

# **Webstatistieken: achtergronden, mogelijkheden en valkuilen**

**Verslag deelonderzoek “Meer Digitale Feiten”  
in opdracht van Digitaal Erfgoed Nederland**



**Henk Voorbij**  
**Koninklijke Bibliotheek**  
**Universiteit van Amsterdam**

**Bauke Freiburg**  
**Video Dock**

**31 augustus 2009**



Op dit werk is een [Creative Commons Licentie](https://creativecommons.org/licenses/by-sa/4.0/) van toepassing.

## Inhoudsopgave

1. Inleiding	3
2. Achtergronden	4
2.1. Inleiding	4
2.2. Logfile analyse	4
2.3. Page tagging	6
2.4. Cookies	6
2.5. Vergelijking logging en tagging	7
3. Webstatistieken	13
3.1. Definities	13
3.2. Voorbeeld statistieken op basis van page tags: Google Analytics	15
3.3. Voorbeeld statistieken op basis van logfiles: AWStats	23
3.4. Categorieën en segmenten	24
3.5. Audio en video	24
4. Key performance indicators	26
5. Vergelijkend onderzoek	29
5.1. Vergelijking van page tagging programma's	20
5.2. Verschillen tussen logging en tagging programma's	31
5.3. Verklarend onderzoek	32
6. Conclusies	33
Literatuur	35
Bijlage 1. Voorbeeld statistieken aan de hand van AWStats	36

## 1. Inleiding

In 2008 is in opdracht van Digitaal Erfgoed Nederland (DEN) het onderzoeksproject “De Digitale Feiten” uitgevoerd. Aan de hand van een uitgebreide vragenlijst werden de stand van zaken rond digitalisering en de daarmee gepaard gaande kosten bij ruim honderd erfgoedinstellingen in kaart gebracht. Het vervolgonderzoek “Meer Digitale Feiten” heeft onder meer tot doel inzicht te verschaffen in de mogelijkheden om het gebruik van digitale voorzieningen te meten. Fysieke bezoekersaantallen geven in steeds mindere mate een goed beeld van het gebruik van een instelling. Ter aanvulling zijn gegevens over het gebruik van de website steeds meer noodzakelijk.

Het onderzoek bestaat uit vier delen.

1. *Literatuuronderzoek.* Dit is er met name op gericht inzicht te geven in de achtergronden van veel gehanteerde begrippen zoals hits, pageviews, bezoeken, (unieke) bezoekers en tijdsduur van het bezoek. Hoe zijn deze begrippen gedefinieerd? Hoe betrouwbaar zijn de resultaten? Hoe dienen de resultaten geïnterpreteerd te worden? In hoeverre zijn statistieken van verschillende programma's vergelijkbaar?
2. *Survey.* Een beknopte vragenlijst is verzonden naar 112 erfgoedinstellingen die in 2008 de zeer uitgebreide vragenlijst ‘De Digitale Feiten’ hadden beantwoord. De vragenlijst is bewust kort gehouden om niet weer een zwaar beroep te doen op deze instellingen. Het doel was vooral te achterhalen in hoeverre men beschikt over statistieken van het gebruik van de website en databases, en welke statistieken men zinvol acht voor een onderlinge vergelijking.
3. *Interviews.* Met een beperkt aantal instellingen zijn gesprekken gevoerd om nader inzicht te krijgen in de achtergrond van de belangrijkste statistieken en het belang dat de instelling aan dergelijke gegevens hecht.
4. *Analyse jaarverslagen.* Een aantal jaarverslagen van erfgoedinstellingen is bestudeerd om vast te stellen in hoeverre deze melding maken van webstatistieken.

Dit rapport bevat de bevindingen van het literatuuronderzoek. De resultaten van de overige drie deelonderzoeken zijn gerapporteerd in een afzonderlijk verslag, getiteld: *Het gebruik van webstatistieken in erfgoedinstellingen.*

## 2. Achtergronden

### 2.1. Inleiding

Webstatistieken geven inzicht in het gebruik van een website. De statistieken laten periodiek zien hoe vaak de site wordt bezocht, welke pagina's het meest zijn bekeken, op welke pagina's men binnenkomt en via welke pagina's men de site verlaat, hoeveel verschillende bezoekers er waren, uit welke landen of domeinen deze komen, hoe de bezoekers op de site zijn terecht gekomen (via een zoekmachine, via een link op een andere site, of rechtstreeks), op welke uren of dagen er het meeste verkeer is, en met welke operating systemen en browsers de bezoekers werken. Om deze statistieken te vervaardigen zijn meer dan honderd softwarepakketten beschikbaar.<sup>1</sup> Een belangrijk onderscheid is dat tussen logfile analyse en page tagging.

- Logfile analyse. Alle transacties worden vastgelegd in de logfiles van de server. Softwarepakketten zoals AWStats en Webalizer leiden uit deze gegevens kant en klare statistieken af. Omdat de pakketten gewoonlijk geïnstalleerd zijn op de server van de website, spreekt men wel van server-based applicaties.
- Page tagging. Deze methode vereist dat een klein stukje Javascript wordt geplaatst op elke pagina van de website. Elk bezoek aan deze pagina's wordt geregistreerd door Google Analytics of een andere page-tagging server, en verwerkt tot statistieken. Omdat de analyse gewoonlijk plaats vindt op een externe server, spreekt men wel van host-based applicaties.

Een derde optie, packet sniffing, is nooit goed van de grond gekomen. Programma's die gebaseerd zijn op packet sniffing analyseren het verkeer op het web en verzamelen daarmee gebruiksgegevens van veel webservers tegelijk. Uniek voor 'sniffers' is dat ze niet alleen inzicht geven in de content (welke pagina's zijn bekeken), maar ook in de delivery (bijvoorbeeld responsetijden, netwerkproblemen) (Peterson, 2005).

Een beperking van logging en tagging is dat het onvoldoende mogelijk is bezoekers van elkaar te onderscheiden. Cookies bieden deze mogelijkheid wel. In de praktijk maken vooral applicaties gebaseerd op page tagging gebruik van cookies.

### 2.2. Logfile analyse

De meeste servers ondersteunen het NCSA<sup>2</sup> Common Log Format, dat in één regel een aantal gegevens vastlegt, telkens wanneer de server bezocht wordt. Het betreft de volgende zeven gegevens: host – rfc931 – username – date:time – request – statuscode – bytes. Het NCSA Combined Log Format voegt daar nog drie elementen aan toe: referrer – user agent – cookies.

Een concreet voorbeeld<sup>3</sup> is:

```
125.125.125.125 - dsmith [10/Oct/1999:21:15:05 +0500] "GET /index.html
HTTP/1.0" 200 1043 "http://www.ibm.com/" "Mozilla/4.05 [en] (WinNT; I)"
"USERID=CustomerA; IMPID=01234"
```

---

<sup>1</sup> Zie onder meer:

<http://www.conversionrater.com/2007/03/10/complete-guide-to-web-analytics-solutions-2007-edition/>

[http://www.business.com/directory/internet\\_and\\_online/site\\_management/log\\_analysis/](http://www.business.com/directory/internet_and_online/site_management/log_analysis/)

[http://www.cbel.com/site\\_management\\_software/](http://www.cbel.com/site_management_software/)

<http://www.antezeta.com/web-analytics-resources.html>

<sup>2</sup> National Center for Supercomputing Applications

<sup>3</sup> Ontleend aan [http://publib.boulder.ibm.com/tividd/td/ITWSA/ITWSA\\_info45/en\\_US/HTML/guide/c-logs.html](http://publib.boulder.ibm.com/tividd/td/ITWSA/ITWSA_info45/en_US/HTML/guide/c-logs.html)

125.125.125.125	Het IP-adres van de bezoeker. De meeste IP-adressen worden aan de hand van het Domain Name System (DNS) omgezet in een domeinnaam, eindigend met de aanduiding van het land (zoals .nl of .de) of het type organisatie (zoals .org of .edu). <sup>4</sup> Vertaling van IP adres naar domeinnaam heet 'reverse DNS lookup'. Niet alle IP-adressen zijn geregistreerd in DNS, er resteren dus vele 'unresolved numerical addresses'.
-	Login naam van de gebruiker, vaak ontbrekend
Dsmith	De gebruikersnaam of ID van de bezoeker. Alleen bekend wanneer identificatie vereist is.
[10/Oct/1999:21:15:05 +0500]	Datum, tijd en tijdzone
"GET /index.html http/1.0"	De opgevraagde pagina, in dit geval de homepage (index.html); de http methode (GET); de http protocolversie(1.0)
200	De antwoordcode: 200 = OK Andere veel voorkomende opties zijn: 206 - Partial request successful (not complete) 301 – Moved permanently 304 – Not modified 400 - Bad request was made by the client 401 - Authorization is required for this document 403 - Access to this document is forbidden 404 - Document not found 500 - Server internal error 501 - Application method (either GET or POST) is not implemented 503 – Server is out of resources Voor een volledig overzicht, zie <sup>5</sup>
1043	De omvang in bytes van de opgevraagde pagina
"http://www.ibm.com/"	De referrer ofwel herkomst. In dit geval kwam de gebruiker via een link op <a href="http://www.ibm.com/">http://www.ibm.com/</a> op de site terecht. Een tweede mogelijkheid is via een zoekmachine. Het voorbeeld <a href="http://www.google.com/search?hl=en&amp;ie=iso-8859-1&amp;q=empire+state+college">http://www.google.com/search?hl=en&amp;ie=iso-8859-1&amp;q=empire+state+college</a> laat zien met welke zoekmachine en met welke termen gezocht is. <sup>6</sup> Een derde optie is rechtstreeks: via het intypen van een url of het aanklikken van een favoriet.
("Mozilla/4.05 [en] (WinNT; I)")	Webbrowser en platform van de bezoeker
"USERID=CustomerA;IMPID=01234"	Cookie

<sup>4</sup> Beter gezegd diens provider. Zo komen nogal wat bezoekers uit het nietige plaatsje Vienna, Virginia. Dit is de vestigingsplaats van de provider America Online (AOL).

<sup>5</sup> <http://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html>

<sup>6</sup> Ontleend aan <http://www.esc.edu/personal/klf/kftest.nsf/webpages/handout.html>

Logfile analyse is de oudste methode om webstatistieken te verzamelen. Oorspronkelijk was de analyse erop gericht om foutmeldingen te ontdekken, al snel daarna zag men in dat men langs deze weg ook inzicht in het gedrag van de bezoekers kon krijgen.

De omvang van de logfiles is vaak bijzonder groot. Raadpleging van een enkele pagina waarop zich vier afbeeldingen bevinden resulteert in vijf 'hits' en daarmee vijf regels in de logfile. Bovendien zijn alle logs min of meer chronologisch gerangschikt. De handelingen van eenzelfde bezoeker tijdens eenzelfde bezoek staan dus niet netjes onder elkaar. Programma's voor webanalyse destilleren uit deze veelheid van ruwe gegevens een aantal kant en klare statistieken.

### **2.3. Page tagging**

Page tagging houdt in dat de pagina's van een website worden voorzien van een tag, veelal in de vorm van een Javascript code. Elke keer dat een bezoeker een dergelijke pagina opvraagt maakt de website verbinding met een page-tagging server, bijvoorbeeld Google Analytics. De page-tagging server registreert onder meer welke pagina's iemand bezoekt en welke browser hij gebruikt. De page-tagging server slaat deze gegevens op in logbestanden en verwerkt ze tot webstatistieken. Uiteindelijk is ook nu sprake van logfile analyse. De terminologie logging versus tagging creëert een schijntegenstelling. In werkelijkheid gaat het om logfile analyse van onbewerkte pagina's versus logfile analyse van pagina's die voorzien zijn van een tag.

De werklast die gepaard gaat met het plaatsen van tags hangt af van de wijze waarop de website is gebouwd. Veel websites bestaan al lang niet meer uit een verzameling statische pagina's. Bij dynamische websites is, om het onderhoud te vereenvoudigen, de inhoud opgeslagen in een database of Content Management Systeem. Een pagina wordt pas gegenereerd op het moment dat een bezoeker deze opvraagt. Een veel voorkomende toepassing is het tonen van een navigatiebar op een vaste positie in elke pagina. Het plaatsen van een tag bij een website die dynamische pagina's genereert vereist vaak slechts een enkele handeling, terwijl in het geval van statische websites elke afzonderlijk pagina van een tag voorzien moet worden.

### **2.4. Cookies**

Een cookie is een klein tekstbestandje dat in de browser van je computer wordt opgeslagen bij bezoek aan sommige websites. De website stuurt het cookie naar de pc van de bezoeker, bij een volgend bezoek aan die site wordt het cookie teruggestuurd. Elke pc kan een groot aantal cookies bevatten.<sup>7</sup>

Cookies worden vooral gebruikt om een bezoeker te herkennen, aan de hand van een uniek nummer. Zij maken het daarom mogelijk bezoekers te wijzen op voor hen interessante producten (Amazon: 'we have recommendations for you'). Ze kunnen verhinderen dat iemand meerdere keren stemt in een poll. Ze zorgen ervoor dat het bestellen van producten sneller verloopt, omdat de bepaalde gegevens van de klant al bekend zijn na een eerdere bestelling. Theoretisch is het ook mogelijk om e-mail adressen, creditcardgegevens en andere vertrouwelijke gegevens in een cookie op te slaan. In de praktijk komt dit vanwege privacy- en beveiligingsbezwaren echter bijna nooit meer voor. Een cookie kan geen virus met zich meebrengen. Het kan alleen worden gelezen door de website die het cookie geplaatst heeft. Na verloop van tijd wordt het cookie automatisch verwijderd of verloopt de geldigheid.

Cookies zijn heel nuttig voor de gebruiker, maar kunnen ook worden ingezet om zijn gedrag op een bepaalde website of op het hele internet te volgen. De beheerder van een website kan met cookies bijhouden hoe vaak een bezoeker terugkomt, hoeveel en welke pagina's hij

---

<sup>7</sup> Enkele van de volgende passage zijn ontleend aan de heldere uiteenzetting op <http://www.hcc.nl/eCache/DEF/10/327.html>

bekijkt en hoe lang hij op de site blijft. Aan de hand hiervan kan de site zo nodig worden verbeterd en aangepast aan de behoefte van gebruiker.

Er zijn twee soorten cookies. Session cookies worden geïnstalleerd aan het begin van een bezoek en weer verwijderd als het bezoek beëindigd is. Zij leggen de activiteiten tijdens het bezoek aan de website vast en registreren begin- en eindtijd van de sessie. Permanent of persistent cookies blijven langere tijd staan in de browser van de bezoeker. Zij houden bij hoe vaak iemand op de site is geweest, wanneer hij er voor het eerst was, en wanneer de laatste keer was. Zonder deze informatie zou het niet mogelijk zijn nieuwe van terugkerende bezoekers te onderscheiden en dus inzicht te krijgen in de loyaliteit van een bezoeker.

