Pergamon

Museum Management and Curatorship, Vol. 19, No. 3, pp. 229–252, 2001

Printed in Great Britain
0260-4779/02/ $ - see front matter

PII:S0260-4779(01)00017-6

Usability Evaluation for Museum Web Sites

DANIEL CUNLIFFE, EFMORPHIA KRITOU AND DOUGLAS TUDHOPE

Overview

Museums, galleries and other cultural organisations have been swift in their adoption of the Web and virtual visits to museum Web sites have become popular. For many smaller museums the cost of developing and maintaining a Web site is difficult to justify, particularly where the economic benefit to the museum is hard to demonstrate. Many sites are therefore developed as unofficial, in-house projects—often without financial support. Evaluation is essential for determining whether a Web site is meeting the needs of its users and should be part of an ongoing process, from initial conception to long-term maintenance and development. This contribution focuses on the evaluation of Web sites as a way of addressing some of the limitations inherent in the non-professional development environment, placing it within the context of current museum Web development practice. The paper presents a case study illustrating Web evaluation issues from a computing perspective, using methods appropriate to the non-professional development environment. Four evaluation methods are examined in detail: direct observation, log analysis, online questionnaires, and inspection methods. The experience of applying each method, the benefits and limitations of each method and the relative effectiveness of the methods are analysed.

Introduction

Museums, galleries and other cultural organisations have been swift in their adoption of the Web as a mode of communication. Dawson and McKenna (1998) report the results of a 1998 survey of 290 United Kingdom museums of which 38% had a Web site and 20% were planning to have one in the near future. National museums were then leading the way, 86% of them already having a Web site. Museums are perhaps ideally placed as content providers for the Web, since they typically possess a large quantity of high quality content, they often own the Intellectual Property Rights to that content, and they have an established reputation for quality of information provision, objectivity, authority and so on. Virtual visits to museums have also become popular. Haapalainen (1999) writes that the number of visitors to the Finnish National Gallery Web site was expected to exceed the number of visitors to the physical museums it represents in 1999. The Museum of the History of Science, Oxford, reports a conservative estimate of three virtual visitors for every one physical visitor (Sphaera, 1999). In 1996 the Web site of the National Museum of American Art was receiving

10,000 visits per day as compared to 125,000 physical visits per year (Dery *et al.*, 1996, cited in Campbell and Wells, 1996). There are also cases in which the Web site launch has predated the opening of the physical museum by several years (Bowen, 2000). In addition to the opportunities offered by the Web, the virtual museum faces a number of challenges. Trant (1999) discusses the role of authenticity and trust in museums and cultural heritage institutions and ways in which this can be maintained in a largely unregulated, distributed Internet environment. Museums are also faced with a conflict between the need to protect and exploit financially their Intellectual Property Rights (IPR) and their mission to disseminate knowledge. In some cases museums are finding that the IPR they have over items in their collection do not permit digitisation or distribution over the Web. An in-depth review of the impact of IPR on cultural organisations and initiatives addressing new IPR models can be found in Rees (1998).

The larger national and international museums can often afford either to employ Web consultants or to develop the necessary skills in staff, but for the many smaller museums the cost of developing and maintaining a Web site is difficult to justify, particularly where the economic benefit to the museum is hard to demonstrate. Many are therefore developed as unofficial, in-house projects—often without specific financial provision (Hertzum, 1998). Those involved in the development are sometimes unaware of the importance of Web site development methods, resulting in an *ad hoc*, unprincipled development process which fails to consider the visitors' needs. Hertzum suggests that museum Web sites tend to suffer from three characteristic problems:

1. The majority of museum sites have been developed without a clear notion of what the site should achieve;
2. The sites have not been evaluated to find out whether they match the users' needs and wishes; and
3. The material on the sites tends to duplicate material in the physical museums rather than to rethink it, given the possibilities provided by the new medium.

The effects of poor design can be dramatic. Forrester Research (cited in Nielsen, 1998b) estimate a loss of 40% of potential repeat visits following an initial negative experience. There is also an increased recognition that simply having a Web site is no longer sufficient, a Web site should be achieving specific business objectives (Day, 1997). This is especially true where Web development is competing for resources with more traditional museum activities.

These problems are not unique to museum Web sites, and several authors (for example Murugesan, 1999; Nielsen, 1999a) suggest that the entire Web is facing an imminent usability crisis, particularly in large scale sites. Murugesan suggests this is in part because developers view Web development as a visual design and authoring problem rather than as an application development problem and many developers are unaware of the real issues and challenges involved in major Web developments. The emerging discipline of Web Engineering seeks to integrate scientific, engineering and management principles into Web development and to define systematic approaches to Web development, deployment and maintenance (Murugesan, 1999). Whilst this Web Engineering approach will be well suited to professional developers and well-funded development programmes, it will be less well suited to the non-professional, under-resourced, in-house developments which characterise many museum Web projects. How-

ever, even these non-professional developments can adopt methods and practices which will lead to a more principled development process and a more user-centred and usable product.

The central element in any development approach which aims to deliver a usable, user-centred product is evaluation, particularly evaluation involving users. This paper reviews evaluation methodologies by drawing on the computing literature and illustrates key issues via a Web site evaluation case study. The second section of this paper discusses the role of the museum Web site, and the difficulties of defining 'success' for a Web site, while the third reviews visitor studies and museum surveys which contribute towards an understanding of the virtual visitor. The particular challenges which the Web poses for developers, and the limitations inherent in the non-professional development environment are examined in the next section, while the fifth section presents a case study evaluation which was conducted under limitations similar to those in the non-professional museum development environment. Within this case study four evaluation methods are discussed in detail, direct observation, log analysis, online questionnaires and feedback, and inspection methods. The experience of applying each method, and the benefits and limitations of each method, are analysed in the context of current Web development practice, particularly in the museum area. The relative effectiveness of the methods are compared and the implications of the case study examined. The concluding section briefly discusses the evaluation issues associated with the long term maintenance and development of Web sites.

The Role of the Museum Web Site

Whilst interactives and 'exhiblets' deployed within the museum are generally seen as enhancing the visitor experience, and are popular with visitors (Booth, 1998a; Economou, 1998; Gammon, 1999), the roles of the museum and the virtual museum, and their relationship is a topic of current debate. Traditionally the foundation of a museum is its physical collection, with the core museum activities of collection, preservation, research and display being focussed around the physical objects in that collection. The growth of virtual museums, some of which have no physical counterpart, may cause unease in the traditional museum on three counts: it undermines the essential value of a physical collection; the physical visitors may become virtual visitors, reducing a valuable income source; the virtual visitor is distanced from the direct experience of the physical object. Teather (1998) and Teather and Wilhelm (1999) analyse this debate and discuss how patterns of visiting might be influenced by virtual museums, both positively and negatively. Schweibenz (1998) identifies virtual museums as part of a move towards a focus on information rather than on objects. Alsford (1997), Bowen *et al.* (1998), Quinn (1998) and Tinkler and Freedman (1998), among many others, suggest a variety of rationales for the creation of a museum Web site, from simply having a 'Presence' (which is often essentially just an electronic brochure, lacking any clear purpose), through supporting exhibitions currently on display, to acting as a social community space, promoting moderated debate and discussion, and allowing visitor participation and interaction (see for example Drennan *et al.*, 1998). Perhaps the principal dichotomy in the role of the virtual museums is between those which seek to

provide a genuine online visitor experience and those which seek instead to encourage and support physical visiting.

