User-Centered Evaluation Model for Medical Digital Libraries

Patty Kostkova and Gemma Madle

City eHealth Research Centre (CeRC), City University
Northampton Square, London, UK
`patty@soi.city.ac.uk`

Abstract. It remains unclear whether the recent explosion of medical Internet Digital Libraries (DLs), enabled by substantial investments into eHealth by national governments and international agencies, has brought the desired improvements. As ultimately life-critical applications, medical DLs play a crucial role in delivering evidence to professionals and empowering patients. How- ever, little attention has been given to impact evaluation with domain experts in real settings to assess whether they actually make a difference to clinical practice.

In this paper we describe a novel evaluation framework – Impact-ED – developed at CeRC to fill the gap in impact evaluation research taking into ac- count the community, content, services and technology dimensions of DLs. We present an account of Impact-ED's application in assessing the impact of the National Resource for Infection Control in the UK (NRIC www.nric.org.uk) – a real-world medical DL used by over 40 000 professionals monthly.

Keywords: Medical Digital Library, impact evaluation, Weblogs, infection control, searches.

1   Introduction

Substantial budgets have recently been spent on eHealth programmes by national governments aiming to change the way healthcare is delivered in the 21st century [1,2]. The Internet has the potential to instantly disseminate the best available evidence, create communities of practice for professionals in widely dispersed geographical locations or for patients with rare conditions, and provide a quality-assured educational vehicle to improve the wellbeing of European and global citizens. As June Forkner- Dunn foresees: "the impact of the Internet has largely been unforeseen, and it may have a revolutionary role in retooling the trillion-dollar health care industry in the United States" [3].

However, with the growing popularity of general search engines there is an increasing need for quality-assured digital library collections which are compiled and kept up to date by experienced medical domain experts, enabling customization, personalization, and profiling of services [4]. Medical Internet DLs, ultimately life-critical applications, play a crucial role in delivering medical evidence to healthcare professionals and

information to the public [4]. However, is their potential fully exploited and their impact realistically evaluated?

This paper presents a novel DL impact evaluation model, the Impact-ED, and illustrates its implementation on a real-world DL, the National Resource for Infection Control (NRIC, www.nric.org.uk), hosted by the National electronic Library of Infection (NeLI, www.neli.org.uk) at the City eHealth Research Centre, City University in the UK.

Applying Impact-ED, NRIC was evaluated as a case study with infection control professionals in the UK using a combination of pre- and post-visit questionnaires, study beginning and end questionnaires, web server logs, and structured interviews.

The paper is organized as follows: Section 2 brings a background to the National Resource for Infection Control (NRIC) and outlines NRIC users' search and information needs. Section 3 discusses an overview of DL evaluation and details the Impact-ED model. In Section 4, we illustrate an application of Impact-ED on NRIC as a use case bringing detailed results while Section 5 presents conclusions.

2 National Resource for Infection Control (NRIC): Background, Web Server Traffic and Users' Information Needs

The National Resource for Infection Control (NRIC) was launched in May 2005 in response to National Audit Office [5,6] (2000/04) recommendations for a national infection control manual. The project, funded by the Department of Health in the UK and endorsed by the UK National electronic Library of Infection (www.neli.org.uk), covers a broad range of infection prevention/control and infectious diseases information and attracts over 40 000 healthcare professionals monthly.

In order to better understand users' information needs, NRIC's web server logs and traffic are evaluated on a monthly basis. In addition to basic statistics, a key area of interest is investigation of users' search terms (both internal keywords searched on the NRIC site using a built-in search engine, and external searches bringing users to NRIC from other search engines). Search keywords are a key mechanism for capturing user information needs, and feed directly into the NRIC content strategy plan.

However, do we understand users' information seeking needs and their underlying online behaviour? While there are a number of definitions of user information seeking and searching behaviour, for our purposes we shall use T. Wilson's definition of the term *information searching behavior*: all user activity on the website with the purpose of finding certain information, as opposed to "surfing" the website without a prior information need [7].

