EVALUATION OF DIGITAL LIBRARIES

10

THE NEED FOR DIGITAL LIBRARY EVALUATION

Digital library evaluation entails a systematic assessment of value and significance. The objective of digital library evaluation is to determine to what extent it meets its objectives and offer suggestions for its improvement (Chowdhury and Chowdhury, 2003). Digital library evaluation includes every aspect of its development and operation (Tsakonas et al., 2013). Recommendations for undertaking periodic evaluations are part of the guidelines for "building good digital collections" (NINCH, 2002; NISO, 2007). However, the exponential growth of digital libraries has not been accompanied by extensive evaluation studies (Saracevic, 2004). Moreover, there is no agreement on the key concepts, assumptions, parameters, and criteria related to digital library evaluation (Fuhr et al., 2007). Just as Tsakonas and Papatheodorou (2011) state, "Digital library evaluation is a complex field, as complex as the phenomena it studies…. However, the community has still to reach a consensus on what evaluation is and how it can effectively be planned" (p. 1577). As digital libraries are also under constant development change, it is important to evaluate digital libraries to ensure the right direction for future development and the acceptance by users and other stakeholders. Since digital libraries are new, complex, and multifaceted entities, researchers and practitioners need a set of guidelines pertaining to why it is important to evaluate, when to undertake evaluation, what to evaluate, how to evaluate, and how to incorporate the results into the development process.

The objectives of evaluation provide a rationale for why we evaluate. The purposes of digital library evaluations can be understood from the constructs, the relationships, and the evaluation (Fuhr et al., 2007). There are a variety of purposes for digital library evaluation including identifying user needs and problems, enhancing the interface design, planning, making budget decisions, etc. Determining the objective(s) sets the foundation for digital library evaluation.

The next question is when to evaluate. Digital library development and management go through a series of phases: planning, prototyping, building, testing, launching, operating, and upgrading. At different phases of digital library development, the objective of its evaluation might be different. Accordingly, different evaluation criteria and measurements might be applied. In addition, phases of digital library development also affect research methods and approaches (Buttenfield, 1999).

Evaluation criteria and measures are associated with what to evaluate. Although those used in assessing library print collections are applicable to a certain extent, they are insufficient for the new dimensions that emerged with digital libraries, such as interface design, system performance, sustainability, effects on users, and user engagement. As Van House et al. (1990) point out and as remains true today, there is no single, best way to do an evaluation; furthermore, the digital library environment, accompanied by rapid changes in users' expectations and behaviors, calls for evaluation from multiple viewpoints with an emphasis on the user's perspective (Carr, 2006; Kani-Zabihi et al., 2006; Nicholson, 2004). Evaluation involves comprehensive activities that compare "what is" to "what ought to be" (Van House et al., 1990). To measure "what is," evaluation frameworks should include

well-defined evaluation criteria and corresponding reliable and valid measures. The key concepts in evaluation consist of criteria and measures. Criteria refer to a standard or set of standards by which something can be judged or decided (Saracevic, 2004). Buchanan and Salako (2009) point out that what to measure and how to measure are the key challenges for digital library evaluation.

The final question focuses on how to evaluate. Since a digital library is a complex and dynamic system, digital library evaluation requires multifaceted approaches (Marchionini, 2000; Marchionini et al., 2003). Digital library evaluation employs multiple data collection methods ranging from online surveys to think-aloud protocols and log analyses. Moreover, it integrates both qualitative and quantitative methods. Most important, user-centered digital library evaluation is drawing more attention. Researchers emphasize the importance of incorporating digital library evaluation criteria and collection instruments to solicit users' perspectives and feedback (Heradio et al., 2012; Xie, 2008; Zhang, 2010).