The precise role or roles of the Web site, and its relationship with the museum itself, are likely to follow a variety of different models, depending on the characteristics of the museum, its collection and its visitors. There will, however, be a need to define the purpose in having a Web site—the information provider needs—as few organisations can afford a Web site which serves to achieve no organisational objectives. The British Columbia Museums Association (BCMA, 1996) found that many museums were unable to measure in an objective way any benefits which might have been gained from being on the Internet. The understanding of the museum's needs must be coupled with an understanding of the visitors and their needs. Contradictions may arise between meeting the visitor needs and the other organisational objectives. For example, Futers (1997, reporting the results of Reynolds, 1997) found that 52% of virtual visitors expect to be able to download images from online exhibitions, the museum on the other hand wishes both to encourage paying visitors to view the physical exhibits and to gain financially from the Intellectual Property Rights it controls. Once these needs have been identified the site should be measured to determine whether the site is 'successful'. Defining success can be difficult, the most widespread measures of success are still relatively simple, number of page impressions, click through rate, the number of hits, etc. Whilst measures such as these are easy to define and to quantify, they are likely to be a largely unsatisfying measure with respect to measuring the achievement of organisational objectives. Bevan (1998) talks of "quality and usability goals which can be evaluated" and lists "demonstrate superiority of organisation to the competition, appropriateness of the Web site to users' needs, professionalism of the Web site, percentage of users who can find the information they need, ease with which users can locate information, number of accesses to key pages, percentage of users visiting the site who access key pages". Others might include the value of online sales, number of catalogue requests, the number of physical visits which result directly from a virtual visit, and so on.

Understanding the Virtual Visitor

If users are going to be at the centre of the virtual museum experience and the Web development process, then it is important that the nature of the user is understood as fully as possible. The importance of conducting 'visitor studies' has long been recognised by the museum community (see for example Booth, 1998b), and there have been several studies of gallery-based multimedia applications (see for example Economou, 1998; Borda *et al.*, 1999; Gammon, 1999). Virtual visitor studies, however, are not always considered as important, and Pierroux (1999) reports that 75% of the art museums surveyed had not conducted visitor research on their Web sites. Indeed, although it may appear tempting simply to assume that virtual visitors will have the same characteristics as physical visitors, this has yet to be established and there are many reasons to doubt such a straightforward equivalence.

A number of studies have reported demographic information about virtual visitors, for example Futers (1997), Thomas and Paterson (1998), and Chadwick and Boverie (1999). Whilst general information about the ages and backgrounds

of virtual visitors will have an impact on the content and design of a Web site, it is not clear whether it is possible to generalise from these studies or whether different museums will have highly individual demographics. Museums and museum Web sites vary in their purposes and intended visitors, and it is likely that the profile of the virtual visitor population will also vary between museum sites. It should be noted that in many of the surveys (particularly those based on the Web) the respondents were self-selecting, so may not form a representative sample of the underlying virtual visitor population.

How is a Virtual Visit Conducted

Chadwick and Boverie (1999) report that 69.8% of virtual visitors were visiting on their own, with family groups accounting for 21.6% and other groups making up the remainder. Group visits typically comprised two or three people. They found several groups with over 20 members, raising some interesting questions as to how exactly a virtual visit with such a large group is carried out. It is not clear whether these represent a group of people all sharing one computer in a collaborative visit or a group of people each with their own computer visiting together—perhaps in a classroom setting. Web design has generally been based around the model of a single user interaction, and the design of Web pages for group use is largely unexplored. Chadwick and Boverie also found a statistically significant difference between the number of pages visited by groups and individuals, with groups tending to access more, but no significant difference in the duration of a visit, averaging 20.78 minutes for individuals and 23.49 for groups. Nordbotten (2000) found that only 57% of virtual visitors looked at more than one topic page. Thomas and Paterson (1998) report that "On average visitors only looked at 1–2 other pages after the home page". It is not clear how to interpret these findings, perhaps some visits were accidental, perhaps the visitors were able to find the information they wanted very efficiently, or perhaps the site failed to engage the visitors interest. Making the connection between observed behaviour and the reasons underlying that behaviour is difficult, particularly where there is no additional information. Nordbotten found that 85% of virtual visits started with a search engine request, and of these 76% started at a content page rather than an index or home page. Only 13% began at the exhibition 'homepage', visits that did were significantly longer than those that did not, and richer in topic content visited. Futers (1997) reports that virtual visitors found out about virtual museum sites in a variety of ways: linked site 43%; by chance 37%; word of mouth 10%; article 8%; museum 3%; mailing/news group 2% (all percentages approximate). The need to support visitors who do not begin their virtual visit at the site homepage is clearly identified.

What do Virtual Visitors Expect

Several studies have investigated the purposes behind virtual museum visits (for example, Thomas and Paterson, 1998; Chadwick and Boverie, 1999) while others have concentrated on what the virtual visitors expect to find on the Web site and the perceived benefits of accessing such information (Futers, 1997). One of the difficulties in building up a clear picture of user expectations is the wide range and ambiguous nature of category descriptions used in different

surveys. Virtual visitors and their information requirements have typically been categorised in a similar way to that presented by Booth (1998a,b):

> Potential visitors—to help to plan visit to the Museum;
> Technical Enquirers—to provide facility for detailed enquiries;
> Schools enquiries—to provide access for schools to project-based information and to plan visits; and
> Virtual visitors—to provide access to those who are unable to visit in person.

Other researchers have provided a finer-grained breakdown of virtual visitor types, see for example Streten (2000). These finer-grained categories can be useful either for deciding the types of information to be included within the broader category or in developing specialised content for these finer grained categories of virtual visitor. When designing Web sites it may not be possible to meet the needs of all virtual visitor groups equally, the museum may need to decide who are its key visitors and focus on meeting their needs. It may also be the case that local and non-local visitors have very different needs, and again the museum will need to consider how best to target its resources.

The Development Process

Many museum Web developments are undertaken with very limited resources in terms of time, money and expertise. Hertzum (1998) identifies three typical models of the museum Web development team: a single person (who may or may not be familiar with Web development), a mixed team of museum professionals plus internal computing people (again who may or may not be familiar with Web development), and a mixed team of museum professionals in co-operation with external partners or consultants with Web development experience. Whilst there are many sets of Web guidelines which developers can follow, early guidelines tended to concentrate on Web *page* design. Many of these were based on personal preference and lacked empirical justification, others were derived from existing Human Computer Interface (HCI) guidelines, often with little consideration of whether or not they held true in a Web environment. Guidelines for the development of gallery-based multimedia applications (see for example Economou, 1998) may also be informative, though the degree to which it is possible to generalise them to the Web must be considered. Adherence to design guidelines is no guarantee of usability, so evaluation, particularly where it involves real users, is necessary (Shneiderman *et al.*, 1989). Usability must typically be considered at two interrelated levels—site-level usability and page-level usability (Nielsen, 1998a). With increased collaboration between museums, and the potential to combine online exhibits from different museums, the need to consider usability at an inter-site level may also arise. More recent publications (e.g. Flemming, 1998; Rosenfeld and Morville, 1998; Nielsen, 2000b) have sought to provide practical advice for Web *site* development, often including evaluation. However, there is still no single accepted approach to Web development and no defined set of methods for ensuring usability.

Based on published case studies of museum Web development practice, it is possible to characterise the non-professional museum Web development environment:

Limited Development Team

Oono (1998) reports that 6.3% of museum Web sites surveyed had no staff working on their site and 57.1% had only one member of staff working on their site, with a further 33.7% having 2–5. Futers (1997) and Streten (2000) recommend involving curatorial, marketing, education and research departments in the development and maintenance. The development team may therefore lack Web development experience and skills in evaluation methods, and may underestimate the importance of user-centred design. James (1997, cited in Teather, 1998) reports a survey of 33 museum Web sites in which 54% performed no testing with virtual visitors.

Limited Access to Virtual Visitor Population

As the site is aimed at external visitors, these will ideally form an important element of the evaluation process. Whilst it is possible to use physical museum visitors, this assumes that they are representative of the virtual visitor population. This assumption may not be correct as certain groups, such as foreign visitors or those with special needs, may be under-represented in the physical visitor population. There are also methodological issues involved in asking people to *act* as if they were virtual visitors.

Limited Resources

Oono (1998) reports that 70.2% of museum Web sites had a budget of less than $1000 per year (excluding staff costs). Hertzum (1998) reports several cases where museum Web development was conducted without being allocated a budget. Where no budget is available the team will not have the time or money to conduct expensive usability studies.