Er is ook een onderscheid tussen first party en third party cookies, afhankelijk van de website die het cookie plaatst. First party cookies (ook wel direct cookies genoemd) worden geplaatst door de website die men op dat moment bezoekt. Er zijn echter ook bedrijven die cookies plaatsen via andere websites, vaak met reclamebanners. Zij hebben er, vanwege reclame- of marketingdoeleinden, belang bij mensen te volgen bij hun bezoek aan meerdere websites. Zo kan op basis van het surfgedrag bijvoorbeeld blijken dat iemand veel belangstelling heeft voor digitale fotografie. Hierop kan men inspelen door hem bij een volgend bezoek een reclame te tonen van digitale camera's. Aangezien de gebruiker de website van de adverteerder niet bewust heeft bezocht wordt dit een third party cookie of indirect cookie genoemd. Third party cookies worden vanwege privacy redenen vaak geblokkeerd of regelmatig verwijderd door gebruikers. Google Analytics en veel andere applicaties versturen daarom altijd first party cookies.<sup>8</sup>

Page tagging applicaties maken vaak, maar niet altijd gebruik van cookies. Ook pakketten gebaseerd op logfile analyse kunnen gebruik maken van (first party) cookies, maar dit vergt wel een speciale –vaak ingewikkelde - configuratie van de webserver.

## 2.5. Vergelijking logging en tagging

Logging en tagging berusten op verschillende grondslagen. Beide opties hebben voor- en nadelen. Deze hebben betrekking enerzijds op het beheer, anderzijds op meer inhoudelijke aspecten.<sup>9 10 11</sup> Onderstaande tabel maakt een vergelijking tussen beide categorieën. Daarbij is ervan uitgegaan dat page tagging applicaties gebruik maken van cookies.

	<i>Logfiles</i>	<i>Page tags</i>
<b>Data verzameling</b>	De logfiles, waarop de analyse berust, zijn reeds aanwezig op de eigen server.	Alle pagina's van de website dienen voorzien te worden van Javascript codes. De werklast hangt samen met de wijze waarop de website is gebouwd.
<b>Opslag webstatistieken</b>	De webstatistieken zijn opgeslagen op de eigen server. Dit kost veel opslagcapaciteit. Zij zijn wel beschikbaar voor hergebruik of nadere analyse.	De webstatistieken zijn opgeslagen op een externe server. Blijvende aanwezigheid is niet gegarandeerd. Een ander bezwaar is dat men geen zicht heeft op wat de leverancier in de toekomst met de verzamelde data gaat doen. De Duitse deelstaat Nordrhein-Westfalen heeft daarom uit privacyoverwegingen de

<sup>8</sup> <http://code.google.com/intl/nl/apis/analytics/docs/concepts/gaConceptsCookies.html#HowGAUsesCookies>

<sup>9</sup> [http://en.wikipedia.org/wiki/Web\\_analytics](http://en.wikipedia.org/wiki/Web_analytics)

<sup>10</sup> <http://www.advanced-web-metrics.com/docs/web-data-sources.pdf>

<sup>11</sup> [http://netinsight.unica.com/products/Tags\\_and\\_Logs.htm](http://netinsight.unica.com/products/Tags_and_Logs.htm)

		Analytics code van zijn website verwijderd. <sup>12</sup>
<b>Data verkeer</b>	Het verzamelen van logfiles heeft geen invloed op de snelheid van het internetverkeer.	Het laden van een pagina kan vertraging oplopen wanneer er problemen zijn met de page-tagging server. Om dat te voorkomen wordt wel aangeraden de tag onderaan de pagina te plaatsen, niet bovenaan. Het laden van de pagina gaat dan vooraf aan het lezen van het script, zodat de snelheid van het verkeer niet hoeft te leiden onder (problemen met) de registratie.
<b>Identificatie bezoekers</b>	<p>De identificatie vindt plaats op basis van het IP-adres. Daarmee gepaard gaat een aantal onnauwkeurigheden. De eerste twee leiden tot een te lage opgave van het aantal bezoekers, de laatste twee tot een te hoge opgave.</p> <ul style="list-style-type: none"> <li>• Het aantal bezoekers kan veel groter zijn dan het aantal IP adressen, indien verschillende bezoekers gebruik maken van eenzelfde pc (bijv op studiezalen).</li> <li>• Het aantal bezoekers kan veel groter zijn dan het aantal IP-adressen indien gebruikers de site bezoeken via een proxy server. In dat geval hebben alle bezoekers eenzelfde IP adres.<sup>13</sup> Vaak vindt daarom identificatie plaats op basis van de combinatie IP adres en user agent (operating system en browser). Dan is een iets beter, maar nog geen afdoende onderscheid mogelijk tussen verschillende bezoekers via eenzelfde proxy server.</li> <li>• Aan de andere kant kunnen internet service providers dynamische IP adressen verstrekken: dit kan zelfs</li> </ul>	<p>De identificatie vindt vaak plaats op basis van cookies opgeslagen in de pc van de bezoeker. De problematiek die gepaard gaat met IP-adressen (proxy servers, dynamische IP adressen) is hier dus niet van toepassing. Wel gelden enkele andere problemen. De eerste drie leiden tot een te laag aantal, de laatste twee tot een te hoog aantal.</p> <ul style="list-style-type: none"> <li>• Bezoekers kunnen cookies tegenhouden of verwijderen. Dit geldt met name voor third party cookies. Via logfile analyse kan worden vastgesteld hoe vaak een site bezocht wordt door bezoekers zonder cookies. Peterson (2005) stelt dat op 15%.</li> <li>• Bezoekers kunnen Javascript deactiveren, al wordt het dan lastig op internet te surfen en komt het niet vaak voor.<sup>14</sup></li> <li>• Net als bij IP adressen geldt dat verschillende bezoekers vanaf eenzelfde pc als eenzelfde gebruiker gezien worden.</li> <li>• Omgekeerd telt, net als bij IP adressen, eenzelfde bezoeker die vanaf verschillende pc's werkt als verschillende bezoekers.</li> </ul>

<sup>12</sup> <http://www.emerce.nl/nieuws.jsp?id=2814241>

<sup>13</sup> "Every single person inside the Ford corporation has the same IP address". Zie <http://www.thinkmetrics.com/inaccuracies-in-website-measurement.php>

<sup>14</sup> Volgens <http://www.googlelytics.net/awstats-log-file-analysis-vs-google-analytics/> gaat het wereldwijd om ongeveer 5% van de internetgebruikers.



	<p>binnen eenzelfde sessie plaatsvinden. In dat geval heeft eenzelfde bezoeker verschillende IP-adressen.</p> <ul style="list-style-type: none"> <li>• Eenzelfde bezoeker die vanaf verschillende pc's werkt, geldt als verschillende bezoekers.</li> </ul>	<ul style="list-style-type: none"> <li>• Ten slotte wordt ook eenzelfde bezoeker die met verschillende browsers werkt als verschillende bezoekers gezien.</li> </ul>
<b>Registreren bezochte pagina</b>	<ul style="list-style-type: none"> <li>• Eerder opgevraagde pagina's worden opgeslagen in het geheugen (cache) van de pc van de gebruiker. Dit kan op nog veel grotere schaal plaatsvinden wanneer de provider de pagina's die zijn opgevraagd door één van de aangesloten gebruikers in zijn cache opslaat. Wanneer een pagina opnieuw wordt opgevraagd, wordt deze uit de cache (van de gebruiker of de provider) gehaald en niet opnieuw verstuurd, om het internetverkeer te ontlasten. Er vindt nog wel controle plaats om te kijken of de pagina niet veranderd is. Dit resulteert in een 304 –antwoordcode.<sup>15</sup> Wanneer ook dit achterwege blijft, bijvoorbeeld wanneer een gebruiker via de backbutton teruggaat naar een zojuist bezochte pagina, vindt geen registratie plaats.<sup>16</sup> Webmasters kunnen op verschillende manieren caching tegengaan ('cache busting').<sup>17</sup></li> <li>• Een dynamische pagina kan bestaan uit meerdere losse HTML pagina's of kan uit meerdere bestanden worden samengesteld. Het oproepen van een enkele, uit</li> </ul>	<ul style="list-style-type: none"> <li>• Elke keer dat een pagina wordt geladen, wordt het JavaScript uitgevoerd. Page tagging programma's zijn dus niet gevoelig voor caching. Wel doet zich het omgekeerde back button probleem voor. Gebruikers verlaten de site soms door (herhaaldelijk) te drukken op de backbutton. Zij 'bezoeken' dan uitsluitend voor navigatiedoeleinden wederom een of meer pagina van de site. Dit wordt - formeel juist maar eigenlijk ten onrechte – als pageviews geteld.</li> <li>• Raadpleging van pagina's die al dan niet met opzet verstoken zijn gebleven van een tag, wordt niet gemeten</li> <li>• Het is lastig niet-html pagina's, zoals pdf en Wordbestanden, van tags te voorzien. Wel kan men een link naar dergelijke pagina's taggen en dus indirect raadpleging van die pagina's meten. Dan nog blijft raadpleging ongeregistreerd indien men direct (bijvoorbeeld via een zoekmachine) op een niet-html pagina belandt.</li> </ul>

<sup>15</sup> Het project COUNTER richt zich op uniformering van de gebruikstatistieken van elektronische tijdschriften van verschillende uitgevers. Een artikel wordt beschouwd als geraadpleegd bij de antwoordcodes 200 en 304. Zie [www.projectcounter.org](http://www.projectcounter.org)

<sup>16</sup> Volgens sommige bronnen wordt de cacheproblematiek sterk overdreven. Zie bijvoorbeeld <http://www.antezeta.com/web-log-file-javascript-embedded-page-tracking.html>

<sup>17</sup> Namelijk door in de HTTP header de metatag 'nocache' op te nemen, door het plaatsen van een script waardoor een pagina voortdurend een nieuwe naam toegewezen krijgt, of door de uiterste datum (expiration time) op nul te zetten. Een goede middenweg tussen snelheid van het internetverkeer en betrouwbaarheid van de statistieken is om alleen het opslaan in de cache van non-content onderdelen, zoals afbeeldingen en navigatie-elementen toe te staan.

	<p>verschillende onderdelen opgebouwde pagina resulteert in de registratie van evenzovele pageviews. Daarom is het aantal pageviews bij logfile analyse vaak hoger dan bij page-tagging (zie ook paragraaf 5.2).</p>	
<p><b>Identificatie dynamische pagina's</b></p>	<p>Programma's gebaseerd op logfile analyse zijn vaak niet in staat onderscheid te maken tussen verschillende dynamische pagina's. Sostre en Le Claire (2008) geven het voorbeeld van een website die dynamische pagina's genereert met aanduidingen zoals <code>article.php?id=178</code>. Programma's zoals AWStats en Webalizer beschouwen artikelen met bijvoorbeeld <code>id=110</code> en <code>id=750</code> als eenzelfde pagina, namelijk <code>artikel.php</code>. Het gedeelte na het vraagteken wordt als het ware veronachtzaamd.</p>	<p>Op page-tagging gebaseerde applicaties kunnen een dergelijk onderscheid wel maken. De tendens is om ingewikkelde tekens zoals vraagtekens in het url weg te laten en gemakkelijk leesbare url's te creëren, teneinde dynamische pagina's te optimaliseren voor zoekmachines.</p>
<p><b>Bezoek door robots</b></p>	<p>Deze bezoeken worden meegeteld. Vaak kan het programma dergelijke bezoeken wel herkennen en in aparte tellingen vastleggen. Uitfilteren van bezoeken van robots is gezien de grote omvang absoluut noodzakelijk. Het voorbeeld in bijlage 1 (Samenvatting) toont dat tegenover 5.853 pagina's bezocht door mensen 29.532 pagina's bezocht door robots staan. Bij een veelbezochte site als die van de Koninklijke Bibliotheek is de verhouding ongeveer 50-50. Programma's kunnen verschillen in de mate waarin zij robots herkennen en het bezoek door robots filteren. Er komen voortdurend nieuwe robots bij, zij worden ook steeds slimmer. Webmasters kunnen ook zelf robots op het spoor komen door af te gaan op bezoekers die zeer veel pagina's in een enkele sessie bekijken, slechts zeer kort op de pagina's vertoeven en op exact dezelfde tijd terugkeren. Logfile analyse geeft</p>	<p>Programma's gebaseerd op page tagging meten alleen het bezoek van gebruikers die JavaScript geactiveerd hebben. Robots van zoekmachines gebruiken geen JavaScript, zodat bezoek door robots niet wordt vastgelegd. De statistieken zijn in dit opzicht dus niet vervuild.</p> <p>In het verleden werd de mogelijkheid om klikpaden van robots te analyseren als een voordeel van logging gezien. Inmiddels hebben de grote zoekmachines zoals Google, Yahoo en Microsoft webmaster tools ter beschikking gesteld die direct inzicht geven in pagina's die door de robots bezocht en geïndexeerd zijn. Daarnaast laten zowel logfile als page-tagging applicaties zien vanaf welke website de bezoeker komt. In het geval dat een bezoeker is binnengekomen via een zoekmachine worden ook de zoekwoorden weergegeven. Daarmee is zichtbaar welke</p>

	gewoonlijk een hoger aantal pageviews dan analyse gebaseerd op page tagging. Dit verschil zal groter zijn naarmate meer robots niet herkend worden (zie ook paragraaf 5.2). <sup>18</sup>	pagina's goed vindbaar zijn via zoekmachines en welke pagina's wellicht aangepast moeten worden vanuit het oogpunt van zoekmachine optimalisatie.
<b>Foutmeldingen</b>	Alle antwoordcodes worden in statistieken verwerkt	Alleen het gebruik van pagina's die voorzien zijn van een tag wordt gemeten. Uiteraard vallen foutmeldingen hier buiten. Daarmee blijft de webmaster verstoken van informatie die van belang kan zijn om de site te verbeteren. Bij dynamische pagina's echter worden foutmeldingen zoals 404 ook meestal dynamisch gegenereerd. Door in de rapporten te kijken naar de 404 pagina en de pagina's die voorafgaand aan die pagina werden bezocht kan achterhaald worden waar foutieve links staan. Deze oplossing geldt niet voor alle typen foutmeldingen.
<b>Inspelen op technologische ontwikkelingen</b>	Logging geeft inzicht in raadpleging via mobiele telefoons die geen Javascript ondersteunen	Tagging geeft inzicht in het gebruik van web 2.0 applicaties gebaseerd op Ajax of Flash. Bovendien kunnen ook andere handelingen dan het opvragen van pagina's geregistreerd worden, zoals het lezen van mouse-over teksten en het scrollen van pagina's.

