_____

Investigating school pupils' information behaviour via intranet usage data

Andrew K. Shenton and Andrew Johnson

Abstract

Despite the popularity of electronic information resources among young people and the prevalence of organisational intranets in schools, few researchers have exploited the potential of intranet usage logs to investigate pupils' information behaviour. Nevertheless, the method is attractive on several counts. As it enables data to be collected without the participants' knowledge, it is unlikely to affect behaviour and highly objective data can be gathered. This material can lead to the use of supplementary methods allowing the elicitation of detailed qualitative data. Researchers are, however, restricted to the types of insights that the system log provides and are further constrained by the purposes for which school managers use their intranet. Ethical questions surrounding the gathering of data that participants are not aware is being generated may also be asked. Notwithstanding these issues, intranet usage logs clearly offer considerable potential for the investigation of information behaviour in a range of contexts.

1 Introduction

This is the last in a series of papers written for *Library and Information Research* dealing with the strengths, weaknesses and potentials of three particular methods of collecting data on information behaviour. Although each technique offers exciting possibilities, none has been exploited as frequently in this regard as the long-established tools of observation, interviews and paper questionnaires. In fact, it was the lack of coverage that has been given to each in LIS literature which inspired the writing of the papers. The first article examined the value of online questionnaires (Shenton and Johnson, 2006); the second considered how suggestions made by library patrons can be employed to investigate their information needs (Shenton, 2007); this final piece addresses the role that the usage data available from organisational intranets can play in uncovering information needs, the action that individuals take in response to such needs and

_____

Author

Dr. Andrew K. Shenton is a former Lecturer in the Division of Information and Communication Studies at Northumbria University.

E-mail: ashenton1@hotmail.com

Andrew Johnson is Network Manager at a school where research of the kind that is described in this article has taken place.

_____

the use made of any retrieved material. The first and third papers are based on the authors' experiences in undertaking actual research incorporating these methods, in a school environment. The library suggestions article is more hypothetical, however, and is principally concerned with exploring ideas and possibilities. It should be understood from the outset that this final article focuses quite specifically on usage logs associated with organisational intranets. The contribution that can be made to research into information behaviour of more general Web tracking software is not the authors' concern and, indeed, this territory has recently been covered by Fourie and Bothma (2007).

2 Intranets

Undoubtedly the most significant single shift in the information-seeking behaviour of young people in the last twenty years has been their growing reliance on electronic sources. Nevertheless, whilst studies of youngsters' use of the Internet and, in the 1990s, CD-ROM software have frequently been reported in LIS journals, their exploitation of materials provided via their school intranets has been subject to much less attention. This deficiency is surprising given the concerns of commentators such as Herring (1999), who have expressed anxieties that pupils may rely too heavily on electronic sources offered through their organisation's network and may never consult the collections of paper materials found in their school libraries. Although a considerable diversity of types of document can be mounted on an intranet, such as, where copyright permits, copies of existing pages from the World Wide Web, new information sheets specially prepared by teachers and digitised versions of older resources previously in paper form, if the only information sources required for a particular assignment are those made available in this way, educators run the risk of giving their charges insufficient practice in the development and application of skills relating to the wider information world. Still, in schools catering for young children particularly, the ability to restrict pupil use of electronic information materials to a limited range prepared or selected specially for the youngsters renders the intranet an environment that is appealing to educators. Despite widespread employment of filtering software aimed at denying users access to unsuitable materials, parents, teachers and librarians alike may feel uneasy when youngsters are given access to the World Wide Web, and the intranet-only route may be welcomed as a safer option. In addition, Basker (1998) recognises how intranets are generally considered easy to exploit since they are often designed using a Web browser, which forms a familiar tool for many searchers, rather than by using means which remain the preserve of specialists. One of the few studies to explore pupil use of intranets is that of Shenton and Johnson (2007) but their project concentrates exclusively on a newly-constructed library section of their school's intranet, rather than all areas of it available to the pupils.

