis of click paths. Please note this is typically **not** supported by web analytics tools and is therefore **not** required to be performed by the website. It is however an important part of the audit and consequently any material amounts of this type of activity will be excluded from certified totals.

### **Additional Exclusions**

46. ABCe reserves the right to exclude website traffic for other reasons than those stated above if they consider such exclusions to meet the overall principles, logic, purpose or spirit of the Audit. ABCe will at all times keep you informed of the reasons for such exclusions.

## Technical Guidance – Mandatory Metrics

Any technical staff involved in preparing for an audit should read this section.

### Proxies and Caching

47. The caching of your website content by third parties such as ISPs or search engines can have an adverse impact on what requests are logged by your web servers, and hence (if using web server logs) what is certified as your traffic. It is, however, possible to prevent the caching of your website content by time-expiring elements of it. However, there will be a consequent increase in the load on your server(s).

### Page Tagging

48. One method of reducing the impact of caching upon your traffic analysis is to insert tag code into the source code of each page – examples being Javascript-generated pixel requests or non-cacheable pixels. In principle, logs containing information written using either of these methods are auditable, provided they conform to the JICWEBS minimum logging standards set out in the Log format section (above). Note that tagging methods reduce, but do not completely eliminate, robotic activity; websites who use tags are still required to comply with all the industry agreed rules regarding robot (and all other) exclusions (described above).

Since the HTTP Status code is not usually reported by tags, websites must ensure that any invalid pages such as redirect pages, admin pages or errors are not counted as valid Page Impressions.

Note: the Site Publisher must implement the tagging solution properly for ABCe to be able to certify the website's traffic. Where the analytics solution is managed by a third party it is **the Site Publisher's responsibility** to notify the analytics provider that all appropriate audit data must be captured, stored and made available to ABCe for the audit month. Note that the analytics provider may charge the Site Publisher for this service.

### Unique User/Browsers

49. What is a Unique User/Browser (UU/B)? A Unique User/Browser is currently defined as:

“The total number of unique combinations of a valid identifier. Sites may use IP+User-Agent and/or Cookie (and any other agreed valid identifier).”

Hence, the minimum acceptable Unique User/Browser definition is IP address+User-Agent, e.g. 193.133.21.96 Mozilla/4.0; (compatible) MSIE 5.5

### Pushed Traffic

50. If the website wishes to measure pushed traffic (subsiting or contextual linking), then technical staff must be aware that such traffic is **invalid** with regard to **ALL** the industry-standard metrics. Therefore, it is incumbent upon

the website to filter this traffic out and hence make it easily identifiable. The ABCe audit sets out to detect undeclared pushed traffic; if material pushed traffic is found, then the claimed metric totals **will not** be certified. ABCe suggests that a set of naming conventions should be used for any pushed URLs such that they can easily be excluded from the log files or differentiated from non pushed requests for the same URL. The recommended approach is therefore very similar to that used for identification of automated traffic when certifying Page Impressions (see below). If the log files cannot adequately distinguish between pushed and non-pushed requests for a particular URL (i.e. a specific naming convention was not used), then the auditor will have to assume the worst case scenario and exclude **all** requests for the URL in question.

## User-Agents

51. Mixed Logging - Where the website to be audited wishes to use IP+User-Agent as the user identifier, the site **must** ensure that the User-Agent is in the same format throughout. Microsoft IIS web servers substitute a plus for the normal space in any User-Agent (e.g.

"Mozilla/4.0;+(compatible)+..." rather than "Mozilla/4.0; (compatible)".

Hence, websites with mixed logs containing IIS and other data e.g. Apache **must** take steps either to record User-Agent strings uniformly in the compliant logs or to normalise them prior to processing. Otherwise, the UU/B claim will be wrong.

52. Media Player User-Agent Identifiers - Where the site to be audited includes any media files, the user will typically open them using media player software, whether or not this is embedded in the browser. Much of this software appends its own sub-string to the User-Agent value; hence, counting Users using IP+User-Agent, all such media player User-Agents must be **discounted**. The material majority of such User-Agents can be identified by the following strings within the User-Agent:

- **play32**
- **plus32**
- **mplayer**
- **nsplayer**
- **realplayer**
- **realproxy**
- **wmfsdk**
- **realoneplayer**

## Auditing Monthly UU/Bs

53. ABCe can audit daily figures for UU/Bs, Page Impressions, Visits and Visit Duration simply by carrying out audit tests on the compliant data for the sample days tested. ABCe can then conclude whether or not the sampled days' data has produced compliant statistics for Page Impressions, Visits and UU/Bs, and if satisfied that it has, can be confident in all the daily figures for the audit month.

However, since there is an aggregation and deduplication process involved in certifying monthly UU/Bs, daily totals do **not** tell ABCe anything about the

monthly number. Hence, the need for either all the months data (see below) or the UserID+Date file, which contains all unique combinations of Date and fields used to generate the UserID (typically Cookie only or IP+User-Agent only or Cookie and IP+User-Agent) found in the Page Impression records. This gives ABCe a very efficient and very sound audit proof of the Total UU/Bs figure.

54. When the website requires separate UU/B breakouts for each domain, the Domain field should also be present in the UserID+Date file – in other words, it should contain all unique combinations of Date, UserID and Domain. ABCe recommends separation of all fields in this file by a delimiter, ideally tab (Field1 | Field2) or else comma with quotes ("Field1","Field2"). Effectively, this file is a record of all distinct combinations of UserID and Date (and Domain if applicable).

55. **Fig. 1 (below) Typical UserID+Date+Domain file (Cookie).** This is an example of a UserID+Date+Domain file where Cookie Only is being used as the UU/B identifier and the regime is Compliant (see the cookie regime table Figure 2 below).

**Fig 1:**

<b>Date</b>	<b>Domain</b>	<b>Cookie</b>
• 2004-06-04	• www.foo.com	• AAABBB
• 2004-06-04	• www.foo.com	• AAABBC
• 2004-06-04	• ie.mysite.com	• AAABBC
• 2004-06-05	• uk.mysite.com	• -
• 2004-06-05	• ie.mysite.com	• AAABBC
• 2004-06-05	• uk.mysite.com	• AAABDD
• 2004-06-05	• uk.mysite.com	• AAABDE
• 2004-06-05	• ie.mysite.com	• AAABDE

etc...

Please Note: Where you wish to use alternative identifiers to gain additional UU/Bs, e.g. Cookie + Unmatched or Singleton Cookie, the UserID+Date file **must** contain **all** identifiers for **all** UU/Bs.

56. **Alternative to the UserID+Date File:** Instead of carrying out the extra processing required to generate a UDF, sites may wish to supply the total filtered data for the month. If there are more than 20 million such records, this option may be acceptable but is subject to additional charges and will require prior agreement with ABCe.

## Cookies

57. If you wish to count UU/Bs, in whole or in part, by Cookie, your cookie must satisfy the following conditions:

- **Persistence**

Cookies must be persistent, permanent cookies (not temporary, session cookies). Furthermore, cookie expiry dates must be set distant. Where the cookie has a fixed expiry date, this must be a minimum of 5 years after the cookie lands; alternatively, where the cookie expires a given length of time since the cookie was last accessed, this length of time must be at least 1 day longer than the required audit period (e.g. 32 days since last access for monthly audits).

- **Consistency**

The same cookie must apply across the entire website, across **all** domains of the traffic being audited with the same cookie recognised throughout on the same basis. The cookie cannot be a first-party cookie in some parts of the site and a third-party cookie elsewhere. Note, however, that clients gathering data across multiple sets of log files for a network of websites may still wish to use a single Cookie as a user identifier. This is possible with the following provisos:

- only one such cookie is used;
- the cookie used is consistent and persistent within one website;
- the cookie is used as the primary identifier of Unique User/Browsers;
- all additional identifiers used (typically IP+User-Agent) must be “unmatched”. In other words, any additional identifier seen with a valid Cookie can **never** be counted elsewhere.

Figures 2 and 4 (below) should enable you to identify which type of cookie configuration you have. Figure 3 will show how you may count UU/Bs and what defaults you may apply to cookie-rejecting or non-cookied UU/Bs.