External Host

Oono (1998) reports that 73.3% of museum Web sites are hosted by some external company rather than on a server owned by the museum, this may have implications for the collection of usage logs, and the range of technologies permitted.

Whilst a number of 'formal' development models have been proposed (see Lowe and Hall, 1999 for an in depth overview), these are typically orientated towards 'industrial scale' applications and are generally inappropriate for untrained developers. A large number of 'informal' development models for Web sites have been proposed by different authors (for a review, see Cunliffe, 2000). These informal models adopt a user-centred, iterative approach, emphasising evaluation and user testing. There are a variety of different evaluation methods, each with advantages and disadvantages and suited to different stages of the development. Evaluation should be conducted throughout the process of Web development, from initial prototyping through ongoing long term development.

Evaluation

Evaluation is essential for determining whether a Web site is meeting the needs of its users. Whilst the application of evaluation methods cannot address all the challenges which the non-professional Web developer faces, the appropriate application of such methods should result in a Web site which is better orientated towards user needs and is less likely to contain severe usability problems. Evaluations can collect quantitative and/or qualitative data. Quantitative evaluation data can include details of time taken, errors noted, or affective measures, and lends itself to statistical processing. Qualitative descriptions of user behaviour are more difficult to analyse but can lead to richer understanding of the context and provide insights into reasons for user behaviour. The setting for the evaluation can vary along a continuum from a controlled usability lab equipped with one-way mirror, video recording and logging equipment to 'contextual' observations in a realistic workplace environment. The controlled setting may be suited to formal testing of specific hypotheses and investigating causal relationships among experimental variables, such as the effect of different types of menu design or typography. A workplace setting may reproduce more faithfully typical situations, such as frequent interruption and task switching, and may sometimes yield greater validity when attempting to generalise beyond the trial subjects to user populations.

Given the development environment outlined, practical constraints such as time and cost are likely to have a strong influence on the evaluations performed. Most of the methods currently being used for Web evaluation do not require specialist facilities and can be performed cheaply and relatively quickly. Shneiderman *et al*. (1989) suggest that "… testing need not be lengthy or burdensome. Four to six hours should be enough to give an indication of the problems and successes patrons have as well as provide ample suggestions for improvement." Nielsen (1998a) suggests that with experience and under certain conditions it takes two days to conduct user testing for a Web site and reports work by Molich and Gram which indicates that "utter beginners could complete a full Web usability project in less than one week." Nielsen's 'discount usability engineering' approach to developing usable systems (which is appropriate for Web developments) is described in detail in his book *Usability Engineering* (Nielsen, 1993) and his bi-weekly column 'The Alertbox: Current Issues in Web Usability' can be found on his Web site (Nielsen, 2000a).

Case Study

This case study concerns the resource-limited evaluation of a Web site providing information about an annual review journal, *The New Review of Multimedia and Hypermedia* (NRHM, 2000). It is a relatively small site, consisting of approximately one hundred and sixty pages, mainly paper abstracts. According to Oono (1998), museum Web sites are likely to contain only a small number of pages, 28.8% having less than 20 pages and almost half having less than 50 pages. No formal development model was used for the site though Bevan (1998) was taken as an informal guide. The development of the site was completed without any formal user input, relying on the intuitions of the information providers and competitive analysis (Nielsen, 1993) of existing Web sites to provide

user needs models. A summative evaluation of the site was then performed using a combination of evaluation methods selected according to the following factors (Kritou, 1998):

- Time was limited due to organisational constraints, since the evaluations would run for approximately two weeks, the methods selected must provide results within that time;
- The users of the Web site were spread across the globe;
- The most common tasks of the users needed to be tested; and
- The evaluation team consisted of one inexperienced person, and there was no access to specialist HCI testing facilities or equipment.

These factors have much in common with those of the non-professional museum Web development environment. Four evaluation techniques were selected: direct observation of usability tests, log analysis, on-line questionnaire, and heuristic evaluation. The following sections discuss the findings from the investigation, identifying the potential benefits and limitations of the different methods with reference to the museum context.

Direct Observation

Within the constraints of the evaluation, the gathering of a test group of real users proved problematic. In order to perform an evaluation with the five 'proxy users' who were available, a set of six task scenarios was constructed, each task scenario presented a specific task (Kritou, 1998):

> Assume that you are an author who wants to submit a paper and looks for an appropriate journal. You have found the NRHM Web-Site. Find out what topics can be submitted in the next issue and the guidelines for submitting a paper to this journal.

During the evaluation subjects were encouraged to 'think aloud', a video camera was used to record the users' utterances and on-screen behaviour. The evaluator also completed an observer's notebook during the evaluation. Ideally, direct observation will record the behaviour of the visitor performing natural tasks, in a natural context and under natural motivation. Museums have experience of observing visitors in exhibition areas. These observations may include subjective interpretation of facial expression, body language, snatches of conversation and interaction among group members, typically resulting in very rich descriptions and codings of visitor behaviour, "… small boy (6) seems awestruck by sounds/colors but soon fidgets as watches … The girl doing computer is quite relaxed, interested …" (Hilke *et al.*, 1988). When considering direct observation of virtual visitors the situation is more complex—it is typically not possible to observe virtual visitors in their natural context of use, which might be their home or office. The recording of virtual visitor behaviour in log files is one way this can be achieved and is discussed in Log Analysis. Generally direct observation of virtual visitors is carried out under the following limitations.

The context of use may be artificial:

> Typically direct observations will be carried out under controlled circumstances in a specially prepared room. In some cases observation subjects are

asked to imagine that they are in a particular environment, Hardman (1989) reporting on the evaluation of Glasgow Online kiosk writes "All the observations were carried out in an office, and readers were asked to imagine that they were currently at Queen Street Station, one of the main-line stations in central Glasgow". The NRHM observations were performed in the subjects' offices to provide a reasonably natural context. Even where the actual context of use is broadly similar to the observation situation, the visitor may behave differently in the knowledge that they are being observed—the 'Hawthorn effect'.

The tasks the subject carries out may be artificial:

Typically the observation subject will be provided with a set of tasks they must carry out. Often these tasks are based on the developers' perceptions of what constitute typical user tasks. One method for reducing the potential for bias in the creation of tasks is to involve users in their creation (Erskine *et al.*, 1997). It is also worth conducting a pilot test of the tasks themselves, for example in the NRHM evaluation a pilot test identified an ambiguity in one of the task scenarios. Heinecke (1995) suggests that the goals and intentions of 'point of information' system users in museums may be highly individualistic and that it is necessary to observe users in order to understand their goals. There is no reason to suppose that this will not also be true on the Web. The tasks will generally focus on information finding rather than browsing, as Hardman (1989) observes, "No observation looked at how readers would use the information available for browsing, since it would be very difficult to keep track of readers' goals and whether they had reached them or not". Since browsing is a natural behaviour on the Web, and one which a museum Web site would seek to encourage, this is an important limitation of such task-based observations. One interesting observation made during the NRHM evaluation was that a number of the subjects started the tasks by general browsing around the site. It may be that strict adherence to information finding tasks places an artificial restriction on their interactions and that browsing behaviour should be incorporated within the observations.

The motivation for the subject conducting the task may be artificial:

Differences in motivation between real users with real information needs and observation subjects with artificial motivations may result in misleading findings (Hardman, 1989). In the NRHM observations the evaluator reported that some of the test subjects experienced problems because they had not paid attention to the task requirements, perhaps reflecting lack of motivation when completing artificial tasks.

The observation subjects may not actually be part of the virtual visitor population:

In the NRHM evaluation few typical users were available, so use was made of 'proxy' users, raising doubts as to the validity of the behaviours observed. Also some of the reported problems appeared to be due to lack of familiarity with the domain rather than actual usability problems. In some cases distinguishing between the two may not be easy.

During a direct observation, as in a physical visitor observation, there are a variety of data which could be collected. This is influenced to an extent by

the facilities and staff available, balanced against the need not to intimidate the observation subjects. A very practical concern is the need to reassure users that they are not the focus of the evaluation but rather the system and that should they feel uncomfortable they may stop at any time.