Met name proxy servers, dynamische IP adressen, caching, dynamische pagina's en niet als zodanig herkende robots staan de nauwkeurigheid van webstatistieken op basis van logfiles in de weg. De belangrijkste problemen bij page tagging zijn het blokkeren of verwijderen van cookies. In beide gevallen is de herkenning van individuele bezoekers niet volmaakt: verschillende gebruikers kunnen als één persoon beschouwd worden, omgekeerd wordt eenzelfde gebruiker soms als verschillende personen gezien. Hybride pakketten trachten de voordelen van beide methoden zo goed mogelijk te benutten en de nadelen zoveel mogelijk op te heffen. In de praktijk zijn hier echter niet veel voorbeelden van bekend.

<sup>18</sup> Sostre en LeClaire wijzen ook op andere niet-menselijke bezoekers, zoals hosting monitoring services, link checkers en e-mail harvesters

De belangrijkste verschillen zijn samengevat in onderstaande tabel.

	<i>Logging</i>	<i>Page tagging</i>
Werklast om ruwe gegevens te verzamelen	Nee	Soms
Statistieken opgeslagen op eigen server	Ja	Nee
Registratie gaat ten koste van snelheid internetverkeer	Nee	Minimaal
Identificatie bezoekers op basis van	IP-adres, vaak in combinatie met user agent (=browser)	Session cookies en persistent cookies
<i>Storende factoren bij meting aantal page views</i> * pagina's bezocht door robots * eerder bezochte pagina opgehaald uit cache * dynamische pagina telt als meerdere pageviews * pagina is niet voorzien van tag * raadpleging niet-html pagina's (bijv pdf, Word) lastig te registreren	Onvolledig uitgefilterd Ja Ja  Nee Nee	Nee Nee Nee  Ja Ja
<i>Storende factoren bij meting aantal bezoeken</i> * bezoeken door robots meegeteld	Onvolledig uitgefilterd	Nee
<i>Storende factoren bij meting aantal bezoekers</i> * verschillende personen gebruiken zelfde pc (zonder in te loggen) * zelfde persoon gebruikt meerdere pc's * zelfde persoon werkt met meerdere browsers  * verkeer verloopt via proxy server * dynamische IP-adressen * cookies tegengehouden of verwijderd * Javascript gedeactiveerd	Ja  Ja Ja, indien identificatie op basis van IP adres en user agent Ja Ja Nee Nee Nee	Ja  Ja Ja  Nee Nee Ja Ja

### 3. Webstatistieken

#### 3.1. Definities

De Web Analytics Association stelde in 2006 definities vast van de drie belangrijkste gegevens: unique visitors, visits / sessions en pageviews. In 2007 verscheen een uitgebreider rapport met definities van 26 begrippen.<sup>19</sup> Sommige daarvan zijn vooral van lokaal belang, zoals entry page (eerst bezochte pagina tijdens een bezoek), exit page (laatst bezochte pagina tijdens een bezoek), referrer (de pagina die men bezocht voorafgaand aan het bezoek van de site) en conversion (het aantal keren dat een specifieke handeling is verricht, zoals een online aankoop). Andere begrippen zijn potentieel van belang voor onderlinge vergelijking van erfgoedinstellingen. Hieronder volgt de essentie daarvan. Voor alle tellingen geldt dat bezoeken van non-human visitors zoals webrobots niet meegerekend dienen te worden.

##### 1. Pageviews (*bekeken pagina's*)

Het aantal bekeken pagina's gedurende de rapportageperiode. De meeste programma's bieden de mogelijkheid om te specificeren welke bestandstypen of antwoordcodes meegeteld moeten worden. Worden bijvoorbeeld Flash, multimediale bestanden of pdf-bestanden meegeteld? En telt men bijvoorbeeld gedeeltelijk geladen pagina's mee (antwoordcode 206)? Al naar gelang de specificaties kunnen verschillende pakketten verschillende tellingen opleveren.<sup>20</sup>

Het aantal *hits* is een verouderde en misleidende maat. In de begindagen van het web bestonden webpagina's vaak uit een enkel html-bestand. Het aantal hits gaf toen een goede indicatie van het aantal bekeken pagina's. Met het toenemen van de complexiteit van webpagina's is deze maat steeds minder waardevol. Een pagina die twee afbeeldingen (bijvoorbeeld een logo) bevat, zal drie hits genereren (een voor de html-pagina, twee voor de afbeeldingen), terwijl in feite maar een pagina bezocht wordt. Op een pagina kunnen zich ook onzichtbare afbeeldingen bevinden, die door de webmaster zijn aangebracht ter wille van de lay-out: html biedt in dit opzicht weinig mogelijkheden. De logs voor afbeeldingen zijn overigens wel herkenbaar (bijvoorbeeld gif bestanden) en kunnen uit de statistieken gefilterd worden. De meeste programma's geven dan ook afzonderlijke tellingen van het aantal hits en het aantal pageviews.

##### 2. Visits / sessions (*bezoeken*)

Een bezoek aan een site bestaat uit een serie opeenvolgende handelingen binnen een bepaalde tijdsduur. Gewoonlijk wordt een sessie als beëindigd beschouwd na dertig minuten non-activiteit. Verschillende definities (andere tijdsduur) leveren verschillende tellingen op. Het aantal bezoeken kan nooit hoger zijn dan het aantal bekeken pagina's

##### 3. Visit duration (*gemiddelde duur van het bezoek*)

De duur van het bezoek wordt berekend door het verschil van het tijdstip van de laatste handeling en het tijdstip van de eerste handeling. De tijd die de bezoeker doorbrengt op de laatste pagina wordt dus niet gemeten. Wanneer tijdens het bezoek maar een enkele pagina wordt geraadpleegd, wordt gewoonlijk geen tijdsduur berekend. Uiteraard betekent de vastgestelde tijdsduur niet noodzakelijk dat de bezoeker de gehele tijd actief met de site is

---

<sup>19</sup> <http://www.webanalyticsassociation.org/attachments/committees/5/WAA-Standards-Analytics-Definitions-Volume-I-20070816.pdf>

<sup>20</sup> Bij programma's gebaseerd op tagging hebben webmasters niet de mogelijkheid zelf te bepalen welke bestandstypen of antwoordcodes meegeteld worden. Deze programma's meten immers alleen het gebruik van daadwerkelijk verzonden html-pagina's. Er valt dus niet veel te kiezen.

bezig geweest; wellicht heeft hij andere activiteiten ondernomen zonder de pagina weg te klikken.

#### 4. *Single pageview visits (bounces)*

Het aantal bezoeken waarin slechts een enkele pagina wordt geraadpleegd. Op basis hiervan wordt de bounce rate (weigeringspercentage) berekend. Dit is het aantal single pageview visits gedeeld door het totaal aantal visits. Single pageview visits moeten niet verward worden met single-page visits. Dit zijn bezoeken waarbij een enkele pagina een of meerdere keren wordt geraadpleegd. Een hoge bounce rate geeft aan dat bezoekers ofwel direct vonden wat ze zochten of al snel in de gaten hadden dat de site voor hen niet interessant was.

#### 5. *Unique visitors (unieke bezoekers)*

Het aantal verschillende personen dat de site bezoekt gedurende de rapportageperiode. Een bezoeker geldt slechts als uniek zolang de rapportageperiode duurt. Een gebruikelijke rapportageperiode is een maand. Iemand die gedurende vijf maanden in een jaar de site bezoekt telt vijf keer als een unieke bezoeker, maar dit zijn niet vijf verschillende personen. Het aantal unieke bezoekers mag dus *niet* gecumuleerd worden tot een jaarlijks totaal. Indien men een rapportageperiode aanhoudt van een dag, kan eenzelfde persoon als maximaal 365 unieke bezoekers gedurende een jaar tellen.

De definitie volgens de Joint Industry Committee for Web Standards (JICWEBS) is zeer verwarrend. Het aantal unieke bezoekers per maand wordt berekend door het aantal unieke bezoekers per dag te tellen, en dat te vermenigvuldigen met 31. Dit levert een aantal op dat veel groter is dan het werkelijke aantal verschillende bezoekers in een maand.<sup>21</sup>

Identificatie van de gebruikers kan plaatsvinden aan de hand van verplichte identificatie, persistent cookies of het IP adres. De meest voorkomende methode is gebruik te maken van cookies. Een probleem is dat gebruikers cookies kunnen verwijderen of blokkeren. Het gevolg is dat iemand die een tweede bezoek aflegt aan de site binnen de rapportageperiode niet herkend wordt als een eerdere bezoeker. Het aantal unieke bezoekers is dan in werkelijkheid lager dan de tellingen aangeven.

Het aantal unieke bezoekers kan nooit hoger zijn dan het aantal bezoeken.

#### 6. *New visitor (nieuwe bezoeker)*

Een nieuwe bezoeker is een unieke bezoeker die voorafgaand aan de rapportageperiode de site nooit eerder bezocht heeft. Identificatie geschiedt op basis van persistent cookies.

#### 7. *Return visitor (terugkerende bezoeker)*

Een terugkerende bezoeker is een unieke bezoeker die voorafgaand aan de rapportageperiode de site al eerder bezocht heeft. Identificatie geschiedt op basis van persistent cookies.

Het aantal nieuwe bezoekers plus het aantal terugkerende bezoekers is gelijk aan het totaal aantal unieke bezoekers.

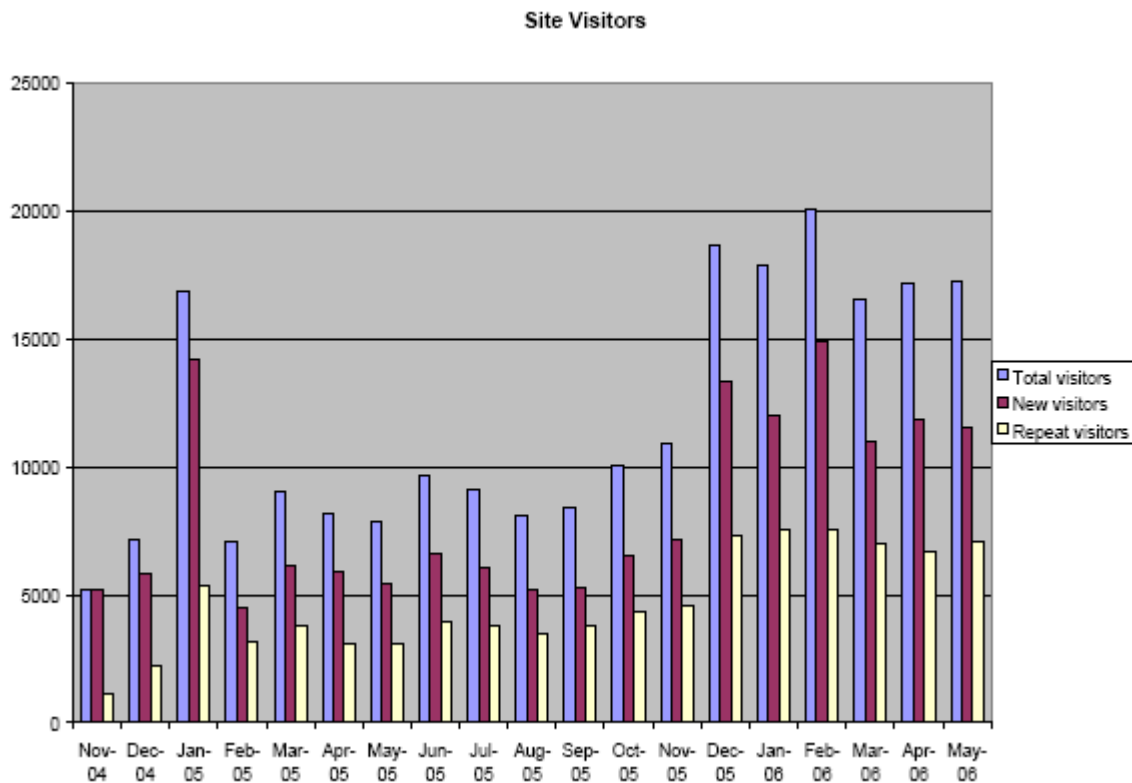
#### 8. *Repeat visitor (herhaaldelijke bezoeker)*

Een repeat visitor is een bezoeker die tijdens de rapportageperiode meer dan een bezoek aflegt aan de site. Anders dan bij het begrip return is het niet nodig om naar de voorgeschiedenis van de gebruiker te kijken. Een bezoeker kan binnen eenzelfde rapportageperiode zowel een new als een repeat visitor zijn. Het is daarom wel mogelijk het aantal new en returning visitors op te tellen (met als resultaat het aantal unieke bezoekers), maar niet het aantal new en repeat

---

<sup>21</sup> Althans volgens <http://www.thinkmetrics.com/inaccuracies-in-website-measurement.php>, de website van JICWEBS zelf (<http://www.jicwebs.org/standards.php>) is weinig informatief.

visitors. Onderstaande grafiek<sup>22</sup> illustreert dat de som van het aantal new en repeat visitors hoger is dan het aantal unieke bezoekers. Het is daarom van groot belang het onderscheid tussen return en repeat goed in het oog te houden.



### 3.2. Voorbeeld statistieken op basis van page tags: Google Analytics

Hieronder zijn ter illustratie enkele schermen afgebeeld van een rapportage van eenzelfde site gedurende een zelfde onderzoeksperiode met behulp van Google Analytics.<sup>23</sup> Deze hebben betrekking op of zijn nauw gerelateerd aan de acht eerder gedefinieerde begrippen. Daarnaast bevatten de rapportages nog tal van gegevens die vooral van lokaal belang zijn.<sup>24</sup>

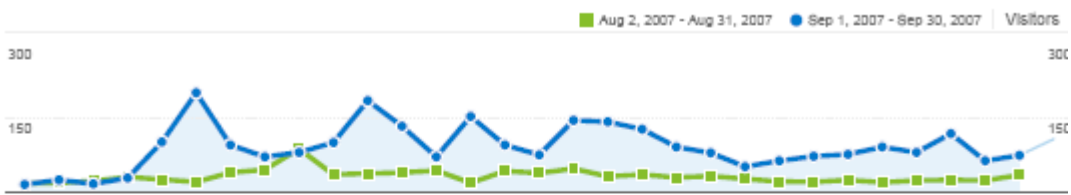
#### 1. Visitors Overview

Het rapport Visitors Overview geeft de meest essentiële informatie. Er waren 7390 pageviews tijdens 3005 bezoeken. Het gemiddelde aantal pageviews per bezoek bedraagt dus 2,46. De gemiddelde tijdsduur van een bezoek was 2 minuten en 49 seconden. De bounce rate was 57,4%, wat betekent dat tijdens 1725 van de 3005 bezoeken slechts een pagina bekeken is. Een niet op het eerste gezicht duidelijk begrip is New Visits. Dit betekent dat 2368 van de 3005 (78,80%) bezoeken afkomstig waren van nieuwe gebruikers, dus van gebruikers die voorafgaand aan de rapportageperiode de site nog nooit eerder bezocht hebben (afgaand op de persistent cookie). Het betekent niet dat 78,80% van de bezoekers nu voor het eerst de site bezocht.