Essential indicators of the probability of knowledge discovery and overall user satisfaction and site usage are determined by

(i)   whether users use the site the way the site designers expect
(ii)   whether they understand the terminology used for site searching
(iii)   how they navigate the site to find what they are looking for

This is of particular concern in the healthcare domain, since failure to locate relevant information, or, worse, the location and use of outdated or poor quality information, can

e

have serious consequences. Navigation can be measured by so-called "disorientation" [8] – feeling lost within the website space. Disorientation can be caused by complexity of site navigation (browsing and searching access points), unclear terminology, and poor knowledge of the domain [9].

The following section 2.1 presents a set of search results from NRIC web server logs conducted in 2008, comparing Top 20 external searches bringing users to NRIC from search engines to Top 20 internal searches performed using the NRIC internal search engine. Monthly reports providing general statistics, including geographical locations, times and dates, navigation pathways, and other important information can be found on the NRIC website[1].

2.1 NRIC Search Results

Internal search phrases are phrases input by users into the NRIC search box and processed by NRIC's built-in search engine. These are invaluable resources for understanding the information needs of users who visited NRIC (either directly or via another site or a search engine) because of a need for information about infection control.

One of the key indicators of user search behaviour is referral. External search phrases bringing people to NRIC from search engines are equally important for un- derstanding users' needs, although their relevance relies on NRIC ranking and search engine indexing. Of all the pages viewed on NRIC, 56% were viewed by users com- ing from other NRIC pages, 27% from Google (google.co.uk and google.com), 0.5% from NeLI, 0.2% from traininginfection.org.uk, 0.2% from the website of the UK's Department of Health, 0.2% from www.infectioncontrol.nhs.uk, and the remainder from other websites/search engines.

External Search engine keywords in 2008. NRIC received 78 286 page views from search engines which directed users to NRIC. This is a percentage increase of 93% from 2007. The following chart, Figure 1, shows the Top 20 search phrases from online search engines. The highest ranking document called "htm 2031" provides essential NHS guidelines on infection control. The full title reads: "Health Technical Memoran- dum (HTM) 2031. Clean Steam for Sterilization, NHS Estates (1997)". This is one of the documents from a series of national infection control guidelines called "Health Technical Memorandum".

Internal NRIC site search keywords in 2008. In contrast, NRIC received 3 812 searches via the internet search facility on the NRIC site. This continues to confirm a trend that we have previously seen, in that browsing using the navigation menus, rather than searching, is the primary mode of finding information on NRIC. The chart on Figure 2 shows the Top 20 search phrases entered using the NRIC search facility. As shown, the Health Technical Memorandum series of resources also scores very high, while there are a number of specific keywords such as "hcai" (Healthcare Associated Infection) which did not feature on the Top 20 search en- gines list, probably due to the sheer volume of high-profile healthcare websites dedi- cated to this topical phenomenon.

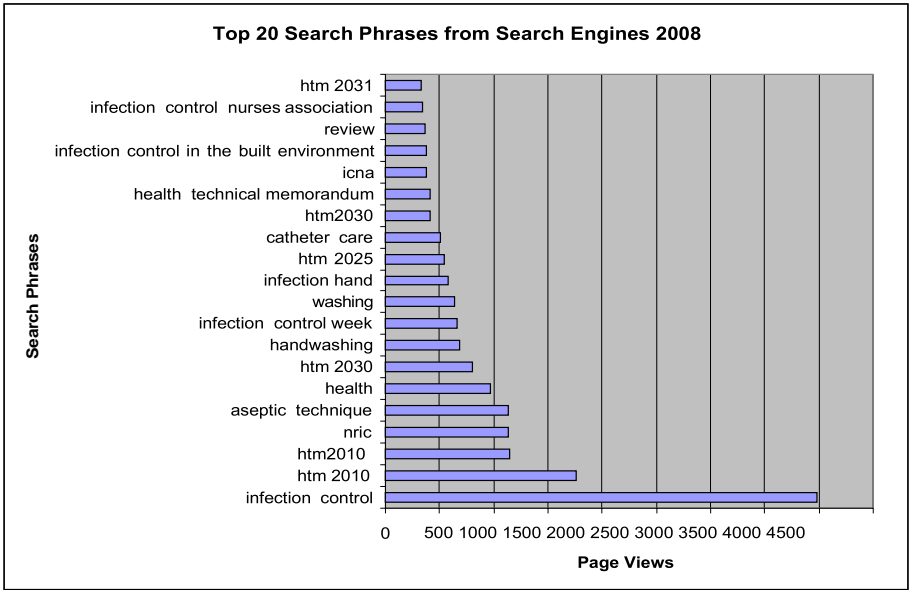[1] http://www.nric.org.uk/IntegratedCRD.nsf/NRIC_SiteStats?OpenForm

