Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Library & Information Science Research 28 (2006) 433–452
Library & Information Science Research

Evaluation of digital libraries: Criteria and problems from users' perspectives

Hong (Iris) Xie

School of Information Studies, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201, USA

Available online 22 August 2006

Abstract

Digital library research has developed over the past decade, but little has been done on the identification of evaluation criteria, especially from users' perspectives. This article identifies users' criteria and applies them to the evaluation of existing digital libraries. Forty-eight subjects were instructed to develop and justify a set of essential criteria for the evaluation of digital libraries. At the same time, they were requested to evaluate existing digital libraries by applying the criteria that they were developing. A compilation of criteria developed by participants show that usability and collection quality were the most important criteria for evaluating digital libraries. Service quality, system performance efficiency, and user opinion solicitation were also deemed essential criteria. The author further compares digital library evaluation criteria identified by users and researchers and applied in previous research. The article reveals problems in current digital library design and development, discusses suggestions for improving digital library design, and presents areas for further research.

1. Introduction

The emergence of digital libraries provides more opportunities for users to access a variety of information resources. There are different definitions in the literature as to what constitutes a digital library. Chowdhury and Chowdhury (2003) place digital libraries into two major categories based on Borgman's (1999) discussion of competing visions. One approach focuses

E-mail address: hiris@uwm.edu.

0740-8188/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.lisr.2006.06.002

on access and retrieval of digital content and the other focuses on the collection, organization, and service aspects of digital resources. The Association of Research Libraries (1995) identifies the common elements of digital library definitions: the digital library is not a single entity; the digital library requires technology to link the resources of many; the linkages between many digital libraries and information services are transparent to the end users; universal access to digital libraries is a goal; and digital library collections are not limited to document surrogates, they extend to digital artifacts that cannot be represented or distributed in printed formats. Digital library research has developed rapidly over the past decade, and millions of dollars have been spent on building digital libraries. However, previous research indicates that many potential users may still not use them (Thong, Hong, & Tam, 2002).

Research on the evaluation of digital libraries is still in its infancy. Researchers are still investigating who should evaluate, when to evaluate, what to evaluate, how to evaluate, and why to evaluate. As Saracevic and Covi (2000) point out, the evaluation of digital libraries is a complex undertaking that is conceptually and pragmatically challenging. Borgman, Leazer, Gilliland-Swetland, and Gazan (2001) further suggest that technical complexity, variety of content, uses and users, and the lack of evaluation methods contribute to the problem. Any evaluation is based on the conceptual model of the evaluators: their understanding of the goals of the system and of users' needs and behaviors. Evaluation itself is a form of sense making, and it is also situated (Van House, Butler, Ogle, & Schiff, 1996).

As to who should evaluate digital libraries, users of digital libraries should have their voices heard. After all, the ultimate goal of the development of digital libraries is to serve users and to facilitate their effective use of information and services. Marchionini, Plaisant, and Komlodi (1998) emphasize that all efforts to design, implement, and evaluate digital libraries must be rooted in the information needs, characteristics, and contexts of the people who may use those libraries. Research on digital libraries has moved from the technical aspects of building digital libraries to how to design digital libraries to satisfy user needs. One way to assess user needs is to investigate digital library evaluation criteria from the user point of view.

2. Research problem

Although published research on digital libraries has increased, it mostly focuses on technical issues and digital library use patterns. In order to improve the design of existing and future digital libraries, there is need to identify what criteria need to be applied in the evaluation process. Little has been done on the identification of evaluation criteria. Moreover, current evaluation criteria are identified by researchers, not by users. Few studies have applied users' digital library evaluation criteria to the evaluation of existing digital libraries. This study investigates digital library evaluation from the users' perspectives, with a focus on the following questions:

- What criteria do users identify as important for the evaluation of digital libraries?
- What are the problems with the existing digital libraries?

3. Literature review

An evaluation is a judgment of worth. A system is evaluated to ascertain the level of its performance or its value. Digital libraries can be judged by their effectiveness (how well does a system or any of its parts perform the roles or tasks for which it was designed?) and efficiency (at what cost?) (Chowdhury & Chowdhury, 2003; Saracevic & Covi, 2000). Marchionini (2000) further points out that evaluation is a research process that aims to understand the meaning of some phenomenon situated in a context and changes that take place as the phenomenon and the context interact. Evaluation specifies what is the research process (metrics and procedures), what is the phenomenon (its mission and salient characteristics), and the context(s) in which the phenomenon occurs. Evaluation of a new phenomenon, such as digital libraries, is affected by the existing criteria for evaluating related institutions or systems. Bertot and McClure (2003) propose a framework to relate traditional evaluation components and terminology of library services/resource assessment to the networked environment. This literature review mainly addresses the research on digital library evaluation criteria.

The majority of research on digital library evaluation focuses on how users use a digital library, essentially usability studies, to either recommend design principles or improve the existing design. Nielsen (1993) point out that usability has multiple components, and he summarizes the five usability attributes: learnability, efficiency, memorability, errors, and satisfaction. These attributes, especially user satisfaction, have been investigated in many digital library usability studies. Van House et al. (1996) discussed the iterative design process for the University of California Berkeley Electronic Environmental Library Project. After observing and interviewing users about design elements (query form, fields, instructions, results displays, and formats of images and texts), they enhanced the design of the digital library. Bishop et al. (2000) presented the nature and extent of a digital library testbed application, which included extent of use, use of the digital library compared to other systems, nature of use, viewing behavior, purpose and importance of use, and user satisfaction. Data were collected from potential and actual users through focus groups, interviews, observations, usability testing, user registration and transaction logging, and user surveys. Kassim and Kochtanek (2003) conducted usability studies of an educational digital library through the use of focus groups, Web log analysis, database usage analysis, satisfaction surveys, remote usability testing, and more. These usability studies attempted to understand user needs, find problems and desired features, and to assess overall user satisfaction. Another type of usability study is to compare an experimental group with a control group. For example, Borgman et al. (2001) evaluated the Alexandria Digital Earth Prototype for use in undergraduate education, using surveys, interviews, and classroom observations. They mainly investigated which components of the simulation would be useful and assessed learning outcomes by comparing a experimental group and control group on a series of tasks.