<sup>22</sup> Ontleend aan [http://www.tmc.state.ri.us/pdf/rhodeways\\_stats.pdf](http://www.tmc.state.ri.us/pdf/rhodeways_stats.pdf)

<sup>23</sup> [http://www.butkaj.com/statistics?id\\_menu=749](http://www.butkaj.com/statistics?id_menu=749)

<sup>24</sup> Zie bijvoorbeeld <http://www.google.com/adwords/learningcenter/text/29518.html>



**2,447 people visited this site**

**3,005 Visits**  
 Previous: 984 (205.39%)

**2,447 Absolute Unique Visitors**  
 Previous: 693 (174.02%)

**7,390 Pageviews**  
 Previous: 2,097 (252.41%)

**2.46 Average Pageviews**  
 Previous: 2.13 (15.40%)

**00:02:49 Time on Site**  
 Previous: 00:02:02 (38.66%)

**57.40% Bounce Rate**  
 Previous: 59.35% (-3.28%)

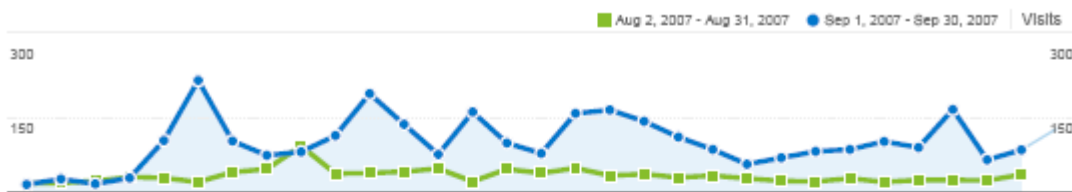
**78.80% New Visits**  
 Previous: 85.26% (-7.58%)

## 2. New vs Returning

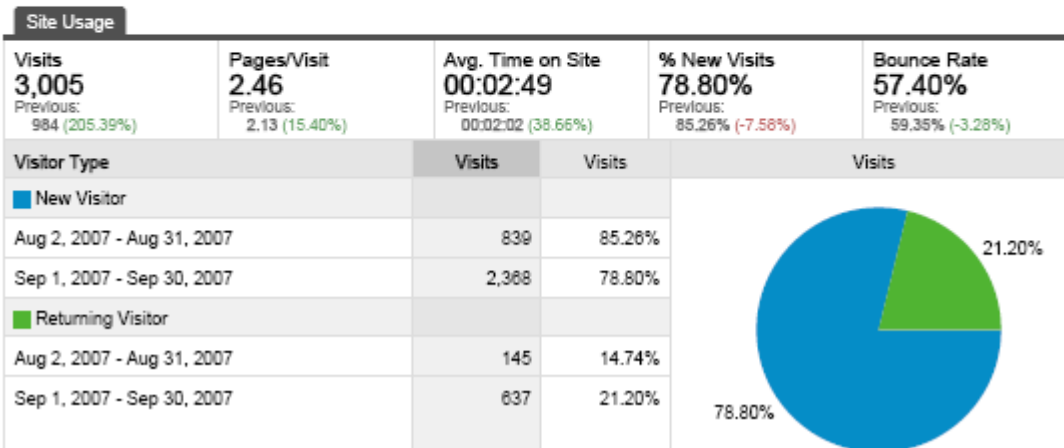
Het rapport Visitors Overview toonde reeds dat 78,80% van de bezoeken afkomstig was van nieuwe gebruikers, en dat er totaal 3005 bezoeken waren. Op basis daarvan kan al berekend worden dat het aantal bezoeken van nieuwe bezoekers in absolute zin  $0,788 * 3005 = 2368$  bedroeg, en dat het restant van de bezoeken (637 ofwel 21,20%) afkomstig was van eerdere bezoekers. Het rapport New vs Returning toont deze waarden expliciet.

Het rapport New vs Returning hanteert dus bezoeken, niet bezoekers als meeteenheid. Het maakt geen onderscheid tussen nieuwe en terugkerende bezoekers, maar tussen bezoeken van nieuwe en terugkerende bezoekers. Google Analytics hanteert dus een afwijkende, niet direct duidelijke maat.





## 3,005 visits from 2 visitor types













### 3. Visitor Loyalty

Het rapport Visitor Loyalty geeft in feite een nadere specificatie van de 637 return visitors, beter gezegd van de 637 visits van return visitors. Zo blijkt bijvoorbeeld dat 49 bezoeken in de rapportageperiode afkomstig zijn van gebruikers die de site nu voor de vierde keer gebruiken. Deze gebruikers hebben de site dus drie keer eerder bezocht voorafgaand aan de rapportageperiode. Hoeveel gebruikers dit zijn vertelt de tabel niet. Het kunnen 49 verschillende personen zijn, in theorie zou het ook kunnen dat eenzelfde persoon de site in september 2007 49 keer bezoekt, nadat hij daarvoor al drie keer op de site was geweest. De gegevens zijn niet gemakkelijk te begrijpen, misverstanden bij de interpretatie liggen op de loer. Het betekent in elk geval niet dat 49 gebruikers de site nu voor de vierde keer bezoeken. Het zou iets duidelijker zijn wanneer de kopjes 'number of visits' en 'visits' zouden luiden: 'number of prior visits' en 'number of visits during reporting period'.

Google Analytics geeft geen inzicht in het aantal repeat visitors: het aantal bezoekers dat de site gedurende de rapportageperiode meerdere malen bezoekt. Het rapport Visitor Loyalty geeft inzicht in het gebruik van de site voorafgaand aan de rapportageperiode en komt daarmee het meest in de richting van repeat visit(or)s

Onderstaande screenshot toont een deel van de rapportage.

**Most people visited: 1 times**








Number of Visits	Visits	Percentage of all visitors
<b>1 times</b>		
Aug 2, 2007 - Aug 31, 2007	839	 85.26%
Sep 1, 2007 - Sep 30, 2007	2,368	 78.80%
<b>2 times</b>		
Aug 2, 2007 - Aug 31, 2007	91	 9.25%
Sep 1, 2007 - Sep 30, 2007	285	 9.48%
<b>3 times</b>		
Aug 2, 2007 - Aug 31, 2007	20	 2.03%
Sep 1, 2007 - Sep 30, 2007	90	 3.00%
<b>4 times</b>		
Aug 2, 2007 - Aug 31, 2007	5	 0.51%
Sep 1, 2007 - Sep 30, 2007	49	 1.63%
<b>5 times</b>		
Aug 2, 2007 - Aug 31, 2007	7	 0.71%
Sep 1, 2007 - Sep 30, 2007	27	 0.90%
<b>6 times</b>		

**4. Visitor Recency**

Het rapport Visitor Recency kijkt op andere wijze dan Visitor Loyalty naar de voorgeschiedenis van de 3005 bezoeken tijdens de rapportageperiode. Stel dat een gebruiker de site tijdens de rapportageperiode drie keer bezoekt, op 2 september, 24 september en nog een keer op 24 september, en dat zijn laatste bezoek voorafgaand aan de rapportageperiode 30 augustus was. De recency waarden zijn dan respectievelijk drie, 22 en nul dagen.

Onderstaande screenshot toont een deel van de rapportage. Onder de 2745 bezoeken afkomstig van gebruikers die de site nul dagen eerder bezocht hebben, bevinden zich in elk geval de 2368 bezoeken van nieuwe bezoekers. Tevens blijkt (niet zichtbaar in de screenshot) dat één bezoek werd afgelegd door een gebruiker die meer dan een jaar geleden de site voor het laatst bezocht.











**Most people last visited: 0 days ago**

Last Visit	Visits	Percentage of all visitors
<b>0 days ago</b>		
Aug 2, 2007 - Aug 31, 2007	895	 90.96%
Sep 1, 2007 - Sep 30, 2007	2,745	 91.35%
<b>1 days ago</b>		
Aug 2, 2007 - Aug 31, 2007	9	 0.91%
Sep 1, 2007 - Sep 30, 2007	46	 1.53%
<b>2 days ago</b>		
Aug 2, 2007 - Aug 31, 2007	5	 0.51%
Sep 1, 2007 - Sep 30, 2007	25	 0.83%
<b>3 days ago</b>		
Aug 2, 2007 - Aug 31, 2007	4	0.41%
Sep 1, 2007 - Sep 30, 2007	30	 1.00%
<b>4 days ago</b>		

*5. Length of Visit*

Het rapport Visitors Overview toonde reeds de gemiddelde tijdsduur van een bezoek. Het rapport Length of Visit geeft nadere specificaties. Zo zijn bijvoorbeeld (1758 + 202 =) 1960 bezoeken (65,22%) binnen dertig seconden beëindigd, terwijl (niet zichtbaar in de screenshot) 62 bezoeken langer dan 1800 seconden duurden.











**Most visits lasted: 0-10 seconds**

Length of Visit	Visits	Percentage of all visitors
<b>0-10 seconds</b>		
Aug 2, 2007 - Aug 31, 2007	610	 61.99%
Sep 1, 2007 - Sep 30, 2007	1,758	 58.50%
<b>11-30 seconds</b>		
Aug 2, 2007 - Aug 31, 2007	79	 8.03%
Sep 1, 2007 - Sep 30, 2007	202	 6.72%
<b>31-60 seconds</b>		
Aug 2, 2007 - Aug 31, 2007	67	 6.81%
Sep 1, 2007 - Sep 30, 2007	200	 6.66%
<b>61-180 seconds</b>		
Aug 2, 2007 - Aug 31, 2007	91	 9.25%
Sep 1, 2007 - Sep 30, 2007	344	 11.45%
<b>181-600 seconds</b>		
Aug 2, 2007 - Aug 31, 2007	83	 8.43%
Sep 1, 2007 - Sep 30, 2007	285	 9.48%

## 6. Depth of Visit

Het rapport Visitors Overview meldde reeds een bounce rate van 57,40%. Dat wil zeggen dat in 57,40% van de bezoeken slechts een pagina is bekeken. Het rapport Depth of Visit geeft nadere specificaties. In ruim 21% van de bezoeken zijn twee pagina's bekeken, etc. Onderstaande screenshot toont een deel van de rapportage.

### Most people visited: 1 pages

Depth of Visit	Visits	Percentage of all visitors
<b>1 pages</b>		
Aug 2, 2007 - Aug 31, 2007	584	 59.35%
Sep 1, 2007 - Sep 30, 2007	1,725	 57.40%
<b>2 pages</b>		
Aug 2, 2007 - Aug 31, 2007	210	 21.34%
Sep 1, 2007 - Sep 30, 2007	644	 21.43%
<b>3 pages</b>		
Aug 2, 2007 - Aug 31, 2007	80	 8.13%
Sep 1, 2007 - Sep 30, 2007	258	 8.50%
<b>4 pages</b>		
Aug 2, 2007 - Aug 31, 2007	42	 4.27%
Sep 1, 2007 - Sep 30, 2007	111	 3.69%
<b>5 pages</b>		
Aug 2, 2007 - Aug 31, 2007	12	 1.22%
Sep 1, 2007 - Sep 30, 2007	70	 2.33%

## 7. Traffic sources overview

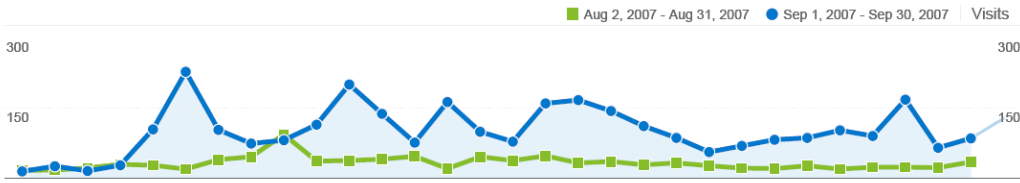
Deze rapportage toont de herkomst van het verkeer, in drie categorieën. Bij meer dan de helft van de bezoeken (57,4%) kwamen bezoekers op de site door het url in te typen of een favoriet of bookmark aan te klikken. Bij 24,8% van de bezoeken was dat via een link op een andere site, bij 17,6% door een treffer in de resultatenlijst van een zoekmachine aan te klikken. Instellingen dienen zelf te bepalen in hoeverre dit een gunstige mix is. Het lijkt erop dat deze site vooral bezocht wordt door mensen die het url al kenden. Inzicht in de herkomst is van belang voor marketingdoeleinden en kan aanleiding geven de site te optimaliseren voor zoekmachines.

Voor elk van de drie toegangswegen zijn gedetailleerde rapporten beschikbaar. Hieronder is de specificatie van het bezoek via zoekmachines afgebeeld. Per zoekmachine ziet men niet alleen het aantal bezoeken, maar ook nadere gegevens van de aldus gegenereerde bezoeken, zoals het aantal bezochte pagina's per bezoek en de gemiddelde tijd die men op de site doorbracht. Juist dergelijke gedetailleerde statistieken zijn bijzonder waardevol.