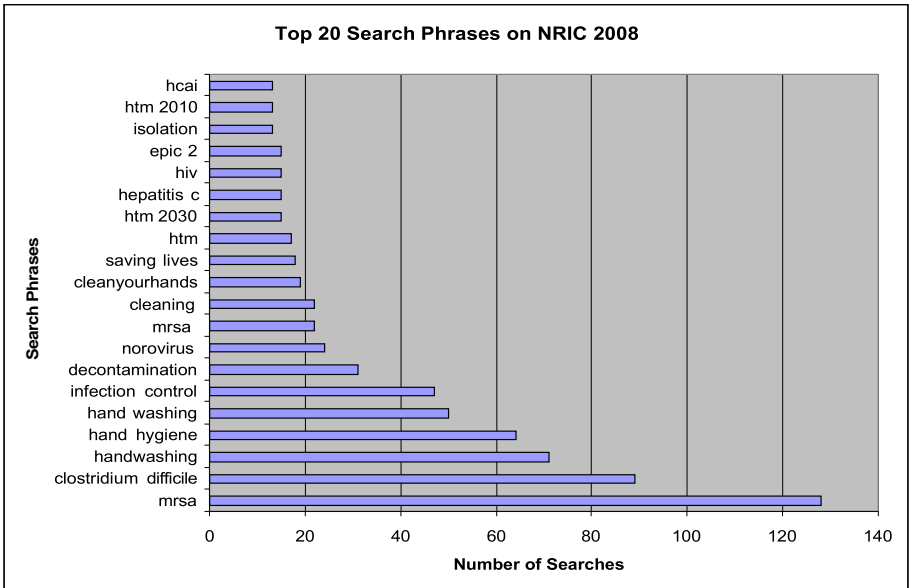**Fig. 1.** Top 20 search Phrases from Search Engines 2008

Fig. 2. Top 20 Search Phrases on NRIC 2008

While the search phrases reveal an invaluable insight into user information needs, they are only the starting point in understanding users and the actual impact of the DL on clinical practice. The need for more in-depth evaluation using multiple data collection methods, in addition to weblogs, was demonstrated by authors who illustrated a disproportion between website developers' aims and users' needs, and also between perceived and actual user behaviour. These were shown on an evaluation of NeLI DL [10] and on an evaluation of a WHO portal, called "Labresources" [11], for microbiologists in developing world.

Therefore, there is a demonstrable need to enhance the weblog results by other data gathering methods, such as questionnaires and interviews, to obtain a more rounded picture of the DL's impact. The Impact-ED evaluation model discussed in the following section 3 provides a general framework fulfilling these requirements.

3 Evaluation of DLs: The Impact-ED Model

Evaluation of Internet digital libraries is a rapidly growing research domain, of interest to the IT community as well as governments and policy makers. It has become clear to all stakeholders that sound evaluation is essential for a realistic assessment of investments into the IT technology. But what do we evaluate and what measures do we apply to assess a DL?

Chowdhury & Chowdhury define evaluation as a judgment of worth to ascertain a level of performance or value [12]. Saracevic takes this further, suggesting that performance can be broken down into two criteria [13]:

• Effectiveness, i.e. how well does a system perform that for which it was designed?
• Efficiency, i.e. at what cost (financial or time/effort)?

