Usability evaluation of an NHS library website
Catherine Ebenezer, Camberwell, London, UK

Abstract

Objectives: To carry out a usability evaluation of the recently launched South London and Maudsley NHS Trust library website.
Methods: A variety of standard methodologies were employed: content and design evaluation of selected comparable sites, focus groups, a questionnaire survey of library and Web development staff, heuristic evaluation, observation testing, card sorting/cluster analysis, and label intuitiveness/category membership testing. All test participants were staff of or providers of services to the trust. Demographic information was recorded for each participant.
Results: Test participants’ overall responses to the site were enthusiastic and favourable, indicating the scope and content of the site to be broadly appropriate to the user group. Testers made numerous suggestions for new content. Usability problems were discovered in two main areas: in the organization of the site, and in the terminology used to refer to information services and sources. Based on test results, proposals for a revised menu structure, improved accessibility, and changes to the terminology used within the site are presented.
Conclusion: Usability evaluation methods, appropriately scaled, can be advantageously applied to NHS library websites by an individual Web editor working alone.

Introduction

Web-based information has become increasingly important in health. There is by now a large body of professional literature in the mental health, community health and social welfare fields relating to Web-based resources and their importance for the development of evidence-based practice.1–8 The Department of Health and other statutory bodies use the Web extensively as a vehicle for official communications. There are also several journals devoted to health informatics and to health information on the Internet. Web pages are now easily produced, but there is a pressing need to evaluate whether the pages are meeting the users’ needs. The evaluation process tends to be ignored in the production of web pages, but it should be core. The purpose of this study was to carry out a usability evaluation of the recently launched South London and Maudsley NHS Trust library website. The scope of the evaluation included, not only design and navigation (the narrower meaning of usability), but also the suitability of the site’s content for its intended purpose and readership.

Background

The South London and Maudsley NHS Trust
During the spring and summer of 2001, a website was built for the South London and Maudsley NHS Trust (SLAM) Multidisciplinary Library in Stockwell, London SW9. The site was officially launched and publicised in late June, with the
The Trust is a large, geographically dispersed organization covering four London boroughs (Lambeth, Southwark, Lewisham and Croydon), with about 4500 staff based at over 100 sites. Mental health and social services are now fully integrated within the borough of Lambeth, and the library is contracted to provide an information service to the staff of Lambeth Social Services. The library also provides a service to the local community health trust, Community Health South London NHS Trust (CHSL). While staff have access to all the King’s College libraries and to the Lewisham Hospital library, SLAM Multidisciplinary Library is the only library managed directly by the Trust, and hence has an important role in the provision of and co-ordination of information services to SLAM. The staff it serves vary widely in educational level, in professional background, and in information and computer literacy.

Health professionals’ access to and use of the Web

During the data collection period for this evaluation, web connectivity within SLAM was very poor. Only about 120 PCs had access to the NHSnet, with most staff having to depend on external access from home, or via NHS, academic or local authority libraries. There is every indication that this connectivity level has improved significantly since then. Web connectivity in CHSL was much better, with most clinical teams having access at their team bases. In Lambeth Social Services, web connectivity within the workplace was available to about 30% of staff.

Anecdotal evidence (conversations between library staff and readers) suggests that many readers (estimated > 50% and increasing steadily) have Internet access at home\(^9\), but that their ability to search for information effectively is limited. A website can be useful also only insofar as it fits with, and relates to, existing patterns of information seeking and use among its intended users; appropriate training and support also needs to be provided to them in respect of information sources and services.\(^10\)

The library website: origins and rationale

Palmer, in her study of information for community mental health workers,\(^11\) suggested the development of ‘signposts’ to information as a method of improving awareness and use of information resources. A signpost should aim to function as:

- a guide to the availability of sources and services of learning resource, library and information providers;
- a way to inform people of their rights to access to these sources and services;
- a channel through which users could be directed to the appropriate initial point of access to resources;
- a device to educate and develop users as independent learning resource users.