# Traffic Sources Overview

Sep 1, 2007 - Sep 30, 2007

previous: Aug 2, 2007 - Aug 31, 2007

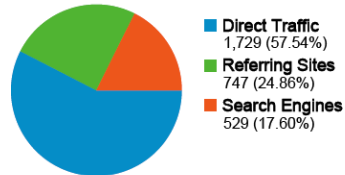


All traffic sources sent a total of 3,005 visits

**57.54% Direct Traffic**  
Previous: 51.42% (11.89%)

**24.86% Referring Sites**  
Previous: 22.15% (12.21%)

**17.60% Search Engines**  
Previous: 26.42% (-33.38%)



# Search Engines

Sep 1, 2007 - Sep 30, 2007

previous: Aug 2, 2007 - Aug 31, 2007



Search sent 529 total visits via 10 sources

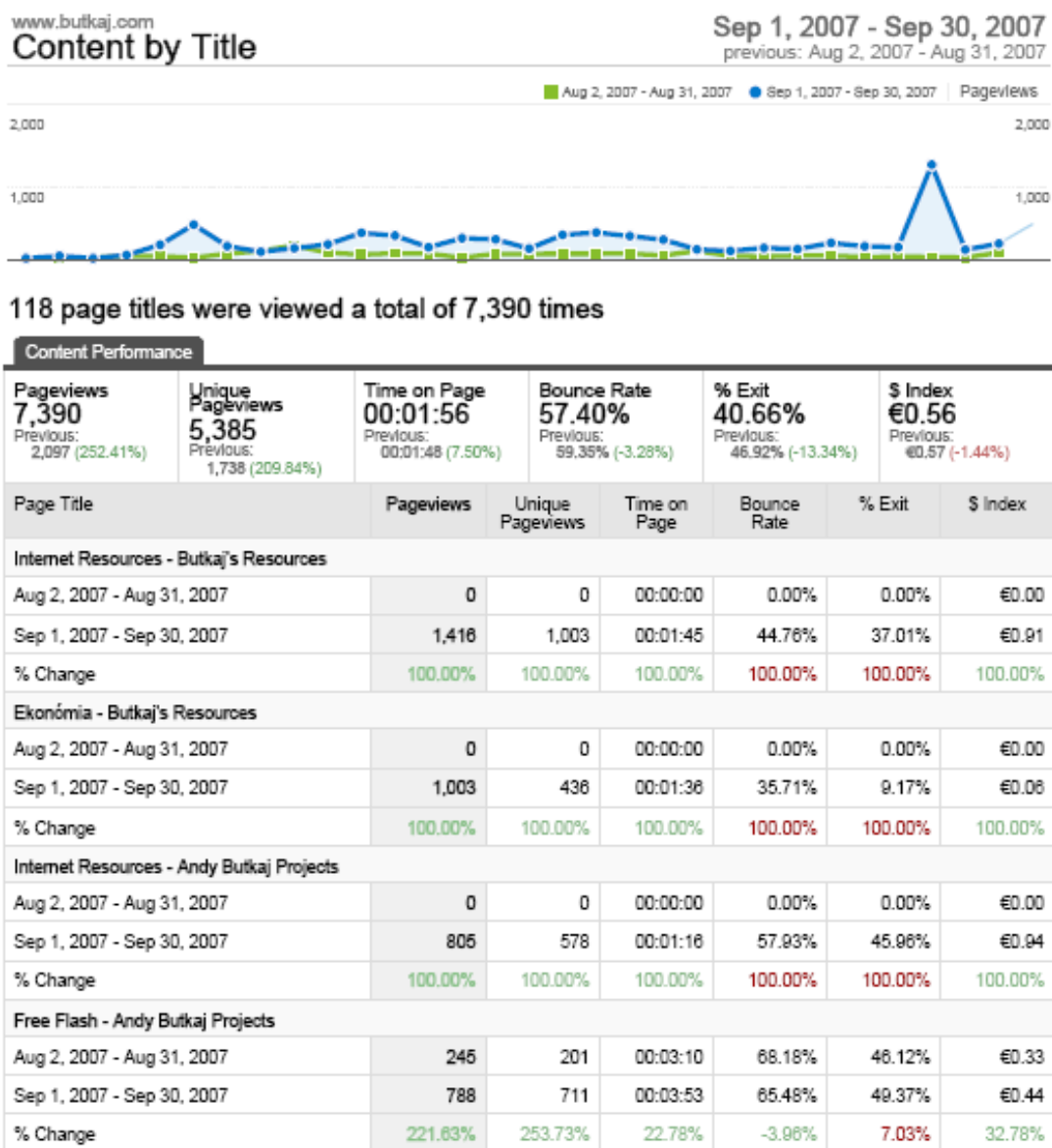
### Site Usage

Visits	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate	
<b>529</b> Previous: 260 (103.46%)	<b>2.13</b> Previous: 2.58 (-17.57%)	<b>00:02:06</b> Previous: 00:02:10 (-2.97%)	<b>85.44%</b> Previous: 82.69% (3.33%)	<b>66.54%</b> Previous: 60.00% (10.90%)	
Source	Visits	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
<b>google</b>					
Aug 2, 2007 - Aug 31, 2007	238	2.64	00:02:15	81.51%	61.76%
Sep 1, 2007 - Sep 30, 2007	489	2.14	00:02:06	85.07%	67.48%
% Change	105.48%	-18.93%	-6.66%	4.37%	9.26%
<b>yahoo</b>					
Aug 2, 2007 - Aug 31, 2007	14	2.07	00:01:16	92.86%	35.71%
Sep 1, 2007 - Sep 30, 2007	19	2.11	00:01:40	89.47%	47.37%
% Change	35.71%	1.63%	31.12%	-3.64%	32.83%
<b>live</b>					
Aug 2, 2007 - Aug 31, 2007	5	2.00	00:00:34	100.00%	60.00%
Sep 1, 2007 - Sep 30, 2007	8	2.00	00:01:39	87.50%	62.50%
% Change	60.00%	0.00%	188.74%	-12.50%	4.17%
<b>msn</b>					
Aug 2, 2007 - Aug 31, 2007	1	1.00	00:00:00	100.00%	100.00%
Sep 1, 2007 - Sep 30, 2007	5	2.20	00:01:49	100.00%	40.00%
% Change	400.00%	120.00%	100.00%	0.00%	-60.00%
<b>altavista</b>					
Aug 2, 2007 - Aug 31, 2007	0	0.00	00:00:00	0.00%	0.00%
Sep 1, 2007 - Sep 30, 2007	3	3.00	00:09:35	100.00%	33.33%

## 8. Content by Title

Dit rapport geeft nadere informatie over de bezochte pagina's. Per pagina ziet men, in volgorde van frequentie, het aantal pageviews. Het herladen van een eenmaal opgevraagde pagina (refresh) telt als een nieuwe pageview. Ook wanneer een bezoeker via de backtoets terugkeert naar een eerder opgevraagde pagina, telt dit als een nieuwe pageview. Bij unique pageviews wordt eenzelfde pagina, die tijdens een sessie meerdere malen is bekeken, ontdebeld. Het aantal unique pageviews is dus gelijk aan het aantal sessies waarin de pagina is bekeken.<sup>25</sup> Uit de tabel blijkt dat de pagina "Internet Resources – Butkaj's Resources" in september 2007 1.416 keer bekeken is bij 1.003 verschillende bezoeken. De cumulatie van het totale aantal unique pageviews is met 5.385 groter dan het aantal bezoeken (3.005) en lijkt daarom weinig zinvol.

Per pagina ziet men ook nog andere gegevens, zoals de gemiddelde tijdsduur. Een hoog aantal pageviews, maar ook een hoge tijdsduur geeft aan dat bezoekers de pagina van belang achten. Ook dit voorbeeld illustreert het belang van gedetailleerde statistieken.



<sup>25</sup> <http://www.google.com/support/analytics/bin/answer.py?hl=en&answer=57164>

### 3.3. Voorbeeld statistieken op basis van logfiles: AWStats

AWStats (een afkorting van Advanced Web Statistics) levert de volgende statistieken.

Samenvatting	
<b>1. Wanneer:</b>	1.1 Maandelijks historie 1.2 Dagen van maand 1.3 Weekdagen 1.4 Uren
<b>2. Wie:</b>	2.1 Domeinen / Landen 2.2 Hosts 2.3 Robots / Spiders bezoekers
<b>3. Navigatie:</b>	3.1 Duur bezoeken 3.2 Bestandstypen 3.3 Pagina's, Binnenkomst pagina's, Uitgang pagina's 3.4 Operating Systems 3.5 Browsers
<b>4. Verwijzing:</b>	4.1 Herkomst, Verwijzende zoekmachines, Verwijzende sites 4.2 Zoektrefzinnen, zoektermen
<b>5. Andere:</b>	5.1 Overigen (bijv add to favourites) 5.2 HTTP foutmelding codes 5.3 Niet gevonden pagina's
<b>6. Extra/Marketing:</b>	List of top level 2 path under /dossiers

Voor een gedeelte zijn dit dezelfde soort gegevens die Google Analytics levert. Kenmerkend voor pakketten die gebaseerd zijn op logfile analyse is dat ze ook inzicht geven in het bezoek van robots en foutmeldingen (zoals code 404 – page not found). Aan de andere kant is het, omdat AWStats niet met persistent cookies werkt, niet mogelijk bezoekers nader te kwalificeren als new, return of repeat. Tevens ontbreekt de bounce rate. In principe zou AWStats wel in staat moeten zijn deze te verschaffen: een applicatie die het aantal sessies kan meten, kan ook de bounce rate bepalen. Waarschijnlijk heeft de afwezigheid te maken met de aard van de applicatie: page tagging applicaties zijn meer gericht op commerciële websites en daarom eerder geneigd bounce rates te rapporteren dan programma's gebaseerd op logfiles.

Voor een onderlinge vergelijking van erfgoedinstellingen lenen zich vooral de samenvatting en de tabel met de gemiddelde duur van de bezoeken. Bijlage 1 toont een vollediger overzicht.

#### 1. Samenvatting

De samenvatting geeft de meest essentiële informatie. Het aantal bekeken pagina's, bezoeken en unieke bezoekers vormen samen de 'Big three'. Tevens is een onderscheid gemaakt tussen pagina's bekeken door robots en menselijke bezoekers. Het verkeer afkomstig van robots is zeer aanzienlijk, in dit voorbeeld vele malen groter dan dat van menselijke bezoekers

Samenvatting					
Rapportageperiode	Maand febr. 2009				
Eerste bezoek	01 febr. 2009 - 00.34				
Laatste bezoek	23 febr. 2009 - 06.38				
	Unieke bezoekers	Aantal bezoeken	Pagina's	Hits	Bytes
Bekeken verkeer *	1374	2617 (1.9 bezoeken/bezoeker)	5853 (2.23 Pagina's/bezoek)	32659 (12.47 Hits/bezoek)	84.63 MB (33.11 KB/bezoek)
Niet-bekeken verkeer *			29532	54691	135.68 MB

\* „Niet bekeken” is verkeer dat gegenereerd werd door robots of worms, of respons met een speciale HTTP-statuscode.

## 2. Duur bezoeken

Aantal bezoeken: 2617 - Gemiddeld: 169 s		Aantal bezoeken	Procent
0s-30s		2196	83.9 %
30s-2mn		156	5.9 %
2mn-5mn		58	2.2 %
5mn-15mn		52	1.9 %
15mn-30mn		49	1.8 %
30mn-1h		88	3.3 %
1h+		14	0.5 %
Onbekend		4	0.1 %

## 3.4. Categorieën en segmenten

Websites van erfgoedinstellingen bevatten meer dan beschrijvingen of reproducties van gedigitaliseerd materiaal. Bijvoorbeeld pagina's met informatie over adres en openingstijden. Om een goed beeld te krijgen van het gebruik van gedigitaliseerd materiaal, zou men moeten afgaan op deelstatistieken van de betreffende onderdelen van de websites. Men spreekt wel van categorieën of 'content groupings'.

Een belangrijk voordeel van een page-tagging applicatie is de mogelijkheid om een of meerdere categorieën toe te voegen aan de page-tagging javascript code op de pagina. Dit maakt het mogelijk afzonderlijke statistieken te vervaardigen voor bijvoorbeeld het bezoek van alle pagina's in de publiekscatalogus, de pagina's van een specifiek genre, de pagina's die in december zijn toegevoegd, of welke groepering men ook maar wenst.

Aan de hand van categorieën kan men niet alleen pagina's, maar ook bezoekers indelen. Daarmee kan men het gedrag van een bepaald type bezoekers analyseren. Het op deze manier opdelen van de bezoekers wordt segmenteren genoemd.

Segmentatie van de bezoekers levert vaak een meer gedetailleerd inzicht op. Bezoekers kunnen op diverse manieren worden ingedeeld, bijvoorbeeld nieuwe versus eerdere bezoekers, intensieve versus incidentele bezoekers, bezoekers die via een zoekmachine dan wel via andere wegen op de site belanden, bezoekers uit Nederland dan wel andere landen. Vergelijking van de statistieken per subgroep kan aantonen of er verschillen zijn in de omgang met de site. Zo zou bijvoorbeeld kunnen blijken dat bezoekers die via een zoekmachine binnenkomen de site eerder verlaten dan andere bezoekers. Of dat terugkerende bezoekers andere pagina's of onderdelen van de site bekijken dan nieuwe bezoekers. In feite houdt een dergelijke exercitie in dat men een kruistabel maakt, met in de kolommen de



subgroepen en in de rijen de activiteiten. Dergelijke rapporten kunnen erg belangrijk zijn bij het optimaliseren van de website.

### **3.5. Audio en video**

Webstatistieken richten zich over het algemeen op het meten van een bezoek aan een pagina of het downloaden van een bestand. Het in kaart brengen van gebruikersinteractie met audio of video is vaak beperkt tot de constatering dat een video gedownload is. Voor sites die veel met audiovisueel materiaal werken is dit vaak onvoldoende. Van een video van een uur zou je bijvoorbeeld willen weten hoe lang een gebruiker naar de video heeft gekeken. Met standaard webstatistieken oplossingen is die informatie niet beschikbaar in het geval dat een gebruiker na een half uur kijken de site wegklikt. Er is geen tweede meetmoment en een (gemiddelde) bezoekduur is dus niet te berekenen.

Indien gebruikt wordt gemaakt van een streaming server voor uitspelen van het audiovisuele materiaal, kan logfile analyse een oplossing zijn. Niet alle loganalyse software kan echter omgaan met het afwijkende logfile formaat dat streaming servers wegschrijven.

Steeds meer page-tagging oplossingen bieden mogelijkheden om video interactie te meten via een aangepaste meettag. Nedstat heeft een product genaamd StreamSense ontwikkeld dat in de videoplayer wordt ingebouwd waardoor alle play, stop en pauze commando's van de videoplayer worden doorgestuurd naar de page-tagging server. Dit levert interessante rapporten op over bijvoorbeeld de gemiddelde kijkduur per video. Een ander rapport geeft informatie over het percentage bezoekers dat een bepaalde video uitkijkt. Ook Google Analytics werkt aan een oplossing om 'events' te meten in plaats van pagina's. Het is wel al mogelijk om de events door te geven aan de page-tagging servers van Google Analytics, maar de bijbehorende rapporten zijn voorlopig nog alleen zichtbaar voor geselecteerde testgebruikers.

## 4. Key performance indicators

De ruwe gegevens vormen de ingrediënten voor de zogenaamde Key Performance Indicators (KPI's). Welke indicatoren men hanteert hangt af van de doelstelling van de website. Zo zullen sites waarop producten te koop worden aangeboden ('business sites') andere indicatoren vergen dan sites die vooral informatie willen bieden ('content sites'). Erfgoedinstellingen vallen in de laatste categorie. Onder meer Peterson (2005) geeft een aantal suggesties voor indicatoren van dergelijke sites. Daarbij gaat het vooral om gemiddelden en percentages. Van belang zijn ook groeipercentages: de huidige scores vergeleken met die van de voorgaande rapportageperiode. Sommige pakketten ondersteunen ook benchmarking: zij vergelijken het gebruik van de onderzochte site met die van qua omvang of sector vergelijkbare sites.<sup>26</sup>

### *1. Gemiddeld aantal pageviews per bezoek*

Deze indicator geeft inzicht in de intensiteit van het bezoek. Op het eerste gezicht lijkt het aantrekkelijk om een hoog gemiddelde te scoren, maar er zijn enkele adders onder het gras. Een hoog gemiddelde kan erop duiden dat de bezoekers de site interessant genoeg vinden om meerdere pagina's te bekijken, maar ook dat de navigatie tekort schiet en bezoekers verplicht om veelvuldig te klikken alvorens bij de beoogde informatie te belanden. Een laag gemiddelde kan erop duiden dat de bezoekers weinig van hun gading vinden, maar ook dat zij direct (bijvoorbeeld via een zoekmachine) datgene vinden dat ze nodig hebben.