Researchers and practitioners have expanded their digital library evaluation efforts from evaluation criteria, measurements, and methods to objectives and phases. However, it remains one of the most confusing areas in the library and information science field (Saracevic, 2004; Tsakonas et al., 2013; Zhang, 2010). Several limitations exist in digital library evaluation research and practices. First, there are few comprehensive frameworks and models available for digital library evaluation. Several researchers note that holistic evaluation studies are conspicuously absent from the digital library field and discuss the potential benefits of comprehensive approaches (Chowdhury et al., 2006; Saracevic, 2000; Xie, 2006; Zhang, 2010). Most of the current evaluation practices have focused narrowly on particular aspects or services. Second, there are still huge gaps between researchers and practitioners. Various digital library evaluation models have been suggested by researchers. However, those models are still mostly conceptual and theoretical, as they have not been tested in digital library practice. Researchers have not provided specific, feasible measures. As a result, digital library practitioners have adopted few of the models suggested by researchers. Third, few evaluation frameworks or models have been validated empirically from the perspectives of stakeholders. These limitations call for further investigation of issues related to digital library evaluation.

The remainder of this chapter provides an overview of the current research on evaluation frameworks as well as evaluation criteria and measures. In addition, it presents the authors' recent study as a detailed example of digital library evaluation within ten dimensions incorporating multiple digital library stakeholders' perspectives.

EVALUATION FRAMEWORKS

Several researchers and research groups developed digital library evaluation frameworks and models to identify the main constructs of such libraries and illustrate their relationships in their assessment. Tsakonas and Papatheodorou (2011) illustrate the ontology development of digital library evaluation at two layers: the upper strategic layer and the lower procedural layer. Although the strategic layer specifies the purpose of the evaluation consisting of classes that define the scope of the evaluation and its relationship to other evaluation studies, the procedural layer focuses on the evaluation activities including their classes describing exact processes, constraints, and requirements.

Saracevic's (2000; 2004) digital library evaluation framework is one of the most widely cited models. His framework comprehensively covers multiple aspects of digital libraries, including content, technology, interface, process/service, user, and context. His framework is the first attempt to measure

context, such as institutional fit, sustainability, and community impact. Additionally, Saracevic (2004) suggests a list of measures for different criteria. Zhang (2010) further investigated and validated Saracevic's evaluation framework. Zhang studied the importance of multiple constructs of Saracevic's six dimensions using survey data from heterogeneous stakeholders. She further developed a holistic model for digital library evaluation that presents specific criteria that can be selected and "tailored for multifaceted and multilevel digital library evaluations" (p. 107).

In Europe, DELOS is a collaborative digital library project that represents joint research activities involving major European teams in this area. The DELOS Network of Excellence has conducted a series of evaluation studies. Fuhr et al. (2001) proposed a digital library evaluation scheme containing four dimensions: data/collection, system/technology, users, and usage. Tsakonas et al. (2004) further examined the interactions of digital library components and suggested key evaluation foci in digital libraries, such as usability, usefulness, and system performance. Fuhr et al. (2007) also developed a digital library evaluation framework by integrating Saracevic's (2004) four dimensions of evaluation activities (construct, context, criteria, and methodology) and key questions in relation to why, what, and how to evaluate. As part of DELOS project results, Candela et al. (2007) established a three-tier digital library model, named the DELOS Manifesto. Even though the DELOS Manifesto is not an evaluation framework, it is a conceptual model that provides useful concepts for digital library evaluation. The DELOS model (Fig. 10.1) posits six core concepts: Content, User, Functionality, Quality, Policy, and Architecture (Candela et al., 2007).