**Fig 2 Type of Cookie Configuration**

• <i>What cookie value do you log?</i>	• <i>What additional tests are done to check cookie integrity or acceptance?</i>	• <i>Your configuration is of the following type:</i>
<ul style="list-style-type: none"> <li>• <b>Outgoing cookies</b> (from a server;</li> <li>• HTTP “Set-Cookie:” response header)</li> </ul>	<ul style="list-style-type: none"> <li>• None</li> <li>• Redirect to check whether cookie sent the first time is logged again</li> </ul>	<ul style="list-style-type: none"> <li>• <b>SINGLETON</b></li> <li>• <b>SINGLETON + REDIRECT</b></li> </ul>
<ul style="list-style-type: none"> <li>• <b>Incoming cookies</b> (from the browser’s HTTP “Cookie:” request header)</li> </ul>	<ul style="list-style-type: none"> <li>• Cookie given special “non-trusted” marker until it is logged again</li> <li>• None (first request for a file from a new “virgin” User has no cookie)</li> <li>• Redirect to check whether first-time Users accept cookies</li> <li>• Cookie sent to a new “virgin” User is logged as a query parameter and treated as the cookie if present.</li> </ul>	<ul style="list-style-type: none"> <li>• <b>SINGLETON + TEST</b></li> <li>• <b>COMPLIANT</b></li> <li>• <b>COMPLIANT + REDIRECT</b></li> <li>• <b>COMPLIANT (NO VIRGINS)</b></li> </ul>

The Netscape cookie protocol and RFCs 2109 and 2965 explain “request header” and “response header”. Some cases outlined above illustrate circumstances which cannot be proven by auditing data for a single month – here, additional audit tests may be required.

**Fig. 3 Acceptable Defaults For Rejected Cookies**

<b>Regime Is:</b>	<b>Acceptable UserID defaults:</b>
<ul style="list-style-type: none"> <li>• <b>1. SINGLETON</b></li> </ul>	<ul style="list-style-type: none"> <li>• IP+USER-AGENT for cookies which made one Page Impression ONLY in the period; any record with a null Cookie field <b>cannot</b> identify Unique User/Browsers.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>2. SINGLETON + REDIRECT</b></li> </ul>	<ul style="list-style-type: none"> <li>• IP+USER-AGENT for all Page Impressions which show rejected cookies; any record with a null Cookie field <b>cannot</b> identify Unique User/Browsers.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>3. SINGLETON + TEST</b></li> </ul>	<ul style="list-style-type: none"> <li>• IP+USER-AGENT for all Page Impressions with non-trusted cookies; any record with a null Cookie field <b>cannot</b> identify Unique User/Browsers.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>4. COMPLIANT</b></li> </ul>	<ul style="list-style-type: none"> <li>• IP+USER-AGENT UNMATCHED in the period for all Page Impressions with a null Cookie field.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>5. COMPLIANT + REDIRECT</b></li> </ul>	<ul style="list-style-type: none"> <li>• IP+USER-AGENT for all Page Impressions with a null Cookie field.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>6. COMPLIANT NO VIRGINS</b></li> </ul>	<ul style="list-style-type: none"> <li>• IP+USER-AGENT for all Page Impressions with a null Cookie field.</li> </ul>

a. “Unmatched” IP+User-Agents are those which were **never** seen with a cookie in the audit month.

b. The “period” for defaulting to additional identifiers is a **month** for Monthly Unique User/Browsers, but a **day** for daily Unique User/Browser breakdowns.

c. You may claim additional UU/Bs by a process of ascription if (and only if):

- your cookie is a Compliant type
- your audit data is gathered from the browser side (e.g. via a tagging solution), and
- a material majority of the Page Impressions in your compliant data have a valid cookie.

In such circumstances you may count the number of Page Impressions which carry a cookie and the number of unique cookies present in these Page Impressions. This allows you to establish the ratio of Page Impressions per cookie UU/B. Applying this ratio to the number of Page Impressions whose cookie is null gives an industry-acceptable estimate of the UU/Bs that created such Page Impressions. This method can also be applied to Visits with the above provisos.

Fig 4 Example of Various Cookie Behaviours

This table shows what happens to the Cookie when four Page Impressions are made under each of the Regime Types described in Figure D-3. The variables to consider are:

- What is the cookie regime?
- Does the user accept or reject cookies?

Cookie Regime	User accepts or rejects?	Contents of Cookie Folder	Cookie Values Logged				Notes
			PI 1	PI 2	PI 3	PI 4	
TYPE 1 (Basic Singleton)	Accepts ✓	(empty)	JOHN	JOHN	JOHN	JOHN	Cookie appears more than once, so it is a valid identifier.
	Accepts ✓	JOHN	This User only made one page during the period	JOHN	JOHN	JOHN	Cookie has appeared once only so is not a valid identifier.
	Accepts ✓	JOHN	JOHN	JOHN	JOHN	JOHN	All these cookies appear once only so are not valid identifiers.
	Rejects ✗	(empty)	PAUL	GEORGE	RINGO	STUART	Page 1 has a "non-trusted" cookie which becomes "trusted" only if the cookie reappears. JOHN then becomes a trusted cookie for all pages it appears in.
TYPE 3 (Singleton + Trust Test: ? is untrusted marker)	Accepts ✓	(empty)	?JOHN	JOHN	JOHN	JOHN	Only one page was made, but this cookie is trusted and therefore can be used as an identifier.
	Accepts ✓	JOHN (or ?JOHN)	JOHN	This User only made one page during the period	JOHN	JOHN	
	Accepts ✓	JOHN (or ?JOHN)	JOHN	JOHN	JOHN	JOHN	
	Rejects ✗	(empty)	?PAUL	?GEORGE	?RINGO	?STUART	All these cookies are untrusted and therefore cannot be counted.
TYPE 2 (Singleton+ Redirect)	Accepts ✓	(empty)	JOHN	JOHN	JOHN	JOHN	
	Accepts ✓	JOHN	JOHN	This User only made one page during the period	JOHN	JOHN	However, the redirect test means that this Cookie is a valid identifier (conversely to "Basic Singleton" above)
	Accepts ✓	JOHN	JOHN	JOHN	JOHN	JOHN	
	Rejects ✗	(empty)	PAUL	GEORGE	RINGO	STUART	
TYPE 4 (a) (Compliant Server-Side)	Accepts ✓	(empty)	(null)	JOHN	JOHN	JOHN	Page 1 has a null value here.
	Accepts ✓	(empty)	(null)	This User only made one page during the period	(null)	(null)	Even though this user nominally accepts Cookies, they have never been on the site before and have only made one Page Impression. So a null is logged.
	Accepts ✓	JOHN	JOHN	JOHN	JOHN	JOHN	
	Rejects ✗	(empty)	(null)	(null)	(null)	(null)	If the browser-side code is requested after the page is loaded, all Cookies are already in the browser; no Virgins can occur.
TYPE 4 (b) (Compliant Browser-Side)	Accepts ✓	(empty)	JOHN	JOHN	JOHN	JOHN	
	Accepts ✓	JOHN	JOHN	JOHN	JOHN	JOHN	
	Accepts ✓	JOHN	JOHN	JOHN	JOHN	JOHN	
	Rejects ✗	(empty)	(null)	(null)	(null)	(null)	Redirection has ensured that no Virgins can occur.
TYPE 5, TYPE 6 (Compliant+ Redirect, Compliant No Virgins)	Accepts ✓	(empty)	JOHN	JOHN	JOHN	JOHN	
	Accepts ✓	JOHN	JOHN	JOHN	JOHN	JOHN	
	Accepts ✓	JOHN	JOHN	JOHN	JOHN	JOHN	
	Rejects ✗	(empty)	(null)	(null)	(null)	(null)	

## Counting Cookies

### 58. Singletons:

Unique singleton cookies (cookies only ever present in one record in the audit period) are **not** automatically deemed invalid for the identification of UU/Bs. However, sites must be able to demonstrate, in an auditable manner, appropriate interpretation. This must be agreed in advance with ABCe as part of the audit plan and may require auditing of UU/Bs at a Page Impression level.

## UserID+Date File Requirements

59. If the cookie logged is that sent by the server (part of the "Set-Cookie:" header) regardless of whether or not the browser has accepted it, and the website wishes to certify a monthly UU/B figure, the website must send ABCe a list of the potential User identifiers (Date, Cookie, and additionally IP+User-Agent if used as a default) present **in each Page Impression** made during the audit period (as per UserID+Date file requirements).
60. If the cookie logged is that **set by the browser** (part of the "Cookie:" header) and the site wishes to certify a monthly UU/B figure, then the website need only send the UserID+Date file – i.e. an appended list of all unique cookies (and any additional identifiers used to calculate non-cooked Unique User/Browsers) found per day in the Audit Period.