Technische beperkingen zijn dat mogelijk niet alle raadplegingen geregistreerd zijn: pagina's met een bepaalde antwoordcode, pagina's van bepaalde bestandstypen, pagina's die uit de cache van de browser van de bezoeker of diens provider worden gehaald, pagina's die niet voorzien zijn van een tag.

### *2. Percentage bezoeken waarin één dan wel meerdere pagina's geraadpleegd is*

Dit is een verbijzondering van de eerste maat. Gemiddelden geven geen inzicht in de spreiding. Vaak wordt daarom ook apart het aantal bezoeken waarin slechts één pagina wordt bekeken als indicator opgevoerd (bounce rate, weigeringspercentage). Men kan natuurlijk ook het tegengestelde benadrukken: het aantal bezoeken waarin meer dan één pagina (of welke limiet dan ook) wordt bekeken. Men spreekt ook wel van klikdiepte of 'depth of visit'.

### *3. Gemiddelde tijdsduur van het bezoek*

#### *4. Percentage bezoeken dat langer duurt dan .. seconden*

Deze maten geven eveneens inzicht in de intensiteit van het bezoek. Voor de interpretatie geldt hetzelfde als wat is opgemerkt bij de eerste indicator. Een hoge tijdsduur wordt over het algemeen als gunstig beschouwd, maar niet uit te sluiten is dat dit een gevolg is van de slechte vindbaarheid van de gewenste informatie. Bovendien is er geen enkele garantie dat de bezoeker de pagina daadwerkelijk zolang heeft bekeken als de tijdmeting aangeeft; wellicht verricht hij inmiddels andere activiteiten en heeft hij alleen maar het venster niet afgesloten. Dergelijke bezoeken tellen dan als minimaal dertig minuten. Hoewel de meeste pakketten bij de berekening van de gemiddelde tijdsduur uitgaan van het rekenkundig gemiddelde, zou de mediaan een betere maat zijn.

Een technische beperking is dat de tijd besteed aan het bekijken van de laatste pagina niet gemeten wordt. Er wordt dus helemaal geen tijdsduur berekend indien het bezoek slechts uit het bekijken van een enkele pagina bestaat. Er zijn echter pakketten die desondanks dergelijke

---

<sup>26</sup> <http://www.googleanalyticsresults.com/2008/03/benchmarking-reports-in-google.html>

bezoeken van ‘nul seconden’ meetellen bij de berekening van de gemiddelde tijdsduur. Dit resulteert in een lagere gerapporteerde gemiddelde tijdsduur dan in werkelijkheid het geval is.

Ook bij rapportages over de gemiddelde raadpleegtijd van een pagina (in plaats van bezoek) moet men zich goed afvragen hoe die berekend is. Peterson (2005) geeft het voorbeeld van een bezoek dat bestaat uit het bekijken van twee pagina's en dat vijf minuten duurt. In feite staat hier dat een bezoeker op pagina A binnenkomt, en na vijf minuten doorgaat naar pagina B. Hoelang hij daarop vertoeft is onbekend. De bezoeker verblijft dus vijf minuten op de pagina, en niet 2,5 minuten (de totale vastgestelde tijd gedeeld door het aantal bekeken pagina's). Zo ook verblijft iemand die in zes minuten tijd zeven pagina's van de site bezoekt niet gemiddeld  $6 / 7$  minuut = 51,5 seconden op een pagina, maar  $6 / 6$  minuut = 60 seconden.

#### *5. Gemiddeld aantal bezoeken per (unieke) bezoeker*

Een hoog gemiddelde is een teken van loyaliteit. Bij de interpretatie moet men scherp in het oog houden welke tijdsduur geldt voor de definitie van unieke bezoekers. Het maakt nogal wat verschil of bezoekers elke dag of elke maand ontdudd worden. Het laatste zal tot een veel hoger gemiddelde leiden dan het eerste.

Een technische beperking is dat het niet altijd goed mogelijk is bezoekers te identificeren: onder eenzelfde IP adres kunnen meerdere bezoekers schuilgaan (proxy server, pc's in publieksruimten), omgekeerd kan eenzelfde gebruiker meerdere IP adressen hebben, of de site vanuit verschillende pc's of met verschillende browsers bezoeken. Bezoekers kunnen ook persistente cookies blokkeren. In alle gevallen is het aantal unieke bezoekers niet betrouwbaar.

#### *6. Percentage bezoekers dat één dan wel meerdere bezoeken aflegt*

Dit is een verbijzondering van de vorige maat. Gemiddelden geven geen inzicht in de spreiding. Tellingen van het aantal repeat visitors tonen hoeveel bezoekers gedurende de rapportageperiode één, twee, drie of meer bezoeken afleggen. Ook nu moet men weer goed rekening houden met de definitie van unieke bezoeker. Kan eenzelfde persoon elke dag of elke maand een unieke bezoeker zijn?

#### *7. Percentage nieuwe of terugkerende bezoekers*

Het aantal terugkerende bezoekers is een teken van loyaliteit. Het aantal nieuwe bezoekers is een teken van potentie. Eigenlijk zijn beide typen bezoekers van belang. Het is lastig om aan te geven wat een ideale verhouding is. Alleen maar nieuwe of alleen maar terugkerende bezoekers is in elk geval niet wenselijk.

Een technische beperking is weer dat het niet altijd goed mogelijk is bezoekers te identificeren en daarmee als nieuw of terugkerend te beschouwen.

#### *8. Aantal pageviews*

#### *9. Aantal bezoeken*

#### *10. Aantal unieke bezoekers*

Gemiddelden verhullen de absolute aantallen. Een gemiddelde van 2 pageviews per bezoek kan gebaseerd zijn op 2000 pageviews en 1000 bezoeken, maar ook op 20.000 pageviews en 10.000 bezoeken. Het is daarom van belang ook oog te hebben voor absolute aantallen.

#### *11. Groeipercentages*

Alle maten lenen zich ook voor trendanalyse: hoe verhouden de huidige aantallen zich tot die van de vorige rapportageperiode, of tot dezelfde periode een jaar eerder? Het meest gebruikelijk zijn groeicijfers van het aantal bezoeken en unieke bezoekers.

*Samenvattend voorbeeld*

Performance indicator	Okt 2009	Sept 2009	% verschil tov sept 2009	Okt 2008	% verschil tov okt 2008
<p><i>Ruwe gegevens</i></p> <p>1. aantal pageviews 2. aantal bezoeken 3. aantal unieke bezoekers</p> <p><i>Bezoek</i></p> <p>4. Aantal pageviews per bezoek 5. Percentage bezoeken waarin meer dan .. pagina's bekeken zijn 6. Gemiddelde tijdsduur van bezoeken 7. Percentage bezoeken langer dan .. seconden</p> <p><i>Bezoekers</i></p> <p>8. Aantal bezoeken per bezoeker 9. Percentage bezoekers dat meer dan .. bezoeken aflegt (repeat) 10a. Percentage nieuwe bezoekers 10b. Percentage terugkerende bezoekers (return)</p>					

## 5. Vergelijkend onderzoek

### 5.1. Vergelijking van page tagging programma's

Verschillende pakketten kunnen, ook al behoren ze tot eenzelfde categorie, aanzienlijk verschillende statistieken opleveren. Deze problematiek is nader onderzocht door Stone Temple Consulting, die in 2007 het Web Analytics Shootout rapport publiceerde.<sup>27</sup> Ten behoeve van het onderzoek is het gebruik van vier websites volgens een vijftal programma's onderzocht: Clicktracks, Google Analytics, IndexTools, Unica Affinium NetInsight en Visual Sciences' HBX Analytics (voorheen WebSideStory). Deze vijf programma's zijn alle gebaseerd op page tagging. Per website vonden twee metingen plaats gedurende twee verschillende perioden. Gemeten werd het aantal pageviews, het aantal bezoeken en het aantal bezoekers.

Onderstaande tabel is afgeleid uit de resultaten van het onderzoek. De tabel is beperkt tot de drie websites en vier programma's waarvan vergelijkbare gegevens bestaan. Zo is Unica niet in de tabellen opgenomen omdat dit programma niet op alle websites geïnstalleerd is. Bovendien zijn de ruwe aantallen vertaald in relatieve scores. De laagste waarde is telkens gesteld op 100, de andere waarden zijn daaraan gerelateerd. Zo geeft WSS voor site 1 in periode 1 het laagste aantal pageviews (relatieve score 100). IndexTools rapporteert 20% meer pageviews (relatieve score 120,1).

Drie belangrijke conclusies zijn:

#### *1. De afwijkingen tussen de vier programma's kunnen oplopen tot 60%.*

In de tabel zijn de laagste en hoogste waarde van een zelfde site in een zelfde periode vet gedrukt, zowel voor pageviews, bezoeken als bezoekers. Op één uitzondering na (bezoekers bij site 3) registreert WSS de laagste waarden. Bij de hoogste waarden is meer variatie. De relatieve verschillen variëren nogal per indicator. Zo is het maximale verschil in het aantal pageviews 61%, het aantal bezoeken 60% en het aantal bezoekers 23%. Hierbij moet aangetekend worden dat alle vier programma's uitgaan van tags en cookies, en niet van logfile analyse. Verschillen tussen programma's van verschillende categorieën zullen nog aanzienlijk groter zijn.

#### *2. De afwijkingen tussen de verschillende programma's zijn stabiel.*

Voor elk van de drie sites geldt zowel voor het aantal pageviews, het aantal bezoeken en het aantal bezoekers dat de relatieve scores van de eerste en de tweede periode zeer goed met elkaar overeen komen. Dit maakt het mogelijk om de waarden volgens het ene programma te vertalen in waarden volgens het andere programma. Dit is van belang voor sites die overgaan op een ander programma. Indien zij gedurende een aantal maanden beide pakketten aanhouden, kunnen zij de verhouding tussen de tellingen van beide pakketten vaststellen en daarmee oude tellingen herberekenen.

#### *3. De afwijkingen tussen de vier programma's verschillen per site.*

Opvallend is dat het maximale verschil sterk varieert tussen de drie websites. Zo is het maximale verschil in het aantal pageviews volgens de vier programma's bij de derde site 6%, maar bij de tweede site 61%. Dit betekent dat het niet mogelijk is om op basis van de verhouding in de scores van verschillende programma's bij de ene site, uitspraken te doen over de verhouding bij andere sites. Een verklaring voor het hoge aantal pageviews volgens

---

<sup>27</sup> Deel 1: <http://www.stonetemple.com/articles/analytics-report-august-2007.shtml>

Deel 2: <http://www.stonetemple.com/articles/analytics-report-august-2007-part2.shtml>

Google Analytics bij site 2 is de locatie van het Javascript, namelijk in de html header en dus bovenaan de pagina. De registratie van een pageview gaat dan vooraf aan het laden van de betreffende pagina. Bij dezelfde site is het script van de andere drie programma's echter onderaan de pagina geplaatst. In die gevallen is het mogelijk dat de pagina getoond wordt en de gebruiker al weer op een link op de pagina heeft geklikt of de pagina heeft weggeklikt voordat het Javascript is uitgevoerd, zodat registratie achterwege blijft. Verder onderzoek naar de achtergronden van de onderlinge verschillen is nodig.

De onderzoekers wijzen erop dat in de praktijk de installatie van het pakket en alle benodigde randvoorwaarden de belangrijkste oorzaak is van de afwijkingen. In dit onderzoek speelde dat geen rol, omdat de installatie in nauw overleg met de beheerders van de softwarepakketten heeft plaatsgevonden.

Site 1	Pageviews		Bezoeken		Bezoekers	
	periode 1	periode 2	periode 1	periode 2	periode 1	periode 2
Clicktracks	114,5	116,1	<b>123,2</b>	<b>126,4</b>	115,0	118,8
Google Analytics	114,1	113,3	114,6	115,0	114,2	114,3
IndexTools	<b>120,7</b>	<b>123,4</b>	117,3	121,6	<b>116,5</b>	<b>120,5</b>
WebSideStory HBX Analytics	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>
Site 2	Pageviews		Bezoeken		Bezoekers	
	periode 1	periode 2	periode 1	periode 2	periode 1	periode 2
Clicktracks	109,8	110,7	125,2	127,8	100,0	100,7
Google Analytics	<b>161,2</b>	<b>159,6</b>	<b>154,2</b>	<b>160,5</b>	<b>112,4</b>	<b>113,1</b>
IndexTools	118,0	116,7	104,6	105,1	100,2	100,6
WebSideStory HBX Analytics	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>
Site 3	Pageviews		Bezoeken		Bezoekers	
	periode 1	periode 2	periode 1	periode 2	periode 1	periode 2
Clicktracks	<b>105,5</b>	<b>106,2</b>	<b>140,2</b>	<b>136,7</b>	119,6	122,7
Google Analytics	103,9	104,8	134,1	131,6	<b>120,0</b>	<b>123,3</b>
IndexTools	101,4	102,1	105,4	104,4	<b>100,0</b>	<b>100,0</b>
WebSideStory HBX Analytics	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	119,3	121,5

*Toelichting: aantallen in deze tabel zijn relatieve getallen. Per onderdeel is de laagste waarde op 100 gesteld. Deze zijn, evenals de hoogste relatieve waarden, vet gedrukt. De tabel bestaat uit achttien onderdelen. Elk onderdeel wordt gevormd door de site (1, 2 of 3), de indicator (pageviews, bezoeken, bezoekers) en de periode (1 of 2) en bestaat uit vier waarden. Zo bevat het onderdeel pageviews van site 1 gedurende periode 1 de waarden 114,5 / 114,1 / 120,7 / 100,0 voor de vier verschillende pakketten.*

TNO presenteerde in maart 2009 een vergelijkbaar onderzoek, getiteld: *Verschillen in Web Analytics: feiten, fabels en verwachtingen*.<sup>28</sup> Daarbij waren zes websites betrokken, die allemaal gebruik maken van Google Analytics en nog een of twee andere pakketten, waaronder vier maal SiteStat en een maal WebTrends. Nagegaan werd in hoeverre de aantallen pageviews, bezoeken en bezoekers van eenzelfde site varieerde per pakket. De bevindingen komen overeen met die van het Shootout rapport:

1. Verschillende pakketten genereren verschillende waarden voor eenzelfde website:
    - De aantallen liggen soms tientallen procenten uit elkaar
    - De verschillen zijn niet identiek voor visitors, visits en pageviews
    - Over de meetperiodes heen zijn de onderlinge verschillen vrij constant
    - Het verschil tussen de maximale en minimale score van twee pakketten is meestal (kans 1 op 2) niet groter dan 9,5%, en zelden (kans 1 op 10) groter dan 23,2%. Bij grotere verschillen is sprake van een verkeerde implementatie van het pakket. De corresponderende waarden bij vier pakketten (zoals in het Shootout rapport) zijn 22,5 en 34,7%
  2. Afwijkingen tussen de verschillende pakketten zijn website specifiek
- De conclusie is ook nu weer dat de pakketten geen absolute waarden opleveren, maar vooral inzicht geven in trends gedurende de tijd.

