

An Exploration of the Potential for studying the usage of Investor Relations Information through the analysis of Web Server Logs

Rowbottom, Nicholas; Allam, Amir; Lymer, Andrew

DOI:

[10.1016/j.accinf.2004.08.002](https://doi.org/10.1016/j.accinf.2004.08.002)

License:

None: All rights reserved

Document Version

Peer reviewed version

Citation for published version (Harvard):

Rowbottom, N, Allam, A & Lymer, A 2005, 'An Exploration of the Potential for studying the usage of Investor Relations Information through the analysis of Web Server Logs', *International Journal of Accounting Information Systems*, vol. 6, no. 1, pp. 31-53. <https://doi.org/10.1016/j.accinf.2004.08.002>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

NOTICE: this is the author's version of a work that was accepted for publication in International Journal of Accounting Information Systems. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in International Journal of Accounting Information Systems, [VOL 6, ISSUE 1, (2005)] DOI: <http://dx.doi.org/10.1016/j.accinf.2004.08.002>

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

AN EXPLORATION OF THE POTENTIAL FOR STUDYING THE USAGE OF INVESTOR RELATIONS INFORMATION THROUGH THE ANALYSIS OF WEB SERVER LOGS

N. Rowbottom (University of Birmingham)*
A. Allam (University of Birmingham)
A. Lymer (University of Birmingham)

* Department of Accounting & Finance
University of Birmingham
Birmingham B15 2TT
UK
N.Rowbottom@bham.ac.uk

International Journal of Accounting Information Systems (2005) 6 pp31-53.

doi:10.1016/j.accinf.2004.08.002

Abstract

This paper introduces the Web server log file and assesses its potential as a research instrument in measuring and interpreting the use of corporate reporting information. Measuring Investor Relations output, including annual financial reports but covering a wider range of corporate reporting and market informing activity, has proven a difficult task in the past due to a lack of truly effective research methodologies to access such activity. This paper highlights the growth in the provision of online Investor Relations information and details how online information can be measured using activity logs taken as a Web server fulfils user requests for information over the Internet. The paper analyses the limitations of this methodology for measuring the use of Investor Relations output and illustrates its possible application by drawing on data from a UK FTSE 100 company. Finally, the paper concludes that this methodology has significant potential in measuring the use of online Investor Relations information, and can therefore make valuable contributions to corporate reporting research and policy making in this area.

Keywords: Investor Relations, Financial reporting, Online reporting, Internet reporting

1. Introduction

This paper assesses the feasibility of extracting valuable information from an analysis of Web log files to gain insights into understanding the use of Investor Relations (IR) information. The discipline of IR has grown significantly in the last 20 years, particularly in economies with strong equity markets (Marston & Empson, 2003). Consequently, the type and quantity of IR information disclosed has increased rapidly. This growth has been greatly aided by the development and widespread availability of the Internet as a communications channel for information in an electronic format. Despite this growth of IR information availability, there is little direct published evidence of what information is actually used. Although several authors have considered the usefulness of online IR against that distributed in more traditional formats (e.g. Ashbaugh et al, 1999, Jones et al, 2001, Dull et al, 2003), no study provides data on what is actually accessed by users. Previously, it has proven difficult to observe the use IR information by the various constituents who have access to this information. Discovering what corporate reporting information is demanded and/or used will be of keen interest to accounting policy makers and researchers, as well as IR practitioners. The distribution of corporate reports online represents an opportunity to directly view and to measure the usage of reported financial information, the real time activity of users interacting with reported data, proposed by Parker (1982).

Several authors have addressed the use of corporate reports and usefulness of other corporate data (e.g. Trueblood Committee Report, 1973, Carsberg et al, 1974, ASSC, 1975, Hansford et al, 1996, ICAEW, 1998, ICAS, 1999, Weetman & Beattie, 1999). Studies generally captured data by employing questionnaires (e.g. Lee & Tweedie, 1976, Bartlett & Chandler, 1997, Beattie & Pratt, 2002), simulation (e.g. Elias, 1972) interviews (e.g. Lee & Tweedie, 1977, MORI, 2003) or observation (Nielson Norman Group, 2003). Although valuable, information usage in each case is measured indirectly and subject to various distortions through the experimental setting. This paper aims to investigate whether Web logs can be used as an alternative research methodology to directly measure the use of corporate reporting information for both IR and accounting research purposes.

Before discussing the content of IR information specifically, the paper defines what Web logs are and investigates the type and validity of the information to be gained from using them. It then continues by assessing the suitability of the information yielded by Web log analysis in studying the usage of online IR information. Finally, the paper draws upon the results

of a pilot study examining the Web logs of online IR Web sites of large UK listed companies to test the validity of our theoretical suggestions with real data. One specific company is used to illustrate the use of Web logs, as proposed in the paper¹.

2. The Investor Relations Function

This section describes the IR function, and specifically identifies how it has changed in response to the recent introduction of new communication technologies. This paper is based primarily on practice in the UK and reports data from companies domiciled in the UK and listed on the London Stock Exchange. However, where possible, comparisons are drawn with international practice and inferences are applicable to other private sector economies operating in developed capital markets.

2.1 Investor Relations

Investor relations can be described as the business function through which companies communicate with their stakeholders, principally their investors. The US National Investor Relations Institute (NIRI²) define IR as “a strategic management responsibility using the disciplines of finance, communication and marketing to manage the content and flow of company information to financial and other constituencies to maximize relative valuation” (NIRI, 2003).

Alternatively, the UK Investor Relations Society (IRS³) describe IR as the “means whereby companies maintain a dialogue with existing shareholders and potential investors. Its purpose is to present an accurate picture of corporate performance and prospects, thus allowing the investment community, through an informed market, to determine a realistic share price. As a result, IR can have a positive impact on a company’s market value and cost of capital relative to its industry sector and the overall economic climate” (IRS, 2003).

Both definitions imply that the objective of IR is not only to service the information demands of current and potential stakeholders (for example, investors, lenders, the general public) but to manage this service in a manner advantageous to the company. The nature of IR information typically provided extends beyond the financial reports including share price data,

¹ Although a UK company is used, the paper argues the methodology is equally applicable for all countries.

² The main professional body of Investor Relations professionals in the USA.

³ The UK equivalent of the NIRI

analyst reports, press releases and financial and non-financial data on future prospects, strategy, intangible assets and management quality.

IR is said to have originated as a response to corporate failures and hostile takeover bids in the 1970's and 1980's where directors were keen to establish relationships with their institutional shareholders (IRS, 2003). Where previously corporate communications were generally private and inert, the post World War II growth in financial institutions and their globalisation in holding international portfolios, had led to the need for more open communications with these generally more proactive shareholders. Geerings et al, (2003) argue that increasing globalisation of capital markets will lead to a strengthening and expansion of IR activities as companies need to attract more foreign investors in the face of greater competition for capital. Furthermore, larger institutions can often not easily sell large tracts of shares as their response to poor performance, without putting downward pressure on the share price causing associated problems for the company, and perhaps wider incidental effects in the broader market. As such, they are more likely to be active in communicating their expectations to company management (IRS, 2003).

In addition to this changing profile of company shareholders, tighter regulation imposed on companies on the disclosure of price sensitive information has focused on the role of IR. The growth of interest in, and influence of, corporate governance has led to companies maintaining more transparent and wide-ranging communications with their stakeholders. A recent survey by Marston & Empson (2003) covering the 500 largest companies in Europe, found that 93% of companies now have a formal IR strategy with an average annual budget of £780,702.

2.2 The Impact of the Internet on Investor Relations

The use of the Internet can assist the IR remit, particularly through its general accessibility as a communications channel, enabling the IR function to have greater reach and by automating the provision of some IR services, in comparison to the traditional IR model. The Internet also facilitates the rapid communication of information at a low marginal cost relative to alternatives available, thereby potentially reducing the costs of the IR function or enabling service improvements without comparable cost increases.