Some digital library evaluation studies go beyond usability and also examine the content and performance of the system. Marchionini et al. (1998) applied multifaceted approaches to the evaluation of the Perseus Project. The evaluation was focused on learning, teaching, system (performance, interface, and electronic publishing), and content (scope, accuracy). Hill et al. (2000) tested the user interfaces of the Alexandria Digital Library through a series of

studies and collected feedback about the users, the interface, system functionality, and content of the digital library. Cherry and Duff (2002) conducted a longitudinal study of a digital library collection of Early Canadiana Materials, focusing on how the digital library was used and the level of user satisfaction with different features of the digital library, such as response time, browsing capabilities, comprehensiveness of the collection, print function, search capabilities, and display of document pages. These studies provide a basis for improvements to the digital library so that it could meet, and possibly exceed, end user needs and expectations.

Digital library service evaluation received relatively little attention in the early phases of the Internet revolution when developments tended to be technology driven (Heath et al., 2003). Recently, digital library evaluation has begun to include to service evaluation. Heath et al. (2003) discussed the benefits of a combined methodology that would provide greater potential for evaluation of digital library services but failed to conduct an actual evaluation study. Monopoli, Nicholas, Georgiou, and Korfiati (2002) evaluated the electronic journals service of a digital library and emphasized understanding the users and their use patterns. The majority of digital library service evaluation has focused on digital reference services. For example, Carter and Janes (2000) analyzed logs of over 3000 questions asked of the Internet Public Library on the basis of questions asked, and how those questions were handled, answered, or rejected.

Another type of digital library evaluation is to examine the impact of digital libraries on users and their communities, typified by the longitudinal study of the Perseus Digital Library, evaluating its impact on users and the humanities community (Marchionini, 2000). Bollen, Luce, Vemulapalli, and Xu (2003) analyzed the usage patterns derived from log analysis and further identified research trends that evolves in an institution's user communities over time. This research has implications for the acquisition and retention of digital content. In addition, research on digital library evaluation explores what determines digital library acceptance. The results showed both perceived usefulness and perceived ease of use were determinants of user acceptance of digital libraries (Thong et al., 2002). Digital library evaluation also involves comparing use of digital libraries and other types of systems. MacCall, Cleveland and Gibson (1999) reported a preliminary evaluation of a classical digital library model by analyzing usage logs and comparing users' experience in using a digital library with their experience in general Internet use.

Although the volume of published research on the evaluation of digital libraries has increased, little of it discusses criteria for evaluation. Currently most of the research on digital libraries evaluation is based on existing evaluation criteria for traditional libraries, performance of information retrieval systems, human–computer interaction, digital technologies, and so on. Marchionini (2000) suggests digital libraries are extensions and augmentations of physical libraries. Researchers might use existing techniques and metrics to evaluate digital libraries, for example, circulation, collection size and growth rate, patron visits, reference questions answered, patron satisfaction, and financial stability. Evaluation criteria for digital technologies can also be useful, such as response time, storage capacity, transfer rate, user satisfaction, and cost per operation. However, digital libraries provide new services, products, and capabilities, making it more difficult to compare them with physical libraries. After reviewing evaluation criteria for libraries by Lancaster (1993), for library and information services by Saracevic and Kantor (1997), information retrieval systems and

human–computer interaction by Su (1992) and Shneiderman (1998), the following list of criteria is presented by Saracevic (2000) and Saracevic and Covi (2000): traditional library criteria: collection (purpose, scope, authority, coverage, currency, audience, cost, format, treatment, preservation), information (accuracy, appropriateness, links, representation, uniqueness, comparability, presentation), use (accessibility, availability, searchability, usability), and standards; traditional information retrieval (IR) criteria: relevance (precision and recall), satisfaction, and success; traditional human–computer interaction/interface criteria: usability, functionality, efforts, task appropriateness, and failures.

As previously noted, most digital library evaluation studies are mainly usability studies. Some evaluation studies also assess collection content, system performance, and services. While Marchionini (2000) emphasized applying traditional library evaluation criteria to digital libraries, Saracevic (2000) and Saracevic and Covi (2000) extended the evaluation criteria to include those developed for information retrieval systems and human–computer interaction. Chowdhury and Chowdhury (2003) added that it is also necessary to measure the overall impact of digital libraries on users and society. However these evaluation criteria were developed by researchers not users. This study explores users' criteria for digital library evaluation.

4. Methodology

4.1. Participants

A total of 48 subjects participated in the study. To enhance the representation of the data, the author collected data twice, in 2003 and 2004. Subjects were recruited from the School of Information Studies, University of Wisconsin-Milwaukee. The recruitment message stated that a study to investigate digital library evaluation would be conducted in order to create digital libraries that truly support effective use. Users' inputs were needed regarding their perceptions of digital library evaluation criteria and their evaluation of the strength and limitation of existing digital libraries. Twenty-three subjects were recruited and their responses were collected in fall 2003 and another twenty-five were recruited and their responses were collected in fall 2004. These subjects were graduate and undergraduate students who were interested in the use of digital libraries. Female subjects account for 69% of the participants. While the majority of the subjects (87.5%) were graduate students, 2.5% were undergraduate students. All of them (100%) had basic knowledge of digital libraries and had used and searched digital libraries.

4.2. Data collection

The objective of this study is to develop criteria for the evaluation of digital libraries. Before data collection, the subjects were informed of the common elements of digital library definitions identified by the Association of Research Libraries (1995). Data were collected mainly from an open-ended self-administrated survey. The same instruction was given to all 48 subjects. All the subjects had approximately one week to respond to the survey. The survey

included two parts. In the first part, the subjects were instructed to develop a set of evaluation criteria for the evaluation of digital libraries. To be more specific, they were asked to identify the essential criteria that were important for the development and use of digital libraries. They were also instructed to justify their choices. For each of the criteria that a subject identified, he or she needed to discuss why that criterion was important.

