Editors’ Note
Changing Business Models of Online Content Services – Their Implications for Multimedia and Other Content Producers
Online Newspapers in the U.S. – Perceptions of Markets, Products, Revenue, and Competition
Evaluating the use of newspaper web sites logs
New Digital Media and Devices – An Analysis for the Media Industry
Exploring the link between culture and strategy in media organisations: the cases of the BBC and CNN

Calendar of Events
Impressum
Order Form

Beat F. Schmid, Peter Gomez, Peter Glotz, Dörte Wittig
Robert G. Picard
Turku School of Economics and Business Administration, Finland
Hsiang Iris Chyi
Chinese University of Hong Kong, George Sylvie
University of Texas at Austin, U.S.A.
David Nicholas
and Paul Huntington,
City University, London, U.K.
Joachim Rawolle and Thomas Hess
University of Göttingen, Germany
Lucy Küng
University of St. Gallen, Switzerland
Dear Reader

Welcome to a new issue of JMM – The International Journal on Media Management.

In this issue we have a distinguished group of authors, whose collected articles mainly cluster around the challenges arising from new digital media for the content industry. Through new services, technologies, and devices, innovative business models and products are required and the potentialities imbedded in the change have to be exploited.

Robert G. Picard leads off this issue with an analysis of business models of online content services, especially how they have changed through changes of technology and audience demand and how current business models resulted from these changes. In his article he explores the implications of these business models for multimedia and other content producers and possible prospects for the future.

In their article, Hsiang Iris Chyi and George Sylvie analyze the development of online newspapers and how they take advantage of the possibilities the online medium offers. Based on a survey with 14 online newspaper-practitioners they show that while most newspapers hope to complement print and online products, these possibilities are not part of every newspaper’s strategy. Most of them are testing several models of earning revenues to become an economically viable medium.

David Nicholas and Paul Huntington assess the use of newspaper website logfiles. The goal was to determine the most appropriate method for evaluating the use of these logfiles and to establish what types of analysis could be drawn. For this, the logfiles of The Times/The Sunday Times Web were examined. Based on these findings the study lays a foundation and identifies new classifications on which more detailed cross-classifications and modeling can be based.

Technology-driven innovations in the area of transport media and new devices pose a challenge for both media companies as well as their customers. The main questions for the media industry is how these new technologies can be exploited, e.g. through new content-oriented products or new services based on these technologies. Joachim Rawolle and Thomas Hess concern themselves with an analysis of attributes of digital contents and an assessment of different combinations of target devices and transport media. Based on this, they deliver a discussion of two emerging concepts.

The influence of corporate culture on the achievement of strategic aims in two leading international broadcasting companies – BBC and CNN – is investigated by Lucy Künz. The article aims to explore how cultural beliefs support the organizations strategic goal is assessed and to uncover senior managers’ unconscious assumptions concerning organization mission, the competitive environment and acceptable strategic responses. The author concludes that culture can act as a restraint to strategic plans and that culture in general is a valuable strategic asset for media organizations.

This issue again concludes with our calendar of events. We hope you find our collection of articles interesting. We look forward to continuing to deliver strong, peer reviewed content to you and to develop our relationships with the (new) media community. You are always welcome to contribute your research or your feedback to the JMM and to take the opportunity to share your ideas with this community. Since we are a journal focused on the possibilities of new media, you will find all our content online under www.mediajournal.org.

Beat F. Schmid
Peter Glotz
Peter Gomez
Dörte Wittig
Evaluating the use of newspaper web sites logs

David Nicholas and Paul Huntington
Internet Studies Research Group, Department of Information Science,
City University, London, U.K.

Introduction

Web sites proliferate, Web users increase by leaps and bounds and the Web becomes an ever more important medium for information, communication, advertising and, now, business. Not surprisingly Web site owners are becoming increasingly interested in obtaining information on who is using their sites, how they use their sites and how long they spend reading their pages. Possibly information ‘use’ data has never been so important – and certainly never had as much financial worth. It was only ever a tiny minority of people who were interested in the transactional logs of OPACs (online public access catalogues) and those of the traditional commercial online hosts, like FT Profile. People need Web site use information for a number of reasons: to assist marketing departments in their planning; to satisfy their sponsors and advertisers — and attract new ones; to demonstrate to their own organisations that the — often huge — investment is worthwhile; and to help them develop and improve their site. Fortunately, much that is done on the web is logged on an ongoing and routine manner. And to help in the analysis there are number of software packages on sale that provide them with additional data, and some others enlist the help of cookies in an attempt to make the data capture more precise.

Aims, objectives and rationale

“Treasure troves of valuable information” are how Gutzman (1999) describes web logs. However, journalists and information scientists have not been in the forefront of their use and evaluation. There have been very few papers published on the topic in the professional literature. This paper, and the research it reports on, attempts to fill this vacuum. To this end the logs from a Web site that had a large amount of traffic and a wide range of users was obtained. These were the logs of The Times/Sunday Times web sites. These two newspapers are British broadsheets of high reputation and published by News International. Our interest in these particular sites lay with the fact that they attracted a very large, heterogeneous and popular audience. Nearly one million people subscribed to The Times/Sunday Times sites at the time of the study.

In an attempt to obtain activity data of a range and quality on a par with that produced for commercial hosts and OPACs the authors have experimented with the web logs of two national newspapers – those of The Times and Sunday Times. These experiments were conducted to provide the owners of these sites with data that would enable them to understand their new customer base.

Summarising then the primary objectives of the study were to:

- establish the kinds of analysis that can be profitably undertaken on Web newspaper logs;
- determine the nature of use of one of the country’s most popular newspaper Web sites.