Of the various facilities provided by the presence of the Internet, the World Wide Web (the Web), is a user friendly information sharing platform particularly ideal for this purpose (for

example, see IFAC, 2002, Ettredge et al, 2001). Electronic mail (email), an additional facility provided by the Internet, is also becoming widely used by the IR community.

Online corporate reporting using the Web and/or email can be available to a user as and when they need it. It can also be provided in a form that is tailored to meet the information needs of the user. Unsurprisingly, there has been a significant and very rapid increase in the use of the Internet as a medium for corporate reporting and wider IR activities since the early 1990's. Several national and international studies have documented increases in the number of companies reporting corporate information via the Internet, and have reported increases in the volume of information disclosed via this medium (see Lymer et al (1999), Trites (1999), Deller et al (1999), Ashbaugh et al (1999), F.A.S.B. (2000), Geerings et al, (2003)). To cite just three studies, Petravick & Gillet (1996) find that 69% of sampled US companies had a Web site, with 54% disclosing some financial information. By 1999, upon examination of an international sample, Lymer et al (1999) find that 84% had Web sites, with 62% disclosing some financial information. Finally, Allam & Lymer (2002) find that 99.6% of sampled, large international companies in five countries had a Web site, all with an IR section disclosing financial and other information. This shows a clear growth in the use of this medium over time, and in the provision of IR Web sites.

An 'IR Web site' is defined as that part of a company Web site that holds information designed to meet the needs of key stakeholders. As noted in Allam & Lymer (2002), nearly all companies demarcate a specific IR section of their website for this purpose, which enables the easy identification of IR information. It typically contains financial reports, corporate press releases, contacts, shareholder information, webcasts, presentations, details of company meetings and links to analyst forecasts.

2.3 The Changing Role of the Internet in Investor Relations

A statutory responsibility exists in all countries for companies to provide financial and accounting information for public inspection, but this is not extended to requiring the provision of an IR Web site. In many cases, this implied that any provision of IR information online was voluntary disclosure on behalf of the reporting company. However, using the UK as an example, under the UK Companies Act (Electronic Communications) Order 2000⁴, limited companies can now meet at least part of their statutory reporting obligations by placing the necessary reports on

⁴ An amendment to the UK Companies Act 1985.

a company Web site, subject to shareholder agreement to opt into this form of distribution, where shareholders have been notified of its location and publication. Similar requirements also exist in the US (see SEC, 2000) and elsewhere around the world for many developed markets. Recent proposals by the UK government to amend company law place more weight on a company Web site as a primary means of information dissemination. For example, a major recent White Paper recommends that all company announcements are placed on the Web site as soon as they are released and that the annual reports of listed companies are placed on the Web site within 4 months of the year end (DTI, 2002).

Furthermore, listing rules often require key information to be placed upon a company Web site as a condition of listing. For example, as part of recent major corporate reporting reforms in the USA, the New York Stock Exchange has proposed that listed companies must disclose specific corporate governance information on their Web sites (NYSE, 2003). The SEC have also mandated the posting (or linking) of beneficial ownership reports on ‘insider transactions’ (SEC, 2003, for other non-UK examples, see Trites, 2002).

The provision of information on an IR Web site may no longer therefore solely represent voluntary disclosure for all companies and companies may be liable for any online disclosures made. For example, in the US, SEC Securities Act Release Number 33-7233 states that liability provisions of federal securities laws apply equally to electronic and paper based media (Gray & Debreceeny, 2001). Likewise, in Australia, AGS 1050 states that the responsibilities of the auditor and management do not change with respect to electronic presentation (AARF, 2002).

An IR Web site may also become a means for compliance with fair disclosure regulations, such as ‘Regulation FD’ in the US, where material, price sensitive information must be released publicly rather than in selective, non-public disclosures. At present, Web site disclosure alone does not satisfy the distribution requirements in Regulation FD, although there have been calls to extend the role of this dissemination medium (SEC, 2001). However, despite the trend to wider disclosure online, few regulations currently exist that require companies to maintain any specific set of general corporate information on a Web site, and the Internet is not regarded as a method of universal public disclosure as yet. This may change in the future.

2.4 Online Voluntary Disclosure

If companies are not required to disclose full sets of accounting and financial information online why do many choose to do so?⁵ Literature provides several theories for explaining voluntary disclosure (see for example Healy & Palepu, 2001, Verrecchia 2001, for a review of theory and Dechow et al, 2002, Ettredge et al, 2002 for empirical tests in an online setting). Agency theory suggests voluntary disclosure can reduce information asymmetry between shareholder and management, thereby reducing monitoring costs. However, this theory suggests that firms will disclose information up to a point where the benefits of disclosure in reducing adverse selection problems outweigh the costs of disclosing proprietary information. Alternatively, in the context of online reporting, signalling theory and institutional theory may give rationales for online disclosure. Signalling theory suggests that higher quality firms will use the Internet to disseminate 'positive' accounting information as widely as possible (Craven & Marston, 1999). Institutional theory suggests that online disclosure may be driven by a desire to conform to social or capital markets expectations (Rowbottom, 2002). That is, larger listed companies are all expected to provide online IR information.

Regardless of the theoretical motivation to voluntarily disclose information, a binding condition is that the benefits of disclosure outweigh the costs. An implicit aim in IR is that it will attempt to raise demand for equity, thereby increasing the price of shares, and, as suggested above, reduce the cost of capital. Further remits may be to spread ownership of equity, increase liquidity or marketability of company shares in order to 'maximize relative valuation' (NIRI, 2003). In demonstrating the marketing function of IR, Marston & Empson (2003) document how one large IR department is involved in obtaining product placements in films.

IR Web sites are increasingly referred to in many corporate press releases as a source of further information. Allam & Lymer (2002) find that 98.8% of large companies in a multi-country study disclose press releases on their Web sites. As such, companies have the potential to manage their information disclosure by boosting the corporate image of a company by emphasising positive and interpreting potentially negative information. The NIRI definition of IR alludes to this role whereby the content and flow of information are managed, activities carefully choreographed to fulfil statutory obligations and commercial concerns.

⁵ See section 2.2

2.5 Other Motivations for Investor Relations Websites

Evidence on the rationale for creating and maintaining IR Web sites is mixed. A survey by Ashbaugh et al (1999) found that those firms engaged in online corporate reporting placed greater emphasis on communications with potential and existing shareholders. Interviews conducted by Ettredge et al (2001) found that IR directors do see the Web site as a means of reducing administrative costs and that online disclosure helps provide a common level of disclosure to all stakeholders. However, those IR directors of companies with a low proportion of private investors viewed it as an expensive extra, indicating that the marginal benefits of online disclosure were minimal. The UK IRS suggested that Web site outlay and maintenance costs of £20,000-30,000 could be quickly recouped where the marginal costs of distributing online financial reports are zero compared to an estimated £5 on average to send a hard copy of the financial report⁶ (Beattie & Pratt, 2001). In the US, NIRI (2002) state that the average cost of the annual report in 2002 was \$3.73, down from \$4.38 in 1999. Furthermore, they report that the average 'hard copy' print run has fallen by 6.4%. However, Jones et al (2001) report that cost savings from distributing information online may not be a motivator for operating an IR Web site because gains can be eroded by the costs of processing greater requests for information that are generated by the Web site.

Evidence has also been published on the quality of IR Web sites. An analysis of UK FTSE-100 company Web sites, reported by Interactive Bureau (2003) found that 72 needed 'substantial attention' in terms of site design, usability and content, and failed to meet the needs of key constituents. A global survey of 100 companies reported usability problems related to poor navigation, formatting, archiving and the use of Flash and other proprietary software unavailable to all users automatically (Jones, 2003).

These findings suggest that there is a need to appraise corporate Web sites and their usefulness as a means of disseminating information but, to date, there is a lack of available evidence on this issue⁷.