In the second part of the survey, subjects were asked to report on their evaluation of existing digital libraries. They were instructed to evaluate an existing digital library by applying the set of evaluation criteria developed by themselves. In order to conduct the evaluation, the participants were provided with a list of existing digital library projects that represented a variety of digital libraries developed by different agencies and organizations, including governments, associations/organizations, universities, museums, and companies. These included such collections as the Library of Congress American Memory Project (http://memory.loc.gov/), the ACM Digital Library (http://www.acm.org/dl/), and theElectronic Poetry Center at SUNY-Buffalo (http://wings.buffalo.edu/epc/). Each subject was instructed to select one digital library with which he/she was familiar and evaluate it. A subject could also select a digital library that was not on the list as long as it satisfied the requirement of a digital library. Table 1 presents detailed information on the types of digital libraries selected by the subjects. Digital libraries developed by associations/organizations, governments, companies, and universities were the major ones evaluated.

Before providing their evaluation, participants were instructed first to explore the digital library thoroughly. To be more specific, they were told to browse through (e.g., whether they can find one item from two paths) and search their own questions in the selected digital library. After that they were asked to discuss problems with the digital library by going through every criterion they identified in the first part. They were also asked to provide examples to illustrate the problems they encountered in the digital libraries they selected to evaluate.

4.3. Data analysis

Both quantitative and qualitative methods were used to analyze the data. Quantitative methods were employed to conduct descriptive data analysis, such as frequency and mean. Content analysis (Krippendorff, 2004) was applied to develop categories and subcategories of evaluation criteria and problems. According to Krippendorff (2004), categorical distinctions define units by their membership in a category by having something in common. In this study,

Table 1
Types of digital libraries (DL) selected for evaluation

Types of DLs	No. of DLs selected	No. of subjects selected DLs
Government	3	7
Association/organization	7	27
University	3	3
Museum	1	1
Company	3	9
International	1	1

each category and subcategory of evaluation criteria and problems with digital libraries is defined and discussed by citing responses directly from the participants. Instead of creating an existing structure, types of criteria and subcriteria and types of problems were derived directly from users.

The data analysis proceeded as follows: (1) Each subject's responses about evaluation criteria were extracted from the survey. (2) All the responses that shared the same properties were classified into one category. For example, all the responses about interface use were classified into one category, and all the responses about content were classified into another. (3) A name for each type of criterion was assigned based on the content of the responses. The types of criteria identified generally used terminology common in the library and information science field (e.g., usability, collection quality, service quality, etc.). (4) After all the responses were classified into general types of evaluation criteria, all the responses for one type of criterion were further analyzed to differentiate subgroups based on their properties. For example, all the responses about different ways of finding information were classified into one subcategory while the responses about the importance of help functions were categorized into another subcategory. (5) A subcategory name was assigned to the subgroup based on the content of the responses and its general type name. For example, in the usability category, search and browse, navigation, help features, view and output, and accessibility emerged as subcategories. Some of the responses discussed the types of evaluation criteria in general; therefore, usability in general, content in general, and IR in general were also identified as separate subcategories. The same procedures were followed for the analysis of problems. Examples of categories and subcategories of evaluation criteria and problems with digital libraries are further discussed in the Results section. In order to provide a better understanding of participants' views of different types of digital library evaluation criteria, the author also calculated the number of subjects and the percentage of the total subjects who identified each criterion. The results of the data analysis of types of evaluation criteria are presented in Table 2.

5. Results

5.1. Types of evaluation criteria

Table 2 presents frequency and percentage of types of criteria that participants developed. Both content and format are important to users. Interface usability and collection quality were considered as important criteria by a majority of the participants. In addition, service quality, system performance efficiency, and user opinion solicitation were also deemed as essential criteria by many of the participants.

5.1.1. Usability

Almost all the participants considered usability as the most important criterion for a useful digital library. One subject stated it, "I think the most important criterion that I listed above is usability. Users need to be able to navigate through a digital library with a certain sense of ease,

Table 2
Types of evaluation criteria (N=48)

Types of criteria	Types of subcriteria	Number of responses	Percentage
Usability	Usability in general	25	52
	Interface usability	26	54
	Search and browse	34	71
	Navigation	24	50
	Help features	20	42
	View and output	6	13
	Accessibility	4	8
	Other	10	21
Collection quality	Content in general	37	77
	Quality	13	27
	Scope	19	40
	Authority	18	38
	Accuracy	17	35
	Completeness	15	31
	Currency	12	25
	Other	7	15
Service quality	Mission	15	31
	Intended user community	12	25
	Traditional library service	5	10
	Unique services	5	10
	Other	6	13
System performance efficiency	IR in general	9	19
	Efficient and effective	4	8
	Relevance	3	6
	Precision and recall	2	4
	Other	6	13
User opinion solicitation	User satisfaction	16	33
	User feedback	7	15
	Contact information	5	10
	Other	2	4

or else they may become frustrated and decide to go elsewhere" (S32). More than half of the participants (52%) discussed usability in general as the key evaluation criterion. Participants cared the most about interface usability. Fifty-four percent of the participants emphasized the importance of interface design. As one subject pointed out, "This is the connection between the user and the system. It is the first thing they see. It is their tool to receive output from the digital library" (S36). A couple of participants emphasized the importance of the interface as a gateway between a user and a digital library. One subject concluded that interface evaluation is important because "if the users can't access the DL effectively or finds it cumbersome, the information is not really accessible to them" (S24).

The majority of participants (71%) identified search and browse functions as essential criteria for the evaluation of usability of digital libraries. As one subject stated (S39), "Search features are important in any digital library, to enable the quick retrieval of information, when certain details of that information are already known. The browse feature is important for

retrieval of information that is not already known in part. Browsing enables people to look through a digital library and discover things they had no previous knowledge of." Participants indicated that the search and browse functions enhanced a user's ability to find relevant information effectively.

For half of the participants (50%), navigation efficiency is another key criterion for interface usability. Being intuitive is the basic requirement for navigation. One subject stated that "a user must be able to navigate and find his/her ways to information quickly and easily" (S47). To ensure effective navigation, "multiple access points" and "site map" were identified as essential from the responses, especially "easily understood access points and good organization structure which must abide by good traditional cataloging and classification methods as well as principles of HCI" (S31).