Think-aloud protocol:

A think-aloud protocol can be used with direct observation, and in the NRHM evaluation a video camera was used to record the on-screen behaviour and spoken comments of observation subjects. This proved useful in understanding *why* the subject had or had not performed a particular action. However, all bar one of the subjects were uncomfortable with the thinking-aloud protocol and reported that it interfered with execution of the tasks.

Video:

The use of a video camera to capture the subjects' on-screen behaviour can be useful for creating an accurate and impartial record of that behaviour (see for example Cleary, 2000). The audio track can also be used to record a subjects' thinking-aloud, providing easy synchronisation with the on-screen behaviour. Whilst subjects are often uncomfortable in front of a camera, a camera directed at the screen is perceived as being less threatening. In the NRHM evaluation video analysis proved time consuming, but did enable the observer to identify things which might otherwise have been missed. Retrospective testing (Nielsen, 1993), in which the test subjects comment on the recorded session, might be used as a way of reducing analysis time.

Observer's note book:

The observer's note book can be used to record a number of different events. Borda *et al.* (1999) use observation sheets which attempt to generate a rich description, requiring the observer to "Describe step by step how visitor navigates and holds interest" and record "What do they say or ask each other/you?" There are issues involved with gathering rich descriptions including accurate recording, subjective interpretation, and subjective recording. In the NRHM evaluation the notebook was used to record the time taken for each task, whether the task was completed, whether the test subject followed the optimal path (see below) and the subject's affective state. In retrospect, the observer's notebook required the evaluator to take down too many details during evaluations.

Optimal path:

One event which was recorded by the observer in the NRHM evaluation (see also Badre and Jacobs, 1999) was deviance from the optimal path when completing a task—that is the path involving the fewest link traversals. It is important to consider whether this is an appropriate measure for a browsing activity, even when the subject is performing an information finding task. Adherence to an optimum path could indicate a failure to engage with the content being passed through.

Time to complete:

It is often tempting to record the time taken to complete a task or set of tasks during a direct observation. Indeed, during the NRHM evaluation, the time to complete a task was recorded. What is less clear is *why* this data would be useful. Firstly the time taken to complete a task is only of value if

there exists some other experimental condition, or some absolute performance target with which to compare it. Secondly, in an environment which is intended to facilitate browsing, the time taken to complete a task becomes relatively unimportant. Where the application is essentially an information retrieval task, speed of task completion may be more relevant. Given that speed of downloading is an issue for Web users, this may be a factor which could be assessed in an observation, but typically the experimental conditions do not replicate network traffic, caching, speed of the server under load, and so on.

Interviews:

Direct observation is often combined with an interview or questionnaire (see, for example, Borda *et al.*, 1999; Cleary, 2000). The NRHM direct observations were followed by a short interview during which the subjects were encouraged to comment generally on their experience and to express their subjective satisfaction. Where a task was not completed the subject was asked if they had expected to find the information elsewhere. If subjects appeared unsatisfied when they had completed a task, they were asked if they were expecting to find more or different information. The interviews identified several missed opportunities in terms of not meeting user expectations and were able to generate suggestions for improvement. All the subjects had previous experience using the Web and were familiar with the sorts of facilities which can be provided. It is not clear whether the feedback from more naive Web users would be so informative. The interviews also provided positive comments which confirmed the appropriateness of some design decisions.

Direct observation sessions can be performed reasonably quickly, depending on the nature of the tasks and whether or not an interview is performed. They are also relatively cheap, if observation subjects and access to computer facilities can be located cost free, the only real cost is the evaluator's time. Video cameras are commonly available, but are not essential for direct observation methods. Direct observation is ideally a method which uses real users and real tasks—real usability problems can be identified with only a small number of test subjects. The method is reusable and can be used to test initial concepts, prototypes and completed systems. The tasks and scenarios generated to support direct observation can also be used to guide inspection methods, see below in 'Heuristic Evaluation'.

Log Analysis

Several museum developers have stressed the importance of gathering Web statistics through log analysis (for example, Streten, 2000). The automatic collection and analysis of access logs appears to be a close analogy to observing visitors in an exhibition area, potentially gathering important information about real users and real patterns of use within a site and avoiding the problems of artificiality faced by direct observation. Heinecke (1995) suggests that methods such as direct observation and interviews are not cost effective for museums, and favours log analysis as a cheap approach. The NRHM evaluation collected Web access logs over a period of sixteen days. In addition to simple 'hits per page' information the evaluator was interested in trying to characterise particular patterns of use. This identified a number of problems with log analysis.

Difficulties in identifying visitors and defining sessions:

A session refers to the set of activities performed by a single visitor during a single visit to the site. However, Web log information identifies the originator of a request for a file by the IP address of the requesting computer. IP addresses are not necessarily unique to a person and may not be unique to a particular computer. This means that, unless there is a controlled experimental environment, it is not possible to generate a record of visitor activity over several sessions as it is impossible to be sure that sessions with the same IP address were conducted by the same user. It is also necessary to determine how long a period of inactivity will be taken to indicate the end of a session. The 'timeout' period for a Web site will generally be longer than that of an interactive, for example, twenty minutes in the NRHM evaluation. There are problems in determining the duration of the timeout period. When a different visitor accesses from the same IP address within twenty minutes of a previous visitor's session finishing, the two sessions will be treated as a single session by a single visitor. On the other hand if a visitor pauses on a single page for more than 20 minutes, perhaps because the content is particularly interesting or complicated, then their subsequent activity will form a separate session and will not be attributed to the same visitor. It is also not possible to determine when a group of visitors is sharing a machine during a session.

Caching:

The caching of Web pages is a part of the normal working of the Internet. However, if visitors retrieve cached pages rather than retrieving them from the server, the log of accesses no longer contains a full record of a visitor session.

Limited range of data captured:

Compared to the rich descriptions generated during the observation of visitors in an exhibition area, in the logging of interactions with an interactive or kiosk, or during some direct observations, the data gathered in a typical Web log is relatively poor. Essentially it comprises the IP address of the requesting computer, the file requested, and the time that the request was made. Some useful indicative data may be gathered, regarding hits on specific pages, country of origin of requests, requests for downloadable files, and so on. Streten (2000) suggests that log files are useful for analysing the most popular entry and exit points and paths through site.

In the NRHM logs it was possible to identify some patterns of behaviour which seemed to confirm the patterns of use hypothesised during the site development. However, given the problems with the log data and the subjective nature of their interpretation as user behaviours, it proved difficult to draw any definite conclusions. The analysis of the data is non-trivial and it is difficult to identify specific usability problems from the analysis of logs alone. The difficulties in mapping from *what the user did* to *why the user did it* are significant. Streten (2000) found that some assumptions made on the basis of access analysis were not backed up by questioning users and recommends that log analysis needs to be augmented by qualitative evidence about the user experience. Shneiderman *et al.* (1989) suggest that observation and interviews be used in addition to log analysis to gather information about problems and potential solutions. For

further discussion of the issues surrounding log analysis, see Berthon *et al*. (1997), Bertot *et al*. (1997) and Drott (1998).

In order to generate sufficient data for analysis, logging must run for some time depending on the number of visitors to the site. The logging can be run continuously with data analysis performed at regular intervals. The analysis of the data can be performed quickly if only simple information is required. More complex analysis of user behaviour is more time consuming. Comparison of data over the long term can be used to capture the dynamic changes in user behaviour over time.

Online Questionnaire and Feedback

Questionnaires are often used in visitor studies, and their familiarity suggests they would make an ideal tool for museum Web developers. In the NRHM evaluation an online questionnaire was created containing questions to gather three types of information: demographic information, technical information, and visit information. Demographic information included occupation, age, locality and Internet experience. Technical information included the type of browser used and the speed of their Internet connection. As the intended audience was expected to be technically literate, this type of question was appropriate. Where the users are technically naive they may not know the answers to them, potentially leading to non-completion of the questionnaire, guessed answers, or missing data. Visit information included the number of times the user had visited the site, their purpose in visiting the site, the page they entered the site on, pages they visited, how useful they found the pages they visited and which pages they bookmarked. These questions were intended to build up a more detailed picture of user behaviour. Users were also asked to rate their general satisfaction with the site and were offered the opportunity make any general comments. The questionnaire was linked to prominently from the NRHM homepage and email was sent to various mailing lists.