⁶ Figures stated in 1998 prices

⁷ See Section 1

3. Web Logs

This section outlines the nature and information content of Web logs: the primary data source for the methodology investigated in subsequent sections of this paper aimed at measuring the demand for online IR information.

Whilst transmitting data over a network, servers monitor and record network communications in various logs. The primary interest of this paper is the Web server log file, or access log⁸, that provides specific data on each file processed by a server. There are several different formats of Web log files that are derived from different Web server configurations. All record data on the Internet Protocol (IP) address of the computer requesting the information, the date and time of the request, the size and Universal Resource Locator (URL) of the file requested.

Log files are written in ASCII format⁹ but the information recorded differs with different Web servers recording slightly different data. Web log files are written in a common log format of which there are several derivations. The primary ones are the NCSA (National Center for Supercomputing Applications) and the IIS (Microsoft Internet Information Server) versions. In the common Web log format, each line represents a single request for data from a 'client' browser (the user) to the Web 'server' (provided by the company in the case of IR data¹⁰). The information captured is illustrated in Table 1 below.

Table 1: Relevant data collected in a common Web log format

<i>Common Log Format</i>
Client IP: 'address' of the entity requesting information
Navigation date / time: when the request was made, offset from GMT
HTTP type of request made e.g. 'get' a particular Web page
Requested URL: name of 'page' being requested
Protocol Version & HTTP status code: the Web servers response to a request e.g. 200 = 'File Transfer OK', 404 = 'File Not Found'
Number of bytes transferred: indicating the size of 'resource' requested

A typical entry from a Web log file in the common format is shown below:

⁸ Referred to hereafter as the web logs

⁹ American Standard Code for Information Interchange

¹⁰ Requests are logged by the suppliers Web server and relayed to the user via their client Web server.

```
61.152.17.128.user_address.net - - [31/Mar/2002:00:01:14 +0100] "GET
http://investor_relations/Web_page/address/ HTTP/1.0" 200 776
```

The text firstly identifies the IP address of the party requesting the information, the exact time and date of the request and the nature of the request, to ‘get’ a particular Web page in HTTP version 1.0. The entry then details the authorisation code (200) and the amount of data transferred (776 bytes).

Many modern Web servers are also now able to record requests for information in extended or combined log file formats that provide additional information in one Web log file. The extra variables available in the combined log file format are displayed in Table 2 (below). The ‘Referer URL’ field indicates the URL of the last location that the user’s browser had accessed before it was directed to the requested page. If present, it records the address of the ‘Web page’ that provides a link to the Web site. The ‘http user agent’ field provides information on the browser’s system (i.e. version and user platform) used to access the page. It describes the software used by the ‘client’ to make the request and can be used to identify the user’s operating system. This may be useful to Web site designers to indicate the technical platform most suitable for the users of that information. It can also be used to identify automated ‘robots’ that search Web pages undertaking tasks such as updating search engines. The ‘http cookie’ field provides details of cookies: tokens that identify an unique user that are passed back and forth between client browser and server when a Web page is requested. If used, they can provide historical browsing information about an user.

Table 2: Additional relevant data collected in a combined Web log format

<i>Combined Log Format</i>
Referer URL: the ‘address’ of the users browser before they visited the site
HTTP User Agent: name & version of the browser being used to access information, and type of user platform e.g. <i>"Mozilla/4.0 (compatible; MSIE 4.01; Windows NT)"</i>
HTTP Cookie: holds contents of tokens that are passed between a browser and server during a session. Holds data on a unique user

Web log files of Web sites can hold very large volumes of data depending on the frequency of requests made to the server being logged. For example, at the very large end of the

scale, Nicholas et al (2000) report that the Times/Sunday Times newspaper Web site generated 200 million lines in its Web log each day as long ago as August 1999.

Web logs were originally designed to measure the volume of traffic passing through a Web site and the subsequent actions of computers and computer networks, not the direct actions of end users (Zawitz, 1998). Consequently, some further analysis of this information is necessary to assess usage. Various metrics have been established by researchers seeking to analyse Web logs in attempting to extrapolate data from the Web logs into variables useful to those analysing Web site usage. Numerous software packages are available that filter and identify these variables within the Web log files, producing customisable reports that can be used to turn raw data logs into more directly useful and usable information based upon these metrics¹¹. The next section will briefly summarise the most widely used of these metrics, outline their uses and their effectiveness in interpreting usage.

3.1 Basic Web Log Analysis Metrics

A key problem with Web log analysis lies in defining the ‘unit of information’. This can be assessed in different ways, interpreted differently and has different meanings based on the chosen unit of analysis. This section reviews the choices for basic metrics to understand these units of information typically made available in standard Web log analysis tools: Hits, Page Views, Visits and Page Views.

In measuring the consumption of online data, the most basic metrics are *use* and *users* (Nicholas et al, 1999). The information unit of interest in analysis of *use* is the Web page, which is stored and transmitted in a number of different *files*. A ‘Hit’ is registered where any *file* is requested from a Web server. As Web site pages will typically be made up of several files including HTML, graphics, style sheets, maybe audio content and other associated files, the number of Hits recorded when a specified page is requested will be greater for richer pages (i.e. with more elements). In addition, if a Web site uses ‘Frames’ where multiple graphics files are linked together to provide a common template for each Web page, several files will be sent to users requesting to view a Web page. For example, Frames constituted 84% of the number of Hits in an analysis of a newspaper Web site in 1998-9 (Nicholas et al, 2000). Therefore, whilst the number of Hits is a useful indicator of Web site traffic, it does not accurately measure the number of actual (whole) Web pages viewed on a Web site as there is no 1:1 relationship

¹¹ For example, WebTrends, Log Analyzer 123, ClickTracks, SiteCatalyst, Sawmill

between Hits and Web pages viewed (Nicholas et al, 1999). This problem is magnified where Web sites offer 'low graphics' versions for users connecting at lower rates of data transfer, by distorting the relationship between the number of Hits and the number of pages viewed.

As a result of the limitations of Hits as a unit of analysis, 'Page Views' are often used as a more effective way of assessing the 'use' metric¹². Page Views measure the number of actual pages viewed by stripping out Frames, and the additional multiple linked files served when a particular Web page is requested. In addition, Page Views can often be subject to further refinement such as the removal of any Hits from automated agents such as 'Robots' or 'Pushed Pages'¹³ by identifying Web log text in the 'http user agent' field. Robots (or 'Spiders') are used by search engines and similar aggregators to automatically scan Web sites to update their databases of current Internet content. For some sites, access by automated agents can account for one third of all Web log entries (Gutzman, 1999).

Page Views are therefore important for appropriate interpretation of Web log statistics. Current Web log analysis software can usually provide reports on the number of Page Views, in addition to the number of Hits, viewed on a Web site.

Another common metric used in Web log analysis is a 'Visit' (or session). This is used to define a series of Page Views served to one user. When users move from page to page when navigating the Web, clear Visits to a specific page are recorded, and are demarcated by an initial request for each page on the server, or a link off the server to somewhere else on the Web. However, this is not always the case as, on occasion, no sign off from a page will be recorded. Therefore a Visit is automatically assumed to end whenever there is a period of inactivity between Page Views. This period is typically set at 30 minutes. Combining these user metrics and assumptions enables Web log analysis software to report the number and average length of Visits¹⁴.

Web logs primarily record the actions of computers and computer networks as they send information across the Internet. However, they can provide some information on the user who is making the requests.

¹² Sometimes referred to as Page Impressions

¹³ For example, adverts which automatically pop up when entering a site.

¹⁴ As used in Project COUNTER Standards on the usage of electronic resources (see www.projectcounter.org).

Each request for information from a Web server must specify the details of where to send that information. This represents the location of the client computer on the Internet. The numbering system used on the Internet, the Internet Protocol (IP) address, provides this destination information. The number of unique IP addresses making requests of the Web server can be counted giving an indication as to the number of unique users or 'Visitors' that accessed the Web site within a chosen timeframe. By converting the numeric IP address into its corresponding text name, it is often possible to identify what organisation a user comes from. For example, many text IP addresses signify the country (e.g. .uk, .de or .pt), the organisation type, (e.g. .ac or .edu) and occasionally an organisation (e.g. .bp or .ibm). Whilst this is not always possible to do, it can be used to provide basic information about the user.