1 Based upon research conducted for News International, the owners of The Times/Sunday Times, during the period August 1998 – August 1999.
Background

At the time of the investigation The Times offered its online readers an almost exact replica of its daily hardcopy paper, so that each news item contained the same text in both versions, and all items in the newspaper were included on the Web site (See Table 1). In addition the site contained a very substantial reference service, plus an archive housing back copies since January 1996, and also a student supplement “4-D”. The Sunday Times site, contained an exact replica of the hardcopy paper as well, and it could be reached via The Times site, and vice versa. Once registered you could access both newspaper sites.

Literature Review

We have obtained most of what we know about Web logs from the less objective research of software vendors, or from the comparative and individual software product reviews produced by the technology press. Much of the interest and debate emanates from Web enthusiasts, however, this does not mean that it is not informed. Stout (1997) in his seminal book lays the foundation for the subject of Web site statistics. He defines the four different types of logs produced and describes the structure of the two most widely used formats, the common log format, and the extended log format. Stout pays particular attention to the analysis of Web logs and the variety of measures that maybe undertaken, explaining visits, tracking users and how to redesign Web sites using information gained from an analysis of the web log. Extensive comparisons are made between several software packages and analysis services. The main products reviewed were WebTrends, Market Focus, netAnalysis, and the log analysis services evaluated include NetCount and I/Pro.

Williams (1999), amongst others, points to the frustrations posed by logs: “it’s a marketers dream and worst nightmare to be able to watch your customers every move, but possessing only limited tools to influence them”. In fact much of the literature is concerned with the problems and pitfalls associated with Web log analysis. Gutzman (1999) in fact believes that the problems are so great that he regards “hit counters to be dead and deservedly so”. In his defence he points out that a third of all traffic is made up by Spiders. Zawitz (1998) makes the very important point that server logs and their measures were designed originally to measure and manage server traffic and not to analyse the use/effectiveness of Web sites. As a result often misquoted or misunderstood measures such as hits, are used to compare sites when in reality they are not measuring like with like. Fieber (1998), highlights another problem area – that of caching, and assesses the effect that browser caching has on the accuracy (under-reporting) of Web server log data, by comparing video taped user sessions with the data recorded in the log. The results showed that, depending on the length of the session, between 32 per cent and 55 per cent of transactions were cached and as a result were not recorded in the Web log. Fieber also points out that the default common log format only logs data the HTTP protocol provides, and that the “opaque nature of the client” and “stateless nature of the protocol” regarding caching leads to unreliable reports being produced by analysis software. Krishnamurthy (1998) of AT&T Research Laboratories investigated discrepancies in the results of log analysis software by comparing them with their own measurement of hits, etc. It was shown that errors were introduced to the log file either at the time of writing the file, or through errors such as embedded “new-line” characters in URL’s, or unrecognised control characters. When offending lines were corrected or removed one of the logs they processed was reduced by five per cent in size. He also attempted to resolve the fact that a single page can be represented by more than one URL.

Backman and Rubin (1997), evaluated a range of eight software products to assess their worth to the two main interested parties - web administrators, who require structural or content analysis and marketers/advertisers, who were looking into demographics and patterns of usage – including entry, exit and path statistics. In their comparisons of the software they expected to see some differences in the
metrics because of the different definitions and default settings for variables, but they did not expect to find differences in the numbers of hits. This research again called into doubt the absolute accuracy of the products. In fact Rahmel (1997) recommends that the best option is to build a custom parsing application, though he claims this to be the most expensive option. This was a recommendation taken up by the authors.

Rubin and Reimundez (1998) noted that, although changes are being made to software, improvements could still be made in error reporting and user demographics. Demographics in the packages do not relate to the user, but come from WhoIs, a service based on DNS and IP contact information registered with Internic, the domain main registry service. So that if nearly all your users seem to live in Vienna, VA, this is probably because AOL’s headquarters are registered there. Nevertheless they still claim that with “web servers’ current configurations and high end log analysis software, you can learn more about a virtual user than about a shopper physically present in a store.”

Stone (1999) sounds a more optimistic note, saying “while server logs can’t identify users by name or e-mail address they can provide information that will drive changes in the site content, marketing strategies and inventory.” Stone cites cases of companies big and small which having analysed their web server logs that have made changes to their sites or the marketing strategy. For example 3Com found that 40 per cent of visitors to their site were from overseas, 3Com then launched an international marketing campaign, something it had never done before. It included targeted mailings, buying international adverts aimed at visitors’ ‘areas of interest’ and the launching of native language Web sites.

**Methodology**

Four month’s worth of logs was obtained from News International and they covered the period January – April 1998. The logs were supplied on four CD-ROM disks. The enormous size of these logs can be gauged by the fact that The Times/Sunday Times web log generated up to 174 million lines a day. In addition personal details of the million or so subscribers were provided on a Dat tape. Data from this file were aggregated to retain anonymity.

Because of its sheer volume this data were sampled in a number of ways:

- For data obtained from the subscriber’s database the full sample was used, except in the case of age data, where a 25% sample was used.
- For log file use (as measured by page impressions and time online) and user (as determined by IP address) analysis was based on a random 1 in 10 lines of the processed log files for seven randomly selected days from April, May and June. This sample included six daily editions and one Sunday edition.
- In addition, a follow up analysis of users was made. 300 users were randomly selected from one day’s log file, and all the use related to these users were analysed. The 300 users generated 4,330 lines of use and an approximate download of 14 pages per user.