## First v Third-Party Cookies

61. Some browsers are set to reject third-party cookies by default. Hence, ABCe recommends that wherever possible any cookie used to identify a user should be first-party cookie – in other words, sent from the same domain as the user's browser is visiting. Furthermore, some Web browsers may not accept any cookie from a domain that does not have a compliant Compact Privacy Policy (P3P). More dangerously, they may turn such a cookie into a temporary "session" one. Third parties must be even more careful when serving cookies to Users of a website. The industry-agreed standards require User-identifying cookies to be persistent, and P3P non-compliance may lead to over-counting of UU/Bs.

## Interactive TV Cookies

62. Websites may have material amounts of Interactive TV Impressions because they have a sizable audience using Interactive TV devices ("set-top boxes"). Any such website choosing to use Cookie as a user identifier should be aware that most current set-top boxes lose their cookies whenever completely powered off (as opposed to being put on standby). This may lead to these cookies becoming non-persistent, and hence to material over-statement of UU/B totals. Therefore, ABCe may need to investigate the behaviour of such cookies, and apply adjustments to claimed UU/B figures if deemed appropriate.

### **Automated content (Web Feed and Podcast) UU/Bs**

63. If a website wishes to measure additionally Unique Web Feed User/Browsers or Podcast Subscribers, then the web feed traffic or podcast notifications must be broken out into a separate inventory, and the Cookies within this inventory counted appropriately (following the appropriate industry agreed rules) to provide the claimed metric totals. Note this traffic is **not** considered valid for Page Impression, UU/B, Visit and Visit Duration measurement. The Cookies used to measure such UU/Bs must, nonetheless, comply with all the standard rules on persistence, consistency and distant expiry.

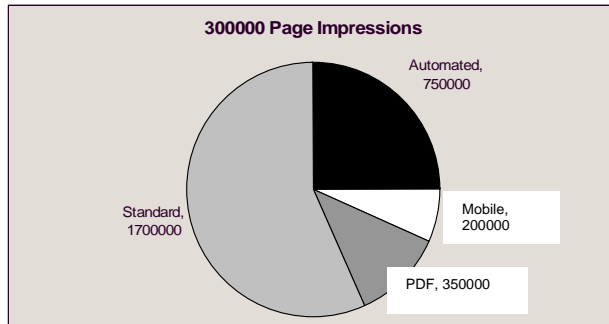
### **Compound UU/B Identifiers**

64. If it is wished to use Registration ID within a compound measurement of UU/Bs (e.g. attributing multiple Cookies or IP+User-Agents to the same Registration ID), then the Site Publisher should be aware that a lower number will be produced than could potentially be certified.
65. Registration ID on its own is **not** a valid UU/B identifier and consequently Registered Users should be measured (and certified) separately.

### **Page Impressions**

66. Breakout of types of Page Impression - There are a number of types of Page Impression which, regardless of whether they are separately stated on the certificate, can form part of the overall total Page Impression count. In all such cases, the simplest way of identifying "special" pages is to use a filter based on the URL name or directory, or to apply a special query parameter. Other approaches may be possible but will need discussion with ABCe. Where Automated Page Impressions, Mobile Impressions or PDF Impressions, form 5% or more of the Page Impression count in the audit month, these **must** be separately stated. Any other Page Impression subtypes may be stated at the site's discretion regardless of their percentage of total Page Impressions certified. If a website has more than 5% of a subtype that requires separate statement but does not wish to state this traffic, it may remove this Page Impression subtype total from the overall Page Impression total.

**67. Fig. 5 Example Site with Page Impression Subtypes.**



Consider a website [www.foo.co.uk](http://www.foo.co.uk) with 3,000,000 Page Impressions. This website has 750,000 Automated Page Impressions, 200,000 Mobile Impressions and 350,000 PDF Impressions, as shown in Figure 5. The Page Impression total is still 3,000,000; however, each Page Impression subtype forms separate and specific totals within the overall number. Only the Automated Page Impression total for this website must be stated on the certificate, since this is over 5% of the total – or a claim could be made for 2,250,000 Page Impressions by excluding the Automated Page Impressions.

**Identifying Automated Page Impressions**

68. Automated Page Impressions must be broken out from the overall Page Impression total where such traffic forms 5% or more of the total Page Impressions. One approved methodology for identification is to construct a refresh for a page (say `thispage.html`) that gives any refreshed page a specific query parameter (e.g. “`&refresh=1`”) as follows:

```
<META HTTP-EQUIV="refresh" CONTENT="5; URI=thispage.html?user=fred&refresh=1">
```

Websites may change the name of the refresh marker (“`refresh=1`”), as long as the name chosen is consistent and so identifiable. With this approach, the counting solution used can easily find any refresh by matching the chosen marker in the URI, and thus break out automated traffic. If this URI were updated every 5 seconds, the following log file entries would appear in an Apache ELF-type log:

```
10.6.13.57 - - [27/Oct/2009:17:22:30 +0100] "GET /Test/thispage.html HTTP/1.0" 200 126 "-"
"Mozilla/3.75 [en] (WinNT; U)" "-"
10.6.13.57 - - [27/Oct/2009:17:22:35 +0100] "GET /Test/ thispage.html?refresh=1 HTTP/1.0"
200 126 "-" "Mozilla/3.75 [en] (WinNT; U)" "-"
10.6.13.57 - - [27/Oct/2009:17:22:40 +0100] "GET /Test/ thispage.html?refresh=1 HTTP/1.0"
200 126 "-" "Mozilla/3.75 [en] (WinNT; U)" "-"
```

## Identifying Interactive TV traffic

69. Interactive TV traffic (Interactive TV Impressions) can be broken out from the overall Page Impression total if required. Normally it is possible to identify such traffic from dedicated URL directory such as /itv/ or /interactive/. However, where the directory naming convention is ambiguous, or other files are contained in the Interactive TV directory, the website may choose to use a combination of URL and User-Agent to identify their Interactive TV Impressions.

## Identifying Page Impressions within Rich Media

70. In June 2008, the Page Impression definition was revised to state that directly attributable user-initiated requests for content (e.g. mouse clicks) can be used to count Page Impressions, whether served in HTML, Ajax, Flash or other environments. This, however, is open to varying interpretations. ABCe therefore offers the following framework of overarching guidance principles for establishing what is, or is not, a valid Page Impression.

71. **Guidance** - If a “Yes” answer can be given to **all** the following questions within a rich media application, then Page Impressions should be able to be certified:

- Does website content in the user’s browser change as a result of a user’s mouse click?
- Is a certain type of log record only ever logged as a direct result of the mouse click?
- Can this record be disaggregated from other logged events within the rich media application and from anything which may be concurrently served with it?
- Would you be able to make a page constructed purely in HTML behave in a similar way?
- Is the average interval between mouse clicks greater than 5 seconds?
- Is the event different (distinct)?

### Examples:

- Online magazine built in Flash which turns a “page” when a user clicks the edge of the “page” they are on – **valid as Page Impressions**
- “Match scorecard” pages built in Flash which update automatically with the latest score – **valid as Automated Page Impressions**
- “Balloon popping” game requiring a mouse-click (or key stroke) to pop each balloon – **each “pop” is NOT a Page Impression.**

The above guidance and examples provide a simple framework for decision-making. Nevertheless, in case of doubt, Site Publisher’s should agree a

position with ABCe **before** the audit period commences otherwise ABCe may not be able to certify the claim.

### Identifying Mobile Traffic

72. **Mobile URLs** - are those created by the Site Publisher specifically to fulfil requests from mobile devices. These URLs will typically have naming conventions such as “/mobile/”, “/pda” etc. or be situated in separate domains.
73. **Mobile Impressions** - Mobile Impressions must be broken out from the overall Page Impression total if they represent more than 5% of the total Page Impressions in the audit month.
74. **Mobile Unique User/Browsers** - Mobile UU/Bs must be broken out from the overall UU/B total if they represent more than 5% of the total UU/Bs in the audit month, and Page Impressions are **not** being certified.
75. **WAP Impressions** - WAP traffic (WAP Impressions) may be broken out from the overall Page Impression total if required. This should be possible by looking at either the filetype (.wml), or where the filetype is a “standard” content-bearing one (such as .asp or .xml) at the URL directory structure.

### Identifying HTML Chat Traffic

76. HTML Chat traffic may be broken out from the overall Page Impression total if required, typically by URL directory structure e.g. “/chat”.