The basic metrics identified above can be amalgamated to provide summary statistics on the use of a Web site, and form the basis of reports exported by popular Web log analysis software.

3.2 Limitations in the Use of Basic Web Log Metrics

This section identifies several constraints, relating to defining 'Visits', IP addresses and memory caching, in interpreting the data provided by Web logs.

Firstly, the recording of a Page View from a user does not mean the information provided on that Web page was actually used or even wanted by the user. For example, it does not characterise mistakes where users clicked on the wrong hyperlink to access an unwanted page. Use statistics must therefore be used carefully if they are to be a surrogate for the demand for a particular Web page. In defining Visits, the assumption of a 30 minute time interval between Page Views as defining when somebody has signed off may be spurious (Fieber, 1997). For example, once this interval limit has been reached, if a user starts to browse again, subsequent Page Views will be recorded as a new Visit even though it is still, in reality, the same user session. A further problem with this limit is determining the meaning of the reaching of the page visit interval. A 30 minute time lapse could be interpreted as satisfaction with the information requested, particularly analytical financial reporting information where users may be expected to spend more time digesting the information presented. Conversely, a long Visit requesting many Page Views could be due to frustration with the design of the site, where the information required cannot easily be found rather than an indicator that the site is of great interest. Alternatively, Nicholas et al (2002) find that typically, most pages are read within one

minute and users will have ceased ‘using’ the information long before the 30 minute cut off point, whatever information it contains. In such cases, the combination of Visit information with timing data can provide some measure to mitigate these problems. For example, pages viewed for only a few seconds are unlikely to have been read if they contain a reasonable amount of information. The use of a 30 minute cut off point will cause most mean averages reported to be unrepresentative where the distributions of timing variables will be heavily skewed as empirically, most Visits are characteristically short (Nicholas et al, 2000). To deal with the skewness of the underlying distribution, it is preferable to calculate a median or other average statistics insensitive to skewness, such as the 5% mean or Huber’s M-estimator (Nicholas et al, 2002).

In addition to these Page View metric limitations a user of Web logs must consider, it is also difficult to relate Visits to Visitors on the basis of Web log data alone. Each Visit is characterised by an IP address. However, one person can have more than one IP address they may use over time and single IP addresses do not always identify individual computers. For example, where one individual user accesses a particular Web site from home and from work, two Visitors would normally be identified as each will almost certainly have different IP addresses when in fact they are the same visitor and would ideally be analysed as such.

Also problematic is that many users can appear to have one IP address in the Web log. This will be the case where organisations, using a firewall to filter all access to the public Internet, will record just one IP address (of the proxy server) rather than the IP address of individual users (Riphagen & Kanfer, 1996). Therefore, only one Visitor will be recorded even if many individuals have accessed a Web page from that organisation within the timeframe.

Those users accessing a Web site through a ‘dial-up’ connection will not have an unique IP address but will be assigned a different one every time they dial up the Internet Service Provider (ISP) to access the Internet. To compound this problem, several ISPs change the IP addresses of individual users during a session.

Therefore, because of a combination of these problems, the number of unique Visits recorded in Web log analysis will not record exactly the number of individual users visiting a site and some additional metrics (even if just rules of thumb) are needed to adjust the plain statistics to overcome these issues.

The information gained from IP addresses, when converted into text names, can sometimes be misleading (Zawitz, 1998). For example, UK based users can have what appear to be 'foreign' IP addresses. This is due to the fact that IP addresses are converted into physical addresses using information contained in WHOIS¹⁵. This service assigns IP addresses based on the domain name server and the ISP contact information registered with the domain name registry service, InterNIC¹⁶. For example, a site with a large user base who access the Internet using AOL ISP will find many of these users commonly reported as having come from Vienna, Virginia, USA. However, this merely indicates the registered address of AOL as their provider – not their specific location when accessing the Web site under review. Furthermore, most Web log analysis software incorrectly initially labels all '.com' sites as from the US and more detailed analysis of the '.com' data is needed to determine if this is in fact true or not where .com addresses can be assigned to servers anywhere in the world. As a result, the information extracted from the client IP address field requires careful interpretation to avoid reporting misleading information.

Perhaps the most significant problem in interpreting Web log files relates to memory caching – local stores of Internet pages used to aid speedy access to the Internet. Where users select a Web page using the navigation buttons on their browser, such as the back or forward buttons during an Internet session for example, often no new line in the Web log is recorded. Instead the computer would simply return the information from its local memory cache, rather than requesting the information again directly across the Internet unless the user explicitly opts for their browser to always revisit the site on each requested viewing – not the default setting for most users. This cuts down on the quantity of data transmitted over the network and as such improves the efficiency of the network. However, it creates problems for the Web log analysis process. A navigation study by Tauscher & Greenberg (1997) found that 30% of navigation actions by users were backward and therefore not recorded as new Page Views, even when a user is returning to this page to extract new information and this fact could be usefully recorded.

Furthermore, even if the user does not specifically use backward navigation, if the Web site has been visited recently it may still be loaded from the local memory cache, rather than directly from the Web server again. Therefore, Hits, and subsequent aggregated Page Views,

¹⁵ A program that enables users to register their names and e-mail addresses and to search a database of people, domains, networks, and hosts using telnet

¹⁶ Internet Network Information Centre

Visits and Visitor statistics will be underestimated to the extent of those files being requested from the cache.

The proportion of files loaded from the cache is dependent upon the configuration settings of the browser as set by the user for their computer, but will typically hold files for a day, week, month, or until the memory cache becomes full. To exacerbate this caching problem, where a browser connects through an organisational connection or firewall, as typical for most non-individual users, files may be stored in a further local site cache to reduce traffic on the local network the user is connected to. Also for those accessing a Web site through a 'dial-up' ISP connection, typical for many individual users, the ISP themselves may also have a memory cache on one of their servers to reduce the amount of traffic being transmitted to the Internet from the ISP. Therefore local cache, site cache and ISP caches will result in further underestimation of the level of information requested from a Web site. A simulation by Fieber (1997) found that 43% of pages were cached at one level or another in this chain of caches suggesting a downward bias on actual Hits of information recorded of between a third and a half.

Memory caching may also impact on the subsequent analyses of the data that are recorded by the Web log. A key influence will be the architecture of the Web site. For example, in a hierarchical structure more upper level pages (e.g. menus or home pages) will be cached than lower level pages. This may underestimate the relative demand for more popular pages such as menus which are fewer levels or 'click downs' through a Web site, where users are able to use cached page impressions and do not request the information directly from the server, thereby remaining unrecorded within the Web log. This will also be affected by the overall popularity of the site, as seldom accessed sites will be less severely impacted in this way than more frequently accessed sites, and the browsing style of the client in using navigational or history buttons. In addition, measures recording the time spent on each Page View will be affected. Times are calculated as the differences between subsequent entries on the Web logs. It will thus overestimate Page View times if cached pages are also viewed in this interval but are omitted from the Web log (Nicholas et al, 2002). However, dynamic Web pages, common in IR Web sites, will be cached less often as the information contained in them is constantly changing (Fieber, 1997). Thus, Web sites with dynamic content, such as share price 'tickers' will be less affected by this problem.

As a response to these limitations, it is accepted that Web log analyses do not provide absolute levels of online information usage. Although they measure what information is demanded online, they do not indicate whether the information has been used or read. As such, they provide a proxy measure of online IR information usage, but alone will not be an appropriate proxy for general IR information usage.