In 1999, Fox and Marchionini [13] presented a model of digital library dimensions, suggesting that there are four dimensions to DL work: *Community, Services, Technology,* and *Content.* DL impact evaluation, particularly in the healthcare domain, should be measuring the impact of a DL's content on its community. Is it changing a clinician's work practice and healthcare outcomes in a tangible and/or measurable way? Evaluation of DL impact needs to define, untangle, and measure the longer-term effects on the user, rather than just short-term changes in decision-making.

The Impact-ED evaluation model developed by Madle [14] applies to each dimension of the DL and aims to assess the impact of the functions and purposes of a DL on the user community that it serves. Therefore, the evaluation data, obtained from different respective methods, are linked for individual users and for the entire study to obtain a more in-depth picture of digital library use and impact.

The review of the literature [12] identified 12 healthcare DL evaluations (of which 2 were of the same DL). Only one of the studies evaluated real-time use of the DL at the point of need in the user's work, and none linked data on an individual basis from different sources. Figure 3 shows the model.

The intention of the Impact-ED (**Impact E**valuation for **D**igital Libraries) model [14] is that a variety of methods are used to collect data, and data is linked to provide a more rounded picture of a digital library's impact.
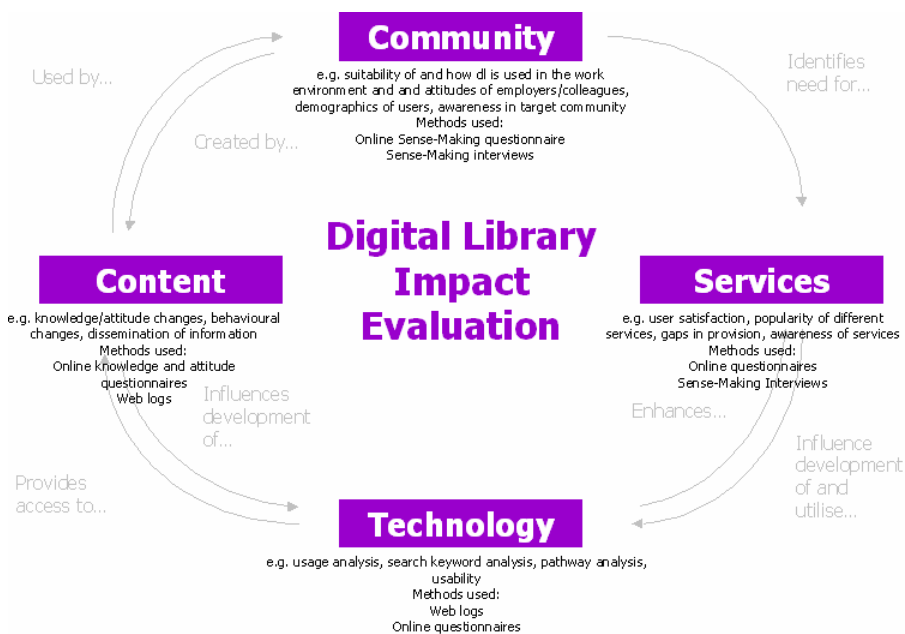
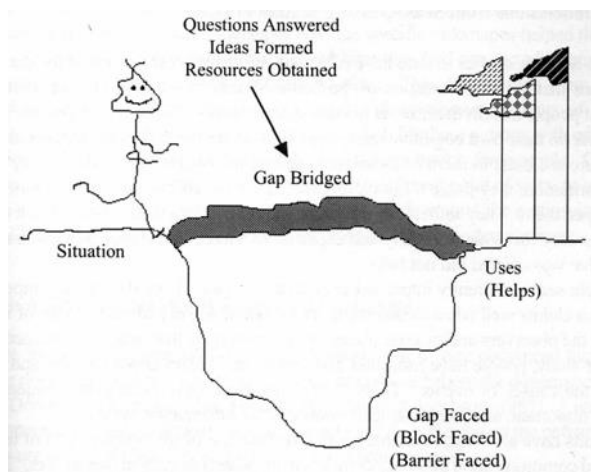**Fig. 3.** The Impact-ED model

