

Evaluating the Online Activity and Searching Behaviour of Users of a Medical Digital Library

Gemma Madle
City University
Northampton Square
London EC1V 0HB
0044 207 040 4136

g.c.madle@city.ac.uk

Patty Kostkova
City University
Northampton Square
London EC1V 0HB
0044 207 040 4084

patty@soi.city.ac.uk

Jane Mani-Saada
City University
Northampton Square
London EC1V 0HB
0044 207 040 8391

j.mani-saada@city.ac.uk

ABSTRACT

The National electronic Library of Infection (NeLI) is a UK-based resource, providing free access to current best-available evidence in Infectious Disease for the public and professionals alike. The content of the NeLI is intended to be developed and maintained by an online community of professionals. The library is currently running as a pilot website and analysis of the behaviour of current users, in particular professionals in the field of infectious disease is necessary to inform development of the new NeLI website. A web log analysis was undertaken for the period January 2002 to June 2003. In particular use by hospital/National Health Service (NHS)-based users was evaluated. Results indicate an increase in activity during the period and an increase in the number of hospital/NHS-based users. Hospital/NHS-based users were more likely to return to the site, spend more time on the site and to view more pages than other users. In addition, a free text search facility was added in July 2003 and an analysis of the queries entered into this search has also informed further library development and will be evaluated in future research. Further research is now necessary to promote the library among professionals and develop the online community.

General Terms

Management, Measurement, Performance, Design, Reliability, Experimentation, Human Factors.

Keywords

Communicable Diseases, Libraries, Information Dissemination, Evidence-based Medicine, Guidelines, Evaluation Studies, Internet.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HDL'04, September 16, 2004, Bath, UK.

Copyright 2004 ACM 1-58113-000-0/00/0004...\$5.00.

1. INTRODUCTION

Medical Digital Libraries are essentially life critical applications enabling professionals to stay up to date. They have the potential to change working culture, creating new types of professional relationships and communities based across distances, hierarchy and other traditional barriers. Recent research indicated that some senior clinicians are reluctant to provide access to evidence-based information to junior staff, one clinician even suggesting that they should learn on the job instead![1]

The National electronic Library of Infection (NeLI) is a Specialist Library of the National electronic Library for Health (NeLH).[2] The NeLI is a freely available Internet gateway to the best available evidence. The key value of the library lies in the quality appraisal of documents by expert reviewers. The library has been operating as a pilot version for three years and is soon due to be re-launched. The new library will include discussion fora and development of an online community of professionals in communicable disease to provide and maintain the quality appraisals of documents.[3]

Web log analysis provides information about the path a user takes through a website.[4] We can find general patterns in use e.g. most commonly visited pages, search terms used, time spent on a page etc. We can also employ a technique known as microanalysis, analysing use of the library by a small number of individual users. This provides a clearer picture of individual user behaviour when in the library, rather than general trends. [5] Previous research has evaluated the impact of the NeLI on user knowledge and attitude. [6] Analysing the web logs of the library will allow us to evaluate current user behaviour, in particular that of the target group for the online community, health professionals. This will help ensure that development of the new website is in-line with user preferences.

In this paper we will discuss the analysis of the web logs of the pilot NeLI for the period January 2002 to June 2003. The next section outlines the methods used and the aim of the web log analysis. This is followed by the results and a discussion of the application of these results to future development of the NeLI and future research identified as a result of this study. This is followed by a brief conclusion.

2. METHODS

This evaluation was conducted by analysing access logs of the NeLI for the period January 2002 to June 2003. As already discussed, during this time the site was operating as a pilot version and the aim was to use the results of this study into user behaviour to inform further design, before launch of the new website.

The access logs provide data about visits to the website in the following fields: (1) The IP or hostname of the origin of the request, (2) date and time of the request, (3) the type of request, (4) the page requested, (5) the returned status of the page and (6) the number of bytes transferred. This raw data must first be cleaned e.g. a request from one user for one page can result in a number of hits as each picture file on the page is recorded as a hit. We also adjusted the time and date format to separate these and, because we were only interested in "real" user behaviour, we removed hits from identifiable robots.