Het is niet uitgesloten dat de verhouding bij andere pakketten weer anders ligt. Verder onderzoek op dit terrein is dan ook van belang.

## **5.2. Verschillen tussen logging en tagging programma's**

Verschillen tussen op logfiles gebaseerde pakketten enerzijds en op page tagging gebaseerde pakketten anderzijds zullen vaak nog groter zijn. De site [www.imediaconnection.com/content/12584.asp](http://www.imediaconnection.com/content/12584.asp) vermeldt een voorbeeld van een verlies van 80 procent pageviews na overgang van Web Log Expert naar Google Analytics. De meest gemaakte vergelijking is wellicht die tussen de meest populaire vertegenwoordigers van beide categorieën, AWStats en Google Analytics. Het aantal pageviews volgens AWStats is niet zelden een veelvoud van dat volgens Google Analytics. Hiervoor kunnen verschillende redenen zijn:<sup>29</sup>

- Een (dynamische) pagina bestaat vaak uit een conglomeraat van bestanden. AWStats telt alle samenstellende onderdelen van zo'n pagina, terwijl Google Analytics alleen het eindresultaat, de voor de gebruiker zichtbare pagina telt.<sup>30</sup>
- Waarschijnlijk zijn pagina's bekeken door robots onvoldoende uitgefilterd. Er komen dagelijks robots bij en niet alle robots geven eerlijk hun user agent naam door, met als gevolg dat AWStats de robot niet als zodanig herkent.
- Niet uitgesloten is dat het bezoek aan sommige pagina's niet door Google Analytics wordt geregistreerd. Dit is het geval bij pagina's die om wat voor reden dan ook niet van een tag zijn voorzien of pagina's die zo snel weer worden weggeklikt dat er geen tijd is om het script uit te voeren.

AWStats rapporteert ook vaak een (veel) hoger aantal bezoeken dan Google Analytics. Een reden kan ook hier zijn het meetellen van bezoeken door robots. Robots keren regelmatig terug, het aantal bezoeken moet niet onderschat worden.

Een andere mogelijke reden is dat gebruikers met een Mozilla Firefox browser No-Script kunnen installeren.<sup>31</sup> Met deze plugin/add-on kan men gemakkelijk Javascript uitschakelen of

---

<sup>28</sup> <http://www.martijnstaal.nl/2009/04/07/online/verschillen-in-website-statistieken/>

<sup>29</sup> Met dank aan Jan van Oorschot (Koninklijke Bibliotheek) en Menno Rasch (Universiteitsbibliotheek Utrecht)

<sup>30</sup> <http://www.webvenue.org/google-analytics-vs-awstats-why-the-difference-in-stats/>

juist toestaan voor één of meerdere sites. Daarmee kan men voorkomen dat kwaadwillige sites schade aanrichten. NoScript zorgt daarmee voor een extra beveiliging tijdens het surfen met Firefox op het internet. Deze plugin heeft in 2006 de “PC World Class Award” ontvangen. Wanneer de gebruiker terecht komt op een site die hij niet veilig of betrouwbaar acht, kan hij ervoor kiezen Javascript tegen te houden. Google Analytics is daarmee, anders dan AWStats, niet in staat het bezoek (noch pageviews, noch bezoek, noch bezoeker) te registreren.

Er zijn echter ook andere ervaringen. Sinds kort meet de Koninklijke Bibliotheek het gebruik van haar website niet alleen met AWStats, maar ook met Google Analytics. De cijfers van juni 2009 laten zien dat het aantal pageviews volgens Google Analytics wel enigszins achterblijft bij dat van AWStats (- 17,1%), maar ook dat het aantal bezoeken (+ 1,4%) en bezoekers (+ 4,8%) bij Google Analytics juist hoger is. Verondersteld mag worden dat de manier waarop de website is opgebouwd en de filters die door de webmaster worden gehanteerd van zeer grote invloed zijn op de uiteindelijke resultaten.

Gezien de afwijkende achtergrond van beide programma's is het niet verwonderlijk dat de tellingen vaak sterk uiteenlopen. Op zich is een vergelijking dan ook niet erg zinvol. Voor instellingen die overgaan van AWStats op Google Analytics is het wel van belang een indruk te krijgen van de onderlinge verhouding. Aan de hand daarvan kan men, met enige slagen om de arm, de oude AWStats cijfers omzetten in Google Analytics cijfers. Dit maakt het mogelijk zonder onderbreking trends te meten..

### 5.3. Verklarend onderzoek

Het onderzoek van TNO was er niet slechts op gericht te verschillen te meten, maar ook om de oorzaken daarvan te achterhalen. De daarbij gehanteerde methode was een test in een gesloten omgeving. De onderzoekers legden de handelingen op een van de buitenwereld afgeschermdde testsite nauwkeurig vast en gingen na in hoeverre deze in Google Analytics, Site Stat en Webtrends gerapporteerd zijn. Interessante bevindingen zijn onder meer:

- WebTrends registreert als enige van de drie pakketten geen nieuwe pageview wanneer men via de backtoets op de (eerder bezochte) pagina terechtkwam.<sup>32</sup>
- WebTrends registreert eveneens als enige pakket geen nieuwe pageview wanneer men op de refresh knop (vernieuwen) gedrukt heeft
- Elk pakket telt automatische redirects (bijvoorbeeld na verplaatsing van de pagina op de website) als één pageview
- Geen enkel pakket rapporteert onvolledige downloads (men klikte alweer op de backtoets voordat de pagina geladen was)
- Geen enkel pakket rapporteert een url met een typefout in de querystring als pageview

Dergelijk onderzoek kan het inzicht in de achtergronden van de tellingen aanzienlijk vergroten. Verder onderzoek op dit terrein is daarom van groot belang.

---

<sup>31</sup> <http://www.freelancer-job.com/2009/02/01/google-analytics-problem-reporting-test-google-analytics-vs-awstats-comparison/>

<sup>32</sup> Schriftelijke navraag bij Bart Gijsen, een van de onderzoekers, leerde het volgende. (1) WebTrends biedt wel de mogelijkheid om door de code te wijzigen pagina's bezocht via de backtoets mee te tellen; (2) WebTrends registreert wel het bezoek aan een eerder bezochte pagina als deze tijdens dezelfde sessie opnieuw wordt opgevraagd, niet via de backtoets maar door het url opnieuw in te typen



## 6. Conclusies

1. Tellingen kunnen per programma verschillen, afhankelijk van eigenschappen van het programma (gehanteerde definities en meetmethoden) en externe factoren (proxy servers, dynamische IP adressen, caching, geblokkeerde of verwijderde cookies). Gezien de afhankelijkheid van externe factoren is geen enkel programma in staat 100% nauwkeurige statistieken te leveren.
2. Om de statistieken goed te kunnen interpreteren, moeten instellingen goed inzicht hebben in de achtergronden ervan. Bestudeer de documentatie of doe navraag bij de leverancier, met name ten aanzien van de volgende zaken:<sup>33</sup>
  - *Pageviews*. In hoeverre wordt het raadplegen van niet-html pagina's meegeteld? In het bijzonder bij logging gelden de volgende vragen. In hoeverre worden pageviews door robots uitgefilterd? In hoeverre worden pagina's met andere antwoordcodes dan 200 (bijvoorbeeld 206 of 304) meegeteld? Worden dynamische pagina's van elkaar onderscheiden?
  - *Sessies (bezoeken)*. Geldt de standaarddefinitie dat een bezoek beëindigd is na dertig minuten non-activiteit?
  - *Unieke bezoekers*. Welke tijdsperiode geldt hierbij? Kan iemand elke dag opnieuw een unieke bezoeker zijn, of slechts elke maand? En worden bezoekers herkend aan de hand van IP-adressen plus user agent of persistent cookies? Maakt men, in het laatste geval, een schatting van het aantal bezoekers dat de persistent cookies verwijderd heeft en dus niet herkenbaar is?
  - *Tijdsduur*. Worden bij de berekening van de gemiddelde tijdsduur ook bezoeken die bestaan uit het bekijken van slechts een enkele pagina, en dus met een gemeten tijdsduur van nul seconden, meegerekend? Gaat men af op het rekenkundige gemiddelde of de mediaan?
3. Bestudeer ook de statistieken zelf om een beter inzicht van hun betekenis te krijgen. Zo kan men bijvoorbeeld ontdekken dat Google Analytics niet het aantal nieuwe dan wel terugkerende gebruikers rapporteert, maar het aantal bezoeken van nieuwe dan wel terugkerende gebruikers.
4. Onderlinge vergelijking van statistieken van instellingen die verschillende pakketten hebben of parameters verschillend hebben ingesteld is niet goed mogelijk. Met enige slagen om de arm zou nog wel een vergelijking mogelijk zijn van statistieken die gebaseerd zijn op hetzelfde uitgangspunt, dan wel logfile analyse, dan wel page tagging. Ga na of de bij punt 2 genoemde definities vergelijkbaar zijn, en hanteer – afgaande op de bevindingen van het Shootout rapport en het TNO onderzoek – veiligheidshalve een marge van 20%. Dat wil zeggen dat aangenomen wordt dat bijvoorbeeld een aantal van 50.000 bezoeken in werkelijkheid waarschijnlijk ergens tussen 40.000 en 60.000 ligt
5. Gezien de verschillende uitgangspunten is het niet mogelijk statistieken gebaseerd op logfile analyse (bijvoorbeeld AWStats) te vergelijken met statistieken gebaseerd op page tagging (Google Analytics).

---

<sup>33</sup> Zie ook de nieuwe conceptversie van de WAA (22-9-2008), waarin per begrip het onderdeel 'Ask your vendor' is toegevoegd.

[http://www.webanalyticsassociation.org/attachments/committees/5/WAA\\_Web\\_Analytics\\_Definitions\\_2008092\\_2\\_For\\_Public\\_Comment.pdf](http://www.webanalyticsassociation.org/attachments/committees/5/WAA_Web_Analytics_Definitions_2008092_2_For_Public_Comment.pdf)

6. Wel mogelijk is een vergelijking van trends.
  - a. Intern. Hoe verhouden zich de huidige aantallen tot de aantallen van de vorige rapportageperiode of die van dezelfde periode een jaar voorheen? Hoe hoog is het groeipercentage?
  - b. Intern. Binnen de eigen instelling. Hoe verhoudt zich het aantal fysieke bezoeken tot het aantal bezoeken aan de website gedurende een reeks van jaren? Een indicator zou dus kunnen zijn het aantal webbezoeken gedeeld door de som van het aantal webbezoeken en fysieke bezoeken (ofwel: percentage webbezoeken). Het aantal bezoeken lijkt een meer geschikte maat dan het aantal (unieke) bezoekers. Het fysieke bezoek immers wordt eerder gemeten aan de hand van bezoeken dan bezoekers.
  - c. Extern. Een optie zou zijn de groei van het gebruik van de website van de eigen instelling te vergelijken met de groei bij andere instellingen.
  - d. Extern. Eveneens een optie zou zijn het aandeel van webbezoeken in het totale bezoek bij de eigen instelling te vergelijken met dat bij andere instellingen.Voorwaarde voor trendvergelijking is dat de betrokken instellingen gebruik blijven maken van hetzelfde pakket.
7. Instellingen die overgaan naar een ander pakket houden bij voorkeur het oude pakket nog een jaar aan om de verhouding te kunnen vaststellen. Aan de hand daarvan kan men de statistieken schatten die het nieuwe systeem gedurende de afgelopen jaren zou hebben opgeleverd. Dit maakt het mogelijk zonder onderbreking trends te meten.
8. Een mogelijkheid om tot vergelijkbare statistieken te komen is om, naast het eigen pakket, een gemeenschappelijk pakket te installeren. Pakketten gebaseerd op tagging verschaffen de meeste informatie en verdienen daarom de voorkeur. Binnen deze categorie ligt Google Analytics het meest voor de hand: het wordt al veel gebruikt, geeft uitgebreide rapportages en is gratis. Uit privacy overwegingen zou men echter kunnen besluiten om, zoals in Nordrhein-Westfalen, verzameling van webstatistieken door een externe instantie tegen te gaan.
9. Meer onderzoek is nodig om verschillen tussen statistieken aangeleverd door verschillende pakketten te meten. Vooral van belang is verder onderzoek dat ingaat op de oorzaken van de verschillen, bijvoorbeeld door tests te verrichten in een gesloten omgeving.

## Literatuur

Boogert, Erwin. *Scoren met Google Analytics*. Amsterdam: Entopic, 2008

Burby, Jason, Angie Brown and WAA Standards Committee. *Web Analytics Definitions: version 4.0* Washington: Web Analytics Association, 2007.

<http://www.webanalyticsassociation.org/attachments/committees/5/WAA-Standards-Analytics-Definitions-Volume-I-20070816.pdf> (Laatst geraadpleegd 23 februari 2009)

Freiburg, Bauke. *Fabulous logs deel 2: de webstatistieken onderbouwing*. Paper voor de opleiding Documentaire Informatiewetenschap, Universiteit van Amsterdam (4 januari 2009).

Kaushik, Avinash. *Web Analytics: an hour a day*. Indianapolis: Wiley, 2007.