Due to caching, they will systematically under-report demand. However, the magnitude of error will be normally distributed and can be assumed, *ceteris paribus*, to be constant over time. Thus, whilst not providing absolute levels of use, Web logs can give an indication of trends in, and changes of, information demand over time. In comparison with other sources, the longitudinal analysis of Web logs can indicate the access intensity of online corporate information and provide insights into what types of IR disclosure are most widely utilised. From a narrower perspective, Web log analysis can also assess in detail specific elements of IR activity in isolation such as the demand for financial reporting and accounting information. This can assist regulators in determining what accounting information is relevant by directly recording what information is accessed, in what sequence and at what point in time. In this way, it has the potential to emulate, enhance and triangulate with previous studies of accounting relevance. For example, in assessing what specific financial statement data is accessed by users. It can be argued that the direct measurement of information demand in this way can provide unique insights into the use of IR information, as distinct from estimating usage indirectly from traditional methods such as surveys or questionnaires, or by observation in experimental settings. Web logs record actual behaviour rather than the simulated behaviour that is captured when estimating usage indirectly using surveys or questionnaires. Therefore, Web logs measure what information users have actually demanded whereas surveys and questionnaires may measure what information users perceive they use or have recalled using. Furthermore, the proportion of user actions that can be recorded by Web logs is generally higher than the proportion sampled by traditional methods. Hence, Web logs can provide unique data that can complement existing research methodologies and hence improve the accuracy of empirical findings on corporate information usage by triangulating with studies undertaken using alternative research methods.

Although Web log data from corporate Web sites is commercially sensitive, and its availability may be restricted accordingly, its potential utilisation in research and subsequent policy making may be possible where reports are reported anonymously (as undertaken in section 4).

In the future, the limitations discussed above may diminish due to technological changes. In particular, caching problems can be alleviated by collecting data directly from the user’s browser (the client) rather than from the Web server. All user actions are recorded and sent directly to a 3rd party server. This client side data tagging, utilised in more recent software such as WebTrendsLive, may be more prevalent in the future and enables more accurate data on absolute access levels to be recorded. Thus, the potential of Web log methodology in studying the demand for IR information may be further enhanced over time.

Having described how to research information demand, the next section proceeds by illustrating how this methodology using common Web log metrics can be used to specifically measure the demand for IR information.

4. The Use of Web Log Analysis of Investor Relations Information

To assess the potential of Web log analysis for studying the usage of IR data, this paper draws upon a study conducted on the Web logs of companies listed on the London Stock Exchange. The results shown below are from a longitudinal analysis of the Web log files of the IR Web site of one large listed company¹⁷, commonly a constituent of the FTSE 100 (an index representing the largest companies listed on the exchange). Due to the sensitivity of this data, our conditions of data access prevent the full disclosure of any company details, however the data covers the period July 2001 to June 2002. The aim is to present the type of information that can be extracted and assess its usefulness for future corporate reporting research. Statistics that can be derived from the Web logs are shown in Table 3¹⁸.

Table 3: Investor Relations Web site: Statistics (July 2001-June 2002)

<u>Metric</u>	July 2001	June 2002	Mean (July 2001-June 2002)
Hits per month	256,885	96,432	188,063
Average Hits per Day	8,027	3,110	6,000

¹⁷ The results comprise the Web log entries for all pages available from the IR section of the main company Web site. All other Web log entries recording access to other parts of the site have been filtered out.

¹⁸ Statistics were generated using WebTrends v7.0.

Page Views per Month	86,512	45,763	61,294
Average Page views per Day	2,703	1,476	1,957
Visits per Month	5,337	5,460	5,697
Average Visits per Day	166	176	182
Average Page Views per Visit	16.2	8.4	10.7
Average Visitors per Month	1,875	1,893	2,012
Average Page Views per Visitor	46	24	30
Average One Time Visitors per Month	1,271	1,330	1,408
Average Returning Visitors per Month	604	563	603
Average Visit Length	25m 35s	31m 40s	24m 27s
Median Visit Length	2m 33s	2m 29s	2m 15s

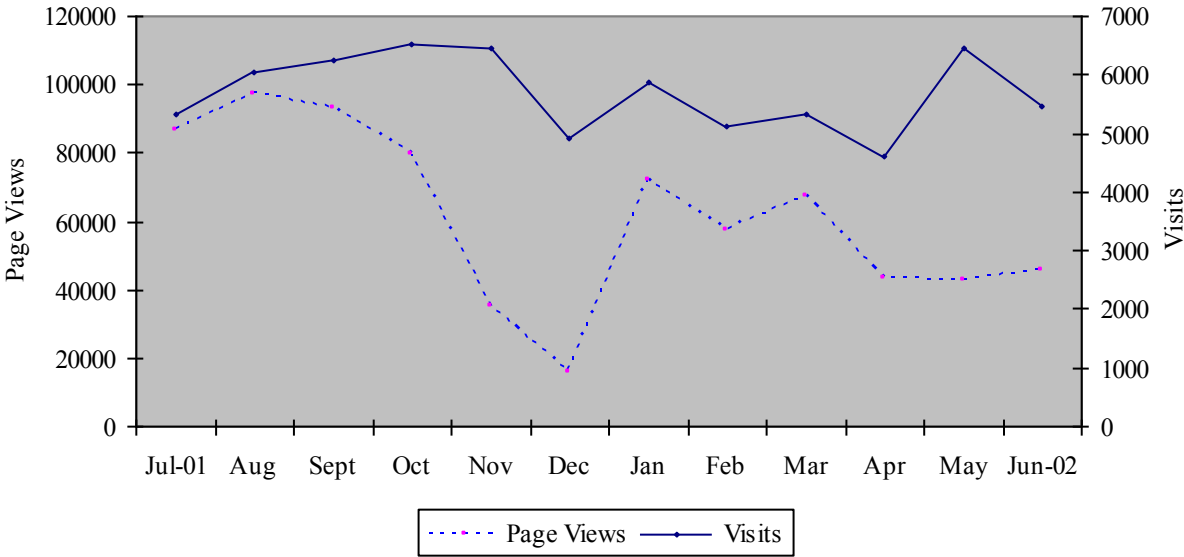
The mean statistics for the year ending June 2002 are disclosed although the limitations in recording absolute usage levels by Web logs constrain our inferences, as discussed in section 3.2. However, useful insights can be made by examining the change in metrics over the year on one Web site, as highlighted by statistics for the months July 2001 and June 2002, and the averages over the year. Where possible, comparisons are drawn with Web log analysis of other types of Internet resources that have been conducted by Nicholas et al (2000) and Thelwall (2001) which can also provide a benchmark for the analysis of IR information.

A comparison of the average Hits and the average Page Views highlights the drawbacks of Hits as a Web log metric as discussed in section 3.1. The mean average statistics indicate that each Web page requested (the Page View) generates approximately 31 Hits (representing the number of different files that are sent).

In assessing access to the IR Web site, the average number of Page Views per month has decreased drastically from July 2001 to June 2002. We note that due to memory caching, actual pages viewed are likely to be 30-50% higher. Figure 1 displays the monthly recorded averages over the period, showing fairly wide fluctuations in demand over the year. The highest levels of Page Views are requested in August 2001, thereafter access declines steadily until December 2001 before rising sharply in January 2002 and remaining relatively stable for the remainder of the sample period. In examining the Times newspaper site, Nicholas et al (2000) find that the volume of Page Views fluctuates by up to 10% each month whereas we find

fluctuations of 59% above the mean and 74% below the mean. This may indicate that the level of demand for company information will be related to current events; economy wide, industry specific and company related¹⁹. Interestingly, the number of Page Views does not respond to the publication of the companies interim or annual reports (in November 2001 and May/June 2002).

Figure 1: Investor Relations Website: Monthly Page Views & Visits (July 2001 - June 2002)



In contrast to the decline in the level of Page Views over the year, the number of Visits has increased slightly. Figure 1 shows that the number of Visits fluctuates between about 4500 and 6500 over the year examined. Whilst the number of Visits has remained fairly stable over the year, the amount of information requested (the number of Page Views) has decreased. In other words, users are maintaining or increasing their visits to the IR Web site but are viewing less pages per visit.