The two major concerns with on-line questionnaires and similar feedback mechanisms are the self-selecting nature of the sample and the response rate required to draw reliable conclusions. For the NRHM evaluation, if the total number of visitors is determined by unique IP address (assuming that each distinct IP address is a distinct visitor and that multiple sessions from the same IP address are the same visitor), then the responses represent approximately 5.9% response rate. However, if the number of visitors is determined by unique sessions (assuming that each distinct session is a distinct visitor, even where the IP address is the same), the responses represent only 2.3%. The true figure is likely to fall somewhere between the two extremes. This low response rate is not unusual for Web based questionnaires. Nordbotten (2000) reports a response rate of 4% and Thomas and Paterson (1998) report 2%. The importance of the absolute number of responses must also be considered, since the 2% response rate reported by Thomas and Paterson was made up of 215 responses, whereas the 4% reported by Nordbotten was made up of only five. In the NRHM evaluation there were only five responses, and it was considered impossible to draw reliable conclusions from the results. It is possible that this low response rate was in part due to the timing of the questionnaire, as it was run for only

a short time in August, a month when many of the target audience would have been on holiday.

Other difficulties with on-line questionnaires are discussed by Zhang (1999), and some useful design pointers can be found in Feinberg and Johnson (1998). Borland and Wongse-Sanit (1997) use both an online comments form and an email feedback link but question their value: "In general unsolicited comments have not been particularly useful as an assessment tool …. Many send us very nice comments and a few have found typographical errors that we missed." Such feedback may also be useful for detecting broken links within a site. One possible way to encourage completion of a questionnaire is to offer some form of inducement. Kiernan (1998, cited in Chadwick and Boverie, 1999) reports a 70% response rate to an online survey conducted by NASA when a free photograph was offered as an inducement. The problem with this approach is that it may simply change the nature of the self-selecting sample and people may visit the site simply to complete the questionnaire and receive the inducement. Questionnaires and feedback are relatively cheap methods, ideally suited to the maintenance stage of the development, though procedures must be put in place to deal with a continuous stream of responses. In the case of a questionnaire it may be beneficial to keep it on-line for a limited period, rather than run it continuously. The NRHM experience suggests that the timing of the release of a limited period questionnaire may have an impact on the response rate.

Heuristic Evaluation

An alternative to testing with users is to use methods that do not require users. Inspection based methods are a class of usability evaluation method which involve a systematic (but not necessarily exhaustive) comparison of a prototype or final design against a pre-determined set of criteria or guidelines (see Nielsen and Mack, 1994 for a detailed discussion). These methods typically require a degree of informed judgement on the part of the evaluator. The use of inspection methods in museum Web development is not widely reported in the literature, possibly because of concerns over the expense involved in using trained evaluators (Heinecke, 1995). However, Nielsen (1993) suggests that whilst evaluators should *ideally* have experience in performing such evaluations, useful results can be achieved by non-experts. It is likely that many developers are performing informal inspections but do not judge them worthy of reporting (for an exception see Hoff, 2000). Museum developments may benefit from a wider appreciation of inspection methods which will enable them to be placed on a more rigorous footing. More generally, inspection methods tend to be popular within Web development. Some of them require less formal training than other methods, they can be used throughout the development process, they do not require test users or specialist facilities, and they find a large number of usability problems (Sears, 1997). Concerns have however been raised regarding the number of false usability problems, i.e. those that actual users would not perceive as problems, which such methods identify (Bailey et al. 1992, cited in Gray and Salzman, 1998).

There are a variety of inspection methods, using different types of guidelines, some involving scenarios of use. For heuristic evaluation, for example, Nielsen (1993) recommends using around ten high level heuristics. The effective use of

high level, abstract heuristics, such as "Speak the users' language" (Nielsen (1993) requires a degree of professional judgement. Less experienced evaluators may find a larger, more detailed set of guidelines more appropriate. HCI research has produced a number of sets of high level heuristics (for example, Nielsen, 1994) and low level guidelines (for example, Smith and Mosier, 1986; Brown, 1998). More specialised sets of evaluation criteria also exist, for example CIDOC have developed a set of 'Multimedia Evaluation Criteria' (CIDOC, 1997) and Barker and King (1993) have developed an evaluation check list and supporting notes for the evaluation of interactive multimedia courseware. However, the direct application of a set of guidelines to Web sites is not always appropriate or sufficient. Even where an existing set of guidelines seems well suited to an application, for example Campbell and Wells (1996) criteria for Museum Web site homepages, the appropriateness of the guidelines for the *specific* site being evaluated must be considered. Levi and Conrad (1996) present an informative case study of the heuristic evaluation of a Web site.

Systematic Usability Evaluation (SUE) is an inspection-based method which has been applied to museum Web sites (Garzotto *et al*. 1995, 1998; Garzotto and Matera, 1997). In the inspection phase SUE uses a set of specific abstract tasks which codify the inspection patterns and behaviours of experienced evaluators in a form that can be applied by non-expert evaluators. It also integrates inspection methods with empirical user testing in order to verify the results of the inspection, thus reducing false usability problems. Whilst the philosophy of the approach is to support novice inspections, currently SUE involves a preparatory phase requiring the use of a formal hypermedia model, typically not appropriate for non-expert evaluators. However, work is ongoing and the aim is to develop a method which yields high levels of inter-evaluator consistency.

In the NRHM evaluation a heuristic evaluation was performed by an untrained evaluator, using a small set of high-level heuristics judged to be appropriate to the site. The evaluator reviewed a number of well-established principles of usability and Web site design in order to derive a set of eight high level heuristics. Each heuristic was described by a heading (shown below) and a short paragraph highlighting key concerns. The evaluator then performed a comprehensive page by page inspection of the entire site.

1. Consistency and conformance to standards.
2. Recognition and predictability.
3. Web pages should stand alone.
4. Flexibility and efficiency of use.
5. Effectiveness.
6. Readability.
7. Every page should express one topic or concept.
8. Consider the global audience.

Generally, inspection methods should be carried out by a number of evaluators, and for heuristic evaluation Nielsen (1993) suggests about five and at least three. The use of fewer evaluators raises a number of issues, including subjectivity in the selection of heuristics, subjectivity in the application of heuristics, and the possibility of usability problems being missed.

The time taken to apply an inspection method depends on the approach taken and the number of pages inspected. One way of addressing this is to use key-

user scenarios to guide a heuristic walkthrough and to identify which usability problems are likely to have the most impact on users. The use of scenarios to guide heuristic evaluation may also reduce the number of false usability problems identified. Other approaches to reducing the number of pages which need to be inspected are to focus on key pages (Nielsen, 1998a) e.g. download pages, registration pages, product purchase pages, or to identify repeated page templates and inspect only one of each type.

Comparison

In the NRHM evaluation a total of twenty-two potential usability problems were identified by the evaluation methods used: 1 by the questionnaire, 2 from log analysis, 12 from direct observation, and 13 from heuristic evaluation (see Table 1).