The data was imported into Microsoft Excel, month by month, and analysed accordingly. IPs and hostnames were looked up for more specific details of the user e.g. domain and country of registration. Accesses were grouped into sessions with a new session starting with a new user (i.e. a new IP or hostname) or after 30 minutes of inactivity from a single IP or hostname address. Of particular interest was the identification of users accessing the NeLI from hospitals and other medical centres and organisations as these health professionals are the target community of the NeLI. These users will be described as hospital/NHS-based users from here on.

The key points to investigate were:

- General use of the library - no. of users, sessions and pages accessed and how these relate to events in the NeLI calendar and current communicable disease issues.
- Users - geographical distribution, those accessing from hospital environments, repeat users. All this information will aid us in the development of the online community around the NeLI.
- Pages - most commonly accessed pages, number of pages accessed by users each month, differences between all users and hospital/NHS-based users.
- Structured search/browsing behaviour - which is preferred by users?

3. RESULTS

The analysis of the web access logs provided information about who was using the site, how often it was used and what pages were most frequently viewed. We were also able to compare use of the electronic library by hospital/NHS-based users with use by all users. Some data for May 2003 is unavailable.

3.1 General use of the library

Figure 1 shows the number of different users accessing the site, the total number of sessions and the number of individual pages of the site accessed during each month of the period.

The number of users and sessions per month is closely related and has increased over the eighteen month period. The first peak in

October and November 2002 follows the presentation of the NeLI at two major UK conferences, the Public Health Laboratory Service (now the Health Protection Agency) annual conference (Sept 9-11) and the Federation of Infection Societies annual conference (27-29th Nov). The second peak occurs around the NeLI project board meeting (17th Mar). Dips are seen at holiday times (Christmas and Easter).

The number of pages accessed has more than doubled from 1396 in January 2002 to 2866 in June 2003. Dips and peaks mirror those of the number of users and sessions.

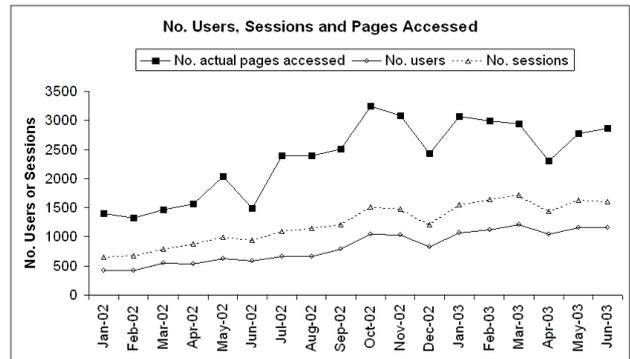


Figure 1. No. of users, sessions and pages accessed

3.2 NeLI Users

3.2.1 Geographical distribution of users

Figure 2 shows the distribution of IP registration of users across regions. Over two thirds of users (70%) are using IPs registered in the UK or US. As this is an English language UK-based library we would expect most users to be from the UK, so it is surprising that users are accessing the NeLI from 118 different countries. The large number of visitors from the US may be partly explained by the possibility of UK users accessing the Internet using US-based ISPs. This is partially supported by the most common time of access to NeLI being during the UK working day. We are currently analyzing the transaction logs from July 2003 to June 2004 and the IP lookup software we are using seems to be more effective at identifying UK addresses than previously so will hopefully produce more accurate estimates.

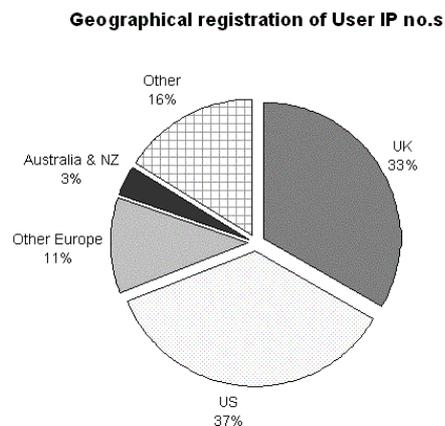


Figure 2. Geographical registration of user IP no.