### Visits & Visit Duration

77. If you require ABCe to certify Visits using log files written across multiple domains or servers, you must ensure that these are all time-synchronised. A true picture of Visits will only emerge if the timestamps of Page Impressions correspond across all log files.
78. Please note when claiming Visits (and working out Visit Duration) that the effective maximum length of a Visit is assumed to be a day (00:00:00 to 23:59:59), since websites claim traffic figures on a daily basis. Websites that can show in an auditable manner that individual Visits last longer than a day may claim on this basis upon agreement with ABCe if required.
79. An alternative acceptable method of calculating and claiming Visits is to use a permanent (non-session) cookie, distinct from any cookie used to identify UU/Bs, which expires 30 minutes after the last page is accessed. This should provide comparable figures to the standard methodology. Note that when verifying totals derived using this methodology, ABCe will continue to use the traditional method of sorting Page Impressions by User identifier and Timestamp.

### Calculating and Claiming Visit Duration

80. Since this is essentially a mathematical calculation, ABCe appreciates seeing the working on the Audit Filtering Rules Return. Please ensure when calculating Visit Duration that you supply us on a daily basis with the number

of multi-page Visits (which therefore have an interval) on that day and the total in seconds of all intervals in such multi-page Visits.

For example, a site might give Visit Duration information for the 7th of August 2009 by declaring 18882372 seconds total time and 99321 multi-page Visits (out of 102192 Visits) on the Audit Filtering Rules Statement. This would give an Average Visit Duration for that day of 190.1 seconds. ABCe can then check the claimed Visit Duration figures from the daily samples supplied.

## **Measuring Visits and Visit Duration in Rich Media Content**

81. Websites based on rich media technology (Flash, AJAX, JSON and similar applications) may wish to measure Visits and Visit Duration from a dataset which includes auditable user-initiated logged events, representing interactions between browser and server that are not Page Impressions e.g. requests launched asynchronously from within the rich media application or events captured by Javascript.

ABCe may be able to support this upon confirmation from the website of the events being measured and substantiation through ABCe's own testing.

Please contact COI at: [transformationalgovernment@coi.gsi.gov.uk](mailto:transformationalgovernment@coi.gsi.gov.uk)

## Technical Guidance – Optional Metrics

Any technical staff involved in preparing for an audit may wish to also read this section, especially if additional metrics are to be included in the scope of the audit.

### Searches

82. Searches are a subset of Page Impressions, and therefore count towards the certified Page Impression total. A Search is defined as “The first Page Impression sent to a valid user as a result of that user’s search request being received by the server”. The objective is to measure completed search events, where a website serves a results page in response to a user’s search request, whether the outcome of that request is positive or not. The two methodologies for measuring Searches below have both been accepted as practicable by the industry. Websites wishing to use alternative methodologies may do so only upon prior agreement with ABCe. Contact COI at: [transformationalgovernment@coi.gsi.gov.uk](mailto:transformationalgovernment@coi.gsi.gov.uk)

83. **Query Parameter** - This is the simplest approach. The website merely adds a parameter to the search query to denote which page in the particular search has been selected. For example

`www.foo.com/search.jsp?town=Berkhamsted&county=Herts&resultspage=1`  
`www.foo.com/search.jsp?town=Berkhamsted&county=Herts&resultspage=2`  
`www.foo.com/search.jsp?town=Berkhamsted&county=Herts&resultspage=3`

Only the first of these (resultspage=1) counts as the Search. Note that this may lead to over- or under-statement of the true Search total depending on whether or not the page is cached when the Back button is pressed; the industry has, however, accepted that this risk is low and therefore that the methodology is acceptable.

84. **Unique Identifier for each Search Event** - In this approach, each Search is assigned a unique identification number (GUID), and the count of such unique IDs is submitted as the Search claim. Consider a website where a search for “doly parton” produces a set of results (a first results Page Impression) and generates a GUID, but will also produce the suggestion for a further search “Did you mean Dolly Parton?” If the user is looking for information on the famous country singer, they will click this second link, and so request another set of results. A second GUID is then generated by the site. Two Searches have therefore been completed, since the user made two distinct submissions of search requests.

### AV Content

85. AV content has distinct characteristics. In many cases, engagement with AV content does not require the user to interact with the HTML pages of the website; in others, AV content such as video clips of news stories may act as optional enhancements to HTML pages.

## Measuring AV Plays

86. Websites may wish to certify their websites' AV inventory, whether live or on demand, as well as their Page Impressions. AV Plays are **not** Page Impressions, and must be counted separately from this inventory. AV content can start to play automatically when a user loads a page of HTML. Such Automated AV Plays must be broken out where these form more than 5% of the total AV Plays (if AV Plays are certified).

## Live and On-Demand AV Plays

87. Live AV Plays are those which follow a true broadcast model – in other words, by requesting the file, a user "tunes in" to constantly available live content. This is in contrast to On-Demand AV Plays, which can be thought of as "clips" (though their duration may be quite long).

## Referrals In

88. For ABCe to certify Referrals In, the website must list ALL referrers which are native, in other words, internal to the domain or domains whose traffic is to be certified. Should ABCe find other referrers which appear to be native to your website, ABCe may need to apply adjustment or reiteration to your Referrals In claim. Where the audit is based on simple pixel tagged data, Referrals In cannot be certified. However, if the technology used captures the referrer of the tagged page, then Referrals In may be certified. Referrals In is often not a standard metric reported by web analytic tools.

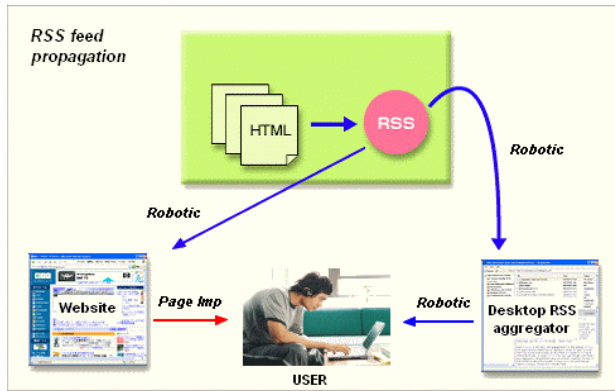
## PDA Traffic

89. Where a website wishes to measure the PDA Synchs metric, it must count the pages served to PDA devices by creating a log of PDA traffic. To create this log, the website must filter on the PDA User-Agent list made available from the Subscribers Area on [www.abce.org.uk](http://www.abce.org.uk) and apply all applicable traffic validity rules for date, HTTP status, method etc. before filtering for robotic activity. Note that these PDA User-Agents are normally counted as robotic, and are therefore part of the industry-standard Robots List used in the filtering process for Page Impressions. PDA traffic should **not** be included in any Page Impression figures, and should only be analysed when the PDA Synch measure is to be audited.

## Web Feed Aggregators and other Automated Content Syndication Agents

90. Agents for automated content syndication such as RSS readers and aggregators are deemed to be robotic, and are therefore part of the industry-standard Robots List used in the filtering process for Page Impressions, since they are automated tools. Web feed traffic is therefore **not** valid for the counting of Page Impressions. ABCe recommend that URLs intended to be visited by web feed aggregators only – typically with the filetypes .rss, .rdf or .xml – be excluded from Page Impression counts. Also, it is strongly

recommended that all websites serve web feeds with the .rss, not .xml, file type to aid filtering in compliance with the industry-agreed standards.



**Fig 6 Web Feed Traffic**  
(In this example RSS is used)

Any page on which web feeds are displayed (the “Website” in the above diagram) **is valid** if requested by a valid user making a valid request with a browser; such pages can be certified as Page Impressions.

## Chat (IRC) traffic

91. A minimum requirement in order to audit is that the chat logs provided must contain all IP addresses, nicknames, time in and time out of users necessary to support the site's claims.

## Clicks and Clickouts

92. The website must have a clearly auditable method of logging Clicks and Clickouts, such as departure (bounce-through) pages. Clicks are subject to the same User validity rules as Page Impressions. In other words, the record must have a valid User-Agent and IP address. Note that it is possible for a UU/B to make a Click or Clickout without making a valid Page Impression (e.g. by clicking on a link in a locally cached page).

## Podcasts and Downloads

93. To enable certification of Requested or Completed Podcasts, websites must adopt a naming convention for their download files that shows a clear distinction between Downloads made through podcast subscription services and those which have been individually requested with no such prior notification. Such individually requested files count only as Downloads **not** Podcasts. Any audit will include tests to gain confidence that this naming convention has been applied in accordance with the above requirements.