3 Usage data in historical context

Whilst the concerns in some quarters that school intranets may pose threats to the use of traditional paper collections should be acknowledged, they also afford opportunities for researchers interested in pupils' information behaviour in schools. Specifically, many intranets are provided by software that offers usage

_____

_____

statistics to staff administering the network. The software associated with the intranet examined in the aforementioned Shenton and Johnson (2007) study gives breakdowns of the numbers of hits associated with, for example, individual pages, named users, each date and the Web browsers employed by those accessing the intranet, as well as figures pertaining to combinations of these variables. The pages of the intranet forming the subject of the Shenton and Johnson paper are accessed through a hierarchical arrangement involving folders, subfolders and hyperlinks to individual documents but, if information is retrieved via the entry of search words, the software allows a researcher to discover the terms that have been employed. All these data are available from the system itself; no additional program is required to generate them. The software supporting the intranet whose library section was scrutinised by Shenton and Johnson was *Microsoft SharePoint* and Figure 1 shows how easily site usage data can be accessed using the latest version of this software package available at the time of writing. The version in question is slightly more advanced than that used in Shenton and Johnson's project. After going to the library area, a researcher can select the "site settings" menu option to view the "site administration" facilities, one of which is a "site usage report". In the figure, the only changes made to the page as it is seen on-screen are those that have been necessary to protect the anonymity of the school. For this reason, the blue title bar at the top has been removed, the URL has been deleted from the address box and various details in the "site information" rectangle have been hidden.
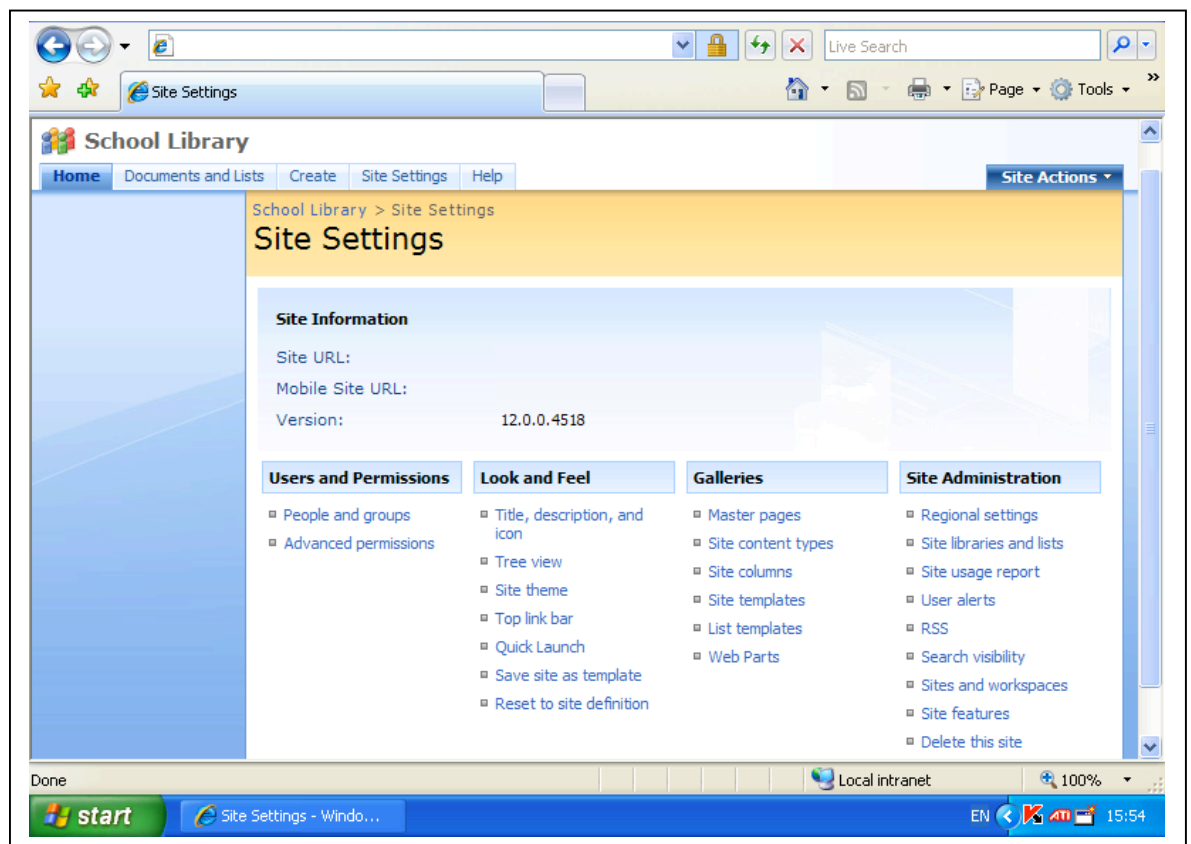
Figure 1: Screen dump showing placement of the "site usage report" feature within the "site administration" facilities offered by Microsoft SharePoint