Fig. 4. Dervin's Sense-Making model (taken from [17])

The model assesses user *knowledge* gain as one of the aims of DLs, enabling users to process the data or information provided. Sharing of what is termed "explicit knowledge" (i.e. knowledge that can be written down) is considered a fundamental aim of DLs [15]. Another variable assessed is an *attitude,* defined by Azjen as "… the degree to which performance of the behaviour is positively or negatively valued …" [16].

e

Are the knowledge gained and the attitudes attained by users influencing their behaviour? The Impact-ED model draws on assumptions from the Theory of Planned Behavior [16], defined by Azjen, and from the techniques of Dervin's Sense-Making model and methodology [17,18]. Dervin's approach allows exploration of *how* users meet their information needs and enables us to unravel the *"how"* of information seeking. The model is shown in Figure 4.

In order to investigate the impact of a DL from the technical, community, content, and services perspectives as defined by the Impact-ED model, appropriate data collection methods were chosen and a triangulation method developed. Reflecting the four dimensions, these are as follows:

1. *Online questionnaires.* Investigating use of the DL within the work environment
2. *Online pre- and post-visit (sense-making) questionnaires.* Investigating real-time, real-world use and how knowledge and attitudes change.
3. *Online tasks.* Investigating how users complete tasks to find information within the library and how this changes knowledge and attitudes.
4. *Weblog analysis.* Showing what users actually did within the DL.
5. *Interviews.* Complementing these other methods by providing more in-depth qualitative data that expands on issues identified in the questionnaires and weblogs.

A triangulation method, linking the pre- and post- and online questionnaires to users' actual behaviour (known from weblogs) to qualitative information revealed in follow-up interviews, provides a much more in-depth picture than previous research has allowed [18] of how a digital library may be impacting its user community and their work. The technical details of the methodology and subsequent calculation of the so-called *impact score* allowing a comparison between impacts of different DLs are too complex for the scope of this paper, and can be found in [19].

4 Case Study: NRIC Impact Assessment Using the Impact-ED

The impact evaluation of the NRIC portal was conducted using the Impact-ED model to test it in a real setting with real domain users [14]. In order to do that, the Impact-ED had to be applied on the NRIC DL to provide a set of criteria, around which questionnaires and interviews were designed to collect appropriate data. The NRIC DL was mapped onto the general Impact-ED model, to obtain insight into the impact of NRIC on the clinical practice of infection control nurses in the NHS (National Healthcare Service in the UK), as illustrated in Figure 5.

The methods used in the impact evaluation were as follows:

- *Study registration (Feb '08) and end questionnaires (May '08).* To find out how, when, why, and by whom NRIC was used, compare answers before and after the study, and provide an opportunity for users to comment on services and suggest improvements.
- *Information seeking/knowledge gain task (May '08).* To examine how well users could complete an information seeking task, i.e. find specific documents and find answers to questions using NRIC.