Another important finding is the identification of the availability of help features as one of the significant criteria. Forty-two percent of the participants chose help feature availability as an evaluation criterion, mainly because they needed support in using a digital library. This relates to their general feelings regarding digital libraries. According to one subject (S1), "Help features allow a user to feel comfortable within a digital library. If a user is experiencing a problem, a help page or a FAQ page will normally provide the answer". Limited by the existing help features, many of the participants only considered the use of help features for technical problems. There is no comment about how help systems can help clarify user need. Here is a typical comment from a subject, "Help system is important to help with technical difficulties. It should be easy to find and easy to comprehend" (39).

Other criteria related to usability were also identified by the participants. For example, 13% of the participants identified view and output issues. A subject explained the reason for the view criterion, "The user needs to have the appropriate player or viewer to access the items in the collection. The user needs to know which player or viewer is needed" (S37). Another one discussed the output criterion, "For instance, can the user print, save, and email the desired information? If the user can't 'take' the information in the form that they want, it has a negative impact on them" (S36). Four participants mentioned accessibility as an issue, indicating that people would not use resources if there were too much required in the way of logging on. Other criteria mentioned by participants were HCI principles, reliability of the digital library site, uniformity, learnability, etc.

5.1.2. Collection quality

As with usability, collection quality was considered another critical criterion. Without a collection, a digital library does not exist. Without a high-quality collection, a digital library does not have much value. One subject explained, "The collection is the most important part of the digital library. It is the reason that patrons visit the site" (S33). Another subject added, "In my mind, content is the most important aspect of any information resource. The quality of content is directly related to how well a digital library supports the needs and demands of the users. The quality of content is a primary factor that sets a digital library apart from the majority of the free Web" (S31). About 77% of the participants discussed the importance of collection quality in general in evaluating digital libraries.

When participants discussed collection quality, they immediately related it to the traditional library evaluation criteria since collection is a major component in both digital and physical libraries. One subject stated, "A further look at the criteria for the evaluation of the content of this DL will be aspects that would also apply to a more traditional library. These would include subject, scope, coverage, and authority" (S9). In addition, the characteristics of electronic materials also lead people to worry about their quality. For example, one subject emphasized that "the quality of the information should be able to be assessed because anyone can put anything on the Web" (S26).

Based on participants' responses, the criteria of quality of a collection can be divided into scope (40%), authority (38%), accuracy (35%), completeness (31%), and currency (25%). It is important for a digital library to clearly define its scope so users can immediately judge whether they have accessed the right digital library. One subject argued that the reason for choosing authority as one of the evaluation criteria was that "Authoritative sources comprise an important aspect of collection development in a traditional library" (S26). In terms of accuracy, it is obvious that "if the information is inaccurate, there is no reason for people to use it" (S17). "How comprehensive is the content of the library, does it cover all subject fields and to what degree?" (S7) One subject further defined "completeness", "A good library will cover its subjects thoroughly and be able to provide information that meets the demands of users with varying levels of information need" (S38). Currency is another key criterion for evaluating the quality of a collection. It relates to the usefulness of the material. One subject stressed its importance, "Collections need to be updated daily, monthly, and yearly in order for them to be any value to people" (S32).

5.1.3. Service quality

Service quality is the third category of evaluation criteria. In this category, the mission of a digital library (31%) and the intended user community (25%) were the main concerns for participants. Respondents suggested that the mission of a digital library needed to be tied closely to the mission of the organization that develops the digital library. "In evaluating [the digital library], we must first look at the mission of the organization—what its purpose of use is, and who their intended users are" (S30), said a subject discussing the relationships between the mission of a digital library and its intended user community. The ultimate goal of digital libraries is to serve users. That is why "digital libraries need to keep in mind their audience in order to present information to meet their needs", according to Subject 10.

Another important criterion for service evaluation is whether a digital library provides traditional library services as well as unique services that only can be offered by digital libraries. Digital libraries need to offer library services since they are libraries. One of the subjects summarized it well: "I believe one of the most important parts of the evaluation process is to look at the digital library from a traditional library standpoint. There are certain features that [the digital library] should contain for it to be granted the name digital library, otherwise every single website on the Internet could be called a digital library. The quality of service is important in both the digital and traditional library" (S33). At the same time digital libraries need also provide special services because they exist in digital forms.

A subject asked, "Does the digital library provide unique services that are not available elsewhere" (S17)?

5.1.4. System performance efficiency and user opinion solicitation

Digital libraries share characteristics of information retrieval (IR) systems so they also need to be evaluated based on the criteria for evaluating IR systems. Nineteen percent of the participants identified efficiency of information retrieval as one of the evaluation criteria since "even the best information has little value if it is impossible to retrieve it" (S38). Efficiency was defined as one of the system performance measures. One subject justified the decision, "It determines how quickly the user could find the needed information" (S18). Three of the subjects mentioned relevance and two subjects specifically identified precision and recall as evaluation criteria.

The ability for users to offer opinion, such as user satisfaction surveys, user feedback forms, and availability of contact information, was deemed critical for the evaluation of a digital library. About one third of the participants would like systems to solicit information about user satisfaction. Some of them further suggested that "this criterion is essential for any digital library that hopes to learn from its users. The feedback should go to the designers and parties responsible for the maintenance of the site. Differing perspectives and interpretations can never be understood without communication in the form of feedback" (S31). Contact information is another way to solicit feedback. "People should be able to find contact information in the digital library", stated Subject 6.

5.2. Problems with existing digital libraries

5.2.1. Usability

Based on the criteria they proposed, participants reviewed and identified problems in the existing digital libraries. Usability is still a major problem, especially with respect to interface design. Navigation efficiency, availability of search functions, and help features emerged as main areas that need to be improved. In regard to navigation, the lack of intuitive interfaces was identified as a problem for some of the digital libraries. Participants found them difficult to navigate. One subject pointed out, "Project Gutenberg's interface, the intermediary between the user and the system, is not as intuitive to navigate and use as it should be" (S42). Another one mentioned a specific problem in one of the digital libraries, "Another omission is that there was no easily discernable homepage" (S9).