Another frequently cited digital library evaluation model is the quality model developed by Gonçalves et al. (2007). Derived from the previous 5S (streams, structures, spaces, scenarios, and societies) digital library model (Gonçalves, Fox, Watson, and Kipp, 2004), Gonçalves et al. (2007) proposed a quality model, consisting of dimensions and measurements of quality. Seventeen types of
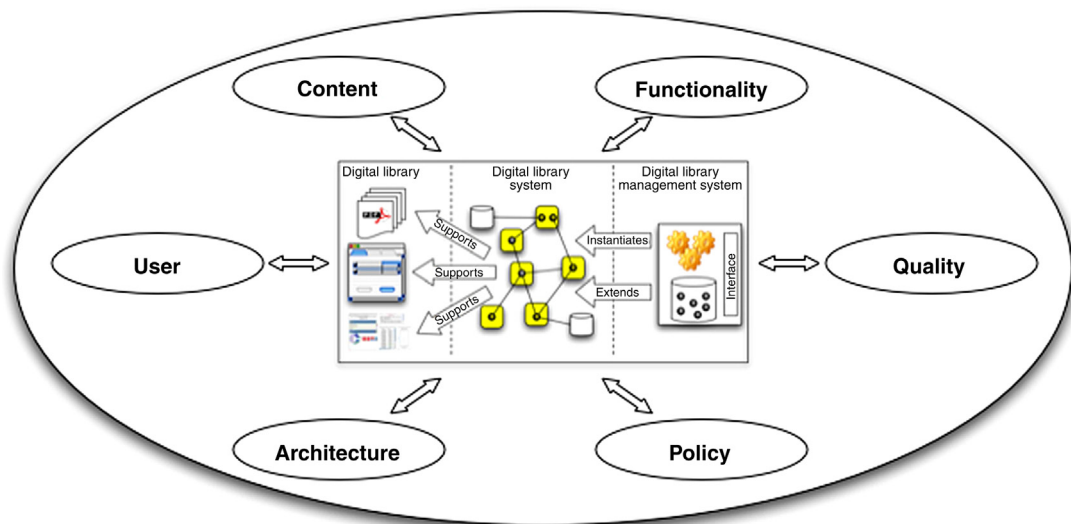

FIGURE 10.1 DELOS Model (Candela et al., 2007)

quality dimensions are identified: accessibility, accuracy, completeness, composability, conformance, consistency, effectiveness, efficiency, extensibility, pertinence, preservability, relevance, reliability, reusability, significance, similarity, and timeliness. The timeliness dimension includes three associated characteristics of response time corresponding to efficiency, cost of migration associated with preservability, and number of service failures linking to assess reliability. Moreover, this model is also connected to different phases of information lifecycle in order to assess, identify, and solve quality problems.

Xie's (2006, 2008) research focuses more onto users and suggests a user-driven evaluation model (Fig. 10.2). She points out that little has been done on the identification of evaluation criteria from the perspectives of users. Her model was derived from her empirical study investigating users' perceptions through diaries and questionnaires. Based on the study, she identifies five dimensions of digital library evaluation and specific criteria consisting of usability, collection quality, service quality, system performance efficiency, and user feedback solicitation. Focusing on a specific type of digital library, Albertson (2015) created a user-centered visual digital library evaluation framework after synthesizing the relevant literature on the topic. This framework is constructed with the following components: user, interaction, system, user–interaction, user–system, interaction–system, user–interaction–system, and domain and topics. The visual context requires more user–system interactions occurring in visual digital libraries.
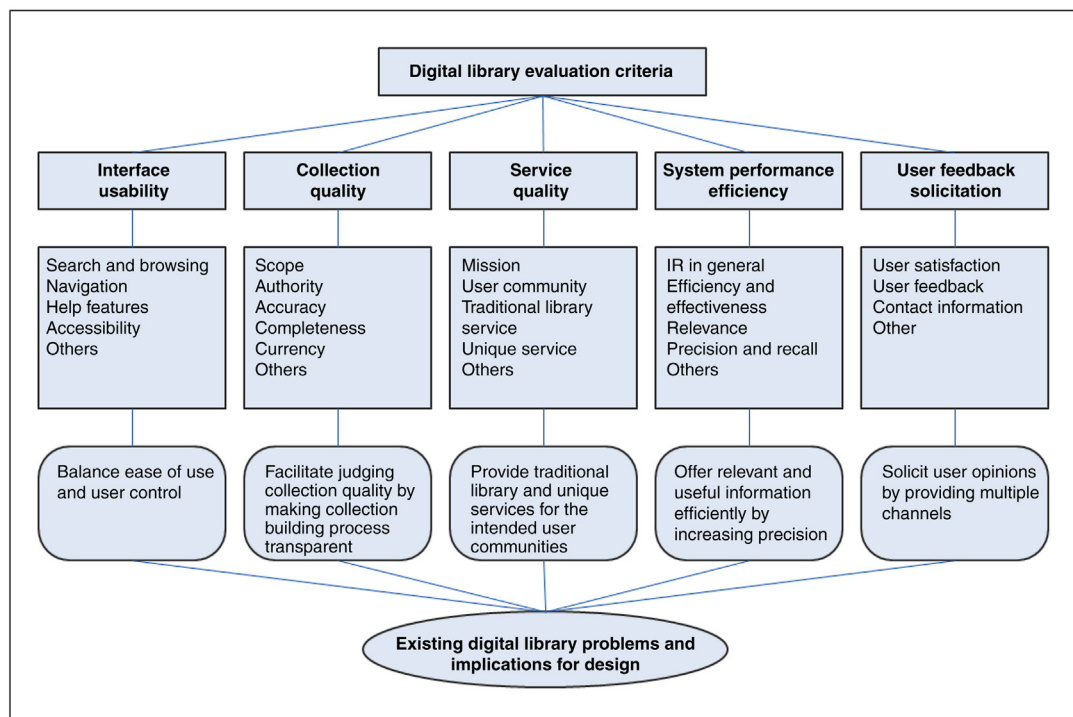