- *Pre- and post-questionnaire (Feb-May '08).* To discover, at the point of use and in their own words, users' reasons for using NRIC and what they know already, to compare with what they think they have learnt from using NRIC and how they will apply this to their work.
- *Web server log collection (Feb – May '08).* To collect data on how the participants actually navigate the library and see how this compares with how they report using it and the impact it has on their work.
- *Interviews (July - Aug '08).* To provide more in-depth information in users' own words about how the site has an impact and how it can be improved.
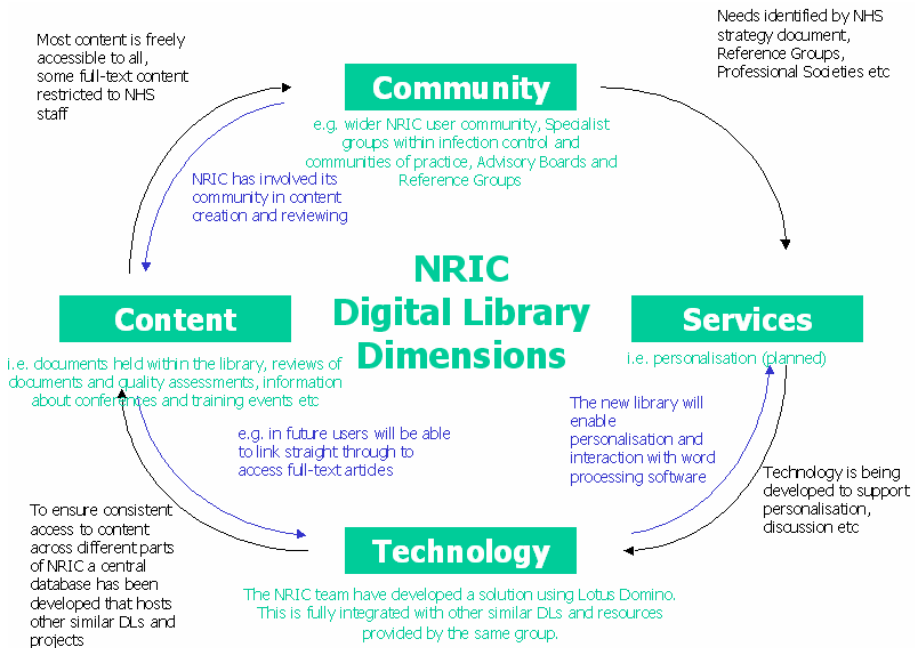

Fig. 5. Mapping the NRIC library onto the Impact-ED model

4.1 Data Analysis and Results

65 NRIC users signed up for the impact evaluation. Of these, 2 officially dropped out, 53 completed the registration questionnaire, 32 completed pre- and post-visit questionnaires of which 72 sets were matched for analysis, and 31 completed the end-of-study questionnaire. In addition, 5 users were interviewed. The study ran from February 2008 to May 2008 with interviews taking place during July and August 2008.

The majority of participants (28 of 52) listed nursing as their profession. The type of information sought at NRIC is illustrated in Figure 6.

At the start of the study most users reported the NRIC library to be either very useful (40.4%) or somewhat useful (38.5%) with only two specifically reporting that it was not useful. There was no significant change in these results at the end of the study
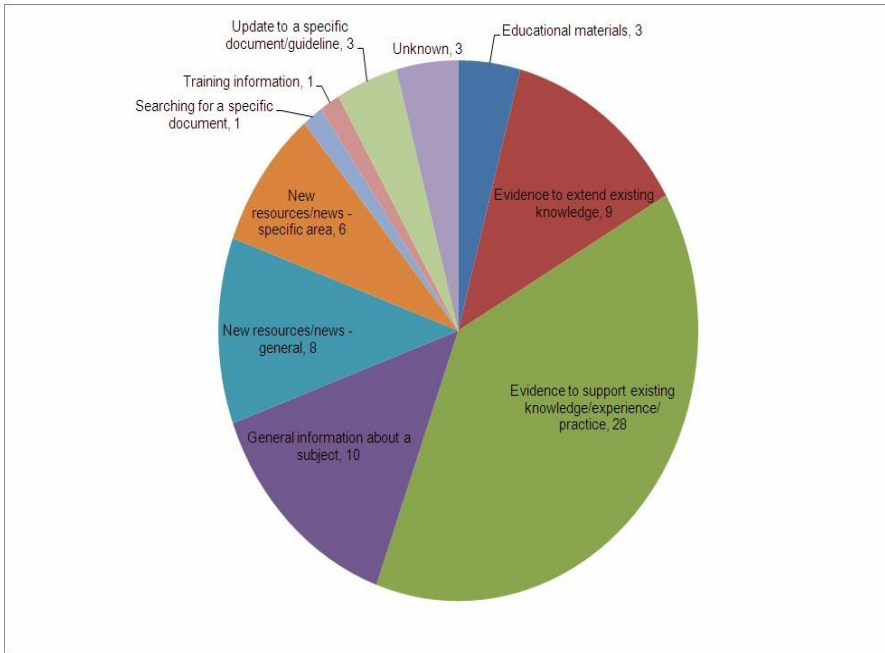
e

Fig. 6. Type of information sought

period. In the 72 visits for which pre- and post-visit questionnaires were collected, users found relevant information in 47 visits (65.3%).