The usability problem identified by the online questionnaire "Search facilities are not efficient enough" was contained as a general comment rather than in answer to a specific question. The two usability problems identified by log analy-

Table 1 NRHM Comparison of evaluation methods

Description of the problem	HE	DO	LA	OQ
Search facilities are not efficient enough	x	x	x	x
Index is not consistent across the Web site	x	x	x	
'Return to top' is not used consistently	x			
There are four broken links in the Web site	x			
There is no Help facility	x			
Search facility not visible Instantly/Grouping of links not effective		x		
Instructions to Authors link not available next to each submission theme		x		
A Volume page in Hypermedia Journal doesn't link back to the Volume	x			
Pages don't provide the creator, date of creation, update and copyright	x			
The title (in ⟨TITLE⟩ tag) is not always representative of the Web site	x			
In a small screen the Home Page is too long	x			
As the Web site gets larger the Index will be very long		x		
The layout is too simple and not inviting		x		
Hypermedia Journal not visible in Volume Contents in a small screen	x			
The distinction between the two journals is not emphasised enough	x	x		
The author's address is not available when clicking on his name		x		
The difference between Editors and Editorial Board is not explained		x		
Subscription information is insufficient		x		
The list of papers under an author's name is not numbered		x		
The purpose of the Web site is not stated	x			
The 'no abstract available' message causes confusion	x			
There are no instructions on how to get a full paper		x		

(HE=Heuristic Evaluation, DO=Direct Observation, LA=Log Analysis, OQ=Online Questionnaire).

sis, — "Search facilities are not efficient enough" and "Index is not consistent across the Web site" — relied heavily on the evaluators' subjective interpretation. Questionnaires can also be useful for determining whether a site is meeting the general needs of its users and whether they are satisfied with the experience. This suggests that their major role may be in the long term development and maintenance of a site.

There was relatively little overlap between the usability problems identified by direct observation and those identified by heuristic evaluation. Direct observation proved useful in identifying usability problems relating to user needs: "There are no instructions on how to get a full paper"; and the user's subjective impression: "The layout is too simple and not inviting". The usability problems identified by direct observation include some probable false positives such as "The difference between Editors and Editorial Board is not explained". This is due to using proxy users as test subjects rather than real users. It is also possible that the use of proxy users resulted in some false negatives: they may have missed usability problems which real users would have encountered. This emphasises the benefits of including real users in the evaluation process.

Heuristic evaluation identified the largest number of potential usability problems, and generally identified problems which were not detected in the direct observations, such as "*Return to top* is not used consistently". Some of the problems identified, such as "The *no abstract available* message causes confusion" were rare and localised and were only found because the inspection was comprehensive. If a scenario-based inspection had been performed it is unlikely that they would have been discovered. The heuristic evaluation also identified potential false positives, such as "There is no Help facility". This reflects the difficulty and potential subjectivity in selecting and applying appropriate heuristics.

The purpose in identifying usability problems is that they can then be rectified. However, in many cases it may not be cost effective to rectify all the problems which have been identified. In order to make an informed choice, some form of severity ranking is necessary (Nielsen, 1999b). Often this involves ranking by, and agreement between experienced evaluators. Methods for placing this on a systematic footing, so that non-expert evaluators can perform this activity effectively, have yet to be developed.

There are a number of other evaluation methods which can be used. Borland and Wongse-Sanit (1997) use online focus groups who respond to detailed questions about various aspects of site. They also monitor contributions to an email discussion group of people interested in the topic area and have occasionally sent questionnaires to the list, though still with a disappointing response rate. For other methods and more general discussion refer to HCI texts dealing with evaluation issues (for example, Nielsen, 1993; Shneiderman, 1998). There is also a large quantity of useful information available on the Web. Instone (2000) maintains 'Usable Web' a searchable, categorised collection of over 1000 links to Web resources focussing on usability and the Web.

Maintenance and Long-Term Development

A Web site is likely to be a dynamic entity, particularly if it wishes to attract return visitors (Haapalainen, 1999). The commitment of resources to mainte-

nance and long-term development is an important area which is often not fully appreciated during the development of a Web site, particularly when that Web site is initially developed by an enthusiastic individual. Lu and Yeung (1998) identify underestimating the cost of maintaining a Web site as a common mistake in Web development and poor maintenance can lead to a decline in site quality and usability over time. The British Columbia Museums Association (BCMA, 1996) mentions "Out of date information" as a common serious fault with cultural heritage Web sites, which is perhaps not surprising in the light of another survey which reports that 6.3% of museum sites had no staff working on them (Oono, 1998). The amount of maintenance required can be reduced through consideration of maintenance issues during the Web site design process and the selection of content with a long 'shelf life' or which is suitable for reuse. Dating of content, technological change, change in visitor needs and expectations, and policy changes within the museum will all drive the continual development of a site and point to a need for continuing evaluation.

Long-term Web site development appears to be characterised by periods of 'evolution' when relatively small changes are made and new material is added within the existing site structure, and periods of 'revolution' when a site undergoes radical change—perhaps in its 'look and feel', or in its organisational or navigational structure—which Haapalainen (1999) refers to as "a sort of paradigm shift". Bowen (1999) recommends that a major overhaul ('revolution') be carried out roughly every two to three years, though the grounds for making this recommendation are not clear. Evaluation has a role in ensuring that such a major overhaul is user-centred, perhaps through the comparative evaluation of new designs against old. Bowen recommends updates ('evolution') on a monthly basis, or more frequently for larger sites. However, this may depend on the type of content provided on the site. Questionnaires, feedback and log analysis can be used during these periods to track longer term changes in user behaviour and to gather user preferences for future major overhauls. Evaluation may also be useful for determining when a major overhaul is necessary, as minor updates can lead to a decrease in usability over time. Oono (1998) has reported on the frequency with which museum Web sites renewed their homepage (it is not clear whether this in fact just refers to the homepage or to the site as there is some ambiguity in the wording of questions): every day 9.0%, once a week 23.9%, once a month 21.3%, every 2 months 24.9%, and once a year 9.0%, and rarely renew 6.3%. No indication is given as to how substantial this renewal is, or what information is actually renewed.

Procedures for checking the currency of content, keeping links updated, adding, deleting and revising materials, and content and presentation quality control must be considered. Additional information on the organisation of files and directories on the server, file and directory naming conventions, standardised layouts, and so on, also need to be recorded and enforced. This is particularly important where the site is maintained by people who were not part of the original development process, or where there may be problems with multiple maintainers (Besser, 1995). The Canadian Museum of Civilization addresses these problems through the use of a set of guidelines developed specifically for their Web site (CMCweb). These guidelines contain three main components (Alsford, 1997):

1. The first provides a common frame of reference by identifying the vision for CMCweb, including the key goals…;
2. The second sets out the process of content development, in terms of roles and responsibilities and the steps involved in obtaining approvals and applying quality control; and
3. The third defines 'house style', providing advice on various design principles (e.g. page length, image size), indicating the common elements that must appear on every page (notably dealing with corporate identification, navigation, and document dating), and indicating coding and file-naming conventions that are important to managing the site and facilitating its maintenance.

Continuing periodic evaluation should form part of the maintenance strategy and procedures for collecting, comparing and analysing data over long periods need to be established and documented. It is likely that the adoption of certain Web technologies (such as Cascading Style Sheets, content reuse through XML, or dynamic generation of Web pages from a database) will reduce some maintenance tasks. They will typically require a higher degree of technical knowledge than is required to produce simple Web pages and are likely to introduce new maintenance problems of their own.

Conclusions

The number and size of museum Web sites can be expected to increase in the future. The importance of the Web for museums and of museums as online content providers will grow. Whilst some Museum sites will be well resourced developments involving people with Web development experience, many will be non-professional in-house developments. Authors in the museum Web development field are calling for a more professional and visitor-centred approach to Web development. At the core of this must be a commitment to evaluation and, in particular, evaluation with users. Non-professional developers can adopt methods and practices which will lead to a more principled development process and a more user-centred and usable product. In many cases these evaluation methods provide useful results while requiring relatively few resources.

This paper has reviewed current museum Web development practice, particularly the non-professional development environment. A case study illustrating Web evaluation issues has been discussed, drawing on research into evaluation methodologies from a computing perspective. Four evaluation methods have been examined in detail: direct observation, log analysis, online questionnaires and feedback, and inspection methods. The experience of applying each method, and the benefits and limitations of each method, have been analysed. The case study suggests that a combination of methods is most effective. Direct observation and heuristic evaluation proved useful in identifying specific usability problems, but there was relatively little overlap between the usability problems identified by the two methods. Direct observation proved useful in identifying usability problems relating to user needs and the users' subjective impressions whilst heuristic evaluation identified more technically orientated problems. The questionnaire and log analysis provided useful information on actual user behaviour as against behaviour under evaluation conditions, but were less suited to detecting usability problems. They can play an important role in the long-term development and maintenance of a site. Whilst each of the

methods used can potentially produce useful information, it is important to recognise that each method also has limitations. The use of an inappropriate method, or of a method under inappropriate conditions or assumptions, is likely to result in misleading or incomplete results. It is worth noting that the evaluation techniques presented are not restricted to new developments; they can be applied equally well to existing sites.