There are several problems related to search features. First, some of the search functions lack user control and are not transparent to users, especially for experienced users. "The Project Gutenberg is somewhat unsophisticated and, because it is unable to be manipulated, is more restrictive than it seems at first glance. It also does not tell the user how it decided that there was a match between the query and the item", one subject (S42) explained in detail. Second, some of the digital libraries do not support the Boolean operators with which some users are familiar. For example, one subject complained, "When it came to running searches, I found that I was able to use natural language but the digital library does not support Boolean searches" (S6).

In terms of help features, the first problem is that not all the digital libraries provide help. One subject said, "There is no defined 'Help' portion to this digital library. Nothing is offered about any of the technical needs to access portions of the site" (S39). Another subject backed up this statement, "Access to information is impeded because there is no site map or help page" (S26). Some existing digital libraries do not provide context-sensitive help; users have to figure our how to find the help and how to ask the question. For example, "The Internet Public Library offers users a link to 'Ask a Question', and this link takes users to a form where they have to fill out their name, email address and a host of other questions including the question that they want to ask. This can be a hassle to many users who just want to do a quick click to find an answer" (S43).

In addition, not all the digital libraries provide output options. One of the subjects complained, "The library does not provide any output features such as print, email or save, since the site contains many large images, this can be a problem if trying to save to a disk due to the large file sizes. The site contains some great images and information but the lack of these features deters the overall user satisfaction" (S10).

5.2.2. Collection quality

Interestingly, the problems identified regarding digital library collections were not about coverage, but more about whether these digital libraries provide any information for users to make judgments about authority, accuracy, currency, and copyrights. First, participants were not able to make authority judgments. Participants found that either "there is no clear statement of authority" (S26) or "one downside to look at the content of this library was the lack of citations or information about who specifically wrote the articles given in the digital library" (S22). Second, in some of the digital libraries, accuracy is not guaranteed. For example, "The only real problem with the collection is that there is very little control over the accuracy and quality of the items in the collection. E-texts are proofread by volunteers, and no attempt is made to make the e-texts perfect since distribution at low cost is more important to the project than accuracy", noted by Subject 42.

The two problems with currency are that some digital libraries do not provide an updating statement, and some update their collections infrequently. One subject pointed out, "Project Gutenberg does not make an update statement. A clear date, including day, month, and year, could help the site appear current. Otherwise users don't know whether items are added occasionally or as often as everyday" (S46). "The currency of the information is fair though digital libraries should be updated more frequently. The last update for the home page was Feb. 2003, therefore it seems ironic that the 'What's New' page was last updated in May of 2002".

Lack of copyright information is another problem for some digital libraries. One subject observed, "The interesting thing to note about this site is that it does not contain any visible copyright information" (S23). In addition, not all the digital libraries have a clear collection policy. One subject commented on one of the digital libraries, "The purpose and coverage are not well defined" (31). That also leads to the problem of creating subject headings. One subject commented on one digital library, "Upon closer examination one notices that the subject headings are somewhat arbitrary and based likely on what the library happens to hold in its collection and not on a pre-determined set of criteria" (S5). Although multimedia collection is

one of the characteristics of digital libraries, many of the digital collections "lack multimedia formats" (S34).

5.2.3. Service quality

One major criticism for the existing digital libraries is lack of a community service, which is one of the important functions for traditional libraries. Community services refer to services that are provided to members of the digital library community, such as allowing members to exchange ideas, make announcements, and more. Subject 7 stated his complaints and offered some specifications for community services of digital libraries, "It does not provide services for users to make announcements or communicate with each other". One subject made some suggestions, "It lacks community, true interactivity, synchronous services, and preservation aspects. It could reach out to local communities, researchers, and staff in dynamic ways, producing more browsing options, and flesh out each page with multimedia resources" (S34). Some of the subjects thought it was not a good idea for digital libraries to link to commercial sites. As one subject noted, "Another critique of the service aspect includes the button for Book Cellar. When the user clicks on Book Cellar, it directs you to a page with a list of books. The titles are linked to Amazon.com. Here the digital library has foregone any service by passing it over to a vendor…" In addition, some of the digital libraries do not state their missions clearly. As one subject observed, "The purpose and coverage are not well defined."

5.2.4. System performance efficiency and user opinion solicitation

Problems with system performance are related to the relevance of the retrieval results and efficiency of the retrieval process. The main problem identified regarding system performance was how to increase precision. In general, users care more about precision than recall since they only need enough information to solve their problems. In addition, precision affects the time a user needs to spend to evaluate and find relevant information to solve his/her problems. Subject 8 stated "I found the amount of information from the search results, not overwhelming, but still enormous, and they are not always relevant". Another related problem with system performance is response time. One subject (S22) complained, "the video also takes several minutes to download using a 56K modem and dial up connection. While the graphics on the digital library are very colorful and nice, the fact that the site is so laden with them makes the digital library taking a while to present the results".

Contact information is essential for users who would like to send feedback. The problems with contact information are either "there is no way to tell who the user should contact" (S33) or "the site does not give an address or other contact information besides the e-mail address. Not all people have e-mail, so it might be beneficial if there was an address or even a P.O. Box where someone could contact this person" (S29).

Some of the issues are interrelated. Sometimes users do not know whether there is a problem in system performance or collection. For example, one participant complained that he could only find one basketball related magazine. According to him, "When I searched [for relevant content] for basketball related magazines, there was only one. This surprised me quite

a bit since basketball is a very popular sport in the United States. I also know of quite a few popular basketball magazines which means either a lot of them are not provided online or the IPL just chose not to put them in their collection" (S32).

In short, a majority of the subjects of this study considered usability and collection quality as essential criteria for digital library evaluation. In addition, service quality, system performance efficiency, and user opinion solicitation were also identified by some of them as important evaluation criteria. After evaluating existing digital libraries, subjects identified some problems with the existing digital libraries. They posed the following challenges for the development of digital libraries: how to design an interface to be easy to use, how to make it easy for people to assess a collection, how to provide both traditional as well as new services, how to increase precision of retrieval results, and how to involve users in digital library development.