Specific comments included:

"NRIC, um, it's a really good resource" (Interviewee A)
"I find it a very useful resource" (Interviewee B)
"Well again it just makes my job easier to do really, I think it makes me, um, it gives me the information I need to perform my role more efficiently." (Interviewee C)

Table 1 shows the basic access statistics for the 72 visits to the NRIC library that were analyzed by pre- and post-questionnaires. The average time spent (excluding the time spent completing questionnaires) was quite high: over 12 minutes, with 1/3 of users spending over 15 minutes and 1/3 under 5 minutes in the library. Users were viewing an average of 13 pages per visit and this included 3 documents. However, the majority of visits accessed either between 0 and 5 pages (41.7%) or 6 to 10 pages (34.7%).

The most popular method of navigating the website was to browse and search (27 visits), with 24 visits only browsing and 17 only searching. Browsing was more effective than searching in terms of whether or not NRIC had an impact, as shown in Table 2, perhaps due to the issues with sorting search results by date as a default, which was not implemented at the time of the study. This feature is now imple- mented to users' satisfaction, illustrating a direct usage of the evaluation study re- sults on the technical improvements of the DL.

Table 1. Basic access statistics

Basic access statistics for the 72 visits analyzed	
Mean time spent per visit	00:12:22
Mean number of different pages viewed per visit	13.74
Mean number of documents viewed per visit	3.07
Median number of documents viewed per visit	2
Total number of reviews available	48
Number of reviews visited	5
% available reviews visited	10.4%

Table 2. Browsing and Searching Behaviour

Category	Confirmed/strengthened or changed knowledge	Gained knowledge	No impact
Browsed only	45.8%	50.0%	37.5%
Searched only	29.4%	23.5%	52.9%
Browsed and searched	33.3%	37.0%	48.1%

User knowledge was confirmed, strengthened, or changed in 36.1% of visits, and knowledge was gained by the user in 37.5% of visits. Most importantly, overall there was an impact on user knowledge in 52.8% of the 72 visits.

To summarize, the NRIC *Community* dimension evaluation showed that NRIC is used for policy development and for keeping up to date with news, but awareness in the community could be improved. The *Services* dimension evaluation demonstrated that NRIC was perceived as a useful resource and provided relevant information in over 65% of visits. The *Technology* dimension evaluation showed a significant amount of time spent per visit (on average over 12 minutes) and visits to 3 documents on average per session. In terms of the *Content* dimension, NRIC had an impact on user knowledge in 52.8% of visits.

As mentioned above, the next step in measuring the impact involves calculation of a single *impact score* indicating the impact of a DL's functions on an interval between 0 and 1; however, this substantial aspect of the work is beyond the scope of this paper and can be found in [19].

5 Conclusion

In recent years there has been an unprecedented explosion of medical websites and Internet DLs for patients and medical professionals. However, these vary widely in quality. Despite the massive investments in IT for healthcare, little attention has been paid to meeting the need for impact evaluation of DLs.

e

In this paper we illustrated the need for a user-centered evaluation of Internet medical digital libraries, taking into account the community, services, content, and technology aspects of DLs. In particular, we presented a novel Impact-ED model which brings together these four dimensions to assess knowledge, attitude, and online information-seeking behaviour to discover a DL's impact using qualitative and quantitative data collected by online and pre- and post-questionnaires, weblogs, and interviews.

The applicability of Impact-ED was illustrated in a case study undertaken on the NRIC portal in the UK in 2008 to demonstrate the impact of this infection control DL on users' knowledge, attitude, and behaviour. NRIC had a particular impact in the "Content" dimension, affecting user knowledge in 52.8% of visits. This illustrates the applicability and suitability of the Impact-ED framework to DLs in the healthcare domain.