## Geographical IP Analysis

94. Results for certification of Page Impressions by Country and UU/Bs by Country may be obtained by standard open-source tools such as DNS

reverse lookup or by the use of proprietary solutions. In **all** cases the methodology used must be clearly specified. Certification at Country level only is endorsed by JICWEBS.

## Certification and Publicity

### Confidentiality

95. All ABCe audits are completed in confidence. The decision to publish rests with the Site Publisher. Subscribers who have published a certificate must publish a certificate at least annually thereafter to keep their ABCe subscriber status.

### Certificate Layout

96. At least the following information must appear on all web certificates released into the public domain:

- Website Name and Website URL;
- Website Inventory Domain(s);
- The period covered by the certificate;
- A description of the content of the website - the Site Publisher's Statement;
- Contact details for the Site Publisher;
- The monthly UU/B total and average daily UU/Bs for the Certification Period;
- If the property is declaring that it operates a registration policy, a breakdown of UU/Bs and/or Page Impressions generated by registered and non-registered traffic;
- Definitions of all metrics cited in Section 3 of the certificate and any from which they may derive, to assist users of the certificate;
- The Counting (web analytics) system used;
- A statement of auditor's opinion.

### Certificate Issue and Release

97. At the end of the audit period, ABCe will confirm any metric totals that are certifiable, and the Site Publisher confirms which statistics it wants to appear on the certificate (subject to the minimum requirements detailed above).

98. Upon receipt of all relevant certificate content information, ABCe will produce a draft certificate for approval by the Site Publisher. When the Auditor has completed and approved the audit, ABCe will issue the draft certificate. **No** information is made public unless the Site Publisher instructs ABCe to do so.

## Declaration of Inventory

99. The certificate must define the perimeter of the inventory certified by stating all material domains (e.g www.site.co.uk) and sub-domains (e.g http://sport.site.co.uk). These must cover at least 95% of the inventory certified (by Page Impressions) but may cover more.

## Comparability in Metrics

100. Different totals for your metrics may result from using different User identifiers (Cookie versus IP+User-Agent). Averages of totals derived from different calculation methods cannot be certified.

## Transparency

101. Websites may obtain certification for any web inventory they feel is appropriate. Consequently, the inventory that is certified can include third-party authored or syndicated content. However, for the purposes of the Audit Certificate, a Page Impression is only valid if it is from an Inventory Domain (or URL) specified on the certificate.

102. Clients may specify the website name they wish to appear on their audit certificate. If this name is a URL or domain name, then that domain name (or URL) **must** be present in the inventory being certified. Additional information such as a website logo or further data analysis (e.g. averages, traffic by date) derived from the metrics certified may be included on the certificate at the Site Publisher's discretion and with the agreement of ABCe.

## Syndicated Content

103. Syndicated content can be included in the audited inventory, since a *caveat emptor* statement is included on all certificates. This statement is worded as follows: "Syndicated content may or may not be included in the traffic certified".

## Syndicated Content - Measurement Guidance Framework

104. However, it is **not** meaningful for a Site Publisher to put a logo into someone else's site and count the requests for it, so inflating certified figures since there is no opportunity to see website content presented on the syndicated page. Syndicated content can be counted towards a Site Publisher's audited web inventory if the following questions can all be answered:

- 1) The user can have an opportunity to see, and interact with, self-contained content presenting information;
- 2) The content makes sense (to a reasonable user) in and of itself;
- 3) The content is not simply a logo, search box or similar branding.

It is always up to the Site Publisher to make the case that their syndicated content meets the above three requirements.

## Breakouts of Inventory

105. Where a set of inventory is broken out from audited totals and declared as pertaining to either a specific website area (a domain or set of domains) or a specific vertical e.g. jobs, all such inventory **must** either:

- be of the selected domain(s) or
- have relevance (typically appropriate content) to that vertical.

## Breakouts of Page Impression Types

106. If Page Impressions are certified, specific types of Page Impression must be broken out (stated separately) on the certificate. These are as follows:

- Automated Page Impressions if they form 5% or more of the overall Page Impressions in the audit month
- Mobile Impressions (based on URL) if they form 5% or more of the overall Page Impression in the audit month
- PDF Impressions (if they are not certified separately as Downloads) if they form 5% or more of the overall Page Impression in the audit month

## Site Publisher's Statement

107. This statement for the certificate should be no more than 150 words in length. The statement should help the certificate reader understand the purpose of the website. Statements must be restricted to a reasonable description of the website itself and must not include other information or claims. The following types of claim must not be made within the statement:

- Claims that use numbers or percentages not appearing elsewhere on the certificate, e.g. "70% of users are male";
- Quotes;
- Comparative claims in relation to the market or other websites (e.g. "The No. 1 website").

The accuracy of Site Publisher statements is not audited. However, ABCe will review the statements to ensure they are reasonable and meet the above requirements. ABCe's decision is final.

## Site Publisher's Logo

108. This can be added to certificates. However, ABCe reserves the right to prohibit or amend a logo if it includes any statements or claims prohibited by the rules described for the Site Publisher's statement above (e.g. a strapline saying "The UK's No. 1 [something] website") or is misleading.

## **Publicity and Use of ABCe Kitemarks and Logo**

109. Site Publishers may choose to make some or all of any certified information public. However, where the Site Publisher chooses to limit what they make public, this must include at least the total certified for the mandatory minimum metric of the UU/Bs total for the audit month. Site Publishers can only disseminate information from an ABCe certificate, or the certificate itself, into the public domain after the certificate has been made public by ABCe. ABCe will post a copy of such public-domain certificates to [www.abce.org.uk](http://www.abce.org.uk).

## **Quoting and Sourcing Figures**

110. Statements or claims relating to data certified by ABCe or attributed to ABCe must be factually correct. Any figures sourced to ABCe and quoted on any material must:

- have been certified by ABCe;
- state the period of certification;
- not include any claim that any further data or system is certified by ABCe.

The source of any figure not certified by ABCe that is published with or near to any ABCe kitemark, or any reference to ABCe, must be expressly and clearly stated so that it is clear whether or not ABCe have certified such a figure. ABCe will decide on the method of correction if appropriate.

## **Use of ABCe Kitemarks**

111. Any Site Publisher who has a published Certificate may use the ABCe Kitemark issued and dated for that month in all forms of publicity relating to the website certified. Any client who has ceased to be an ABCe subscriber must cease to use any ABCe kitemark on any material.

## **Use of the ABCe Logo**

112. This is and remains the exclusive trade mark of ABCe. No subscriber is entitled to use this logo.

## **Un-audited ABCe Subscribers**

113. Any ABCe subscriber who does not have a published ABCe audit certificate may (if they wish) make reference to their status as a Registered ABCe Subscriber in publicity statements or on their website using the following form of words only:

"As of <Date>, <Site Publisher> is Registered with ABCe and supports industry agreed standards for electronic media measurement."  
They may NOT use any logo or kitemark relating to ABCe.

## Complaints and Appeals

### Publicity

114. Complaints about:

- a Site Publisher (Media Owner) or Associate Subscriber's use of ABCe certified figures
- ABC Group logos or kitemarks
- other claims which are believed to infringe the requirements of the industry-agreed JICWEBS standards and ABCe rules may be made to ABC Group's Head of Compliance by post at:

211 High Street,  
Berkhamsted,  
Herts  
HP4 1AD

or by email to: [compliance@abce.org.uk](mailto:compliance@abce.org.uk)

Complaints will be investigated by contacting relevant parties. If a complaint is upheld, appropriate action will be taken and communicated. There will be a right of appeal by either party to the Chairperson of JICWEBS as detailed below. In cases of doubt the case will be referred to JICWEBS for a decision.

### Conduct

115. In cases where one customer of ABCe complains about the conduct of another customer, in so far as the Managing Director of ABCe believes the complaint falls within the remit of the ABCe certification process, that complaint will be reviewed by the Chairperson of JICWEBS.

### Appeals against an ABCe decision

116. Where there is a difference of opinion between ABCe and a subscriber on a point of judgement, the subscriber will have the right to appeal to JICWEBS for a review of the decision. JICWEBS will consist of representatives of both buyers and sellers of electronic advertising and their decision will be binding.

## The JICWEBS Standards

### About JICWEBS

117. JICWEBS ([www.jicwebs.org](http://www.jicwebs.org)) is a body created by the UK and Ireland media industry. Its purpose is to ensure independent development and ownership of standards for measuring on a site-centric, census basis audience reach, frequency and activity levels including the use of advertising on electronic media.