FIGURE 10.2 Xie's User-Driven Evaluation Model (Xie, 2008)

Resulting from comprehensive digital library evaluation research, Vullo (2010) developed an integrated LIS-oriented evaluation framework from both the user and system perspectives. Four entities constitute the model: organization, content, services, and users. Although the user perspective starts from users, the system perspective begins with organizations. Based on expert interviews, Lagzian et al. (2013) built a model for the identification of successful factors in a digital library. Six dimensions emerged from the data: motivation, resource, people, process, location, and time. Accordingly, 36 criteria are classified into six dimensions.

At the same time, evaluation models also emphasize different aspects of digital library dimensions. For example, Dillon (1999) created a multileveled framework for evaluating the usability of digital libraries ranging from physical to social-cultural. He identifies four key factors affecting usability with the acronym TIME. Specifically, T refers to tasks that reveal users' needs and uses for the materials; I refers to an information model that represents users' mental models to organize the information space; M refers to manipulation skills, in particular the design of manipulation facilities to support the users' use of documents; and E refers to the visual ergonomics that assists reading electronic text effectively. In accomplishing different tasks, the sequence of TIME might be different. Kim and Kim (2008) developed an evaluation framework tailored to digital collections, which covers four dimensions and 18 criteria. Those four dimensions are as follows: content; management and policy; system and network; and use, user, and submitter. Their model was built empirically from interviews with digital collection managers. More important, they suggest specific measures to numerically represent the evaluation criteria.

In the area of operational libraries, several practical tools have been proposed to expand evaluation efforts, including DigiQUAL, eMetrics, and EQUINOX, but they are often limited to specific elements or services. DigiQUAL was created as a tool for assessing service quality (Kyrillidou and Giersch, 2005). Based on the analysis of focus group data, the DigiQUAL research team identifies 250 items related to digital library service evaluation. These items are further classified into the following dimensions: accessibility/navigability, interoperability, community for users, developers and reviewers, collection building, federations' role, copyright, resource use, collection evaluation, and digital library sustainability (Kyrillidou, Cook, and Lincoln, 2009). The ARL's e-metrics project and the COUNTER and SUSHI protocols were devised for the purpose of assessing the outputs of digital libraries using the usage statistics, but they focus more on vendor-provided data.