_____

_____

One of the principal attractions of the data provided by such logs is that, since the pupils are likely to be oblivious to the fact that the software is recording the moves that they make, the trustworthiness of the research will not be undermined by changes in behaviour which result from an awareness that their interactions with the computer are under scrutiny. The employment of unobtrusive measures is not, of course, new in LIS. Over forty years ago, Eugene J. Webb, Donald T. Campbell, Richard D. Schwartz and Lee Sechrest wrote a classic book advocating methods of this type (Webb et al, 1966), and their work has proved influential in research in many areas of the social sciences. A feature of their argument is a wariness of traditional methods such as questionnaires and interviews. The authors write that these

intrude as a foreign element into the social setting they would describe, they create as well as measure attitudes, they elicit atypical roles and responses, they are limited to those who are accessible and will cooperate, and the responses obtained are produced in part by dimensions of individual differences irrelevant to the topic at hand.

(Webb et al, 1966, 1)

For many years researchers intent on investigating young people's information behaviour have employed a variety of techniques incorporating data that are not collected directly from participants in contrived "research" situations. Investigators have seized opportunities to draw on information provided by users for some other purpose or to generate data unobtrusively at the time of an actual information encounter. In terms of the former, works listed by pupils in bibliographies at the end of assignment work might be subjected to citation analysis techniques so as to ascertain the types of information sources used, their age and the collections from which they have been obtained. Another technique has involved the logging of queries put to information professionals in reference interviews. The records created can subsequently be examined to gain insights into the information needs of library users. With regard to data deliberately *generated* for research purposes, rather than simply information *disclosed* in the course of normal information-seeking activities and then recorded, library issue figures have been scrutinised, again to infer patrons' information needs. Records of book borrowing are, in fact, one of several unobtrusive measures that Webb et al (1966) note can be used to determine the use of information materials. The authors' concern, however, lies with methods to assess the popularity of individual books, rather than the inference of information needs. In addition to issue figures, they highlight the degree of wear of the volumes, especially on the corners where pages are turned, and the amount of dust that has accumulated on the items. Neither measure, however, has been widely adopted by researchers examining information needs and, today, book issue data are frequently associated in the minds of commentators with an older and, it may be argued, outdated research framework for the investigation of information behaviour. Dervin and Nilan (1986) are among the most prominent in detecting a shift of emphasis. They observe a movement away from measurable, external behaviour pertaining to the use of information systems to the exploration of individuals' internal cognitions.