The average Page Views per Visit has fallen by half over the year from 16.2 to 8.4, reasserting the decrease in Page Views over the year. This metric may offer conclusions about the ease of navigation through a site or satisfaction with the information disclosed. The number of Visits also shows some correlation with the reporting calendar of the company. Those periods when the company released interim and annual reporting reports (and subsequent press releases) on the Web site show some correlation with the number of Visits.

¹⁹ Subsequent analysis would enable the researcher to examine the nature of this month by month change as part of their regular review of usage patterns for a specific company.

In assessing the number of different users accessing the information, we can draw upon the number of Visitors (separate visits). This fluctuated between approximately 1600 and 2300 per month. From this dataset, an average of 2,012 Visitors made an average of 5,697 Visits to the Web site per month. Thus, each Visitor made an average 2.8 Visits to the Web site per month, suggesting a fairly stable (i.e. frequently returning) user community. However, some caution should be exercised in attempting to quantify Visitors too specifically due to the frequency of proxy server caching prevalent in most commercial organisations (as discussed in Section 3.2).

The average Page Views per Visitor, can inform about the breadth of information requested by each individual user on one or more Visits, indicating how much of a Web site's information may be of use to an individual. Each Visitor viewed an average of 30 different Page Views in one or more Visits, although in line with other access statistics, this fell significantly during the year.

Important information may be given in the number of 'Return Visitors' to a site, especially when analysed over multiple time periods. This metric indicates where user information needs have been met by the Web site to the extent that they return to reuse it as needed or to monitor it changing over time (if the data is dynamic), suggesting some satisfaction with the information disclosed. As such, it is a powerful and reliable measure of site 'stickiness' (Nicholas et al, 2002). The average number of Return Visitors who have visited the site more than once within a month was 603 and remained fairly static ranging from 520 and 680, perhaps indicating a recurrent user community. However, due to memory caching, this could mean that different users from the same organisation are returning to the Web site rather than all being unique Visitors. The number of Visitors who made only one Visit to the site on average was 1,408, and also remained stable over the year with lower levels for the UK summer months of July 2001 and June 2002. One time Visitors made up 70% of the total Visitors to the Web site. This is lower than comparable statistics, 84% and 89%, collected on healthcare Web sites by Nicholas et al (2002) supporting the assertion that the IR Web site has a relatively regular community of core users.

The difficulties in interpreting the identities of Visitors, detailed in section 3.2, are highlighted by examining the IP addresses of Visitors. For the sample period, 78.2% of users are classified as US users, whilst only 20.5% are classified as international users (1.3% are of unknown origin). This may misrepresent the origin of many users accessing the IR Web site of a

UK company due to the misinterpretation of IP addresses. It would be expected that the international numbers should be somewhat higher.

The average length of time spent on each Visit and each Page View can indicate how quickly users navigate through the site. Although, the time-spans will be influenced by connection speeds²⁰, these metrics can provide some useful insights and help distinguish between those accessing the data and potentially using it. The mean average Visit is calculated at 24 minutes 28 seconds but analysis of the median time (2 minutes 15 seconds) demonstrates that the distribution is skewed by a number of long Visits. This is caused by the assumption of a 30 minute sign off point as indicating the end of a visit. Hence, the median time spent on visiting the site of just over 2 minutes is the more reliable measure and illustrates the lack of robustness of the mean in this context. Coupled with the average Page Views per Visit of 10.7, Visitors spend an average of 12.6 seconds on each Web page. Given that this metric is likely to be depressed by memory caching, online IR information tends to be scanned very quickly. This is similar to the findings of Jones (2002) who report that IR site visits last approximately 3 minutes.

Further insights can be made from analysing the type of information that is actually requested by the users to the IR Web site. For this analysis, a subsection of the IR Web site is used in isolation from the rest of the site, namely the financial reports and accounts. Over the sample period, financial reporting information was broken down into several sections; the Annual Financial Report and Accounts, available in HTML and PDF formats, and an Annual Review available in HTML and PDF. The Annual Review contained summarised information from the Annual Financial Report such as summary financial statements and a selection of narrative and visual material. Due to the possible duplicity of this information, the analysis here is focused on the more detailed Annual Financial Report and Accounts section in keeping with the illustrative nature of this analysis.

Table 4 shows the number of Page Views and Visits to the Annual Financial Report and Accounts and the proportionate use of each format. It clearly indicates that users favour viewing financial reporting information in a PDF format, where an exact replica of the paper document is downloaded to the user's computer and can be subsequently printed or stored. It is important to note that once financial reports are downloaded in the PDF format, they may be used more than once, so PDF figures will generally show minimum levels of information usage. The preference

²⁰ Connection speeds will be affected by the time of connection, type of Internet access chosen by the user and number of files requested

for PDF may be due to the strain of viewing financial reports on screen compared to viewing a hard copy document, or perhaps the conservatism of financial reporting users who are unaccustomed to viewing on-screen reports where hyperlinks connect the different sections. Nevertheless, the lack of demand for financial reporting in a HTML format in this case is clearly indicated. However, the popularity of this format may increase in the future as HTML tagging mechanisms such as XBRL (extensible business reporting language)²¹ improve its usability by allowing the rapid extraction of data from online HTML financial reports.

Table 4: Investor Relations Web site: Annual Report – Average Monthly Page Views & Visits

<u>Web Page Requested</u>	<u>Page Views</u>		<u>Visits</u>	
	Number	%	Number	%
Annual Report & Accounts: HTML	283	8.2	265	11.5
Annual Report & Accounts: PDF	3184	91.8	2041	88.5
Total	3467	100	2306	100

We finish our illustrative analysis by focusing on the number of Page Views and Visits made to individual sections of the Annual Report & Accounts disclosed in HTML format. Although the data is generated by only 265 Visits, we can assess the relative demand for specific types of financial reporting information. These metrics are shown in Table 5.

Table 5: Investor Relations Web site: Annual Report Page Views & Visits (July 2001-June 2002)

<u>Web Page Requested</u>	<u>Page Views</u>	<u>Visits</u>
	%	%
<i>2001 Annual Report & Accounts</i>	<u>100</u>	<u>100</u>
Corporate Governance	10.5	10.6
Remuneration Report	11.9	12.0
Directors' Report	7.8	7.8
Report of the Auditors	5.3	5.2
Group Profit and Loss Account	14.4	14.7

²¹ An XML derivative created to support online financial reporting information dissemination. See <http://www.xbrl.org> for further details.

Balance Sheets	13.2	13.2
Group Cash Flow Statements	6.8	6.9
Statement of Total Recognised Gains & Losses	5.4	5.3
Notes to the Financial Statements	14.2	13.7
Five Year Financial Summary	10.5	10.6

The most requested financial reporting information, as indicated by Page Views and Visits is the Profit & Loss Account (Income Statement) with 14.4% of all annual report Page Views. This supports survey findings for private shareholders and finance directors in Beattie & Pratt (2002) and professional users in Barker (2001) who find that the Profit & Loss Account is ranked as the most useful source of corporate financial information. This is followed by the Notes to the Accounts (somewhat comforting for accounting regulators) and the Balance Sheet. The least requested part of the annual report is the Auditors Report, although a possible explanation for this result may be that it is only requested where there are concerns over the reliability of the financial information presented. Nevertheless, few users appear to check whether the information viewed elsewhere in the annual report is considered to be a true and fair reflection of the organisations activities. This supports Hodge's (2001) assertion that online users are often unaware of whether the information presented has been audited and also Lymer & Debreceeny (2003) regarding the need to address the nature of online audit reports.

Of the narrative sections of the report, the Remuneration Report is most frequently requested along with the statement on compliance with the Combined Code for Corporate Governance. This may be a reflection of increasing recent stakeholder interest in corporate governance and, particularly in the UK, management compensation levels. Surprisingly, the 2 other financial statements, the Cash Flow Statement and the Statement of Total Recognised Gains & Losses are amongst the least requested parts of the report. In contrast, the Cash Flow Statement is the most useful financial statement by expert users and audit partners in the Beattie & Pratt (2002) study.