**Working definitions**

The definitions used by the project for the key metrics were:

- A user was a login from an external client to the server identified by a – not necessarily unique – IP number (address).
- Use was the activity of downloading files from a server and is measured by the number of pages viewed on screen.

The time to view a page was calculated by the lapse in time between one page and another. Clearly no view time can be calculated for the last page visited nor for those users that enter the site and leave having only viewed one page. The problem arises because nobody logs off from the Web.

Session continuity was determined by an active hyperlinking between pages on the site. A page session was assumed to be over if the user remained on a single page for 5 minutes or over, the next page downloaded then started the new page session.

**Results**

**Characteristics of subscribers**

The Times/Sunday Times subscriber’s database contained 966,346 entries in 1998. Double log-ins (i.e. people who have forgotten their password) reduce that figure a little. Also there were a number of bogus entries that were deleted – Micky Mouse and Tony Blair figured more than once. Plainly, too, not everybody who registered went on to use the service – although the probability must be high that they did, and not everyone who used the web sites were registered users – work colleagues or family members might have used the registered user’s password.

The average Times/Sunday Times subscriber turned out to be 38 years old (Table 2). A study of journalists conducted by the authors came up with a very similar figure (Nicholas et al, 1998). Men (38) turned out to be slightly older than women (35), the difference is significant at the 1% significance level: few signs of teenage surfers here.

---

2 At the time of our survey users of the Web sites had to register their personal details before they could use them.
Over three-quarters of subscribers were men, with women forming a much higher proportion (28%) of US subscribers (Table 3). According to their e-mail addresses — not a foolproof method of course, the vast majority (66%) of subscribers were from commercial organisations. Net providers — the majority of whom, one would suspect, were home users, accounted for 17% of registered users, and educational establishments 14% (Table 4). There were some geographical differences with commercial organisations figuring very highly in the EU (80%), academic establishments in the UK (18%) and Net providers in the US (24%).

As mentioned previously registry entitled users to access both The Times and Sunday Times Web sites. It was quite common for Times people to switch to The Sunday Times site, especially on a Sunday and Monday, and very common for Sunday Times users to switch to The Times site during the working week. Most of the analyses furnished below refer to The Times’ site, unless stated otherwise.

**Volume and pattern of traffic.**

The volume of traffic can be measured in terms of both the number of users (identified by an IP address) visiting the site and the amount of use these people made of the site as determined by pages downloaded. Very high volumes of use were recorded. Of course, the very fact that the sites attracted one million subscribers speaks legions. So too does the fact that The Times site attracted 75,000 visitors per day, who downloaded more than 165,000 pages a day. These are big numbers but, interestingly, still relatively small by hard copy standards. On the whole data tends to be very volatile. In more detail:

- **Monthly page impressions averaged just over 4.9 million pages. Monthly totals were quite variable, swinging up and down by as much as 10% (Table 5). Over the survey period totals fell by 5.3%. Figures for The Sunday Times were about 10% of those for The Times (Table 6).**

- **Day of the week appears to be a significant3 search characteristic, with some days being significantly more busy than others (Table 5, last column). Overall midweek was the busiest for the number of pages downloaded and Saturday the quietest (interestingly the busiest day for the hard-copy paper). However, this can change from month to month and is plainly connected with the importance of the news for that day.**

---

**Table 2: Subscribers by gender and age**

<table>
<thead>
<tr>
<th>N</th>
<th>Mean Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>937871</td>
</tr>
<tr>
<td>Women</td>
<td>213179</td>
</tr>
<tr>
<td>Men</td>
<td>724691</td>
</tr>
</tbody>
</table>

**Table 3: Subscribers by gender and country of origin**

<table>
<thead>
<tr>
<th>GENDER</th>
<th>US</th>
<th>UK</th>
<th>Other EU</th>
<th>Can, Aust, NZ</th>
<th>Main Asia</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>180253</td>
<td>327573</td>
<td>63820</td>
<td>68939</td>
<td>22474</td>
<td>80850</td>
</tr>
<tr>
<td>Female</td>
<td>69264</td>
<td>86038</td>
<td>17325</td>
<td>21036</td>
<td>6449</td>
<td>19095</td>
</tr>
<tr>
<td>Total</td>
<td>249517</td>
<td>413611</td>
<td>81145</td>
<td>89975</td>
<td>28923</td>
<td>99945</td>
</tr>
</tbody>
</table>

**Table 4: Subscribers by type of organisation and country of origin**

<table>
<thead>
<tr>
<th>Type of User</th>
<th>Education</th>
<th>Commercial</th>
<th>Govern.</th>
<th>Military</th>
<th>Net Provider</th>
<th>Non-Commercial</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>13.0%</td>
<td>59.6%</td>
<td>.8%</td>
<td>.4%</td>
<td>23.8%</td>
<td>2.3%</td>
<td>100%</td>
</tr>
<tr>
<td>UK</td>
<td>17.7%</td>
<td>69.7%</td>
<td>.9%</td>
<td>.2%</td>
<td>10.4%</td>
<td>1.1%</td>
<td>100%</td>
</tr>
<tr>
<td>Other EU</td>
<td>3.2%</td>
<td>80.0%</td>
<td>.1%</td>
<td>.4%</td>
<td>12.0%</td>
<td>3.5%</td>
<td>100%</td>
</tr>
<tr>
<td>Can, Aust, NZ</td>
<td>8.4%</td>
<td>63.3%</td>
<td>3.7%</td>
<td>.1%</td>
<td>22.8%</td>
<td>1.7%</td>
<td>100%</td>
</tr>
<tr>
<td>Main Asian</td>
<td>13.4%</td>
<td>62.9%</td>
<td>.7%</td>
<td>.0%</td>
<td>22.4%</td>
<td>.6%</td>
<td>100%</td>
</tr>
<tr>
<td>Other</td>
<td>7.9%</td>
<td>63.3%</td>
<td>1.2%</td>
<td>.1%</td>
<td>25.3%</td>
<td>1.8%</td>
<td>100%</td>
</tr>
<tr>
<td>Total</td>
<td>14.0%</td>
<td>65.7%</td>
<td>1.1%</td>
<td>.2%</td>
<td>17.2%</td>
<td>1.6%</td>
<td>100%</td>
</tr>
</tbody>
</table>

---

3 The hypothesis of no association is rejected at the 1% significance level.
For The Times Weekends were especially quiet with use down one third on that for mid-week.