_____

_____

Despite this overall trend, as electronic information systems started to become increasingly widely used in the latter half of the 1980s and 1990s, researchers were not slow to exploit methods for logging the keystrokes and commands of users interacting with these systems and to subject the data to a battery of statistical tests. Before the arrival in classrooms and school libraries of the Internet, early projects with youngsters included the use of such methods in relation to remote online encyclopedias and databases (Eastman and Agostino, 1986; Edyburn, 1988) and CD-ROM resources (Marchionini, 1989). As in more recent times electronic information systems have multiplied in their range and number, in addition to becoming ever more prevalent, attention to techniques for examining the use made of them has similarly increased. Just a year before the writing of this paper, Williams and Gunter (2006) noted the considerable interest within information science in the use of logs pertaining to the Web and other electronic systems, although none of the particular papers that they highlight relates specifically to intranets. Perhaps the biggest difference between log data and the older unobtrusive measures is that, whilst the former still provide an external perspective and give little insight into an individual's thinking behind his or her actions, some systems may at least allow a user's search to be reconstructed retrospectively. Thus, whereas citation analysis, for example, is concerned with what Shenton has termed the "where" and the "what" of information-seeking, logging software can give more insight into the "how" (Shenton, 2004, 185). Another useful construct is suggested by Wilson (1999). He defines "information searching behaviour" as that "sub-set of information-seeking… particularly concerned with the interactions between information user… and computer-based information systems" (Wilson, 1999, 263), and clearly intranet usage logs offer particular possibilities for studies into this area.

5   Strengths and applications of the method

As a research technique, the use of logs recording details of pupil interactions with an intranet offers many attractions. One of the principal strengths is that the inaccuracy that may result from the problem of "human instrument error" recognised by Webb et al (1966, 142) is avoided, although the data generated by the system must, of course, be trustworthy and the researcher must have an accurate grasp of what is being recorded. Just as frailties in the data gathering abilities of the researcher are evaded, so too are deficiencies that result from the participants' mental weaknesses. Nicholas et al. (2005) write that logs form

a direct and immediately available record of what people have done: not what they say they might, or would, do; not what they were prompted to say; not what they thought they did.

(Nicholas et al, 2005, 251)

In addition, the data may be scrutinised on a range of different levels. The investigator may draw on them, for example, to reach tentative conclusions about the overall extent of intranet use by pupils within a certain period of time. The degree to which individual pages have been consulted and how they have been accessed may also be explored. As Nicholas et al. (2005) point out, since a log records the use made by everyone who has interacted with the system, the

_____

_____

construction of a sample, which often forms a challenging research task, is unnecessary. If a longitudinal perspective is taken, the investigator may wish to look at how visits to pages fluctuate depending upon the assignment and the information materials made available through the intranet. Attention to this area would give the research a dimension which Boyd (2004) believes is often missing in investigations into information-seeking. Specifically, he identifies that they generally concentrate on a "snapshot in time" and pay no attention to the "frequency of encounters" (Boyd, 2004, 81). If the usernames of individual users of the different pages are logged, insights into the exploitation by each pupil may be gained, too. It may then be possible to trace patterns of use that emerge in terms of gender, age and class groups.

Where a detailed understanding of the circumstances surrounding a pupil's accessing of certain pages is to be developed, however, or the researcher is to appreciate the role of the intranet in that individual's wider information universe, the log data will be insufficient in itself. A second stage in the research may involve identifying, through their username details, those who have gone to information resources made available via the intranet and contacting these people with a view to their recounting their experiences in questionnaires or individual interviews. In a school with a well developed ICT infrastructure, the provision of an online questionnaire may present possibilities for such follow-up inquiries. If the log records reveal that there are too many intranet users for them all to be approached individually, the researcher may choose to concentrate on a small number, such as those in a particular class or year group. Alternatively, he or she may opt to construct a sample that is balanced in terms of including youngsters of different genders, ages and ability levels. Another possibility is that of using the log to identify only those pupils who are seen to have been regular users of the intranet over a prolonged period of time, rather than those who have merely exploited the resource for one particular recent assignment.

If the researcher acts on usage data by tracing pupils who have accessed information resources provided via the intranet and carries out interviews with them, the "help chain" approach discussed by Dervin (1992, 72) may be adopted. Dervin explains how this strategy involves asking the informant how the resource or material in question – in this case the intranet pages consulted – helped the individual to resolve the situation motivating its use.

The researcher may, however, wish to widen the discussion beyond this central question to address such matters as

- why the pupil needed to take information-seeking action;
- why the intranet, in particular, was chosen for exploitation;
- other information-seeking options that were available;
- which of these alternatives were actually pursued;
- the strategies employed by the individual to access, and then use, the information available from the sources;
- any problems encountered in the work;

_____

_____

- the degree to which the individual was satisfied with the eventual outcome.