3.2.2 Hospital/NHS-based users

Figure 3 shows the number of and percentage of users each month accessing the site from hospitals or NHS computers. The number of these users has increased and ranges from 14 in February 2002 to 33 in March 2003. Notable increases between July and August 2002, following the NeLI Advisory Board meeting and between February and March 2003 around the NeLI Project Board meeting (Mar 17). However, the percentage of total users each month accessing from hospital/NHS-based computers has decreased, most likely due to the increase in total number of users each month over the eighteen months. The number of users here seems quite low but there may be reasons for this. The limitations of matching IP addresses to individual users means that several users may be using the same IP address. Combined with unidentified access by health professionals from home or medical organisations/centres using ISPs the figures below probably significantly underestimate the actual number of health professionals using the site each month. In addition, a recent survey of a medical decision support tool, Isabel, found that despite over 800 registered users only 50 of these were core users, using it on average once a week, thus suggesting that the medical profession may need further persuading to adopt such tools into everyday work practices[7].

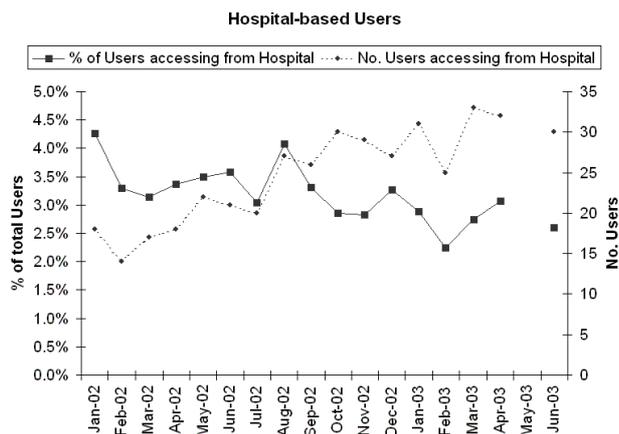


Figure 3. % of hospital/NHS-based users accessing NeLI

Hospital/NHS-based users are spending more time in a session and are returning to visit the site more often in a month than other users. The average time spent by hospital/NHS-based users in a session was 8 minutes and 23 seconds (standard deviation = 2 minutes 18 seconds) compared with 4 minutes 52 seconds (standard deviation = 2 minutes 48 seconds) for all users and hospital/NHS-based users visit the NeLI an average of 6.67 (standard deviation = 1.2) times per month compared with 1.52 (standard deviation = 0.1) times for all users.

3.2.3 Repeat users

The proportion of users returning to the website within each month has slightly decreased over the period with an average of 14.6% (standard deviation = 2). The decrease is probably due to the increase in users overall. For hospital/NHS-based users these figures are much higher. The average percentage of these users returning within a month is 56.2% (standard deviation = 11.8). This measure is not wholly reliable due to the difficulty resolving IP addresses to individual users.

3.3 Pages

3.3.1 Most commonly accessed pages

As would be expected, the most commonly accessed page is the home page, however we are more interested in which topics are most commonly accessed. We were only able to view topics accessed by those users browsing the site or identifying the topic pages directly rather than those using the structure search facility due to a lack of data collected by the server at this time. The search was previously using a cgi script and the keyword chosen was not recorded by the server. As a result of this study we have changed the structured search to consist of hyperlinks allowing capture of this keyword data (see section 4.5). The tables below therefore do not include data from searching NeLI.