For The Sunday Times, not surprisingly, Sunday use was heavy although Monday was in fact the site’s busiest day of the week and Saturday was easily the quietest day. Numbers of visitors varied enormously, too, from month to month and day to day, although the distributions sometimes differed from those of pages downloaded.

Use varied enormously during the day. Table 7 relates this to the country of the user. The first peak was reached at 3am GMT and this can be attributed to use from the Americans, mainland Asian, and Australian users. Asian block countries’ main use occurred at between 2am and 7am. Use then peaked again at 12am GMT. UK use was heaviest between 8am and 1pm GMT. Other European users appeared an hour earlier than those in the UK – at 7am – and then their use fell off after 12am GMT. There is a subsequent increase in US traffic from about 1 to 6pm GMT. It seems that, world over, most Internet traffic on this major UK newspaper, whatever the time zone, occurred in the mornings. First the day dawns in the Asian countries and California, then Europe and finally the Eastern States of America. We all think local time and news is only current in the mornings.

Half a percent of use was accounted for by users who visited the site and left without spending any time at the site at all. These “bouncers” who viewed only one page accounted for 5% of all users. The idea of grouping users by their pattern of use is believed to be particularly important and is an area that is being further researched. Arguably this may prove to be a qualitative metric on a site in so much as a site with a relatively high proportion of ‘bouncers’ would suggest that users are voting with their feet and not taking time to go beyond the one page.
Session – a session was said to have ended if the user had remained on the same page for over 5 minutes. On this basis each user or IP address roughly completed 2.35 sessions. The largest number of sessions conducted within the day by an IP address was 31 sessions (possibly, not all by the same user) and suggests a shared use of connected computer. About 23% of users completed 3 sessions or more and 41% of users completed 2 or more sessions. However, most users (59%) completed just 1 session. Users coming via net providers (possibly home users) were more likely to complete just the single session. 79% of these users completed 1 session, which compares to a figure of 46% of commercial users who completed no more than a single session. Repeat sessions in a day may also prove to be a qualitative metric in that returnees indicate more than just casual use and suggests that these users may have a good working knowledge of the site. Again this is an area for future research; in particular the idea of comparing sites by the percentage number of returnees.

Two IP numbers recorded sessions lasting longer than 15 hours. About 13% of users undertook sessions that lasted longer than one and a half-hours. These users did not necessarily spend all their time on the site – they might have visited other sites during their log-in time or simply left their machine on when they went to a meeting etc. The distribution of session time is skewed and a robust M-Estimator of the mean was calculated at approximately 15 minutes.

Page viewing. In relating hour of use to the average number of minutes that UK users took to view a page, we find this peaked at 1.27 minutes between 1 and 2 o’clock. Table 8 refers.

User categorisation. The huge differences between the number of pages downloaded by visitors to the site and the huge numbers of visitors involved made a form of user classification inevitable – although no standard/industry classification currently exists. A variety of classifications were tried during the project and there is much development work still to do. For this study users were either classified as light, moderate or heavy users. Light users downloaded less than about 10 pages in a day; moderate users between 11 pages and 150 pages, and heavy users over 150 pages in a day. On this basis light users accounted for the vast majority of users (88%) and a smaller, but still significant, proportion of pages downloaded (65%). Table 9 refers. Heavy users, although only accounting for 4.3% of users, accounted for 25% of use. Future research will examine how web sites can be compared by user categorisation. In particular this research may indicate certain types of users by cross classifying type of user by time of day and geographical location.

The three different types of user differed in the frequency of their visits to the site. Light users were least likely to make repeat visits during the week – 85% of light users only made a single visit during the working week, meaning that light users not only make little use of the site when they are there but also make few visits to the site. Heavy users often visited every day – 43% of heavy users made visits on all the week-days examined and 71% of heavy users made visits on 3 or more of the week-days examined. An earlier exploratory study indicated that heavy users were likely to be consistent repeat users over the long term, while both moderate and light users were less site loyal and, furthermore, their site use decayed fairly rapidly over a two week period and then subsequently increased. This research is ongoing and will be reported in a future paper.
Table 10: Geographical location of users according to various criteria

<table>
<thead>
<tr>
<th>Location</th>
<th>Subscribers</th>
<th>EMAIL</th>
<th>IP - The Times</th>
<th>IP Sunday Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>10</td>
<td>9</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Other EU</td>
<td>26</td>
<td>12</td>
<td>62</td>
<td>59</td>
</tr>
<tr>
<td>Can, Aust, NZ</td>
<td>26</td>
<td>41</td>
<td>62</td>
<td>59</td>
</tr>
<tr>
<td>Main Asia</td>
<td>12</td>
<td>6</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>UK</td>
<td>10</td>
<td>12</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>USA</td>
<td>10</td>
<td>5</td>
<td>45</td>
<td>43</td>
</tr>
</tbody>
</table>