The data provided in the electronic log thus forms the starting point for a broader investigation in which the researcher aims to learn more about the totality of the youngster's information-seeking experience with regard to the situation that precipitated the activity. Such an approach is consistent with Williams and Gunter's (2006) observation that qualitative data collected at this stage may be

concerned not only with simple questions about the use of and opinions about the particular system or service, but also with the full context in which an information need arises and is met.

(Williams and Gunter, 2006, 131)

The incorporation of these qualitative data takes the research beyond the systems-oriented perspective offered by the log data and into a more explorative realm revolving around the user's perspective. Furthermore, the framework suggested here can be applied in any situation where named information-seekers taking a particular course of action can be identified. A similar approach can be used, for example, in connection with electronic library management software systems. Here borrowers of non-fiction may be traced and the degree of assistance provided by the works in question in meeting clients' particular needs can be probed.

Further strength may be added to the information use element of a study if the researcher is able to gain access to the assignment work that was ultimately submitted by the pupil, together with any early drafts that are still available. The researcher may examine these alongside the appropriate pages of the intranet to determine, for example, how far the information sources provided have been plagiarised, legitimately used or simply ignored. A comparable strategy has been employed by McGregor and Streitenberger (1998). The authors report how they ascertained the amount of copying by pupils tackling assignment work in their study "by comparing the students' final papers with the original sources listed in their bibliographies".

Although the focus of this paper lies solely in the exploitation of data pertaining to the use of an organisation's own intranet, it is possible that the researcher may wish to employ this material in concert with data provided via other software which records the details of user interactions with the World Wide Web. Such a dual approach is especially appropriate if the pupil begins with resources available through the intranet but then breaks out of its confines by following links to sites on the World Wide Web or by entering a Web address, such as that for *Google*. If researchers want to maintain their insights into the user's searching actions in these circumstances, additional software, beyond the intranet's logging system, is obviously needed.

6 Weaknesses

Perhaps the biggest disadvantage of the intranet usage figures presented by logs is that the researcher is constrained by the types of data available. If he or she is

_____

_____

unable to introduce other methods that provide further illumination, the danger arises that the foci of the inquiry are determined by the data that the system is able to provide, rather than what the investigator deems worthy of study. This disadvantage proved a significant constraint in the work of Shenton and Johnson (2007) in their investigation of the use made of a new section of a school intranet devoted to the organisation's library.

In the Shenton and Johnson project, the researchers were initially hopeful that it might be possible to learn whether, on balance, pages tended to be consulted at school, rather than elsewhere, or vice versa. Ultimately, this proved problematic, however. The Web browser associated with the pupil's computer was recorded by the usage log, yet it was only where this differed from the browser used by the school equipment that it was evident that the page had been accessed in another environment, most likely the home but possibly the youngster's local public library; if the Web browser was listed as the same as the school's, the researcher had no way of knowing whether the interaction had taken place at school or not. Johnson later overcame this problem by adding an extra value to the user agent string on computers within the school. All Web browsers have a user agent string which is sent to the Web server upon connection. The string identifies what browser the client is using. Thus, by adding a unique extra value, Johnson found that he was able to distinguish between internal and external access to the site.

When working with some systems, the researcher may find it difficult to reconstruct a user's search retrospectively. Even if the intranet is structured in such a way that searching takes place simply by following the convergent, hierarchical progression already described, again much depends on the type of log data that is available. It may be, for example, that it is apparent that a particular pupil has accessed certain pages on a particular day but unless the times when these pages were accessed are given, the researcher has no indication of the order in which they were consulted or even whether the pages were seen in the same session. Data regarding times can also, of course, enable the researcher to see whether the search took place within a lesson or during free times in the school day, such as before or after classes, over break or during the lunch hour. In addition to being revealing in themselves, such details can help to jog the memories of youngsters if they are used at a later date in interviews about their information-seeking action. Unfortunately, in the Shenton and Johnson study, the intranet log recorded dates, but not times, when pages were accessed.