There are additional areas for evaluation which will become more important as museums attempt to open up their collections databases to the Web public. Dawson and McKenna (1998), in a survey of 290 United Kingdom Museums, found that whilst 73% had collections databases, only 6% had an on-line database. Opening up collections databases is non-trivial, many of the tools used by museum professionals will not be appropriate for members of the general public. In order to meet the needs of this new audience, it is necessary to understand their needs and the forms in which those needs are likely to be expressed (Morrison, 1998). There is also a need to understand the dynamics of the search process, only part of which is on-line, and which may include a series of search activities mixed with browsing activities. The ability to support imprecise queries with poorly defined information needs and new forms of query—such as query by similarity, querying across non-text media, and the provision of browsing interfaces—information retrieval applications are likely to form an important part of the new tools for public access (see for example, Cunliffe *et al.*, 1997). These new forms of interaction will require new forms of evaluation.

Evaluation should be a central part of the ongoing process of Web development, from initial conception to long term maintenance and development. Evaluation is essential for determining whether a Web site is successfully meeting the needs of its users. It therefore plays an important role in ensuring that museum Web sites are providing a genuinely visitor-centred experience.

Acknowledgements

Thanks are due to both Stephen Alsford, Webmaster at the Canadian Museum of Civilization, and the National Museum of Science and Industry (NMSI) for providing access to internal reports. Work supporting this paper was conducted in the HEFCW funded Internet Technologies Research Lab in the School of Computing, University of Glamorgan. An early version of this work was presented at the mda/CIDOC conference, 1999.

Bibliography

Alsford, S. (1997) *From pilot to program: organizing content creation as a web site grows*. In Proceedings of Museums and the Web '97, Los Angeles, USA, 1997. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Badre, A. and Jacobs, A. (1999) *Usability, aesthetics and efficiency: an evaluation in a multimedia environment*. In Proceedings IEEE Multimedia Systems, ICMCS'99, Florence, Italy, 1999:103–106.

Barker, P. and King, T. (1993) Evaluating interactive multimedia courseware—a methodology. *Computers in Education*, **21**(4), 307–319.

BCMA (1996) *Report on cultural heritage institution marketing on the internets*. Report by BC Museums Association (Dogwood Regional Network) and the Canada-BC Agreement on Culture and Communications, 1996. Available at <www.museumassn.bc.ca/report.html> (consulted 24 August 2000).

Berthon, P., Pitt, L. and Prendergast, G. (1997) Visits, hits, caching and counting on the world wide

web: old wine in new bottles? *Internet Research: Electronic Networking Applications and Policy*, **7**(1), 5–8.

Bertot, J., McClure, C., Moen, W. and Rubin, J. (1997) Web usage statistics: measurement issues and analytical techniques. *Government Information Quarterly*, **14**(4), 373–395.

Besser, H. (1995) Multimedia and networks teach about museums. In *Multimedia computing and museums—selected papers from ICHIM'95 and MCN'95*, ed. D. Bearman, Archives and Museum Informatics, Pittsburgh, PA, pp. 124–140.

Bevan, N. (1998) *Usability issues in web site design* (version 3, April 1998). Available via <www.usability.serco.com/netscape/index.html> (consulted 24 August 2000).

Booth, B. (1998) Information for visitors to cultural attraction? *Journal of Information Science*, **24**(5), 291–303.

Booth, B. (1998) Understanding the information needs of visitors to museums. *Museum Management and Curatorship*, **17**(2), 139–157.

Borda, A., Romans, J. and Payne, J. (1999) *Evaluation of prototype exhiblets*. Science Museum Report, London: Science Museum, 1999.

Borland, C. and Wongse-Sanit, N. (1997) *ArtsEdNet: assessing an arts education web site*. In Proceedings of Museums and the Web '97, Los Angeles, USA, 1997. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Bowen, J. (1999) *Time for renovations: a survey of museum web sites*. In Proceedings of Museums and the Web '99, New Orleans, USA, 1999. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Bowen, J. (2000) The website of the UK museum of the year, 1999. In *Proceedings of Museums and the Web '00*, Minneapolis, USA (2000) also available via >www.arhimuse.com/> (consulted 24 August 2000).

Bowen, J., Bennett, J. and Johnson, J. (1998) *Virtual visits to virtual museums*. In Proceedings of Museums and the Web '98, Toronto, Canada, 1998. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Brown, C. (1998) *Human–computer interface design guidelines*. Intellect Books, Bristol.

Campbell, H. and Wells, M. (1996) Assessment of museum world wide web home page format. *Visitor Studies: Theory, Research and Practice*, **9**, 216–226.

Chadwick, J. and Boverie, P. (1999) *A survey of characteristics and patterns of behaviour in visitors to a museum web site*. In Proceedings of Museums and the Web '99, New Orleans, USA, 1999. Also available via <www.archimuse.com/> (consulted 24 August 2000).

CIDOC (1997). *Multimedia evaluation criteria*. ICOM CIDOC Draft Report, 1997. Available at <www.archimuse.com/papers/cidoc/cidoc.mmwg.eval.crit.html> (consulted 24 August 2000).

Cleary, Y. (2000) *An examination of the impact of subjective cultural issues on the usability of a localized web site—the Louvre museum web site*. In Proceedings of Museums and the Web '00, Minneapolis, USA, 2000. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Cunliffe, D. (2000) Developing usable web sites—a review and model. *Internet Research: Electronic Networking Applications and Policy*, **10**(4), 295–307.

Cunliffe, D., Taylor, C. and Tudhope, D. (1997) *Query-based navigation in semantically indexed hypermedia*. In Proceedings 8th ACM Conference on Hypertext - Hypertext '97, 1997:87–95.

Dawson, D. and McKenna, G. (1998) mda survey of information technology in museums 1998. *mda Information*, **4**(1), Available at <www.mdocassn.demon.co.uk/info41.htm> (consulted 24 August 2000).

Day, A. (1997) A model for monitoring web site effectiveness. *Internet Research: Electronic Networking Applications and Policy*, **7**(2), 109–115.

Drennan, C., Nash, C. and Bonaventura, P. (1998) *The web as art space, museum and forum in Tumblong*. In Proceedings 4th International Conference on Virtual Systems and Multimedia, VSMM 98. Gifu, Japan, 1998:275–280.

Drott, M. (1998) *Using web server logs to improve site design*. In Proceedings ACM SIGDOC'98, Quebec, Canada, 1998:43–50.

Economou, M. (1998) The evaluation of museum multimedia applications: lessons from research. *Museum Management and Curatorship*, **17**(2), 173–187.

Erskine, L., Carter-Tod, D. and Burton, J. (1997) Dialogical techniques for the design of web sites. *International Journal of Human-Computer Studies*, **47**, 169–195.

Feinberg, S. and Johnson, P. (1998) *Designing and developing surveys on WWW sites*. In Proceedings 6th ACM Annual Conference on Computer Documentation, 1998, pp. 38–42.

Flemming, J. (1998) *Web navigation: Designing the user experience*. O'Reilly, Sebastopol, CA.

Futers, K. (1997) *Tell me what you want, what you really, really want: a look at internet user need*. In EVA'97 Conference, 1997. Available at <www.open.gov.uk/mdocassn/eva-kf.htm> (consulted 24 August 2000).

Gammon, B. (1999) *Visitor's use of computer exhibits: general findings from 5 years of watching visitors getting it wrong*. Science Museum, Welcome Wing Exhibition Development Report, London, 1999. Also available in *Informal Learning Review*, **38** September/October 1999.