Table 11: Type of user's organisation by various criteria

<table>
<thead>
<tr>
<th>Type of User</th>
<th>Subscribers</th>
<th>IP - The Times</th>
<th>IP Sunday Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic</td>
<td>14</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>Other</td>
<td>46</td>
<td>45</td>
<td>43</td>
</tr>
<tr>
<td>Non-profit or organisation</td>
<td>17</td>
<td>30</td>
<td>34</td>
</tr>
<tr>
<td>Net provider</td>
<td>66</td>
<td>45</td>
<td>43</td>
</tr>
<tr>
<td>Government</td>
<td>30</td>
<td>33</td>
<td>27</td>
</tr>
<tr>
<td>Commercial</td>
<td>33</td>
<td>30</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 12: Type of user's organisation by country

<table>
<thead>
<tr>
<th>Type of User</th>
<th>US</th>
<th>UK</th>
<th>Main Asia</th>
<th>Aus, Can, NZ</th>
<th>Other European</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net</td>
<td>38</td>
<td>55</td>
<td>33</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Commercial</td>
<td>47</td>
<td>40</td>
<td>34</td>
<td>45</td>
<td>46</td>
</tr>
<tr>
<td>Military</td>
<td>47</td>
<td>40</td>
<td>34</td>
<td>45</td>
<td>46</td>
</tr>
<tr>
<td>Government</td>
<td>47</td>
<td>40</td>
<td>34</td>
<td>45</td>
<td>46</td>
</tr>
</tbody>
</table>
Geographical origin of users/use

There are various ways of determining the geographical origin of users and use. We can use the postal address given by users when they registered, their e-mail address, and the IP address of users. These definitions of location are problematical; the IP address in particular is a crude measure as some users will be registered users of an ISP in another country for economic or other reasons. However, whatever metric is used The Times and Sunday Times web sites attracted a truly global audience. The comparison between the hard copy and Web versions of the papers could not be more marked. The former is read almost wholly exclusively in the UK, whereas in the case of the Web product the UK audience is a minority one. The key features of the geographical analysis are that:

- 57% of registered Times users were foreign; however, the proportion rose dramatically to 84% when the IP address of users was taken into account. The geographical distribution of users for The Sunday Times was very similar. The percentages change little if we calculate geographical distribution according to page impressions downloaded (Table 10).

- The United States appears to be the biggest source of users/use, accounting for around 60% of both users and use (Table 10). And this despite the fact that the US accounted for only 26% of the registered users – the majority (48%) of registered users gave their address as Britain. An analysis of the email addresses furnished by the registered users showed that a significant proportion (37%) of British subscribers had US mail addresses (the opposite hardly ever occurred). This is also likely to have been the case with IP addresses too, so the US’s share of use/users is likely to be inflated, and is more likely to be around 40%.

- Leaving the UK and US aside, users and use were distributed widely and evenly over the rest of the world. Canada and Australia, with their large ex-pat communities were the only individual countries to figure significantly (i.e. merit more than a few % points). Europe (outside the UK) accounted for just 7% of both Times’ users and use.

- The geographical distribution of users varied according to the day of the week. The UK presence was most marked on a Wednesday and Thursday (22% of users on those days were British). The US presence increased at the weekend: 64% of users at the weekend were registered in the US compared to 59% on weekdays. UK registered users were most in evidence during the weekdays: 16.8% compared to 12% of weekend users.

Type of user (organisation)

Whatever metric employed most Times’ users came from commercial organisations. On past experience it might have been expected that academic institutions would have figured most prominently, but plainly times are changing. Two-thirds of Times subscribers and 45% of IP addresses came from commercial organisations (Table 11). The difference between the two figures suggests that at least some businesses preferred to use a net provider rather than register as their own IP address. The picture for The Sunday Times was very similar.

Net providers were the next most important source of users. Net providers accounted for 17% of subscribers, and between a quarter and a third of users and use. Most home users would fall into this category. This view is probably borne out by the fact that The Sunday Times, which saw its’ use peak at a weekend, had a higher proportion of Net provider users. Net users made up 40% of users at the weekend and 27% on weekends. Academic institutions

<table>
<thead>
<tr>
<th>% within COUNTRY</th>
<th>News</th>
<th>Sport</th>
<th>Sunday Times</th>
<th>Foreign News</th>
<th>Business</th>
<th>Opinion</th>
<th>Feat</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>15.4%</td>
<td>9.1%</td>
<td>10.6%</td>
<td>9.4%</td>
<td>3.1%</td>
<td>2.6%</td>
<td>2.5%</td>
<td>47.3%</td>
</tr>
<tr>
<td>UK</td>
<td>15.0%</td>
<td>9.8%</td>
<td>6.1%</td>
<td>5.8%</td>
<td>4.3%</td>
<td>2.2%</td>
<td>2.8%</td>
<td>54.1%</td>
</tr>
<tr>
<td>Main Asian</td>
<td>13.7%</td>
<td>15.5%</td>
<td>9.8%</td>
<td>8.5%</td>
<td>2.7%</td>
<td>2.4%</td>
<td>2.0%</td>
<td>45.3%</td>
</tr>
<tr>
<td>Aus, Can, NZ</td>
<td>16.9%</td>
<td>11.6%</td>
<td>11.3%</td>
<td>7.1%</td>
<td>1.9%</td>
<td>3.1%</td>
<td>2.2%</td>
<td>45.8%</td>
</tr>
<tr>
<td>Other European</td>
<td>15.0%</td>
<td>13.1%</td>
<td>10.8%</td>
<td>8.0%</td>
<td>2.9%</td>
<td>2.4%</td>
<td>2.2%</td>
<td>45.4%</td>
</tr>
<tr>
<td>Other</td>
<td>13.0%</td>
<td>11.7%</td>
<td>9.9%</td>
<td>10.1%</td>
<td>2.9%</td>
<td>1.8%</td>
<td>1.8%</td>
<td>48.9%</td>
</tr>
<tr>
<td>Total</td>
<td>15.3%</td>
<td>10.1%</td>
<td>9.8%</td>
<td>8.5%</td>
<td>3.2%</td>
<td>2.5%</td>
<td>2.5%</td>
<td>48.2%</td>
</tr>
</tbody>
</table>
were the third largest source of users, accounting for 14% of subscribers and about a fifth of users/use. Their relatively lowly place could be due to traditional dominance of The Guardian in this market or, maybe, the result of The Telegraph’s active targeting of this group?