Another limiting factor is the use that staff make of the intranet in terms of the types of information that they choose to make available through it. It is entirely possible for schools to use their intranet purely for administrative purposes and no information materials to help with pupils' assignment work may be offered at all. It would certainly seem that most schools are more likely to use their intranet for academic purposes than to help pupils pursue their leisure interests. In these circumstances, any researcher wanting to exploit the resource's log data to uncover the youngsters' personal information needs will be disappointed. In order to learn about decisions made by pupils within even their school-related information-seeking activity, clearly a good range of information resources has to be made available and the youngsters must be set independent learning tasks that

_____

_____

allow them to make their own choices in terms of which materials they will consult and the environments from which they will obtain them. The employment of usage data is thus poorly suited to situations where heavily prescribed assignments are set, with the materials to be exploited predetermined.

However detailed are the usage data provided by the system, it would be unwise to attempt to develop an understanding of pupil information needs inferentially, purely on the basis of the system's data. Indeed, many of the criticisms made of library issue figures in information needs research can be repeated in this context, and its systems-oriented, rather than person-based, perspective renders the method more consistent with the old research framework than with the new. Nicholas (1997) notes that any measurement of needs involving library use entirely disregards the needs of non-users of the library and a similar charge can be made here in terms of usage logs. Notwithstanding the greater trustworthiness of data which stems from the machine-as-instrument, rather than human-as-instrument, it is entirely possible that a record in the system is made when a user has simply clicked on a link to a document accidentally and there is no guarantee that a page which has been visited has even been properly read, let alone used. This danger is akin to a problem identified by Webb et al (1966) in relation to book issue data, i.e. items may be borrowed but, again, not read. In short, an intranet usage log gives no indication of how much attention has been paid to a page that has been logged as visited, and instances of accidental access may come to light only in a later "help chain" interview.

The final area of concern is that of ethics. Even if the name of the user is left undisclosed, researchers may well question the morality of employing a method that may allow them to reconstruct a sequence of the pages viewed by a particular individual and the search terms used to access these, and then report the course of the search in published research. In a previous paper in this series, Shenton (2007) has raised a similar issue in terms of library suggestions. He writes,

It may be argued that, if reports of studies of user needs are to be compiled and then disseminated beyond the library, via conference papers or journal articles, for example, and such work refers to the personal situations of participants, even if all the data are anonymised contributors should be made aware of what is intended in advance of their offering suggestions.

(Shenton, 2007)

In an era in which the need to secure informed consent from participants is frequently stressed, the use of log data – which the youngster may not even know is being generated – may raise anxieties among those concerned with the increasing prevalence of surveillance and "Big Brother"-style intrusions into everyday life. A possible solution is to employ only data that have been specifically declared admissible by pupils, either in e-mails or in follow-up questionnaires/interviews. If such statements are solicited in advance of the log data being captured, however, the method can no longer be considered truly unobtrusive.

_____

_____

7 Conclusions

Although much underused in this capacity, the data available in usage logs associated with organisational intranets offer many possibilities for researchers interested in information behaviour. The approach affords a level of objectivity that is lacking in many techniques in which investigators collect data personally, and its unobtrusiveness renders it unlikely to cause changes in the participants' behaviour. Still, as with many methods whose use goes unnoticed by participants, questions may be asked of the morality of drawing on data which participants are not aware is being collected and, clearly, the extent to which the log data is useful in a particular project depends to a considerable degree on the nature of what the system can provide. Nevertheless, even if the data provided by the usage log are limited, the system is at least likely to be able to show the identities of pupils who have accessed pages. This knowledge can lead to the development of "help chain"-type interviews, the scope of which can be extended to address matters as diverse as the need prompting the search and what was done with the information that was collected, as well as the information-seeking action taken. Furthermore, with regard to the information use dimension, an important element of triangulation can be introduced by comparing the finished work with the material within the information source.