References

1. National Program for IT, `http://www.npfit.nhs.uk/default.asp`
2. National Knowledge Service, `http://www.nks.nhs.uk/`
3. Forkner-Dunn, J.: Internet-based Patient Self-care: The Next Generation of Health Care delivery. Journal of Medical Internet Research 5(2), e8 (2003)
4. Muir Gray, J.A., de Lustignan, S.: National electronic Library for Health (NeLH). BMJ 319, 1476–1479 (1999)
5. Improving patient care by reducing the risk of hospital acquired infection: A progress report (A report by the Comptroller and Auditor General). National Audit Office HC 876 Session 2003-2004, July 14 (2004), `http://www.nao.org.uk/publications/nao_reports/03-04/0304876es.pdf` (Summary), `http://www.nao.org.uk/ publications/ nao_reports/03-04/0304876.pdf` (Full Report)
6. The Management and Control of Hospital Acquired Infection in Acute NHS Trusts in England (2000), HC 230 1999/2000, `http://www.nao.org.uk/pn/9900230.htm` (NAO Press Notice), `http://www.nao.org.uk/publications/nao_reports/9900230.pdf` (Full Report) `http://www.nao.org.uk/publications/nao_reports/9900230es.pdf` (Executive Summary)
7. Wilson, T.D.: Human Information Behaviour. Information Science, Special Issue on Information Science Research 3(2) (2000)
8. Juvina, I., Herder, E.: The Impact of Link Suggestions on User Navigation and User Perception. In: Ardissono, L., Brna, P., Mitrović, A. (eds.) UM 2005. LNCS, vol. 3538, pp. 483–492. Springer, Heidelberg (2005)

9. Draper, S.W.: Supporting use, learning, and education. Journal of Computer Documentation 23(2), 19–24 (1999)
10. Roy, A., Kostkova, P., Weinberg, J., Catchpole, M.: Do users do what they think they do?-a comparative study of user perceived and actual information searching behaviour in the National electronic Library of Infection (2009) (submitted)
11. Madle, G., Berger, A., Cognat, S., Menna, S., Kostkova, P.: User information seeking behaviour: perceptions and reality. An evaluation of the WHO Labresources Internet portal. Informatics for Health and Social Care (in press)
12. Chowdhury, G.G., Chowdhury, S.: Introduction to Digital Libraries. Facet Publishing (2003)
13. Saracevic, T.: Digital Library Evaluation: Toward an Evolution of Concepts. Library Trends 49(2), 350–369 (2000)
14. Madle, G., Kostkova, P., Roudsari, A.: Impact-ED – A New Model of Digital Library Impact Evaluation. In: Christensen-Dalsgaard, B., Castelli, D., Ammitzbøll Jurik, B., Lippincott, J. (eds.) ECDL 2008. LNCS, vol. 5173. Springer, Heidelberg (2008)
15. Rowley, J.: The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science 33(2), 163–180
16. Azjen, I.: The Theory of Planned Behavior. Organ. Behav. Hum. Dec. 50, 179–211 (1991)
17. Dervin, B.: Audience as Listener and Learner, Teacher and Confidante: The Sense-Making Approach. In: Dervin, B., Foreman-Wernet, L., Lauterbach, E. (eds.) Sense-Making Methodology Reader: Selected Writings of Brenda Dervin, pp. 215–232. Hampton Press, Cresskill (2003)
18. Dervin, B.: A Theoretic Perspective and Research Approach for Generating Research Helpful to Communication Practice. In: Dervin, B., Foreman-Wernet, L., Lauterbach, E. (eds.) Sense-Making Methodology Reader: Selected Writings of Brenda Dervin, pp. 251–268. Hampton Press, Cresskill (2003)
19. Madle, G.: Impact-ED: A New Model of Digital Library Impact Evaluation. Ph.D thesis (2009)