This data is potentially very valuable in accounting research as a direct measure of what accounting information is demanded. Whilst absolute levels of demand are subject to error, proportional levels of demand, disclosed above, are more resistant to the limitations discussed in

section 3.2 and provide a valuable indicator of the relevance of accounting information. This methodology has the potential to complement and further existing methods of determining the use of accounting information (e.g. Lee & Tweedie, 1976, Beattie & Pratt, 2002). Similar statistics can be collected for the remaining sections of the IR Web site such as share price information and company profile information, also providing a direct measure of the use of other corporate information sources.

We complete this section by considering what other Web log analyses may yield further metrics of potential use for those monitoring the use of online corporate information. The navigation of a site may be assessed by collecting data on the sequence of Page Views. IR information may be browsed on a regular or ad-hoc basis. For ad-hoc browsers unfamiliar with a site, an overview site map may help them locate the demanded information quickly (Lai & Yang, 2000). The sequence of Page Views may help in the most efficient organisation of a site map, particularly in placing the most requested information at fewer 'click downs' within the site, or even direct linkages from a home page to the most accessed information. The most common combination of Page Views will also indicate what type of information is most commonly requested, allowing it to be placed at easy to access points. For example, for June 2002, the most popular combination of pages for our sample company, representing 8% of Visits that month, was a Page View of the 'what's new' link on the IR Web site followed by a request for the company's current share price. This suggests that the IR Web site is an important updating source for a set of regular online users and this fact could usefully be reflected in future site redesigns, as well as suggesting the need to regularly refresh this page.

The referring Web pages listed by the 'referrer URL' field will show which Web page linked to the Web site or whether the Web site was found by entering a query term on a search engine. The nature of the query terms used, can provide insights into what information users were seeking and how they discovered it. Due to confidentiality constraints, we cannot disclose the most popular search terms although for June 2002, 31% of Visits had no identifiable referring page. This could indicate a large proportion of users have typed in the URL directly into their browsers or bookmarked the page. Where users found the IR Web site using a search engine, the top operators for June 2002 were dominated by Google with 82% of searches followed by Yahoo with 16%. This information could prove useful in future 'marketing' strategies for updating the IR Web site.

5. Summary & Conclusions

As the level of online IR information disclosed by companies is increasing, the need to assess what information is used becomes more important. As the level of corporate reporting information supplied outstrips the level of information demanded, it is important to know what reporting data is used and what is therefore relevant to users. This paper illustrates that Web log analysis has the potential to enable the study of users directly interacting with corporate financial information and to observe sections of users that have not been studied before. In the present corporate reporting environment, characterised by an increasing volume of published information and the feasible production of customised information, this methodology, in conjunction with others, can inform research and policy making. It can also provide insights into the most effective organisation of IR Web sites to enable corporate financial information to be disclosed and presented in such a way that reduces the level of information asymmetry between organisations and stakeholders. Web log analysis can assist the IR function in terms of online content management, justifying Web site investment and deciding on future IR strategy (Adams & Frost, 2003). Whilst, technical barriers exist that reduce the accuracy of estimates of the information requested online, Web logs can provide a unique insight into trends in usage over time. As such, Web log analysis can be seen as the first step to develop direct observation strategies for improved understanding of usage of corporate and financial reporting information.

The methodology outlined in this paper can be developed further where Web sites require users to register their details before accessing IR information as recommended by IR industry groups (e.g. irbestpractice.org, 2003). This would ensure all users would sign on, when accessing the site, enabling Web logs to identify individual users and enable far greater accuracy in profiling demand. However, requiring users to register and sign-on before accessing information may impose an obstacle to using the Web site and may reduce short term demand (Nielson Norman Group, 2003, Neely et al, 2002). Web sites can also be designed with client side data tagging, where information is recorded directly from the users browser thereby eliminating the underreporting problems caused by memory caching inherent in traditional Web sever log files. Information collected by cookies, which record the actions of a user interacting with a Web site, can also extend the analysis of information demand in future work.

As the development of HTM tagging highlights the benefits of online disclosure, more companies are beginning to disclose annual reports in separate parts in HTML or in pdf files,

rather than as a large single pdf file. This will enable Web logs to provide greater detail on the most widely used, most relevant parts of the annual report by specifying which sections are requested. This can be viewed as a precursor to providing customised financial information for users, a key enabling feature of the online disclosure compared to offline disclosure. The schema for tagging financial information, XBRL, will provide tags on individual corporate figures, enabling Web log analysis to report on the use of specific units of corporate information in the future.

Better techniques for reviewing collective trends or comparative trends - than the single company review used to illustrate the methodology in this paper - may be investigated. Web log analysis can be complemented by visualisations techniques to aid investigation of the data. For example, star field visualisations are suggested by Hochheiser & Shneiderman (2001) for complementing Web log analytics and providing richer analysis.

Finally, Web logs also enable us to compare the use of online corporate information with the use of offline corporate information to assess whether reporting data is used differently when it is communicated electronically as opposed to communication via distribution in a 'hard copy' format. Evidence on this issue can contribute to the debate over whether online reporting is a substitute or complement to offline reporting.

This paper has sought to assess the suitability of Web log analysis to gain insights into the usage of corporate reporting and other IR information. Whilst limitations are highlighted, opportunities exist to further pursue this novel methodology in future research.

Acknowledgements

We gratefully acknowledge the financial support of the UK Economic and Social Science Research Council (ESRC) in undertaking this project: E-Society Research Programme grant number RES-335-25-0038. We also are thankful for the comments made by 4 anonymous ESRC referees. For further details of the wider project from which this project is drawn, see www.ir-research.com.

References

- Accounting Standards Steering Committee. *The Corporate Report*. London: ASC, 1975.
- Adams, C.A. & Frost, G.R. The Development of the Corporate Website as a Medium for Communication on Ethical, Social and Environmental Issues: A Comparative Analysis of Australia, Germany and the UK. *EAA Conference Paper*, 2003.
- Ashbaugh, H., Johnstone, K.M. & Warfield, T.D. Corporate Reporting on the Internet. *Accounting Horizons*, 1999;13/3:241--257.

- Allam, A. & Lymer, A. *Developments in Internet Financial Reporting: Review & Analysis Across 5 Developed Countries*. University of Birmingham Working Paper, 2002.
- Australian Accounting Research Foundation. *AGS 1050: Audit Issues Relating to the Electronic Presentation of Financial Reports*. AARF, 2002.
- Barker, R. *Institutional Investors, Accounting Information and the ASB*. Edinburgh: ICAS, 2001.
- Bartlett, S.A. & Chandler, R.A. The Corporate Report and the Private Shareholder. *British Accounting Review*, 1997;29/3:245--261
- Beattie, V. & Pratt, K. *Business Reporting: harnessing the power of the Internet for users*. Edinburgh: ICAS, 2001.
- Beattie, V. & Pratt, K. *Voluntary Annual Report Disclosures: what users want*. Edinburgh: ICAS, 2002.
- Carsberg, B., Hope A. & Scapens, R. The Objectives of Published Annual Reports. *Accounting & Business Research*, 1974;4/13:162--177
- Craven, B.M. & Marston, C.L. Financial Reporting on the Internet by Leading UK Companies. *European Accounting Review*, 1999;8/2:321--333.
- Debreceny, R., Gray, G.L. & Rahman, A. The Determinants of Internet Financial Reporting. *Journal of Accounting and Public Policy*, 2002;21:371--394.
- Deller, D., Stubenrath, M. & Weber, C. A Survey on the use of the Internet for Investor Relations in the USA, the UK and Germany. *European Accounting Review*, 1999;8/2:351-364.
- DTI. *Modernising Company Law*. UK: Department of Trade & Industry, 2002.
- Dull, R.B., Graham, A.W. & Baldwin, A.A. Web-Based Financial Statements: Hypertext links to Footnotes and their Effects on Decisions. *International Journal of Accounting Information Systems*, 2003;4:185--2003.
- Elias, N. The Effects of Human Asset Statements on the Investment Decision. *Empirical Research in Accounting*, 1972; Supplement to the Accounting Review:215--233.
- Ettredge, M., Richardson, V.J. & Scholz, S. The Presentation of Financial Information at Corporate Web sites. *International Journal of Accounting Information Systems*, 2001;2:149--168.
- Ettredge, M., Richardson, V.J. & Scholz, S. Dissemination of Information for Investors at Corporate Web Sites. *Journal of Accounting and Public Policy*, 2002;21:357--369.
- F.A.S.B. *Electronic Distribution of Business Reporting Information*. US: Financial Accounting Standards Board, 2000