Thus, a full-scale library website might be able to function as an ‘electronic signpost’ in this way, and also:

- publicise the library and its services beyond its historical user base and immediate geographical catchment area (Lambeth and North Southwark);
- provide considerably enhanced access to library services and other electronic sources from home for staff who are geographically isolated or who have difficulty accessing conventional library services during their working day;
- provide a window to selected, quality-filtered information geared to the specific needs and interests of practitioners working in mental health, and in community and primary care, which is not readily available elsewhere;
- offer an effective, customised means of access to the large variety of electronic journals which the library provides;
- make available internal resources and products, such as bibliographies and search guides.\(^12,13\)

The initial design of the SLAM library site was based largely on intuition and first-hand knowledge of the intended user group, with some reference to the websites of similar libraries. Some content was drawn from existing material, such as guides to services, forms, new book announcements, and lists of web-based resources; the rest was created de novo.

---

\(^9\) Since the writing of this paper, three PCT’s have taken the place of CHSL. These are Lambeth, Southwark and Lewisham.
Objectives of the usability study

In developing the website a literature review was carried out and a number of references covering: website management, design, content, evaluation usability and navigation were retrieved. These are listed on the author’s web page. All these references were studied in detail and evaluated for use and applicability to this site.

Some further literature on usability was also considered. According to Fowler, ‘usability is the degree to which a user can successfully learn and use a product to achieve a goal.’ It is often assessed in terms of a range of aspects: ease of learning, retention of learning over time, speed of task completion, error rate, and subjective user satisfaction. A web designer aims to create a site that is useful (enables users to achieve their particular ends and meets their needs), easy to use (enables users to move around the site rapidly and with few errors), visually attractive, and popular. A site should be user centred, in that it is based on knowledge of the site’s users, in particular their technological and physical capacities, their cultural context, and their information needs.

A brief usability evaluation of the site was carried out as a means both of evaluating the appropriateness of its design and content, and of establishing what modifications might be necessary. It is suggested by Veldof and by Marmion that information on library websites is typically identified and structured in a librarian-orientated fashion which does not accord with the users’ needs, preconceptions and mental maps; according to Gullikson et al., their organizing principle can be inappropriate. Matylonek identifies five typical sources of bias for the creators of websites:

- discipline jargon—professional terminology not understood by users;
- hierarchical bias—confusion of the organization’s structure with customer services;
- expert proficiency: design of layout that presumes comfort in a web environment;
- ‘folk classification’: naturally preferred terminology sets among various cultures;
- preferred and novel services: developers often use certain services and over-emphasize them in their designs.

Methods

This project aimed, via a combination of appropriate usability testing methodologies, to answer the questions:

- Is the site:
  - readily intelligible, i.e. not confusing to the reader;
  - intuitive and easy to navigate (with respect to overall structure, navigation, labelling, searching/browsing, general features);
  - visually attractive;
  - consistent in design and terminology?
- Are the readers able readily to locate information about library services?
- Are the readers readily able to locate through it the sources of mental health and community health information they need? Is its scope and content, as far as is possible, adequate and appropriate to the needs of the readership?
- Does the way in which it presents information about library resources accord with the reader’s mental maps? In particular, is the division clear between locally networked and web-based resources?

Overall, the aim was to generate a set of evidence-based proposals for redesigning or modifying the site, not focusing narrowly on design and navigation aspects of usability, but going some way towards assessing the usefulness, value and appropriateness of the site content in relation to the perceived role of the library within the Trust.

Nielsen and Landauer in 1993 put forward the view that the best results in usability evaluations come from carrying out as many small tests as possible. While their conclusions have been disputed this work involved several phases and combined several different methodological approaches. In the preliminary phase, a succinct content and design checklist was developed and used on a selection of websites of NHS libraries similar to SLAM Multidisciplinary Library, as a benchmarking and evaluation tool and as a source of new ideas. Focus group meetings with different groups of staff were also conducted in order to identify key issues from the users’ perspective, to gather ideas for development of the site, and to determine if there were any issues requiring more detailed investigation.
The main phase of the project consisted of formal observation testing, card sorting, and a combined label intuitiveness and category membership test.