6. Discussion

6.1. Similarities and differences of evaluation criteria between users' and researchers' perspectives

The evaluation criteria developed by participants of this study echo the criteria that researchers have proposed (Chowdhury & Chowdhury, 2003; Marchionini, 2000; Saracevic, 2000; Saracevic & Covi, 2000) and the criteria that have been applied in previous studies to evaluate digital libraries (Bishop et al., 2000; Bollen et al., 2003; Borgman et al., 2001; Carter & Janes, 2000; Cherry & Duff, 2002; Heath et al., 2003; Hill et al., 2000; Kassim & Kochtanek, 2003; MacCall et al., 1999; Marchionini et al., 1998; Monopoli et al., 2002; Thong et al., 2002; Van House et al., 1996). Table 3 presents types of digital library evaluation criteria identified by subjects of this study and researchers and applied in previous studies. Here researchers refer to those researchers who have worked on and identified evaluation criteria for digital libraries. Previous studies refer to those studies that have evaluated digital libraries or prototypes of digital libraries.

Table 3 does not cover all the criteria identified by researchers and applied in previous studies, but it does highlight the focus of digital library evaluation research and studies. Researchers connect digital library evaluation criteria with previous research in evaluating a variety of information retrieval systems, including libraries. Digital library evaluation studies focus on examining the criteria from selected categories, mostly to do with interface usability. Participants of this study associate their evaluation criteria with their experience in using physical libraries and other types of information retrieval systems. There is a commonality in the overall categories of the evaluation criteria. To be more specific, the same major categories (e.g., usability, collection quality, service quality, system performance, and user opinion solicitation) have been identified by participants of this study and researchers and applied in previous evaluation studies.

However, participants emphasize more the usefulness of the digital libraries from the users' perspectives and less from the perspectives of developers and administrators. For example,

Table 3
Types of evaluation criteria from this study, researchers, and previous studies

Types of criteria	Types of subcriteria	This study	Researchers	Previous studies
Usability	Usability in general	X	X	
	Interface usability	X		X
	Functionality		X	X
	Search and browse	X		X
	Navigation	X		
	Help features	X		X
	View and output	X		X
	Accessibility	X		
	Failures		X	
Collection quality	Content in general	X	X	X
	Quality	X	X	X
	Scope	X	X	X
	Authority	X	X	
	Accuracy	X	X	X
	Completeness	X	X	X
	Currency	X	X	
	Cost		X	
	Format		X	X
	Treatment		X	
	Preservation		X	
	Uniqueness		X	
Service quality	Mission	X		
	Intended user community	X	X	X
	Traditional library service	X	X	X
	Unique services	X		
System performance efficiency/Use	IR in general	X	X	X
	Efficient and effective	X	X	X
	Relevance	X	X	X
	Precision and recall	X	X	X
	Success		X	
	Usage		X	X
User opinion solicitation	User satisfaction	X	X	X
	User feedback	X		
	Contact information	X		
Impact	Social impact		X	
	On user community			X
Other	Standard		X	

cost, treatment, and preservation are some of the important criteria for digital library evaluation from researchers' perspectives, but the participants in this study did not identify these criteria as important ones. Participants placed more emphasis on the ease of use of the interface and high quality of the collection in evaluating digital libraries. Compared with previous research and studies, this study offers not only detailed information about evaluation criteria but also why they are important to users. In participants' discussions of the justification of usability criteria, ease of use and more options were the main concerns. They

also mentioned the importance of the availability of help features in digital libraries. Limited by the existing digital libraries, their criteria for evaluation were still quite basic and tended to focus on whether a help feature existed. They did not further consider the effectiveness of these features. For example, participants emphasized the importance of a digital library providing help for technical problems but did not further discuss the need for help with non-technical problems, such as a help feature that could assist them in clarifying their need. As to collection criteria, participants discussed evaluation criteria in multiple dimensions from coverage, authority, and accuracy to currency. Many participants considered authority and accuracy of a collection more important than comprehensiveness and currency of a collection. The understanding that everything on the Web might not be authoritative and accurate affected their perceptions of the importance of these evaluation criteria.

The participants of this study did identify some of the criteria that researchers and previous studies have neglected. For the service quality criteria, they considered not only traditional services used for evaluating traditional libraries but also unique services for evaluating digital libraries. Participants suggested that digital libraries needed to take advantage of the opportunity to provide services that traditional libraries cannot provide. They also considered user opinion solicitation as a critical element for digital library evaluation, a criterion that researchers and previous studies have not discussed. Participants further suggested that user feedback needed to be collected actively by tracking user satisfaction as well as passively by providing clear contact information. However, participants of this study failed to discuss the social impact of digital libraries on a society and communities as researchers and previous studies have emphasized. Their perspectives again were limited by their emphasis on the use of digital libraries and their approach to digital libraries from the point of view of an individual user rather than that of a researcher or an institution.

6.2. Implications for digital library design

The problems identified by the participants represent their concerns with existing digital libraries. Fig. 1 presents the digital library evaluation criteria, problems, and suggestions to solve the problems. According to participants, ease of use is a main problem for existing digital libraries. While the majority of the participants were mainly concerned with the ease of use of the interfaces in digital libraries, others, especially experienced ones, sought more control in finding the information they wanted. They complained that some of the digital libraries were designed too basically and were not sophisticated enough, therefore they were unable to manipulate the systems. To be more specific, they could not conduct more accurate and specific searches. Digital library design needs to balance ease of use and user control. The basic requirement for ease of use requires providing search, browse, and help functions as well as a variety of output options. Some existing digital libraries do not have these features. At the same time, experienced users also want to have more control, such as more advanced features for them to customize the way they search, access, view, and disseminate information.