The organisational origin of the user varies significantly according to the geographical location of the user (Table 12). Thus the UK (42%) and Japan (43%) had a higher proportion of academic institutions among their users; and the US had a higher proportion of businesses amongst its (68%).

**Pages/subjects used**

Perhaps the most interesting characteristic of use from an information science perspective is the subjects that users seek. The Times site did not have a search engine at the time, so we have to determine subject interest from the names of pages the user downloads. Because of the large number of pages involved – The Times site contained approximately 540 at the time of the survey, related pages were grouped. Thus all sports pages were counted together. The metrics employed for this analysis were IP addresses for establishing the number of users of a page, pages downloaded as an indication of use and average time spent reading a page as an indicator of the depth of use. The key features of this analysis were that:

- Use is not evenly distributed. Regarding the number of pages downloaded just seven subjects on The Times web site accounted for half (52%) of all content pages downloaded. Not surprisingly, News pages were the most popular (accounting for 15.3% of all pages viewed). News was followed by sports pages (10.1% of pages) and then came accessing The Sunday Times from The Times site (9.8%). Generally features pages and opinions were very lightly used and each accounted for just 2.5% of pages used. Future research will look at grouping users by the type of pages viewed and cross classifying users by grouping, time and country. A country analysis of individual page use shows that the Australians, as befits their stereotype, are especially interested in the sports pages and the Canadians and British particularly interested in the general news pages (Table 13). American users were least interested in business pages but were more likely to spend time reading the foreign news pages and the opinions page.
- On average a Times page took 1.05 minutes to read. The fact that people print or download pages for use later would suggest that these figures might underestimate the amount of time people spend actually reading the information.
- A more detailed analysis of time spent on a page shows that users spent more time reading some pages than others. Rolling news is a case in point – read for 2.10 minutes on average in The Times (Table 14). Suggesting, perhaps, that currency is high on the user’s agenda of what they want from a web site. In comparison, the business pages are consumed fairly rapidly – 0.90 minutes on average. Other Europeans tend to take more time to read some pages – this is the case with features, opinion and all sport. UK users take longer to read the Rolling News.
- Page use varies quite considerably according to the day of the week (table 15). Thus Friday sees the greatest use of the news pages (21.1%), Tuesday sees the highest use of sports pages (12.8%), and on Thursday the use of foreign news pages peaks (12.7%).

**Conclusions**

The use distributions that have been described above are quite unlike any that we have witnessed during many years of evaluating online use – admittedly in a pre-Web world. This can be variously attributed to the method of recording the data, the global nature of the audience, the imprecision of the measurement of use/user, the novel na-

---

**Table 14: subject of pages viewed by country and time spent on a page (Minutes)**

<table>
<thead>
<tr>
<th>All Countries</th>
<th>US</th>
<th>UK</th>
<th>Other European</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>1.0618</td>
<td>.9625</td>
<td>1.0208</td>
</tr>
<tr>
<td>All Sport</td>
<td>1.4937</td>
<td>1.4970</td>
<td>.9028</td>
</tr>
<tr>
<td>Foreign news</td>
<td>1.1135</td>
<td>1.2351</td>
<td>.9192</td>
</tr>
<tr>
<td>Business</td>
<td>.8975</td>
<td>.8330</td>
<td>.7063</td>
</tr>
<tr>
<td>Features</td>
<td>1.4129</td>
<td>1.5902</td>
<td>1.0246</td>
</tr>
<tr>
<td>Opinion</td>
<td>1.0054</td>
<td>.9658</td>
<td>.5448</td>
</tr>
<tr>
<td>Puzzle</td>
<td>.7674</td>
<td>1.0144</td>
<td>.4333</td>
</tr>
<tr>
<td>Rolling News</td>
<td>2.1035</td>
<td>1.7433</td>
<td>3.7540</td>
</tr>
<tr>
<td>Other</td>
<td>.7049</td>
<td>.8239</td>
<td>.5921</td>
</tr>
</tbody>
</table>

* The hypothesis of no association is rejected at the 1% significance level.
ture of web searching and the vast numbers of new recruits that have been drawn to the web’s ranks. What really stood out though were the huge volume of use (due no doubt to people surfing about), its volatility, and the international and twenty-four hour nature of use.

It has to be admitted that the data is hardly problem free, and is rather less robust than that obtained from commercial online hosts – although that data also has its flaws. But where web log data has considerable attraction is in its sheer volume, availability, and internationality. Also, probably for the very first time, journalists and information professionals can obtain an insight into how the general public interacts with a genuinely popular information system.

The web logs and subscriber databases provide us with data in breathtaking volumes and on an endless basis, but, not as they stand, a complete picture of the information seeking behaviour of web users. Web logs enable us to follow the progress of packs of users rather than individuals and describe the broad outlines of their information fingerprints. It is a large but fuzzy picture. To get a much clearer picture it is necessary to find out why people search in the ways described by the logs, how satisfied are they with what they found, and how accurate a portrayal are the maps drawn by the logs. This is what the authors will be doing in a study of health logs, funded by the Department of Health.