EVALUATION METHODS
DIGITAL LIBRARY EVALUATION METHODS

Evaluation is not just product testing; instead, it can be viewed as a research process. Collecting multiple data and triangulating the results are essential for digital library evaluation (Marchionini, 2000).

After reviewing evaluation studies, Saracevic (2004) summarizes data collection methods: questionnaires, surveys, interviews, focus groups, observations, task accomplishment, think aloud, case studies, transaction log analyses, experimentation, records analysis, usage analysis, document analysis, and economic analysis. He concludes that almost all evaluation methods have been used for digital library evaluation except the historical method. Ethnography is also proposed as an approach to evaluating digital libraries (Crabtree et al., 1997). Bollen and Luce (2002) analyzed usage patterns to evaluate the impact of a digital library and to determine the structure of a given user community. In

addition, focus groups and online surveys have been used to evaluate digital library services (Choudhury et al., 2002). Albertson (2015) suggests mixed data collection methods for user-centered visual digital library evaluation. Survey, log, observation, interview, and results analysis are the suggested data collection methods to assess user, system, and interactions between them.

Multifaceted approaches are also applied to digital library evaluation at various stages of its development. Marchionini et al. (2003) used three case studies to assess user needs in order to design a prototype of the interface for the Library of Congress National Digital Library Program. Multiple types of data collection methods were used to collect data: reading room visits and interviewing staff, questionnaires of prospective users, and document analysis of reading room handouts and brochures, user study reports and email correspondences between users and librarians. Adams et al. (2005) also applied case studies to assess how social context and system design influence the empowerment of users' use of digital libraries. In-depth interviews and focus groups were used to collect data. Multifaceted approaches help researchers and practitioners have a better understanding of the digital library.

For usability testing, questionnaire, interview, focus groups, think aloud, and logs are the most applied data collection methods (Jeng, 2005a, b; Kengeri et al. 1999; Park, 2000). For example, Ferreira and Pithan (2005) used interviews, think alouds, observations, and logs to conduct a usability study. Blandford et al. (2004) highlight two usability approaches: empirical and analytical. Although the empirical approach involves users in the testing, an analytical approach only involves usability experts in the testing by applying established theories and methods. They employ four analytical techniques to evaluate the usability of digital libraries: heuristic evaluation, cognitive walkthrough, claims analysis, and concept-based analysis of surface and structural misfits. In addition, they discuss strengths and limitations of each technique. Buttenfield (1999) identifies different methods for usability evaluation at different stages. In the system design phase, the method of ethnographic evaluation is the most frequently applied for user needs and requirements. Task analysis is another type of evaluation method. Heuristic evaluation is the third method applied at the design and the development phases. At the system development phase, cognitive walkthroughs and interviews are frequently employed to solicit information about interface design. At the system deployment phase, usability testing and usability inspection are performed, in particular the application of transaction log analysis and think-aloud protocols, and pre- and postsurveys. Van House (2003) also emphasizes the reliance on an ethnographic method to study users' behaviors in natural settings. Hilary et al. (2007) performed comparative studies with children searching and browsing using two types of category browsers in the International Children's Digital Library. Their results suggest that a flat, simultaneous interface provides advantages over a hierarchical and sequential interface for children in both Boolean searching and casual browsing. According to Khoo et al. (2012), "the usability and design literature already recognizes that ambiguity and misunderstanding can occur in user studies" (p. 1623). Based on a comparison of usability studies conducted between users and evaluators, they propose support for users to perform digital library evaluation and use their own terms to collect evaluation data. In this way, evaluation data are not lost in translation between researchers' analysis and the presentation of usability data.