Peterson, Eric. *Web site measurement hacks: tips & tools to help optimize your online business*. Beijing et al: O'Reilly, 2005

Sostre, Pedro en Jennifer LeClaire. *Webanalytics voor dummies*. Amsterdam, Pearson Addison Wesley, 2008

Web analytics. [http://en.wikipedia.org/wiki/Web\\_analytics](http://en.wikipedia.org/wiki/Web_analytics) (Laatst geraadpleegd 23 februari 2009)

## Bijlage 1. Voorbeeld statistieken aan de hand van AWStats

Enigszins gecondenseerde versie gebaseerd op AWStats demo van het gebruik van [www.destailleur.fr](http://www.destailleur.fr):

<http://www.nltechno.com/awstats/awstats.pl?month=02&year=2009&output=alldomains&conf=destailleur.fr&framename=index>

Samenvatting	
<b>1. Wanneer:</b>	1.1 Maandelijke historie 1.2 Dagen van maand 1.3 Weekdagen 1.4 Uren
<b>2. Wie:</b>	2.1 Domeinen / Landen 2.2 Hosts 2.3 Robots / Spiders bezoekers
<b>3. Navigatie:</b>	3.1 Duur bezoeken 3.2 Bestandstypen 3.3 Pagina's, Binnenkomst pagina's, Uitgang pagina's 3.4 Operating Systems 3.5 Browsers
<b>4. Verwijzing:</b>	4.1 Herkomst, Verwijzende zoekmachines, Verwijzende sites 4.2 Zoektrefzinnen, zoektrefwoorden
<b>5. Andere:</b>	5.1 Overigen (bijv add to favourites) 5.2 HTTP foutmelding codes 5.3 Niet gevonden pagina's

Samenvatting					
<b>Rapportageperiode</b>	Maand febr. 2009				
<b>Eerste bezoek</b>	01 febr. 2009 - 00.34				
<b>Laatste bezoek</b>	23 febr. 2009 - 06.38				
	<b>Unieke bezoekers</b>	<b>Aantal bezoeken</b>	<b>Pagina's</b>	<b>Hits</b>	<b>Bytes</b>
<b>Bekeken verkeer *</b>	<b>1374</b>	<b>2617</b> (1.9 bezoeken/bezoeker)	<b>5853</b> (2.23 Pagina's/bezoek)	<b>32659</b> (12.47 Hits/bezoek)	<b>84.63 MB</b> (33.11 KB/bezoek)
<b>Niet-bekeken verkeer *</b>			<b>29532</b>	<b>54691</b>	<b>135.68 MB</b>

\* „Niet bekeken“ is verkeer dat gegenereerd werd door robots of worms, of respons met een speciale HTTP-statuscode.

1.1 Maandelijke historie					
Maand	Unieke bezoekers	Aantal bezoeken	Pagina's	Hits	Bytes
Jan. 2009	1985	3057	9288	53435	148.31 MB
<b>Febr. 2009</b>	1374	2617	5853	32659	84.63 MB
Totaal	3359	5674	15141	86094	232.94 MB

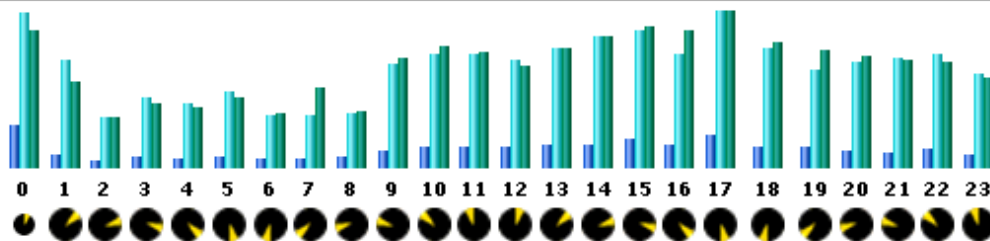
## 1.2 Dagen van maand

Dag	Aantal bezoeken	Pagina's	Hits	Bytes
01 febr. 2009	98	205	1246	3.32 MB
02 febr. 2009	92	204	1608	3.82 MB
03 febr. 2009	143	344	1387	4.89 MB
21 febr. 2009	70	157	1091	2.39 MB
22 febr. 2009	180	288	1140	3.75 MB
Gemiddeld	113.78	254.48	1419.96	3.68 MB
Totaal	2617	5853	32659	84.63 MB

## 1.3 Weekdagen











Dag	Pagina's	Hits	Bytes
ma	301.75	1542.50	3.85 MB
di	259.33	1600	4.48 MB
wo	238.33	1619.67	3.68 MB
do	332.67	1664	4.42 MB
vr	200	1136	3.25 MB
za	219	1116.67	3.13 MB
zo	224.50	1270	3.09 MB

## 1.4 Uren



Uren	Pagina's	Hits	Bytes	Uren	Pagina's	Hits	Bytes
00	585	2083	4.71 MB	12	287	1449	3.55 MB
01	195	1442	2.97 MB	13	306	1621	4.18 MB
02	97	665	1.71 MB	14	294	1780	4.55 MB
03	150	952	2.19 MB	15	382	1847	4.91 MB
04	107	868	2.11 MB	16	309	1528	4.68 MB
05	141	1033	2.42 MB	17	422	2199	5.62 MB
06	99	685	1.81 MB	18	281	1596	4.33 MB
07	113	701	2.76 MB	19	274	1324	4.00 MB
08	134	735	1.96 MB	20	239	1434	3.83 MB
09	241	1400	3.77 MB	21	209	1478	3.68 MB
10	274	1516	4.19 MB	22	253	1536	3.63 MB
11	275	1533	3.96 MB	23	186	1254	3.11 MB

## 2.1 Landen (Top 10) - Volledige lijst

	Landen		Pagina's	Hits	Bytes
	France	fr	1668	8607	18.56 MB
	Sweden	se	1082	1121	11.03 MB
	Onbekend	unknown	1004	4988	11.51 MB
	United States	us	854	12200	27.33 MB
	Morocco	ma	222	903	1.58 MB
	Germany	de	138	646	2.07 MB
	Switzerland	ch	137	306	827.42 KB
	Canada	ca	118	639	1.63 MB
	Great Britain	gb	80	371	834.85 KB
	Italy	it	74	509	2.75 MB
	Overige		476	2369	6.56 MB

## 2.2. Hosts (Top 10) - Volledige lijst - Laatste bezoek - Niet-herleide IP-adressen

Hosts : 0 Bekend, 1585 Onbekend (niet-herleide IP-adressen) 1374 Unieke bezoekers	GeoIP Country	Pagina's	Hits	Bytes	Laatste bezoek
83.198.34.213	France	298	610	808.91 KB	23 febr. 2009 - 00.48
62.2.98.122	Switzerland	79	79	0	22 febr. 2009 - 21.00
131.107.151.157	United States	75	472	1.10 MB	16 febr. 2009 - 21.27
92.133.223.120	Onbekend	60	469	146.72 KB	08 febr. 2009 - 01.51
83.233.30.101	Sweden	56	56	505.53 KB	22 febr. 2009 - 02.44
85.238.78.126	Hungary	44	84	221.00 KB	03 febr. 2009 - 15.34
192.188.57.30	Ecuador	41	71	48.88 KB	03 febr. 2009 - 00.50
190.18.49.128	Onbekend	40	211	248.03 KB	12 febr. 2009 - 17.49
194.204.217.78	Morocco	38	66	79.76 KB	03 febr. 2009 - 12.58
90.45.181.48	France	35	118	154.02 KB	16 febr. 2009 - 18.20
Overige		5087	30423	81.37 MB	

## 2.3 Robots/spiders (Top 10) - Volledige lijst - Laatste bezoek









48 Bezoekende robots*	Hits	Bytes	Laatste bezoek
<a href="#">MSNBot</a>	7352+814	21.61 MB	23 febr. 2009 - 06.29
<a href="#">Googlebot</a>	5000+35	12.01 MB	23 febr. 2009 - 03.06
<a href="#">Yahoo Slurp</a>	4586+209	14.20 MB	23 febr. 2009 - 06.35
Unknown robot (identified by 'crawl')	4777+15	50.56 MB	22 febr. 2009 - 23.59
<a href="#">MJ12bot</a>	1799+7	5.02 MB	04 febr. 2009 - 08.11
<a href="#">MSNBot-media</a>	436+1250	1.56 MB	23 febr. 2009 - 04.58
.....			

\* Robots geven „niet bekeken” door bezoekers, daarom worden ze niet in andere tabellen getoond  
Getallen achter '+' zijn geslaagde hits op „robots.txt” bestanden.

### 3.1 Duur bezoeken

Aantal bezoeken: 2617 - Gemiddeld: 169 s		Aantal bezoeken	Procent
0s-30s		2196	83.9 %
30s-2mn		156	5.9 %
2mn-5mn		58	2.2 %
5mn-15mn		52	1.9 %
15mn-30mn		49	1.8 %
30mn-1h		88	3.3 %
1h+		14	0.5 %
Onbekend		4	0.1 %

### 3.2 Bestandstypen - Webcompressie





Bestandstypen	Hits	Procent	Bytes	Procent
 css Cascading Style Sheet file	17476	53.5 %	32.50 MB	38.4 %
 png Image	5215	15.9 %	7.72 MB	9.1 %
 php Dynamic PHP Script file	4910	15 %	29.85 MB	35.2 %
 js JavaScript file	2573	7.8 %	4.35 MB	5.1 %
 gif Image	1136	3.4 %	385.14 KB	0.4 %
 htc	817	2.5 %	1.03 MB	1.2 %
 jpg Image	406	1.2 %	232.50 KB	0.2 %
 html HTML or XML static page	126	0.3 %	8.58 MB	10.1 %

**Webcompressie**








### 3.3 Pagina's/URL's (Top 10) - Volledige lijst - Binnenkomst - Vertrek

37 verschillende pagina's	Bezoeken	Gemiddelde grootte	Binnenkomst	Vertrek
<a href="#">/phpwebgallery/picture.php</a>	1951	5.55 KB	1244	1279
<a href="#">/phpwebgallery/</a>	1376	9.83 KB	457	442
<a href="#">/</a>	1282	3.82 KB	689	620
<a href="#">/phpwebgallery/template-common/tooltipfix.htc</a>	452	424 Bytes	28	28
<a href="#">/phpwebgallery/template-common/csshover.htc</a>	218	3.38 KB	25	27
<a href="#">/phpwebgallery/template-common/inputfix.htc</a>	147	881 Bytes	8	25
<a href="#">/perso_doc.php</a>	107	1.06 KB	12	42
<a href="#">/dolibarr/cvschangelogbuilder_dolibarr.html</a>	96	74.80 KB	84	85
<a href="#">/phpwebgallery/comments.php</a>	86	10.19 KB	26	16
<a href="#">/awstats/cvschangelogbuilder_awstats.html</a>	29	54.90 KB	17	23
Overige	109	3.14 KB	27	26

### 3.4. Besturingssystemen (Top 10) - Volledige lijst/Versies - Onbekend

Besturingssystemen		Hits	Procent
 Windows		30686	93.9 %
 Linux		935	2.8 %
 Macintosh		703	2.1 %
 Onbekend		238	0.7 %

### 3.5 Browsers (Top 10) - Volledige lijst/Versies - Onbekend

Browsers		Grabber	Hits	Procent
 MS Internet Explorer		Nee	24479	74.9 %
 Firefox		Nee	6509	19.9 %
 Opera		Nee	446	1.3 %
 Google Chrome		Nee	400	1.2 %
 Safari		Nee	335	1 %
 Mozilla		Nee	282	0.8 %
 Onbekend		?	113	0.3 %

### 4.1 Verbinding naar site vanaf

Herkomst	Pagina's	Procent	Hits	Procent
<b>Directe adressering / bladwijzers</b>	2305	57.9 %	2823	57.6 %
<b>Vanuit internetzoeksystemen - Volledige lijst</b>	1151	28.9 %	1152	23.5 %
- Windows Live	571			
- Google	524			
- MSN Search	29			
- Yahoo!	10			
- AOL (fr)	8			
- Unknown search engines	7			
- Google (Images)	2			
<b>Vanuit externe pagina's (andere sites, m.u.v. zoeksystemen) - Volledige lijst</b>	522	13.1 %	922	18.8 %
- <a href="http://www.nltechno.com/awstats/awstats.pl">http://www.nltechno.com/awstats/awstats.pl</a>	223			
- <a href="http://awstats.sourceforge.net">http://awstats.sourceforge.net</a>	87			
- <a href="http://www.itlocation.com/en/software/prd42223,..htm">http://www.itlocation.com/en/software/prd42223,..htm</a>	40			
- <a href="http://www.itlocation.com/en/software/prd40157,..htm">http://www.itlocation.com/en/software/prd40157,..htm</a>	39			
- <a href="http://www.dolibarr.org/component/option,com_fireboard/Itemi..">http://www.dolibarr.org/component/option,com_fireboard/Itemi..</a>	8			
- <a href="http://www.chedong.com/blog/">http://www.chedong.com/blog/</a>	7			
- <a href="http://localhost:81/dolibarr/about.php">http://localhost:81/dolibarr/about.php</a>	6			
- Overige	494			
<b>Herkomst onbekend</b>				



4.2 Gebruikte trefzinnen (Top 10) Volledige lijst		
262 verschillende trefzinnen	Zoeken	Procent
dged	97	8.5 %
gallerie	83	7.2 %
photos	70	6.1 %
disponible	60	5.2 %
vacances	34	2.9 %
photo volee	32	2.8 %
decembre	32	2.8 %
site web personnel	32	2.8 %
almirida	29	2.5 %
gwada	28	2.4 %
Andere zinnen	644	56.4 %

4.2 Gebruikte trefwoorden (Top 10) Volledige lijst		
224 trefwoorden	Zoeken	Procent
photos	126	6.3 %
photo	115	5.8 %
gallerie	97	4.8 %
dged	97	4.8 %
vacances	68	3.4 %
web	62	3.1 %
disponible	60	3 %
site	59	2.9 %
personnel	57	2.8 %
saintes	53	2.6 %
Andere woorden	1186	59.8 %

5.1 Diversen		
Diversen		
Aan favorieten/bladwijzers toegevoegd (relatieve indicatie)	411 / 1374 Bezoekers	29.9 %
Javascript uitgeschakeld	-	
Browsers met ondersteuning Java	-	
Browsers met ondersteuning Macromedia Director	-	
Browsers met ondersteuning Flash	-	
Browsers met ondersteuning Real audio	-	
Browsers met ondersteuning Quicktime audio	-	
Browsers met ondersteuning Windows Media audio	-	
Browsers met ondersteuning PDF	-	

5.2 HTTP-foutcodes				
HTTP-foutcodes*		Hits	Procent	Bytes
404	Document Not Found	23938	91.6 %	7.68 MB
302	Moved temporarily (redirect)	1844	7 %	35.62 KB
400	Bad Request	259	0.9 %	293.63 KB
301	Moved permanently (redirect)	37	0.1 %	7.84 KB
500	Internal server Error	29	0.1 %	560 Bytes
403	Forbidden	8	0 %	2.87 KB
401	Unauthorized	6	0 %	3.43 KB
206	Partial Content	2	0 %	100.56 KB

\* Deze codes geven „niet bekeken” door bezoekers, daarom worden ze niet in andere tabellen getoond