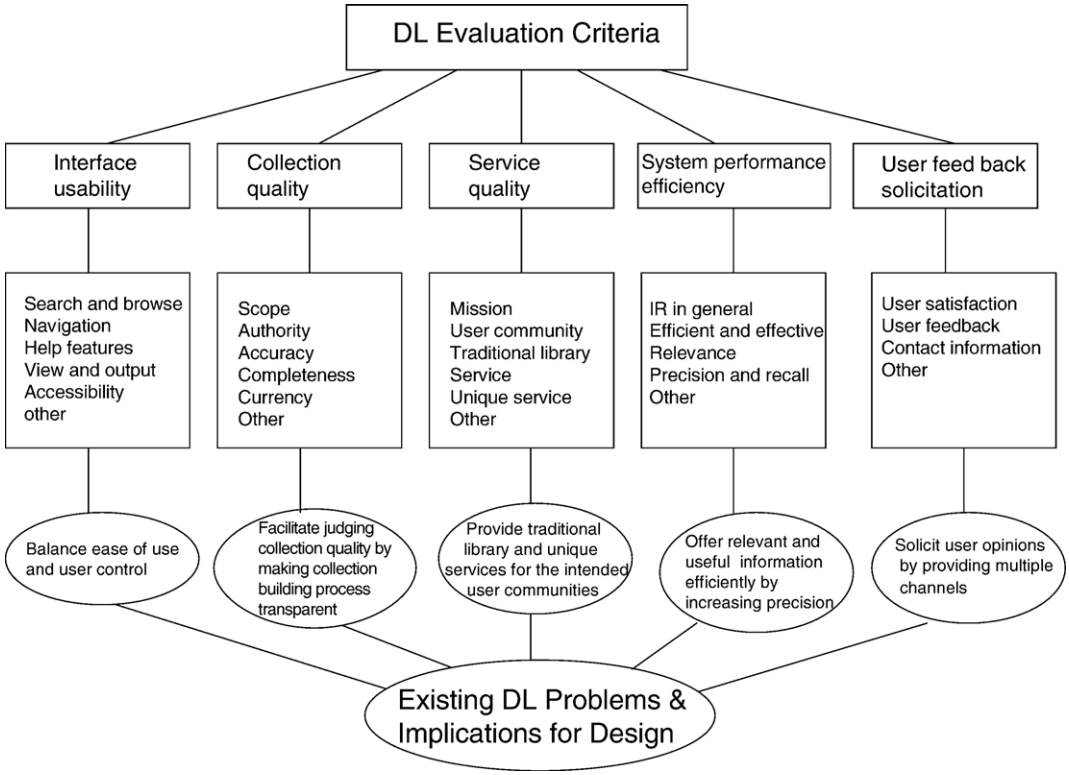

Fig. 1. Digital library evaluation: Criteria, problem, and implications for design.

One unique finding of this study is that collection quality is less an issue. The major problem identified by participants was how digital libraries can assist users in making judgments about quality of the collection. Specifically, they need information to help them make judgments about currency, copyright, authority, and accuracy of the collection. Currency and copyright concerns can be easily solved by consistently providing this information in digital libraries. Users also need assurance on the authority and accuracy of the collection. Digital libraries need to implement mechanisms to assure users that their collections are authoritative and accurate. Transparency of collection building might be one way to help users understand how the collection has been selected and processed and what measures have been taken to ensure the accuracy of the information. Furthermore, the procedures for collection selection should be innovative and convincing and could involve both human intelligence and technology. For example, technologies used to test the integrity of Web items can be enhanced to ensure users of the accuracy of a document in a digital library. Standards can also be developed and used to ensure the authority of a document.

Many participants value the role that traditional libraries play as community centers, and they wonder if the unique technologies offered in digital libraries can be used to not only enable them to play the same role but also to enhance that role. However, existing digital

libraries fail to provide community services that users prefer, such as making announcements about community activities, and promoting the exchange of ideas about topics of interest to the community. Synchronous interactions and services are needed. Some of the Web technologies in video conferences and online chats can be adapted to digital libraries. In addition, more publicity is needed to inform the community of the existence of digital libraries. Collaboration between the community and digital libraries for a variety of activities is also an effective approach to promote community services and uses of digital libraries.

It is essential to offer relevant and useful information to users efficiently. Users generally care more for precision than recall. Subjects of this study were not satisfied with the precision of the search results from some digital libraries. One way to solve this problem is to use some mechanisms to increase precision, such as PageRanking used by Google. Another way to solve the problem is to facilitate the evaluation of results (e.g., highlight keywords in the display). Users like to apply the least effort principle to finding useful information to solve their problems. They are also more sensitive to response time in digital libraries than other types of information retrieval systems because the collection contains multimedia materials. It is important to explore ways of enhancing multimedia storage and retrieval.

Since the ultimate goal of digital libraries is to serve their users, it is essential to have a mechanism to solicit opinions. Based on the responses from the subjects, there are digital libraries that have not provided channels for users to send their feedback. More important, user opinions should not be solicited only for evaluation purposes. It is critical to involve users in the process of design and development of digital libraries. Multiple channels need to be opened to facilitate communications between designers/developers and users along the way. Moreover, these channels should support the active solicitation of information (e.g., distribute user surveys to request user feedback) as well as the passive solicitation of information (e.g., provide clear contact information for users to send their feedback).

6.3. Limitation of the study

The contribution of this study is limited by its sampling. The selection of this group of subjects has both benefits and problems. On one hand, the subjects of this study are users of digital libraries, and they have basic knowledge and skills in the use of digital libraries. They are the ideal subjects for the study. On the other hand, they do not represent a variety of user groups since the subjects are all students of the School of Information Studies. Their knowledge and experience have an impact on their perceptions of evaluation criteria for digital libraries. For example, some of them might have been exposed to evaluation criteria for traditional libraries or IR systems. Furthermore, it is a convenience sample since the subjects are volunteers and they were not randomly chosen. Another limitation of this study is that the evaluation of the digital library is largely based on the subjects' own report on the open-ended survey. This study asked subjects to use the digital libraries, but did not record their use of digital libraries. It would be better if in a further study their use were captured and obtained. Finally, the subjects of this study each selected the digital library he or she would like to evaluate. While this could provide a more comprehensive evaluation of existing digital libraries, it does not offer a focused view on one specific digital library or one type of digital

library from a group of users. In short, this can be considered as a preliminary study of digital library evaluation from users' perspectives. Further research is needed.