In the final phase of the project, proposals were put forward, based on the findings, for revisions of the site. The activities of the preliminary phase cannot be described here in detail for reasons of space. A complete account of the project is available from the author.

Thirty-two participants were recruited to act as testers for the preliminary and main phases via Trust-wide e-mail within SLAM and through personal contacts. It was not, unfortunately, possible to recruit CHSL or Lambeth Social Services participants via e-mail circulars, as access to these organizations' address books was not available. Despite the provision of an incentive (the offer of a free lunch in the staff canteen), it proved difficult to recruit participants for activities which required extended time to be spent in the library, such as the focus groups and usability tests. In practice, most of those who took part were based on site or in premises nearby.

Demographic information was recorded for each tester, as recommended by Davis. Among the volunteers there was a preponderance of medical staff and of professional non-clinical staff; also there were considerably more women than men. It proved impossible to recruit participants from certain groups (social workers, health care assistants). There was, however, a reasonable spread of backgrounds and a wide variation in levels of information and computer literacy. Only three participants reported making extensive use of the site before taking part in usability testing activity.

1 A small sample of six NHS library sites was selected for the benchmarking/content evaluation. The libraries were chosen deliberately for the range of approaches to navigation and design they represented, and the range of their online content.

2 Nine participants in all were recruited for three separate focus groups run at lunch times; each group had three members. All were given about 15 min to ‘play’ with the site before each of the sessions. Each group was facilitated by the author and lasted about 45 min.

3 Seven testers were recruited for the observation test. At the start of each test the participants were given a script and list of tasks. The 15 tasks, some of which had a number of different components, were designed to address anticipated usability problems. The usability metrics derived were: percentage of tasks completed, number of false starts for each task, longest time taken for each task, number of prompts required per task per user, and user satisfaction ratings. Volunteers for the card-sorting test were recruited via a Trust-wide e-mail. Sets of paper slips were created, one slip for each item on each of the menus. Menu category headings were also included among the slips. Subjects were asked to sort the slips into categories, using either one of the menu headings as a label for the category, or devising their own heading if they preferred. The cluster analysis software USort/EZSort, as described by Dong and co-workers, was used to record and analyse the results.

4 Respondents were asked to complete a detailed label intuitiveness/category membership questionnaire. This provided screen shots illustrating the main menu and sub-menus; respondents were asked what they would expect to be included in each main category, and what sort of information they thought each of the links would indicate.

Results and discussion

Content and design evaluation

The content evaluation was not intended to be comprehensive, merely indicative of current practice and trends in the sites of the libraries that were selected.

There is considerable variation in the primary and additional navigation systems used, in the categories of information given on the home page, in scope and content, and in the services provided. Sites tend to group information about library services on a single long page. Most of the sites make little use of interactive features, typically providing only one or two; these, however, varied considerably. Glenfield has a message board, a chat room, and online polls. Exeter has an online membership registration form, while Brighton has online book and journal article request forms.
Typically, print and electronic journals are listed together on one indexed alphabetical list; only one site (Exeter) had separate lists for print and electronic formats. Lists of electronic journals typically include links to the relevant publisher’s or aggregator’s website. One site (Chichester) did not list any journals. Provision of book catalogue access varied: one site provided a link to a university Web OPAC, two others to a consortium union catalogue, another provided access only within the Trust network to the catalogue, and the others had no access at all. Databases available within the library were generally listed and described in varying levels of detail under the general heading of ‘databases’. Direct log-in facilities were provided where applicable.