The web is not short of users but is short on metrics in particular metrics, that say something about the quality of a site. This investigative study has identified two possible new metrics: a classification of users by volume of use and by their repeat use of a site. Clearly this is an area for future research. However both metrics appear to give an indication of user satisfaction and lend themselves both to a between site and an over time comparison. This study has laid the ground and identified new classifications on which more detailed cross-classifications and modelling can be based.

Ironically, the real worry about the logs is that their ready availability and hypnotic numbers will be used wholesale to fill in gaps in our knowledge of the hard copy environment. And this will be done without any recognition that this is as dangerous as making assumptions about the digital environment on the back of data pertaining exclusively to the hard copy world.

**Recommendations**

Currently most newspapers in the UK rely too heavily (and naively) on proprietary software to give them a picture of the users of their site. Generally this software is insufficiently understood by the people that operate it (particularly in regard to the filters that are being adopted), really only provides very broad-brush pictures of use and invariably produces overwhelming quantities of data that are rapidly archived. As this article has illustrated it is possible to produce data of a kind that could prove invaluable in understanding and tracking newspapers’ cyber readers. But we need to go much further than the base data and start producing bespoke analyses.

---

**Table 15: Subject of pages viewed by day of the week (Row %)**

<table>
<thead>
<tr>
<th>% within Day</th>
<th>News</th>
<th>Sport</th>
<th>Sunday Times</th>
<th>Foreign News</th>
<th>Business</th>
<th>Opinion</th>
<th>Features</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mon</td>
<td>14.1%</td>
<td>12.0%</td>
<td>20.6%</td>
<td>7.6%</td>
<td>2.5%</td>
<td>2.9%</td>
<td>1.3%</td>
<td>39.1%</td>
</tr>
<tr>
<td>Tues</td>
<td>14.9%</td>
<td>12.8%</td>
<td>4.8%</td>
<td>11.1%</td>
<td>3.4%</td>
<td>2.5%</td>
<td>2.8%</td>
<td>47.6%</td>
</tr>
<tr>
<td>Wed</td>
<td>17.2%</td>
<td>11.5%</td>
<td>3.8%</td>
<td>6.3%</td>
<td>4.2%</td>
<td>2.3%</td>
<td>3.6%</td>
<td>51.2%</td>
</tr>
<tr>
<td>Thurs</td>
<td>16.6%</td>
<td>8.7%</td>
<td>3.3%</td>
<td>12.7%</td>
<td>4.6%</td>
<td>2.7%</td>
<td>3.0%</td>
<td>48.3%</td>
</tr>
<tr>
<td>Fri</td>
<td>21.1%</td>
<td>9.9%</td>
<td>1.4%</td>
<td>11.0%</td>
<td>3.5%</td>
<td>2.8%</td>
<td>4.5%</td>
<td>45.7%</td>
</tr>
<tr>
<td>Sat</td>
<td>16.1%</td>
<td>12.3%</td>
<td>1.0%</td>
<td>6.0%</td>
<td>4.0%</td>
<td>2.7%</td>
<td>.4%</td>
<td>57.4%</td>
</tr>
<tr>
<td>Sun</td>
<td>4.8%</td>
<td>3.3%</td>
<td>37.6%</td>
<td>2.3%</td>
<td>1.4%</td>
<td>1.1%</td>
<td>.4%</td>
<td>49.2%</td>
</tr>
<tr>
<td>Total</td>
<td>15.4%</td>
<td>10.3%</td>
<td>9.5%</td>
<td>8.6%</td>
<td>3.5%</td>
<td>2.5%</td>
<td>2.5%</td>
<td>47.7%</td>
</tr>
</tbody>
</table>

---

*The hypothesis of no association is rejected at the 1% significance level.*

*The Web, the kiosk, digital TV and the changing and evolving face of consumer health information provision: a national impact study. 2000-2001.* [Department of Health]
References

http://www.nwc.com/811/811cn2.html (visited 8/7/99)


http://www.ecommerce.internet.com/solution/techno_advisor/article/0,1281,9561-186011,00.html (visited 17/10/99)

http://www.w3.org/1998/11/05/WC-workshop/Papers/bala2.html (18/9/99)


Randall N. (1998) “Web Site Analysis Tools: The Results are In”

http://www.nwc.com/917/917r1.html (visited 20/11/99)


http://www.net2b.com/cgi-bin/print_articles.pl/article/940 (visited 17/10/99)


Williams D. (1999) “Data Plays a Key Role in Net Success”
www.adage.com/interactive/articles/19990802/article3.html (visited 17/10/99)

http://www.ojp.usdoj.gov/bjs/pub/ascii/wsmua.txt (visited 9/10/99)

The Author

David Nicholas (nicky@sot.city.ac.uk) is Head of the Department of Information Science, City University and Director Internet Studies Research Group (ISRG).