Fieber, J. *Browser Caching and Web Log Analysis*. 1997. Available at <http://ella.slis.indiana.edu/~jfieber/papers/bswla/bcwla.html>

Geerings, J., Bollen, L.H.H. & Hassink, H.F.D. Investor Relations on the Internet: a Survey of the Euronext Zone. *European Accounting Review*, 2003;12/3:567--579.

Gray, G.L. & Debreceeny, R. *Financial Reporting on the Internet*. 2001. Available at <http://www.ifac.org/news/speecharticle.tmpl?nid=979235133150990>.

Gutzman, A. *Analyzing Traffic on your E-Commerce Site*. 1999. Available at http://www.ecommerce.internet.com/news/insights/ectech/article/0,,9561_186011,00.html

Hansford, A. Hussey, J. & Hussey, R. *Corporate Communications: Shareholders' views of the Annual Review*. London: Deloitte & Touche, 1996.

Healy, P.M. & Palepu, K.G. Information Asymmetry, Corporate Disclosure and the Capital Markets: A Review of the Empirical Disclosure Literature. *Journal of Accounting & Economics*, 2001;31/1-3:405--440.

Hochheiser, H. & Shneiderman, B. Using Interactive Visualizations of WWW log data to Characterize Access Patterns and Inform Site Design. *Journal of the American Society for Information Science and Technology*, 2001;52/4:331--343.

Hodge, F.D. Hyperlinking Unaudited Information to Audited Financial Statements: Effects on Investor Judgements. *The Accounting Review*, 2001;76/4:675--691.

ICAEW. *Performance Measurement in the Digital Age*. London: Institute of Chartered Accountants in England & Wales, 1998.

ICAS. *Business Reporting: the Inevitable Change?* Edinburgh, Institute of Chartered Accountants in Scotland, 1999.

irbestpractice.org. *Setting the Standard for Best Practice in Online Investor Relations*. 2003. Available at www.irbestpractice.org.

Interactive Bureau. *FTSE 100's Web sites are still wallowing in mediocrity*. 2003. Available at www.iablondon.com/news/news.cfm?newsID=150

IFAC. *Financial Reporting on the Internet*. IFAC, 2002.

IRS. *A Brief Introduction to Investor Relations*. UK: Investor Relations Society, 2003.

Jones, D. The Five Drivers of Online Credibility. *IR Web Report* 2002; 6 Nov.

Jones, D. Don't Publish Financial Statements as Images. *IR Web Report* 2003; 3 Feb.

Jones, M., Xiao, J. & Lymer, A. *Trends in Internet Financial Reporting*. UK: ICAEW, 2001.

Lai, H & Yang, T. A System Architecture for Intelligent Browsing on the Web, *Decision Support Systems*, 2000;28:219--239.

Lee, T.A. & Tweedie, D.P. The Private Shareholder: His Sources of Financial Information and his Understanding of Reporting Practices, *Accounting & Business Research*, 1976;6/22:304--314.

Lee, T.A. & Tweedie, D.P. *The Private Shareholder and the Corporate Report*, London: ICAEW, 1977.

Lymer, A. & Debreceeny, R. The Auditor and Corporate Reporting on the Internet: Challenges and Institutional Responses, *International Journal of Auditing*, 2003;7/2:103--120.

Lymer, A., Debreceeny, R. Gray, G.L. & Rahman, A. *Business Reporting on the Internet*, IASC, 1999.

Marston, C. & Empson, J. *A Survey of Investor Relations among the Top 500 European Companies*. EAA Conference Paper, 2003.

MORI. *The City Opinion Survey*. 2003. Available at www.mori.com/products/city.shtml

National Investor Relations Institute. NIRI Releases Annual Report Survey, *NIRI Executive Alert* 2002;5 Aug.

National Investor Relations Institute. 2003. Available at <http://www.niri.org/about/index.cfm>

Neely, A., Adams, C. & Kennerley, M. *The Performance Prism*. UK: Pearson, 2002.

New York Stock Exchange. *Amendment No.1 to the NYSE's Corporate Governance Rule Proposals*. NYSE, 2003.

Nicholas, D., Huntington, P., Williams, P., Lievesley, N., Dobrowolski, T. & Withey, R. Developing and Testing Methods to Determine the Use of Web Sites: Case Study Newspapers, *ASLIB Proceedings*, 1999;51/5:144--154.

Nicholas, D., Huntington, P., Lievesley, N., Wasti, A. Evaluating consumer Web site logs: A case study of the Times / Sunday Times Web site, *Journal of Information Science*, 2000;26/6:399--411.

Nicholas, D., Huntington, P. & Williams, P. Evaluating Metrics for Comparing the Use of Web sites: A Case Study of 2 Consumer Health Web sites, *Journal of Information Science*, 2002; 28/1:63--75.

Nielson Norman Group. *Usability of the Investor Relations Area of Corporate Web sites*. 2003. Available at www.nngroup.com/reports/ir

Parker, L.D. Corporate Annual Reporting: A Mass Communication Perspective, *Accounting & Business Research*, 1982;12/48:279--286

Petravick, S. & Gillett, J. Financial Reporting on the World Wide Web, *Management Accounting (US)*, 1996;78/1:26--29.

Riphagen, J. & Kanfer, A. *In Search of the Elusive User: Gathering Information on Web Server Access*. NCSA, 1996. Available at <http://archive.ncsa.uiuc.edu/edu/trg/webstats/>

Rowbottom, N. The Application of Intangible Asset Accounting and Discretionary Policy choices in the UK Football Industry, *British Accounting Review*, 2002;34/4:335-357.

S.E.C. *Interpretation: Use of Electronic Media*. 2000. Available at <http://www.sec.gov/rules/interp/34-42728.htm>

S.E.C. *Regulation Fair Disclosure Revisited*. 2001. Available at <http://www.sec.gov/news/studies/regfdstudy.htm>

S.E.C. *Final rule (33-8230): Mandated Electronic Filing and Website Posting for Forms 3,4 and 5*. 2003. Available at <http://www.sec.gov/rules/final/33-8230.htm>

Tauscher, L. & Greenberg, S. How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *International Journal of Human Computer Studies* 1997;47:97--137.

Thelwall, M. Web Log File Analysis: Backlinks and Queries, *ASLIB Proceedings*, 2001;53/6:217--224.

Trites, G. *The Impact of Technology on Financial and Business Reporting*, Toronto: CICA, 1999.

Trites, G. *Business & Financial Reporting on the Web*. 2002. Available at <http://iago.stfx.ca/people/gtrites/webrep/webrep.htm>

Trueblood Committee Report. *Objectives of Financial Statements*. New York: AICPA, 1973.

Verrecchia, R.E. Essays on Disclosure. *Journal of Accounting and Economics*, 2001;32:97--180.

Weetman, P. & Beattie, A. *Corporate Communication: Views of Institutional Investors and Lenders*. Edinburgh: ICAS, 1999.

Zawitz, M.W. *Web Statistics - Measuring User Activity*. US Dept. of Justice, 1998. Available at www.ojp.usdoj.gov/bjs/pub/pdf/wsmua.pdf