Sites did not generally have any significant uniquely developed content. One exception was the Knowledgeshare clinical knowledge management website at Brighton. Several, however, had put together extremely useful lists of resources not found elsewhere, e.g. MCQs (Brighton), library training guides (Glenfield), electronic medical textbooks (Exeter), guides to health statistics and sources for tracing practice guidelines (Exeter). All the sites except Chichester provided some selected links to external websites. The Exeter and Glenfield lists are comprehensive and highly developed.

Focus groups

There were some things focus group members liked:

- ‘Library on the desktop’ aspect.
- Navigation and design: clear and intuitive use of frames, one item per page, little scrolling needed, page content fits well within frame body.
- Local map.
- Union catalogues.
- Use of language.
- Links to web logs—these are very useful as they incorporate information circulated via e-mail into a conveniently accessible format.
- Principle of selecting key sites rather than having long comprehensive lists.

Some things they disliked:

- URL too long and cumbersome.
- Date on introductory page in American format.
- Lack of full web OPAC functionality.
- Not clear what is password protected and what is not.
- ‘Search this site’ page is much too plain; would like prompts on search techniques, and examples. Focus group members also gave suggestions for development and additional content:
  - Union catalogue of serials across SLAM libraries.
  - List of professions allied to medicine links.
  - Update clinical governance material to include links to CHI and NICE.
  - Additional specialist bibliographies of relevance to Trust, e.g. mental health informatics, cognitive-behavioural therapy, transcultural psychiatry.
  - Register of Trust research.
  - Database of clinical audits.
  - More community health information, especially community profiling.
  - More on boroughs.

According to the canons of focus group methodology, the groups were (a) too small, and (b) should not have been conducted by the site designer. In this instance, however, combining the two roles in the same person meant that the facilitator knew the participants and had a close knowledge of the site and of the library service, hence, could set individual comments in context within the discussion.

Observation test

The results of this test are analysed, see Appendix 1.

A learning effect was apparent as the test was conducted; as they worked through the tasks, the testers learnt their way around the site, and by the end their comments often indicated that they were able to find things relatively easily. Some basic gaps in people’s knowledge of basic browser functionality became apparent, even with experienced searchers: one of these was not aware of the Find feature available in Internet Explorer 5’s Edit menu. Some testers would go straight to ‘site search’ to locate an item of information, whereas others clearly preferred to work their way through the menus. The ‘menu browsers’ among the testers often went to the more general category headings ‘facilities for readers’ or ‘general information’ if uncertain where to look for something. Twelve failures to complete tasks and 15 prompts were
recorded out of a total of 126 test events. The average unprompted completion rate per participant was 92%.

Testers rated the site highly on comprehensibility of terminology, ease of use and likelihood of subsequent use, somewhat less highly on structure and organization.

The test results were difficult to record with total accuracy; it was difficult to take notes while conducting the tests, and the tapes that had been recorded could not always be deciphered. Again, it was not good practice that the author of the site should have been conducting and recording the tests.\(^{36}\)

It was impossible to refrain entirely from prompting testers. These tests were successful, however, in identifying a number of significant usability issues.

### Card sorting test

Card sorting and cluster analysis is a usability evaluation method that is often adopted early on in the design of a site. The rationale for using it in this project was that it was felt to provide a clear idea of users’ categorization and association of information sources, this providing a corrective to the (possibly) librarian-centred organization of the site as it currently exists.

The ‘hard’ end, which emphasize the quantitative data that is the subject of cluster analysis.\(^{34,37}\)

The method adopted in this study of administering the test by post and using EZSort to analyse and graphically represent results was of the ‘hard’ variety. It had evident advantages in terms of time and convenience, but precluded any informative contact on my part with the subjects. It became apparent that the results were being affected by user uncertainty caused by lack of intuitiveness of the item labels; with hindsight, fuller descriptions should have been given of the item contents.