### How are the Standards developed and changed

118. The needs of ABCe subscribers, and the industry as a whole, change constantly. In order to meet those needs and stay abreast of industry developments, the industry-agreed standards are continuously evolving. It is the industry itself, via JICWEBS (the Joint Industry Committee for Web Standards) fed into by the ITG (the Internet Technical Group) that sets these standards.

119. JICWEBS ([www.jicwebs.org](http://www.jicwebs.org)) is a body created by the UK and Ireland media industry. Its purpose is to ensure independent development and ownership of standards for measuring on a site-centric, census basis audience reach, frequency and activity levels including the use of advertising on electronic media. If you would like JICWEBS to consider any aspect of these rules please contact them at [info@jicwebs.org](mailto:info@jicwebs.org).

120. The ITG meetings are held every other month and are open to all ABCe Subscribers. The purpose of the ITG is to discuss and find solutions to industry raised technical issues - to agree (practical and feasible) proposals to put forward to JICWEBS.

121. Changes to the industry standards must be approved by JICWEBS. ABCe will inform COI of changes and update this guidance accordingly.

## JICWEBS web traffic metrics

### About the standard metrics

122. ABCe, working with the IFABC (International Federation of Audit Bureaux, <http://www.ifabc.org/>) international standards working group, has developed a set of definitions that are the effective world-wide standard for Web audits. Definitions for metrics specific to the internet industry in the UK and Ireland are controlled and developed by JICWEBS, the Joint Industry Committee for Web Standards. The metrics seek to give measurements of reach, volume, frequency and level of interaction. Each set of metrics below falls into one of these general categories.
123. **IMPORTANT:** These definitions can change at any time. Please check that you have the latest version of this document by checking the version number and date against those of the version published on the website.
124. The JICWEBS standard definitions are given in **bold**.
125. **All total numbers and averages defined in the metrics apply to an Audit Period, i.e. a calendar month.**

### Unique User/Browser Metrics (Reach)

#### Unique User/Browser

126. **A unique and valid identifier. Sites may use (i) IP+User-Agent and/or (ii) Cookie.** This metric does NOT measure a person. Instead, it is a measure of a device through which a person interacts with a website or network, in common with all measurement software. Where a Unique User/Browser is calculated by IP+User-Agent, this definition may overstate or understate the real number of individual users concerned due to dynamic IP address allocation (for example by a dial-up Internet Service Provider) or to significant levels of uniformity in IP and browser configurations operating through a proxy. Certain classes of traffic are not valid users and must be excluded. The Visit metric will be calculated on the basis of the Unique User/Browser metric used. Note that other device identifiers may be allowed as Unique User/Browser identifiers when they can be proved in an auditable manner to be persistent and consistent across the domains being measured. This metric can also be known as a “(Unique) Visitor”, “(Unique) Browser”, or “Unique”.

#### Repeat Unique User/Browser

127. **A Unique User/Browser that has made more than one Visit.** This metric should be expressed as a percentage of the total Unique User/Browser figure when shown on certificates. Single Unique

User/Browsers+Repeat Unique User/Browsers = TOTAL Unique User/Browsers.

### **Single Unique User/Browser**

128. **A Unique User/Browser that has made only a single Visit.** This metric should be expressed as a percentage of the total Unique User/Browser figure when shown on certificates. Single Unique User/Browsers+Repeat Unique User/Browsers = TOTAL Unique User/Browsers.

### **Interactive TV Unique User/Browser**

129. **A Unique User/Browser that has made one or more Interactive TV Impressions.** This is typically each Unique User/Browser who has contributed to the total Interactive TV Impressions.

### **AV Unique User/Browser**

130. **A Unique User/Browser who has made at least one AV Play.**

### **Search Unique User/Browser**

131. **A Unique User/Browser that has made one or more Searches.** This is typically each Unique User/Browser who has contributed to the total Searches.

### **Mobile Unique User/Browser**

132. **A Unique User/Browser that has requested a Mobile Impression.**

### **WAP Unique User/Browser**

133. **A Unique User/Browser that has made one or more WAP Impressions.** This is typically each Unique User/Browser who has contributed to the total WAP Impressions. Due to the lack of variation in WAP proxy IP+User-Agent combinations, **this metric must be calculated using cookies.**

### **Chat Unique User/Browser (HTML Chat ONLY)**

134. **A Unique User/Browser that has made one or more Chat Impressions.** This is typically each Unique User/Browser who has contributed to the total Chat Impressions.

### **Web Feed Referral In Unique User/Browser**

135. **A Unique User/Browser that has made a Web Feed Referral In.**

### **Web Feed Unique User/Browser**

136. **A valid Unique User/Browser (measured by Cookie) that has requested a web feed from the source site.** This produces a count of the total Unique User/Browsers who request web feeds (e.g. RSS, Atom) through browsers. The Cookie identifier is agreed to be the only identifier

that provides the required level of granularity. No defaults are currently acceptable. The simple number of web feeds requested is not considered meaningful, since the number of requests relies so much on automated processes.

### **Web Feed Article Impression Unique User/Browser**

137. **A Unique User/Browser generating at least one Web Feed Article Impression.** Note that this can be measured by IP+User-Agent and/or Cookie, as per the Web Unique User/Browser definition.

### **Requested Podcast Unique User/Browser**

138. **Any valid Unique User/Browser who has made a Requested Podcast.**

### **Completed Podcast Unique User/Browser**

139. **Any valid Unique User/Browser who has made a Completed Podcast.**

### **Registered User Account**

140. **An account set up for a user to request to receive access to a website, service or network.** All such Registered User Accounts should be contactable for registration purposes.

### **Active Registered User Account**

141. **A Registered User Account that has accessed the website, service or network in the certification period.** A Registered User Account that has accessed the website, service or network in the audit period. Auditable evidence of activity in the audit period for ALL Active Registered User Accounts is required. Contactability is not mandatory where such evidence of activity is available.

### **Podcast Subscriber**

142. **A valid Unique User/Browser (measured by Cookie) that has requested a notification of podcast availability from the source site.**

### **Unique Host**

143. **A unique IP address.**

### **Registered User Account**

144. **An account set up for a user to request to receive access to a website, service or network.** All such Registered User Accounts should be contactable for registration purposes.

### **Active Registered User Account**

145. **A Registered User Account that has accessed the website, service or network in the audit period.**

## Impression Metrics (Volume)

### Page Impression

**146. A file, or combination of files, sent to a valid user as a result of that user's request being received by the server.** In effect, one request by a valid user should result in one Page Impression being claimed. The counted Page Impression may not necessarily be in focus or fully visible in the user's browser. In most cases, a single request from a user causes the server to send several files to satisfy the request. For example, the server may send an html file followed by several associated graphics images and audio files. A single request from a user may also cause the server to send several additional HTML files to build a frameset. The site must ensure that all additional, non-requested files are filtered out and excluded when counting the claimed number of Page Impressions. Generally, subject to the guidance principles issued by the auditor, directly attributable user-initiated requests for content (mouse clicks) can be used to count Page Impressions, whether served in HTML, Ajax, Flash or other environments. This metric can also be known as a "Page View" or "Page Request". Please note that files that contain specific types of advertising creative, such as banners or skyscrapers, and files that represent Streams are not valid for the counting of Page Impressions but should be used separately to identify Ad Impressions or Streams (defined in Section 2.6 below). Page Impressions must contain textual content beyond simple advertising.

### Automated Page Impression

**147. A Page Impression sent to a valid user as a result of an automatic process.** If a valid user (i.e. a connection to the site from a valid browser) requests a page and subsequently the page is refreshed, or another content-bearing page is sent, to that same User, then both the original page request and all subsequent refreshed pages are deemed to be valid Page Impressions. The fact that the subsequent pages result from an automated process rather than an actual mouse click does NOT make those pages invalid. The User therefore has, by default, made a valid request for all subsequent pages. Hence, the resulting Page Impressions are deemed to be valid and can therefore be claimed. Automated Page Impressions are therefore valid log file records that represent pages normally requested automatically by the browser, without the need for human action. (For example: automated price/news/score updates, text tickers, slide-show sequences, automated tours, etc). All Audit Certificates must carry a breakdown of the totals of Automated Page Impressions from the overall Page Impression total where such traffic forms 5% or more of the overall Page Impression count.