Table 1. Top ten topics accessed by all users

Topic page	% all users
Meningitis	2.82%
HIV	2.72%
Tuberculosis	2.68%
Tinea	2.55%
Antimicrobial Resistance and Healthcare Associated Infection	1.93%
Staphylococcus aureus	1.85%
Chlamydia	1.83%
Parvovirus	1.73%
Helicobacter	1.69%
Salmonella	1.68%

Tables 1 and 2 compare the most commonly accessed topic pages of the NeLI by all users with those topics accessed by hospital/NHS-based users. A higher percentage of hospital/NHS-based users are accessing the topic pages of the NeLI than of all users. The six topics in bold are common to both groups.

Table 2. Top ten topics accessed by hospital/NHS-based users

Topic page	% all hospital/NHS-based users
Antimicrobial Resistance and Healthcare Associated Infection	10.27%
Tuberculosis	9.54%
Meningitis	9.47%
HIV	8.95%
Chlamydia	6.31%
E.coli	5.54%
Staphylococcus aureus	5.26%
Adenovirus	4.84%
Blood borne infections	4.44%

Asplenic Patients	4.12%
-------------------	-------

3.3.2 Changes in access to specific topics

There are indications of several patterns around use of specific topic pages of the NeLI. Around the time of wide media coverage of the outbreak of Legionnaire's Disease in the North of the UK, in August 2002, there was a notable increase in the number of users visiting the NeLI pages on Legionella. The number of users visiting the Influenza pages of the NeLI peaks in Autumn and Winter, reducing during the Spring and Summer months. There is also an increase in the number of users viewing the Chlamydia pages in May 2003. This may be linked to the publication at this time of research (highlighted on the NeLI) into the prevalence of the disease in Scotland.

3.3.3 Number of pages accessed per session

Figure 4 shows the average number of pages accessed in each session. The overall average was 2.11 for all users (standard deviation = 0.23) and 4.38 for hospital/NHS-based users (standard deviation = 4.56). The number of pages accessed by all users remains fairly consistent whilst for hospital/NHS-based users it varies. The graph shows an initial peak for the hospital based users perhaps due to first experiences with the library and users familiarizing themselves with its content. This figure of 2.11 sounds low but actually when taken in the context of the NeLI it is not so low. There will always be a group of users who accidentally visit the site and stay no longer than one page. It takes a user as little as 3 clicks from the home page to get to content in NeLI. Also only than 37% of sessions begin with the home page, 7% begin with a navigation page and 56% start with other pages including topic pages, and administration pages. This suggests that users may be bookmarking content or retrieving content directly from search engines rather than regularly entering via the home page, therefore reducing the number of pages required in each session to find the information they need. We do not have enough control over the server at this time to collect data on the referring page although this will be investigated in future analyses. Further qualitative research is now underway to investigate users' navigation techniques and how well NeLI meets information needs.

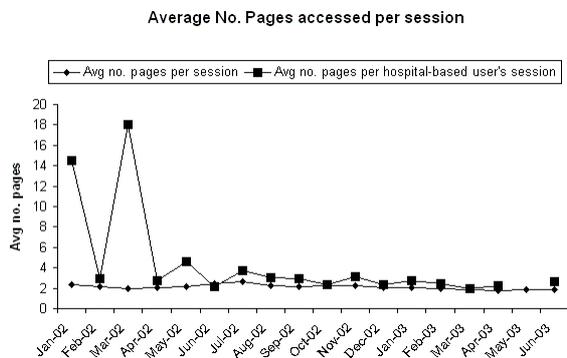


Figure 4. No. pages accessed per session

3.3.4 Structured search and browsing behaviour

Figure 5 shows the percentage of users completing a search on the NeLI i.e. when a results page is returned compared with the percentage browsing. As already discussed in section 3.3.3 a large proportion of sessions are not beginning with the home page or navigation pages, users going directly to content, therefore it may appear below that relatively low numbers of users are using the navigation tools.