The following results were evident across all three charts. ‘Interlibrary loans’ is associated strongly with ‘other libraries’ and with ‘South Thames Libraries’ rather than with journal or book categories. ‘Photocopying’, ‘telephones’ and ‘refreshments’ are only loosely associated with other library facilities. ‘Computing’ is associated with online search facilities, rather than with general library facilities. Participants appeared to have difficulty classifying ‘our holdings’; it does not convey a clear meaning apart from the context of information about books. ‘Current awareness’ is associated with ‘using the literature’ and with ‘subject guides’ and ‘search requests. Of the 13 people who initially volunteered to take part, only seven returned results. Card sorting is considered to be more effective and accurate with 20 users or more. The results did, however, provide some clear pointers for restructuring the menu system.

### Category membership/label intuitiveness test

Eight completed questionnaires were received. This test revealed significant lack of clarity around the main category headings ‘general information’ and ‘facilities for readers’. Four respondents expected that ‘facilities for readers’ would include book loans, three mentioned interlibrary loans, two mention user support, two mentioned journals, two mentioned Internet access, and one mentioned other libraries within the trust. Under the heading ‘general information’, one respondent mentioned other libraries, two mentioned books and book loans, and one mentioned journals. Both headings evidently tend to be interpreted as referring comprehensively to all aspects of the library service. Three respondents expected ‘journals’ to include information on literature searching.

Some of the item labels also appeared to be ambiguous or problematic. With ‘links’, three people expected this to cover specifically links to trust sites, three expected it to provide links to other libraries, and one assumed it would lead to basic contact information. The heading ‘subject guides’ led people to expect more general guides to subject information, rather than guides specifically to web-based resources. 'ATHENS' mystified people who were not familiar with the ATHENS authentication system for databases and journals. People seemed to be unsure of what ‘current awareness’ might refer to. With ‘current titles’, which refers to journals currently subscribed to, three respondents expected this to include information on new books, and one expected it to lead to a book recommendation form. Three people expected ‘our holdings’, which refers to books, to lead to information about journals and other
library materials as well. Three respondents highlighted an ambiguity in ‘recommend a book’; it was interpreted as meaning ‘recommendations by the library’, ‘reviews’, ‘other readers’ recommendations’ as well as an invitation to suggest a book for purchase. ‘Site search’ led two respondents to expect a web search facility, rather than a facility for searching the site; this label was actually amended in the course of testing to ‘search this site’. Three people professed not to understand the meaning of ‘locally served databases (CD-ROM)’, and four were evidently unsure of the meaning of ‘commercial online databases’ (by which had been meant fee-based services available via an online host). The other significant labelling ambiguity seemed to be occurring with ‘request a search’; although this actually leads to a search request form for library staff to carry out a mediated search, four people expected this to lead to information on carrying out their own literature searches, or directly to search facilities mounted on the site.

The test also highlighted some expectations for content which are not currently available. Under ‘book collections and loans’, six respondents expected self-service Web OPAC circulation functionality to be available. One expected there to be a facility to search their own interlibrary loan records online, while another mentioned the desirability of an integrated union catalogue and interlibrary loan request facility.

As a result of usability testing, detailed lists of the proposed changes to the site were drawn up and a new structure for the website proposed. The changes covered the following areas: navigation/readability, HTML validity, content and language.

Conclusion

Relatively few people were involved in each of these tests. It is a commonplace of the usability testing literature that useful and valid results can be obtained with eight users or less; indeed that 80% of usability problems will become apparent with five users.\(^{38-39}\) It became clear, particularly in the formal observation testing, that later testers were identifying a high proportion of repeat, rather than new, problems. This suggests (pace Spool\(^ {40}\)) that, although additional issues would probably have been identified with a larger group of testers, the small sample sizes do not seriously invalidate the results as they stand.

It appeared that the main usability issues had been correctly anticipated.

The major problems encountered by the testers appeared to involve two main areas: (a) the specialized terminology used in referring to information sources and services, and (b) the organization and structure of some of the information about library services.