Garzotto, F., Mainetti, L. and Paolini, P. (1995) User interaction styles in museum hypermedia. In *Multimedia computing and museums—selected papers from ICHIM'95 and MCN'95*, D. Bearman (ed.), Archives and Museum Informatics, Pittsburgh, PA, pp. 217–234.

Garzotto, F. and Matera, M. (1997) A systematic method for hypermedia usability inspection. *The New Review of Hypermedia and Multimedia*, **3**, 39–65.

Garzotto, F., Matera, M. and Paolini, P. (1998) *To use or not to use? Evaluating usability of museum web sites*. In Proceedings of Museums and the Web '98, Toronto, Canada, 1998. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Gray, W. and Salzman, M. (1998) Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction*, **13**, 203–261.

Haapalainen, R. (1999) Revising the Finnish National Gallery's web pages. In *Proceedings of Museums and the Web '99*, New Orleans, USA, 1999. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Hardman, L. (1989) Evaluating the usability of the Glasgow online hypertext. *Hypermedia*, **1**(1), 34–63.

Heinecke, A. (1995) Evaluation of hypermedia systems in museums. In *Multimedia computing and museums—selected papers from ICHIM'95 and MCN'95*, ed. D. Bearman, Archives and Museum Informatics, Pittsburgh, PA, pp. 67–78.

Hertzum, M. (1998) A review of museum web sites: in search of user-centred design. *Archives and Museum Informatics*, **12**, 127–138.

Hilke, D., Hennings, E. and Springuel, M. (1988) The impact of interactive computer software on visitors' experiences: a case study. *International Laboratory for Visitor Studies Review*, **1**(1), 34–49.

Hoff, K. (2000) *The small museum web site: a case study of the web site development and strategy in a small art museum*. In Proceedings of Museums and the Web '00, Minneapolis, USA, 2000. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Instone, K. (2000) *Usable Web*. <www.usableweb.com/> (consulted 24 August 2000).

James, S. (1997) *Museum web page survey result*. MSc Dissertation, (Toronto: Department of Museum Studies, University of Toronto, 1997).

Kritou, E. (1998) *A case study of web usability evaluation*. MSc Dissertation, (Pontypridd: School of Computing, University of Glamorgan, 1998).

Levi, M. and Conrad, F. (1996) *A heuristic evaluation of a world wide web prototype*, ACM Interactions, III. 4 July and August 1996:50–61.

Lowe, D. and Hall, W. (1999) *Hypermedia and the web: an engineering approach*. Wiley, Chichester.

Lu, M. and Yeung, W. (1998) A framework for effective commercial web application development. *Internet Research: Electronic Networking Applications and Policy*, **8**(2), 166–173.

Morrison, I. (1998) SCRAN and its users. *The New Review of Hypermedia and Multimedia*, **4**, 255–260.

Murugesan, S. (1999) Web engineering. *ACM SIGWEB Newsletter*, **8**(3), 28–32.

Nielsen, J. (1993) *Usability engineering*. AP Professional, London.

Nielsen, J. (1994) Heuristic evaluation. In *Usability inspection methods*, J. Nielsen and R. Mack (eds), John Wiley & Sons, Chichester, pp. 25–62.

Nielsen, J. (1998a) *Cost of user testing a website*. Alertbox for May 3, 1998, <www.useit.com/alertbox/980503.html> (consulted 24 August 2000).

Nielsen, J. (1998b) *Failure of corporate websites*. Alertbox for October 18, 1998, <www.useit.com/alertbox/981018.html> (consulted 24 August 2000).

Nielsen, J. (1999) User interface directions for the web. *Communications of the ACM*, **42**(1), 65–72.

Nielsen, J. (1999b) *Severity ratings in heuristic evaluation*. 1999. <www.useit.com/papers/heuristic/severityrating.html> (consulted 24 August 2000).

Nielsen, J. (2000a) useit.com: Jakob Nielsen's Website, <www.useit.com> (consulted 24 August 2000).

Nielsen, J. (2000b) *Designing web usability: the practice of simplicity*. New Riders, Indianapolis, IA.

Nielsen, J. and Mack, R. (1994) *Usability Inspection Methods*. John Wiley & Sons, Chichester.

Nordbotten, J. (2000) *Entering through the side door—a usage analysis of a web presentation*. In Proceedings of Museums and the Web '00, Minneapolis, USA, 2000. Also available via <www.archimuse.com/> (consulted 24 August 2000).

NRHM (2000) *The new review of hypermedia and multimedia*. <www.comp.glam.ac.uk/NRHM/> (consulted 24 August 2000).

Oono, S. (1998) *The world wide museum survey on the web*. Internet Museum, Japan, 1998. <www.museum.or.jp/IM—english/f-survey.html> (consulted 24 August 2000).

Pierroux, P. (1999) *Art in networks: information and communication technology in art museums*. Department of Media and Communication Technical Report, Oslo: University of Oslo, 1999. Also available at <www.media.uio.no/internettiendring/publikasjoner/tekst/Pierroux/02Contents.html> (consulted 24 August 2000).

Quinn, C. (1998) *Why are we here?* In Proceedings of Museums and the Web '98, Toronto, Canada, 1998. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Rees, J. (1998) On the subject of intellectual property rights (IPR), including copyright—a help and a hindrance to cultural heritage? *The New Review of Hypermedia and Multimedia*, **4**, 215–244.

Reynolds, R. (1997) *Museums and the internet: what purpose should the information supplied by museums on the world wide web serve?* MSc Dissertation, (Leicester: Department of Museum Studies, Leicester University, 1997).

Rosenfeld, L. and Morville, P. (1998) *Information architecture for the World Wide Web*. O'Reilly, Sebastopol, CA.

Schweibenz, W. (1998) *The 'virtual museum': new perspectives for museums to present objects and information using the internet as a knowledge base and communication system*. In Proceedings of ISI'98, Internationalen Symposiums für Informationswissenschaft, Prague, 1998.

Sears, A. (1997) Heuristic walkthroughs: finding the problems without the noise. *International Journal of Human-Computer Interaction*, **9**(3), 213–234.

Shneiderman, B. (1998) *Designing the user interface*. Addison Wesley, Reading, MA.

Shneiderman, B., Brethauer, D., Plaisant, C. and Potter, R. (1989) Evaluating three museum installations of a hypertext system. *Journal of the American Society for Information Science*, **40**(3), 172–182.

Smith, S. and Mosier, J. (1986) *Guidelines For designing user interface software*. Mitre Corporation Report MTR-9420. Mitre Corporation, 1986.

Sphaera (1999) Real visits and virtual visits. *The Newsletter of the Museum of the History of Science*, Oxford, issue 10 Autumn (1999), available at <www.mhs.ox.ac.uk/sphaera/index.htm?issue10/articl10> (consulted 24 August 2000).

Streten, K. (2000) *Honoured guests—towards a visitor centred web experience*. In Proceedings of Museums and the Web '00, Minneapolis, USA, 2000. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Teather, L. (1998) *A museum is a museum is a museum … or is it?: exploring museology and the web*. In Proceedings of Museums and the Web '98, Toronto, Canada, 1998. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Teather, L. and Wilhelm, K. (1999) *Web musing: evaluating museums on the web from learning theory to methodology*. In Proceedings of Museums and the Web '99, New Orleans, USA, 1999. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Thomas, N. and Paterson, I. (1998) *Science museum web site assessment*. Research Report, Solomon Business Research, 1998. <www.nmsi.ac.uk/eval/index.htm> (consulted 24 August 2000).

Tinkler, M. and Freedman, M. (1998) *Online exhibitions: a philosophy of design and technological implementation*. In Proceedings of Museums and the Web '98, Toronto, Canada, 1998. Also available via <www.archimuse.com/> (consulted 24 August 2000).

Trant, J. (1999) When all you've got is 'the real thing': Museums and authenticity in the networked world. *Archives and Museum Informatics*, **12**(2), 107–125.

Zhang, Y. (1999) Using the internet for survey research: a case study. *Journal of the American Society for Information Science*, **51**(1), 57–68.