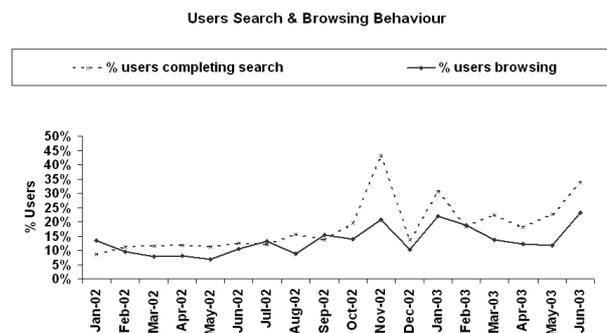


Figure 5. Comparison of use of structured search and browse

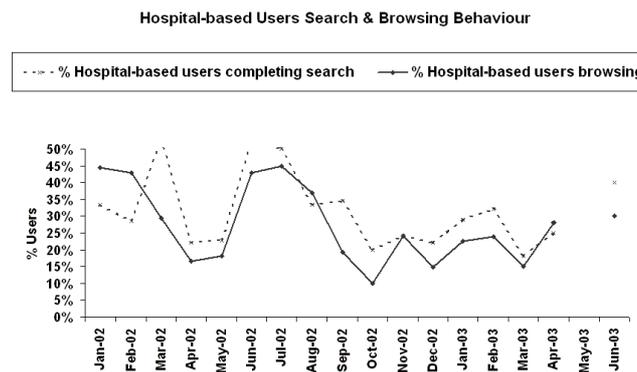


Figure 6. Hospital/NHS-based users search/browse behaviour

Figure 6 shows the percentage of hospital/NHS-based users completing a search compared with the percentage browsing by month. For both groups there is little difference between the proportion of users searching and browsing. There are notable increases in November 2002 and early 2003 in the percentage of all users searching. This matches increases in number of sessions, users and pages accessed (see figure 1). The peaks in use shown in figure 6 may be related to the number of pages accessed by hospital/NHS-based users each month (see figure 4). We would expect that if more users are performing searches then more pages would be viewed per user each month.

4. DISCUSSION

There are some key points to take from this evaluation to aid in the development of the new library website.

4.1 General Use of the Library

Use of the library has increased over the eighteen month period with more users, more sessions and more pages being viewed in

more recent months. This is probably partly due to an increased visibility of the NeLI on the world-wide-web over time e.g. being indexed by search engines. However, further promotion of the library is now essential if new users are to be attracted and if the library is to become a key resource in providing health professionals with current best-available evidence as and when they require it. Since completing this evaluation we have promoted the NeLI at the Federation of Infectious Societies Conference, the Health Protection Agency Conference, the European Congress of Clinical Microbiology and Infectious Diseases and at various smaller meetings around the UK. The current log evaluation underway will indicate the response to these promotions.

4.2 Users

Users are accessing the NeLI from a variety of countries although most access is from the UK or US. This is not unexpected due to the language of the library being English. There has been an increase in the number of hospital/NHS-based users over the period but the number of these users is still relatively small. Users accessing from hospitals are spending longer on the site and are more likely to return to the site within the month than other users. Work needs to be done to promote the library among health professionals and encourage the development of an online community to develop and maintain the content of the library. Community building tools can be attached to the library such as discussion areas and personalization tools. The challenges of developing and maintaining online communities are discussed in a related paper[8].

4.3 Pages

Hospital/NHS-based users are accessing more pages per month than other users. This is partly explained by the higher percentage of repeat users among the hospital/NHS-based group. There is variation in the number of pages accessed by hospital/NHS-based users across the period. Possible reasons for decreases in number of pages accessed by hospital/NHS-based users per session could be familiarity of regular users with the website. As users become more familiar with the NeLI they need to view less pages in order to find what they require, perhaps bookmarking relevant pages rather than entering via the home page and clicking through submenus. Alternatively an increase in new users (see figure 3 for details of the general increase in the number of hospital/NHS-based users during this period) could either reduce the average number of pages viewed per user as they sample just a few pages of the site in their first few visits, or increase the average as they familiarise themselves with the site, viewing lots of pages per visit.

There is little difference in the most commonly viewed topics between all users and hospital/NHS-based users. However, hospital/NHS-based users are more likely to view topic pages than other users. Possible links between communicable disease issues in the news and activity on the NeLI site were identified. More detailed analysis is necessary to confirm this.