This paper has been written from the perspective of two researchers keen to exploit the possibilities of intranet log data for investigating the information behaviour of pupils in the school in which they have undertaken past studies. The principles behind many of the ideas expressed in this paper can, however, be applied to other organisations in which intranets have been established and where the potential of these networks for providing information materials has been exploited. In other educational contexts, it would, for example, be possible to study the information-seeking behaviour of teachers using staff areas of a school intranet, or to examine the use made by teachers or pupils of wider intranets provided by the appropriate pyramid of schools or the local education authority. The knowledge that an intranet of this kind existed in a particular area of north-east England in fact influenced the choice of geographical area made by Shenton (2002) in his doctoral research into how the information-related attitudes, needs and behaviour change during the years of childhood.

References

Basker, J. (1998) Intranets: who's doing what? *Records Management Bulletin,* 88, 7-14.

Boyd, A. (2004) Multi-channel information seeking: a fuzzy conceptual model, *Aslib Proocedings,* 56(2), 81-88.

Dervin, B. (1992) From the mind's eye of the user: the sense-making qualitative-quantitative methodology. In: Glazier, J. D. and Powell, R. R. (eds) *Qualitative research in information management.* Englewood: Libraries Unlimited. 61-84, ISBN 0872878066.

Dervin, B. and Nilan, M. (1986) Information needs and uses, *Annual Review of Information Science and Technology,* 21, 3-33.

_____

_____

Eastman, S. T. and Agostino, D. E. (1986) Commanding the computer: functions and concepts of videotex technology for eighth-grade students, *Journal of Research and Development in Education,* 19(2), 49-57.

Edyburn, D. L. (1988) Examining the successful retrieval of information by students using online databases, *School Library Media Quarterly,* 16(4), 256-59.

Fourie, I. and Bothma, T. (2007) Information seeking: an overview of web tracking and the criteria for tracking software, *Aslib Proceedings,* 59(3), 264-84.

Herring, J. (1999) Developing curriculum services via the Internet, *Library Association Record* 101(4), 223-25.

Marchionini, G. (1989) Information-seeking strategies of novices using a full-text electronic encyclopedia, *Journal of the American Society for Information Science,* 40(1), 54-66.

McGregor, J. H. and Streitenberger, D. C. (1998) Do scribes learn? Copying and information use, *School Library Media Research* [online], 1. URL: http://www.ala.org/ala/aasl/aaslpubsandjournals/slmrb/slmrcontents/volume11998 slmqo/mcgregor.cfm [accessed 13.12.07].

Nicholas, D. (1997) The information needs interview: a long way from library-use statistics, *Education for Information,* 15(4), 343-49.

Nicholas, D., Huntington, P. and Watkinson, A. (2005) Scholarly journal usage: the results of deep log analysis, *Journal of Documentation,* 61(2), 248-80.

Shenton, A. K. (2002) *The characteristics and development of young people's information universes*, PhD thesis, Northumbria University, Newcastle.

Shenton, A. K. (2004) Information-seeking research in schools: opportunities and pitfalls, *Aslib Proceedings,* 56(3), 180-86.

Shenton, A. K. (2007) Library suggestions and typologies of information needs, *Library and Information Research,* 31(98), 3-15. Available at http://www.lirg.org.uk/lir/ojs/index.php/lir/article/view/21/45 [accessed 13.12.07].

Shenton, A. K. and Johnson, A. (2006) The employment of online questionnaires within user studies research in school libraries, *Library and Information Research* 30(95), 25-31.

Shenton, A. K. and Johnson, A. (2007) Pupil use of a school intranet's library pages, *Aslib Proceedings*. 59(6), 512-27.

Webb, E. J. et al (1966) *Unobtrusive measures: nonreactive research in the social sciences*. Chicago: Rand McNally. ISBN 0528686941.

Williams, P. and Gunter, B. (2006) Triangulating qualitative research and computer transaction logs in health information studies, *Aslib Proceedings,* 58 (1/2), 129-39.

Wilson, T. D. (1999) Models in information behaviour research, *Journal of Documentation,* 55(3), 249-70.

_____