7. Conclusion

Although researchers have developed digital library evaluation criteria, and previous studies have evaluated existing digital libraries or prototypes of digital libraries, researchers are still investigating what to evaluate. Furthermore, there is not much input from users in terms of what are the essential evaluation criteria. The major contribution of this study is that it is one of the few studies that examine digital library evaluation criteria based on users' perceptions. By applying their own criteria, participants evaluated five types of digital libraries developed by different organizations. Their evaluation reveals some of the problems in current digital library design and development. As digital libraries normally target different levels of user groups, it is essential for digital libraries to support ease of use as well user control. It is also critical for digital libraries to be transparent in design. Digital libraries not only need to provide high-quality collections but, more importantly, need to assist users in their efforts to make judgments about the quality of the collection. It is also important for digital libraries to provide both traditional and unique services.

Considering the limitations of this study, further research might focus on the evaluation of one or two types of digital libraries, and the participants should be the targeted users of these digital libraries. Digital libraries need to be evaluated in their social and organizational context. Further research also needs not only to identify evaluation criteria but also measure the degree of importance of each criterion to participants. In addition, multiple methods can be used to extend the study from the identification of evaluation criteria to the evaluation of actual use of digital libraries.

Acknowledgments

The author would like to thank Esha Datta for her assistance in data analysis and to the anonymous reviewers and Candy Schwartz for their constructive comments.

References

Association of Research Libraries. (1995). *Definition and purposes of a digital library.* Retrieved July 17, 2002, from http://www.arl.org/sunsite/definition.html

Bertot, J. C., & McClure, C. R. (2003). Outcomes assessment in the networked environment: Research questions, issues, considerations, and moving forward. *Library Trends*, *51*, 590−613.

Bishop, A. P., Neumann, L. J., Star, S. L., Merkel, C., Ignacio, E., & Sandusky, R. J. (2000). Digital libraries: Situating use in changing information infrastructure. *Journal of the American Society for Information Science*, *51*, 394−413.

Bollen, J., Luce, R., Vemulapalli, S.S., Xu, W. (2003). Usage analysis for the identification of research trends in digital libraries. *D-Lib Magazine, 9*(5). Retrieved March 15, 2005, from http://www.dlib.org/dlib/may03/bollen/05bollen.html

Borgman, C. (1999). What are digital libraries? Competing visions. *Information Processing and Management*, *35*, 227−243.

Borgman, C. L., Leazer, G. H., Gilliland-Swetland, A. J., & Gazan, R. (2001). Iterative design and evaluation of a geographic digital library for university students: A case study of the Alexandria Digital Earth Prototype (ADEPT). *Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries* (pp. 390−401). London: Springer-Verlag.

Carter, D. S., & Janes, J. (2000). Unobtrusive data analysis of digital reference questions and service at the Internet Public Library: An exploratory study. *Library Trends*, *49*, 251−265.

Cherry, J.M., Duff, W.M. (2002, January). Studying digital library users over time: A follow-up survey of Early Canadiana online. *Information Research, 7*(2). Retrieved March 12, 2005, from http://informationr.net/ir/7-2/paper123.html

Chowdhury, G. G., & Chowdhury, S. (2003). *Introduction to digital libraries*. London: Facet Pub.

Heath, F., Kyrillidou, M., Webster, D., Choudhury, S., Hobbs, B., Lorie, M., et al. (2003). Emerging tools for evaluating digital library services: Conceptual adaptations of LibQUAL+ and CAPM. Article No. 170, 2003-06-09, *Journal of Digital Information 4*(2). Retrieved November 15, 2005, from http://jodi.tamu.edu/Articles/v04/i02/Heath/

Hill, L. L., Carver, L., Larsgaard, M., Dolin, R., Smith, T. R., Frew, J., et al. (2000). Alexandria Digital Library: User evaluation studies and system design. *Journal of the American Society for Information Science*, *51*, 246−259.

Kassim, A. R. C., & Kochtanek, T. R. (2003). Designing, implementing, and evaluating an educational digital library resource. *Online Information Review*, *27*(3), 160−168.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage Publications Inc.

Lancaster, F. W. (1993). *If you want to evaluate your library*. (2nd ed.) Urbana-Champaign, IL: University of Illinois.

MacCall, S. L., Cleveland, A. D., & Gibson, I. E. (1999). Outline and preliminary evaluation of the classical digital library model. *Proceedings of the American Society for Information Science Annual Meeting*, *36*, 33−46.

Marchionini, G. (2000). Evaluation digital libraries: A longitudinal and multifaceted view. *Library Trends*, *49*, 304−333.

Marchionini, G., Plaisant, C., & Komlodi, A. (1998). Interfaces and tools for the Library of Congress National Digital Library Program. *Information Processing and Management*, *34*(5), 535−555.

Monopoli, M., Nicholas, D., Georgiou, P., & Korfiati, M. (2002). A user-oriented evaluation of digital libraries: Case study: The electronic journals service of the library and information service of the University of Patras, Greece. *Aslib Proceedings*, *54*(2), 103−117.

Nielsen, J. (1993). *Usability engineering*. San Diego, CA: Academic Press.

Saracevic, T. (2000). Digital library evaluation: Toward an evolution of concepts. *Library Trends*, *49*, 350−369.

Saracevic, T., & Kantor, P. (1997). Studying the value of library and information services: II. Methodology and taxonomy. *Journal of the American Society for Information Science*, *48*, 543−563.

Saracevic, T., & Covi, L. (2000). Challenges for digital library evaluation. *Proceedings of the ASIS Annual Meeting*, *37*, 341−350.

Shneiderman, B. (1998). *Designing the user interface strategies for effective human–computer interaction*. (3rd ed.) Reading, MA: Addison-Wesley.

Su, L. T. (1992). Evaluation measures for interactive information retrieval. *Information Processing and Management*, *28*, 503−516.

Thong, J. Y. L., Hong, W. Y., & Tam, K. Y. (2002). Understanding user acceptance of digital libraries: What are the roles of interface characteristics, organizational context, and individual differences? *International Journal of Human–Computer Studies*, *57*, 215−242.

Van House, N.A., Butler, M.H., Ogle, V., Schiff, L. (1996, February). User-centered iterative design for digital libraries: The Cypress experience. *D-Lib Magazine, 2*. Retrieved February 3, 2005, from http://www.dlib.org/dlib/february96/02vanhouse.html