According to France,\(^ {41}\) library users can be characterized as ‘chronic beginners’; this, he says, is in large part due to their uneven demand for library services, which can involve periods of intense research separated by long gaps. This presents particular problems of usability for library applications, in that users may combine domain expertise, and serious and sophisticated information needs, with considerable naïveté in information-seeking behaviour.

Many researchers have highlighted the classification of information systems, and the labelling of resulting categories, as a problem of information design generally and of web information services in particular; a review of the literature is provided by McGillis and Toms.\(^ {42}\) With respect to libraries, Spivey\(^ {43}\) discusses in detail the problem of how librarians should communicate on the Web with people outside the profession, and with an increasing number of remote end-users who may have diverse cultural backgrounds and little previous exposure to academic information seeking.

According to Spivey, experienced library users become familiar with library jargon, but can be confused by new systems and terminology, or by the availability of multiple platforms and interfaces for a single resource (e.g. MEDLINE), leading to frustration and a sense of helplessness. Library jargon can include: short descriptions and nouns for library resources and services, e.g. ‘circulation’, ‘ATHENS’; library acronyms (OPAC, ILL); vendors’ trade names (e.g. SilverPlatter, OVID); and what he terms ‘embedded explanations’, such as phrases in apposition, examples or descriptions, category headings, or prepositional phrases. He found that terms such as ‘reference’, ‘reserves’, ‘indexes’, ‘citations’, as well as more obvious jargon such as ‘proximity operators’, ‘implicit
Boolean’, ‘user authentication’, were obstacles to readers\textsuperscript{44}. Naismith and Stein\textsuperscript{45} in their detailed study of student comprehension of technical language used by librarians, found that readers misunderstood library terms in reference interviews and library handouts about 50\% of the time. Unsurprisingly, given the rapid changes taking place in the information market, most readers do not have in their minds a clear taxonomy of electronic information sources.

Naismith and Stein suggest that a continuum of strategies, such as the use of explanatory phrasing, the provision of glossaries, etc. should be employed in written and verbal communication to bridge the gaps in understanding. These have obvious application to library website design. A library website needs to be seen in the context of the library’s user education and support strategy, and indeed of its ‘information architecture’ as a whole, with appropriate guidance provided to users on identifying appropriate information tools and sources.

Regarding the provision of guides to web resources, this study had sought to answer the pressing question, ‘how much is too much?’ While librarians frequently compile web resource guides, perceiving them to be an extension of library service functions,\textsuperscript{46} there are evident limits to what can reasonably be included on the website of a small specialist library without undue duplication of effort. Interestingly, the participants in the study appeared, in their own use of the Web, to be adopting an ‘anchor strategy’ of making regular use of a few, authoritative sites, as observed by Westberg and Miller.\textsuperscript{47}

Participants in the focus group discussions appeared to think that, for them, the key role of a library website in relation to external web-based resources is not only to act as a form of quality filter, but also to provide readers with jumping-off points for their information seeking. The focus group participants emphasized, as well, the value of local content or content of immediate local relevance; this appears to be an appropriate niche for the SLAM Multidisciplinary Library site.

Usability testing, being limited to what can be readily observed and measured, is necessarily somewhat artificial. It is also limited in the type of questions it can answer; it is unrealistic to expect small-scale usability testing to answer questions about a site’s overall quality and effectiveness\textsuperscript{48} or to establish objective standards of usability.\textsuperscript{42} This project identified a number of significant usability problems within the site; it also afforded, to a limited extent, the opportunity of evaluating emendations. The questionnaire and demographic data obtained in the study gave indications of existing habits of professional information seeking on the web. Focus group data and testers’ comments provided indications of the potential value and usefulness of the site to professional staff within the trust, and yielded many valuable suggestions for improving it and developing it further.

References

9 Anthony, D. The Internet ‘haves’ and ‘have nots’ in nursing. \textit{Presentation given at Healthcare Computing Conference, Harrogate, March 20–22, 2000.}