### WAP Impression

**148. A file, or combination of files, intended for a WAP device, sent to a valid user as a result of that user's request being received by the**

**server.** In effect, one request by a valid user should result in one WAP Impression being claimed. Certain WAP traffic can be excluded on the basis of information derived from the server log files. In most cases, a single request from a user causes one or more WML files (a so-called "WAP deck") to be sent by the server. The site must therefore ensure that any additional, non-requested files are excluded from the claimed WAP Impression total. A WAP Impression does not guarantee that a user actually viewed a page of WML; it only measures the opportunity for the user to view such content. This means that a WAP Impression recorded as valid by the server will be valid even if the content does not load to completion. WAP Impressions can be included within the headline Page Impression total on a certificate; they may be broken out if desired.

### **Mobile Impression**

149. **A Page Impression of a URL created by the site specifically to fulfil a request from a mobile device.** This metric allows the measurement of the total number of pages intended for mobile device consumption.

### **Chat Impression (HTML)**

150. **A file, or combination of files, sent to a valid user while the user is involved in an interactive Chat session (i.e. a Page Impression generated by a Chat URL.). Chat URLs show the input of one or more concurrent Users, visible to each other, updated frequently, so as to enable a text based conversation.** In effect, one request by a valid user should result in one Chat Impression being claimed. This metric is not applicable to Forums, where concurrent users are not typically visible to each other. The metrics available to a site to measure Chat depend on the technology employed by the site to support Chat. Chat based technologies include HTML, IRC, Java and proprietary applications. The Chat Impression metric is **only available for HTML-based chat.** Chat Impressions can be included within the headline Page Impression total on a certificate; they may be broken out if desired.

### **Interactive TV Impression**

151. **A file, or combination of files, intended for an Interactive TV device, sent to a valid user as a result of that user's request being received by the server.** In effect, one request by a valid user should result in one Interactive TV Impression being claimed. Certain Interactive TV traffic can be excluded on the basis of information derived from the server log files. The site must therefore ensure that any additional, non-requested files, such as graphics, audio files or frames, are excluded from the claimed Interactive TV Impression total. An Interactive TV Impression does not guarantee that a user actually viewed the page requested; it only measures the opportunity for the user to view such content. This means that an Interactive TV Impression recorded as valid by the server will be valid even if the content does not load

to completion. Interactive TV Impressions can be included within the headline Page Impression total on a certificate; they may be broken out if desired.

### **PDF Impression**

152. **A file, or combination of files, containing PDF content, sent to a valid user as a result of that user's request being received by the server.** This metric allows the measurement of the total number of PDFs sent to a user within the Page Impression total.

### **Job Exposure Page Impression**

153. **A valid Page Impression that contains headline details of job vacancies including, as a minimum, the job title and at least one other relevant parameter e.g. location or salary.**

### **Job Details Page Impression**

154. **A valid Page Impression that contains the particulars for a single vacancy only.**

### **Proprietary Page Impression**

155. **A valid Page Impression on a recruitment site where the Site Owner (which may be different to the Site Publisher) is materially the same as the Site Advertiser.** Where such Page Impressions represent more than 5% of the total Page Impressions in the audit period, these MUST be broken out.

## **Other Inventory Metrics (Volume)**

### **Search**

156. **The first Page Impression sent to a valid user as a result of that user's search request being received by the server.** In effect, one search request by a valid user should result in one Search being claimed. This requires that the site, and hence the audit, can identify the first Page Impression served in response to a search request from a valid user and differentiate this first results page from any others. The Search total for a site is distinct from its Page Impression total – Searches are a subset of valid Page Impressions.

### **Job Search**

157. **The first valid Page Impression sent to a valid user as a result of that user's search request from a dedicated job search form being received by the server.**

### **AV Play**

158. A file request by a valid user for AV content. This can be measured in either of the following ways:

*AV Play Event* - A client-side play event (such as pressing of the Play button) made by a valid user which is not recorded concurrently with an event of the same type.

*AV Request* - A server-side indicator of a media file successfully served to a valid user. (Success is defined as transfer of content, so where the bytes sent are greater than zero).

### **Automated AV Play**

159. **An AV Play started by a valid user as a result of an automatic process.** Where Automated AV Plays form more than 5% of the total AV Play number, they MUST be broken out.

### **Web Feed Article Impression**

160. **A web feed article requested by a valid user within their browser or web feed reader.** This measurement requires the counting of a serving of a 1x1 tracking pixel within the <description> element of each article in a given web feed. These pixels will only be shown as a result of a definite user action to view the feed article. Since each article request will be counted, more than one article (and pixel) may be requested at once in a reader or Web browser. This metric should therefore not be considered to be comparable with a Page Impression.

### **Web Feed Ad Impression**

161. **A file sent to a valid user as an individual advertisement as a result of that user's client-initiated request for a Web Feed Article Impression being received by the server.** Any client-initiated method of counting Ad Impressions (per the IAB v6.0b standard) is acceptable.

### **PDA Synch**

162. **A synch by a PDA device with the site (or channel).** This is measured by requiring one URL per site (or channel) to be non-cacheable. This then acts as the identifier for counting PDA Synchs. Measuring how many times this URL was requested in the Audit Period will therefore allow a count of total PDA Synchs. Break-outs by channel can also be supported by this approach. This creates a comparable and indicative measure of a site's PDA audience. The agreed metric is not "PDA Users", as ABCe are not identifying the receiving device. This may become possible in future.

### **Requested Download**

163. **A request for a non-HTML file executable offline.** This typically includes audio files, video files, games and document files (e.g. .doc, .xls, .mp3, .ppt, .pdf). By definition, this excludes Streams, since they require a live connection to a server in order to execute (see Stream definition above).

### **Automated Requested Download**

164. **A request for a non-HTML file executable offline as a result of an automated process.** This must be broken out where it forms at least 5% of the total of Requested Downloads.

### **Completed Download**

165. **A completed request for a non-HTML file executable offline, shown by the number of bytes transferred being at least 95% of the stated size of the file.** When counting Completed Downloads, sites offering progressive download facilities need to demonstrate auditable adherence to this 95% rule over the course of the progressive download to a single User.

### **Automated Completed Download**

166. **A completed request for a non-HTML file executable offline, shown by the number of bytes transferred being at least 95% of the stated size of the file, as a result of an automated process.** This must be broken out where it forms at least 5% of the total of Completed Downloads.

### **Requested Podcast**

167. **A Requested Audio or Video Download made by a valid user following receipt of an automated notification of availability.** These are assumed to be automatically requested (i.e. Automated) unless the site can provide auditable evidence to the contrary.

### **Completed Podcast**

168. **A Completed Audio or Video Download made by a valid user following receipt of an automated notification of availability.** These are assumed to be automatically requested (i.e. Automated) unless the site can provide auditable evidence to the contrary.

### **Online Job Application**

169. **The submission to a server of an application form or a CV by a valid user.** The content of the application form or CV submitted is NOT reviewed as part of the audit. The site must provide auditable evidence of submission, such as a "thank you" Page Impression or a redirect to a "success" URL.

### **Publication Opened**

170. **A file, or combination of files, that represents the Page Impression for the first (arrival) page of an online publication served in response to a valid request by a valid user.** This metric applies only to properties which present themselves online in a print-style format typically built in rich media and requiring the user to "turn" virtual pages.

### **User-Initiated Logged Event**

- 171. Any logged event that can be attributed to a particular Unique User/Browser.** Such events are typically captured by browser-side measurement. Examples include mouse-overs, link views, menu selections or filling out of form fields. The use of such events allows more granularity in the measurement of Visit and Duration metrics. To be clear, the number of such events captured is **not** a certifiable metric itself.

## Visit and Duration Metrics (Frequency)

### Visit

- 172. 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.** A Visit is effectively a near-continuous burst of activity by a valid user. Note that, in addition to Page Impressions, the Site Publisher can use User-Initiated Logged Events to calculate this metric if desired.

### Visit Duration

- 173. 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.** In order to measure Visit Duration, a first and last Page Impression record must exist for each Visit. Therefore, Visits of only one page are excluded, since no interval can be established. Note that, in addition to Page Impressions, the Site Publisher can use User-Initiated Logged Events to calculate this metric if desired.

### Unique User/Browser Duration

- 174. The total time in seconds for all Visits of two or more Page Impressions, divided by the number of Unique User/Browsers making such Visits.** In order to measure Unique User/Browser Duration, a first and last Page Impression record (or other auditable logged event) must exist for each Visit. Therefore, Unique User/Browsers only making Visits consisting of only one Page Impression are excluded, since no interval can be established. Note that, in addition to Page Impressions, the Site Publisher can use User-Initiated Logged Events to calculate this metric if desired.