4.4 Problems with web log analysis

There are several problems associated with web log analysis. Crawlers and spiders can distort patterns of use, it is difficult to know how exactly where users are accessing the site from as repeated IP numbers do not necessarily represent the same user

and one user may access the site using more than one IP number. The time spent on the last page of the site is not recorded, use of the back and forward browser buttons is not recorded and many ISPs use proxy servers to retrieve web pages for their users. However, we are able to gain an overall picture of use of the NeLI from the data collected and future evaluations can attempt to address some of these problems to improve the quality of data collected.

4.5 Future Research

This evaluation has identified many areas for further research. More log data needs to be collected in future, for example, referring websites and search keywords. More recently (July 2003) we began to collect information about the keywords used in the full-text search and we will be able to compare this to the most common pages viewed by browsing in the previous months. Collection of query data has already identified topics that the NeLI should be covering and helped to inform development of a navigation tree for the new website. We have also adapted the structured search page so that keyword data that was previously unavailable, can now be captured from user activity. Log data can be linked to qualitative data collected from users e.g. place of work, job title, area of expertise, self-reported impact of the NeLI on the user's work, changes in knowledge and attitudes of users' after using the NeLI. This qualitative data can help to fill in the gaps identified by log analysis. We can identify from which country people are accessing and how often they return to the NeLI, two areas that are left unclear following this evaluation.

By analysing patterns of use by different types of users and using microanalysis techniques as discussed earlier we can profile different members of the online community to help develop and extend this community. We can also personalise the site through user self-selection areas of interest but also potentially automatically, matching patterns of previous user behaviour to current user activity.

5. CONCLUSION

This paper has shown how people were using the National electronic Library for Infection between January 2002 and June 2003. Web logs were analysed to provide an overview of use of the NeLI and were compared with use by hospital/NHS-based users. Results indicate that in general, activity on the library website has increased but qualitative evaluation and further promotion of the library is necessary to increase use and develop the online community that will help to provide and maintain the library content. This study provides a platform for further research into the development of the NeLI.

6. REFERENCES

- [1] Adams A, Blandford A, Acceptability of Medical Digital Libraries, *Health Informatics Journal*, v8:no.2 (2002), pp58-66
- [2] Gray M, de Lusignan S, National electronic Library for Health, *BMJ*, vol319 (1999), pp.1476-1479
- [3] Kostkova P, Mani-Saada J, Weinberg J, Agent-based Up-to-date Data Management in National electronic Library for Communicable Disease. In the Proceedings of Agents

applied in Health Care, 15th European Conference on Artificial Intelligence, ECAI-2002, Lyon, France: 2002.

- [4] Rozic-Hristovski A, Hristovski D, Todorovski L, Institute J. S, Users' information seeking behaviour on a medical library website, *Journal of the Medical Library Association*, v90: no.2 (2002), pp210-17
- [5] Nicholas D, Huntington P, Williams P, Micro-mining log files: a method for enriching the data yield from Internet log files *Journal of Information Science*, v29; no.5 (2003), pp391-404
- [6] Madle G, Kostkova P, Mani-Saada J, Weinberg JR. Evaluating the changes in knowledge and attitudes of digital library users, In the proceedings of the 7th European conference on Research and advanced technology for digital libraries - ECDL 2003, Trondheim, Norway, Koch T, Torvik Sølvsberg I, eds. *Lecture Notes in Computer Science*, Berlin: Springer Verlag, 2003
- [7] Briggs JS, Fitch CJ, The Isabel User Survey, May 2003. Available at: <http://WWW.disco.port.ac.uk/hcc/pubs/ISABEL/report.pdf>
- [8] Mani-Saada, J, Madle G, Williams P. Initial experience with developing communities of practice around the National electronic Library for Communicable Disease. In the proceedings of the 1st Healthcare Digital Libraries Workshop (HDL 03) at ECDL 2003, Trondheim, Norway.