Paul Huntington is a research fellow and statistician working for the ISRG. The ISRG has been actively working with a number of British newspapers in obtaining data from their web logs, including The Times, Sunday Times and The Independent.
## Calendar of Events

### September
- **09/06/2000 – 09/08/2000**
  - **LUMIS 2000:**
    - Second International Workshop on Logical and Uncertainty Models for Information Systems
    - Greenwich, London, UK
    - [http://www.dcs.qmw.ac.uk/~mounia/LUMIS.html](http://www.dcs.qmw.ac.uk/~mounia/LUMIS.html)

- **09/12/2000 – 09/14/2000**
  - **USM 2000 3rd IFIP/GI Int. Conf. on Trends towards an Universal Service Market**
    - Munich, Germany
    - [http://usm2000.informatik.uni-muenchen.de/](http://usm2000.informatik.uni-muenchen.de/)

- **09/11/2000 – 09/15/2000**
  - **SAB2000:**
    - The Sixth International Conference on the Simulation of Adaptive Behaviour
    - Paris, France

  - **Web Site Content Management Summit**
    - San Francisco, U.S.A
    - [http://www.iqpc.com/cgi-bin/templates/0/index.html](http://www.iqpc.com/cgi-bin/templates/0/index.html)

### October
- **10/15/2000 – 10/18/2000**
  - **Strategic Management Society Annual Conference 2000**
    - Vancouver, Canada
    - [http://www.smsweb.org/Pages/Frames/00/00main.html](http://www.smsweb.org/Pages/Frames/00/00main.html)

- **10/19/2000 – 10/22/2000**
  - **7th Annual – Human Resources Management and Organizational Management/Behaviour (HRMOB) Conference**
    - Charlotte, North Carolina, U.S.A.

  - **Measuring & Analyzing Online Customer Behaviour**
    - Chicago, U.S.A.
    - [http://www.iqpc.com/cgi-bin/templates/0/index.html](http://www.iqpc.com/cgi-bin/templates/0/index.html)

  - **Online Intellectual Property**
    - Washington D.C., U.S.A.
    - [http://www.iqpc.com/cgi-bin/templates/0/index.html](http://www.iqpc.com/cgi-bin/templates/0/index.html)

- **10/30/2000 – 11/04/2000**
  - **WebNet 2000**
    - San Antonio, Texas, USA

### November
  - **ACM Conference on Universal Usability: Solutions, Systems, and Methods**
    - Washington D.C., USA
    - [http://www.acm.org/sigchi/uu/](http://www.acm.org/sigchi/uu/)

### December
- **12/04/2000 – 12/05/2000**
  - **2nd IFIP / MASSYVE Working Conference on Infrastructures for Virtual Enterprises**
    - Managing Cooperation in Virtual Organizations and Electronic Business
    - Florianopolis, SC, Brazil

### January
- **01/03/2001 – 01/06/2001**
  - **Hawaii International Conference on System Sciences 2001**
    - Maui, Hawaii, U.S.A.
    - [http://www.hicss.hawaii.edu/](http://www.hicss.hawaii.edu/)
Title
JMM – The International Journal on Media Management
Edition: Vol. 2 – No. 2 – 2000
ISSN (printed edition): 1424-1277
ISSN (electronic edition): 1424-1250

Editorial Office
mcm – Institute for Media and Communications Management,
University of St. Gallen/Switzerland
Müller-Friedberg.Str. 8
CH – 9000 St. Gallen
Phone: +41-71-224 34 28
Fax: +41-71-224 27 71
Email: media.editors@netacademy.org
URL: http://www.mediajournal.org

Published by
mcm – Institute for Media and Communications Management,
University of St. Gallen/Switzerland
Müller-Friedberg.Str. 8
CH – 9000 St. Gallen
Phone: +41-71-224 34 28
Fax: +41-71-224 27 71
Email: media.editors@netacademy.org
URL: http://www.mediajournal.org

Editorial Team
BEAT F. SCHMID – Editor-in-Chief
PETER GLOTZ – Co-Editor
PETER GOMEZ – Co-Editor
DÖRTE WITTIG – Executive Editor

Layout
BELAU, Duisburg, Germany, info@b-e-l-a-u.de

Printer
Niedermann Druck, St. Gallen, Switzerland

Remarks
This publication is part of the NetAcademy on Media Management at the mcm – Institute for Media and Communications Management, University of St. Gallen/Switzerland. Articles from the contributors do not necessarily reflect the opinion of the editors.

Copyright
© mcm – Institute for Media and Communications Management
Partial or total reproduction will be permitted upon application, subject to acknowledging the source and sending two copies to the editorial office. The editors wish to thank Bertelsmann Foundation, Germany and Heinz Nixdorf Foundation, Germany, for their support with the publication of this journal.

The journal will observe the following issues closely
- Strategic, managerial and organizational aspects of the media sector and the media industry
- Economics of traditional and new media
- Evolution of the media industry and media industry segments
- Technology, infrastructure, user behavior related to the changes in the media sector
- Effects of new media on economy, society, politics, law and culture

Our goal is the close analysis of new industry structures, organizational forms, and critical competencies developing as a result of reconfigurations in the media value chain. We want to bring together academics and industry figures to explore the transition from “classic” to “new” media and identify the factors, which will determine organizational success and economic efforts in a fast changing and converging environment.

We are interested in receiving your submission to the JMM. Articles, which deal with one of the above issues, are more than welcome.

Orderform
Annual Subscription (4 Editions)
- Institutional: US$ 300*
- Individual: US$ 64*
*(plus tax and postage)

Single unit price:
- Regular: US$ 22*
*(plus tax and postage)

☐ I wish to subscribe to JMM
☐ This is an institutional subscription
☐ This is an individual subscription

Name
Company
Department
Street
City
ZIP Code
Country
Phone
Fax
Primary Activity
Comments

Please return to
mcm – Institute for Media and Communications Management,
University of St. Gallen/Switzerland
Müller-Friedberg.Str. 8
CH – 9000 St. Gallen
Phone: +41-71-224 34 28
Fax: +41-71-224 27 71
Email: media.editors@netacademy.org
URL: http://www.mediajournal.org