### Chat Duration (HTML only)

- 175. The total elapsed time in seconds between the first and last time stamp recorded for each valid user's Chat Impressions.** In order to measure Chat Duration, a first and last Chat Impression record must exist for each Unique User/Browser. Therefore, Users making single-record Chat Impressions, and any such Chat Impressions, are excluded. This metric should be used alongside the Chat Impression and Unique Chat User metrics.

## **AV Play Duration**

176. The number of seconds of content served to a valid user in an AV Play. Client-side, this would be calculated by measuring “end” events terminating the AV Play Event (such as Pause, Fast Forward or Stop) and calculating the gap between the Play event and the “end” event. Hence, AV Duration can only be measured when there is such an “end” event. Server-side, the Duration field should be used. Where this is not available, an approximation can be calculated from valid AV Requests by dividing the bytes sent to the user by the average bit-rate per second. Ideally, buffering time should be excluded from AV Play Duration calculations, but this may not always be practical.

## **Click Metrics (Interaction)**

### **Click**

177. **The activation of a hypertext link by a valid user.** Note that the site must have an auditable way of logging Clicks. The User must be a valid (non-robotic, non-internal) User.

### **Search Click**

178. **A Click originating from a set of Search results.**

### **Clickout**

179. **A Click to an external (non-native) destination.** This metric can also be known as “Referral” or “Departure”.

### **Job Clickout**

180. **A Click to a job listing on another recruitment site.**

### **Click Visit**

181. **A series of one or more Clicks, served to one valid user, which ends when that user has not made a Click for a 30-minute period.**

### **Online Job Referral**

182. **A Click on a link within a job listing to a URL specified by a Recruiter.**

### **Referral In**

183. **A Page Impression representing an arrival at a property by a valid user from another identifiable property.** Effectively, the event that needs to be detected is the arrival at the identified property by a user who has come from another property. To do this requires testing whether this Page Impression has a Referrer field that is native to the property or is Null; any non-null, non-native values may be counted as Referrals In. This metric can also be known as a Clickin.

### **Web Feed Referral In**

184. **The first Page Impression served to a valid user as a result of redirection from a link in a web feed clicked on by that user.**

### **Web Feed Job Referral In**

185. **A Page Impression to a URL specified by a Recruiter served to a valid user as a result of redirection from a link in a web feed job listing clicked on by that user.**

### **Email Referral In**

186. **A Page Impression representing an arrival at the property by a valid user from a hyperlink embedded in an email.** The destination property must be able to present auditable evidence of such arrivals.

### **Email Job Referral**

187. **A Click on a mailto link within a job listing intended for the submission of applicant information and/or a CV to a recruiter.**

### **CHAT metrics (non-HTML)**

Note that it is currently **not** technically possible to introduce a meaningful Chat Impressions metric.

### **Unique Chat Host**

188. **A unique IP address found within the valid chat log records.**

### **Unique Chat User**

189. **A unique combination of an IP address + a Nick name + a 5 second minimum stay (logged time in minus logged time out).** Sites are also required to show a break out from the above headline total, for Active Chat Users (Chat Users who log on and participate by contributing to the discussion) and Non-Active Chat Users (Chat Users who log on but do not actively participate, also known in the sector as “lurkers”).

### **Unique Active Chat User**

190. **A unique combination of an IP address + a Nick name + a 5 second minimum stay (logged time-in minus logged time-out) + 1 or more logged lines.**

### **Unique Chat User Duration**

191. **The total duration in seconds (measured by taking time joined from time left) of all valid Unique Chat Users, divided by the total valid Unique Chat Users.**

## **Geographical IP-based metrics**

### **Page Impressions By Country**

192. **The percentages by country of all Page Impressions produced by all resolved IP addresses during the Certification Period.** Note: These results were correct at the time of testing. Due to the constant changes in IP address ranges and registration, repeating the tests at a different time may not give exactly the same result. These results may appear in a tabular or pie-chart format – e.g. 15.24% UK, 6.36% US, x% unresolved.

### **Unique User/Browsers By Country**

193. **The percentages by country of all valid Unique User/Browsers identified during the Certification Period.** Note: This requires calculation of an IP address recorded in all Page Impressions made by each valid Unique User/Browser. These results were correct at the time of testing. Due to the constant changes in IP address ranges and registration, repeating the tests at a different time may not give exactly the same result. These results may appear in a tabular or pie-chart format – e.g. 15.24% UK, 6.36% US, x% unresolved.

## Defined terms used in this document

### **193. ACCESS LOG**

See Raw Log.

### **194. AUDIT PERIOD**

The period during which the auditor examines the supporting data (typically the web server or page tag log files) for traffic validity. The minimum Certification Period is a calendar month.

### **195. CERTIFICATION PERIOD**

See Audit Period.

### **196. COMPLIANT LOG**

A file of valid records that support the traffic claim for the certification period.

### **197. COOKIE**

A small text file automatically sent by the website to the user's browser upon visiting a page, which then continues to reside on the user's computer. Cookies must be persistent if they are to be used as Unique User/Browser identifiers.

### **198. COUNTING (aka Web Analytics)**

The process of analysing log file data to establish website traffic statistics preparatory to making a claim.

### **199. FILTERED FILE**

See Compliant Log.

### **200. FIRST-PARTY COOKIE**

A cookie set by the website domain currently being visited by the user.

### **201. GEOGRAPHICAL IP ANALYSIS**

Process used to determine as far as possible the country of origin of each Unique User/Browser or Page Impression. Typically involves, as a minimum, carrying out reverse DNS and Whois lookups to determine provenance.

### **202. INVENTORY DOMAIN**

(also INVENTORY URL where applicable)

An Inventory Domain is any domain, subdomain, URL tree or alias which is included in the traffic being audited, whether owned by the Site Publisher or a third party. Inventory Domains must be stated on the certificate to provide transparency and meaning for certificate users.

### **203. JICWEBS**

The Joint Industry Committee for Web Standards – the UK industry body which regulates the development of electronic media measurement. Members of this body are drawn from associations that represent the new media, publishing and advertising industries. The website is [www.jicwebs.org](http://www.jicwebs.org)

**204. PAGE TAGGING**

The embedding of an asset (typically Java, Javascript, CGI or a simple pixel) on a website, the rendering of which sends data to a collection point (e.g. a log file) every time a page is loaded into a browser, such that one execution of code is intended to equal one Page Impression.

**205. RAW LOG FILE** (also known as ACCESS LOG)

The original text-based record maintained by a web server(s) of all traffic which involves it.

**206. REFERRER**

The fully-qualified URL of the page from which, via a direct hyperlink, a user reached the current requested page (URL). It is typically logged including the http:// string.

**207. RSS**

“Really Simple Syndication” – the most popular web feed technology which allows headline or headline-and-short-summary feeds syndicated from other websites to be incorporated into a web page or read by a web feed reader.

**208. SUBSITED TRAFFIC**

Any Page Impression served for a website within a separate browser window (typically a pop-under) automatically generated by another website as a result of a request by a valid user for the content of the other website.

**209. SYNDICATED CONTENT**

(aka THIRD-PARTY HOSTED CONTENT)

Content served by a third party into the certified website’s Page Impressions, or content served by the certified website into a third party’s Page Impressions.

**210. THIRD-PARTY AUTHORED CONTENT**

Any content which is not authored by employees of the Site Publisher.

**211. THIRD-PARTY COOKIE**

A cookie set by any (website) domain other than that currently being visited by the Unique User/Browser.

**212. USER**

A human interacting with the Internet and leaving an audit trail of logged data. It is assumed that all compliantly filtered metric totals derive from valid users, who

are not robotic or internal. Note that this is expressly **not** the same as a Unique User/Browser.

**213. USERID+DATE FILE**

A data file listing either all dates on which each Unique User/Browser identifier (UserID) has visited the website during the certification period, or all identifiers present on the website during each day of the certification period.

**214. USERID+DATE+DOMAIN FILE**

As per the UserID+Date file but also including the separate domains visited by each Unique User/Browser identifier.

**215. VALID TRAFFIC**

Traffic which is not excluded according to the JICWEBS standards.

**216. WEBSITE (aka Site)**

The inventory, consisting of a single domain or a set of domains, defined by the Site Publisher for the purposes of an ABCe web audit.

**217. WEBSITE NAME**

The brand, name or trademark used by a website for marketing purposes.

**218. WEBSITE URL**

The top-level or brand identifying URL stated on the certificate, typically a homepage URL.