After analyzing conference presentations at the Joint Conference on Digital Libraries (JCDL) and the European Conference on Digital Libraries (ECDL), Tsakonas et al. (2013) find that instruments mainly consist of test collections, prototypes, and algorithms. These evaluation studies are commonly conducted in laboratories. Survey and comparison studies are other preferred methods for evaluation. Experiments are employed to compare different digital libraries, different interface designs, and different organization approaches. For example, Meyyappan et al. (2004) designed an experiment to compare

different information organization approaches—alphabetical, subject category, and task-based—by performing a series of task scenarios. The effectiveness and usefulness of digital libraries' information organization approaches were compared. Both quantitative and qualitative data were collected and analyzed. A task-based information organization approach was considered more useful than the traditional approaches.

MULTIFACETED EVALUATION OF DIGITAL LIBRARIES (MEDaL): STUDY METHODOLOGY

This chapter is written based on a review of the previous research as well as our own research on digital library evaluation. The authors conducted a comprehensive study on digital library evaluation, titled "Multifaceted Evaluation of Digital Libraries (MEDaL)." The study employed document analysis and Delphi surveys. Detailed discussion of our study methodology is presented later including sampling, data collection, and data analysis.

In order to identify digital library evaluation dimensions and associated criteria, Joo and Xie (2013) first conducted document analysis focusing on keywords "digital library," "evaluation," "criteria," "assessment," and other associated terms in different combinations. Google Scholar and digital library-related online databases offered by EBSCO were selected to search for relevant documents published primarily between 2000 and 2010. Two criteria were used to make relevance judgments: (1) whether the paper covers any evaluation theories, frameworks, criteria, indicators, or measures or (2) whether the paper consists of actual evaluation studies or pilot tests. Finally, 85 relevant documents and five digital library evaluation project web sites (EQUINOX, DigiQUAL, LibQUAL+, eVALUEd, DELOS) were chosen and further analyzed for digital library evaluation dimensions and corresponding evaluation criteria.

Strauss and Corbin's (1990) open coding technique was selected for data analysis, which is the process of breaking down, examining, comparing, conceptualizing, and categorizing. Ten dimensions emerged from the analysis of the selected literature and the websites, including collection, information organization, interface design, system performance, effects on users, user engagement, services, preservation, sustainability/administration, and context of use. In addition, associated criteria identified in previous works were also incorporated into the next stage of the Delphi survey.

In the empirical phase of the MEDaL study, two-round Delphi surveys were conducted to identify the importance of evaluation criteria and the appropriateness of measures from different stakeholders of digital libraries. We partnered with five academic libraries in the data collection stage. Ninety subjects were recruited with 30 subjects for each of the three groups. The scholar group includes international researchers who have conducted digital library research with high citations or professors who have taught digital library courses. Scholars were identified based on search results from Web of Knowledge or Google Scholar, as well as from web sites of library and information science schools. Digital librarians were randomly selected from the top 200 US colleges (according to US News Rank: www.usnews.com/rankings) that have operating digital libraries, as well as librarians from the partner libraries. User group subjects, which consisted of faculty members and graduate and undergraduate students, were recruited from five partner academic libraries across the country.

The scholar subjects had well-balanced proportions in professor rank and gender. Digital libraries were their major research areas, and other related research areas include the following: information retrieval, metadata, HCI, and preservation. On average, the digital librarian subjects have had about 8.48 years of experience in digital library-related services. Their official titles contained digital librarian,

digital initiative librarian, digital collection librarian, etc. User subjects, consisting of 20 students with 8 undergraduate students and 12 graduate students, as well as 10 faculty members, participated in the survey. This group contained more female subjects than male subjects. Students and faculty members came from different majors/disciplines.

Two rounds of Delphi surveys were administered to different groups of stakeholders of digital libraries including scholars, digital librarians, and users. In the first round, the importance of evaluation criteria was investigated by using a seven-point Likert scale in which 7 is rated as extremely important and 1 as not at all important. The objective of the first round is to determine which evaluation criteria would be important from the perspectives of different stakeholders. To help subjects understand the meaning of evaluation criteria, researchers presented definitions of each criterion to the subjects. Also, the first-round survey instructed subjects to enter additional dimensions and criteria they perceived to be important that were not included as part of the list. Moreover, subjects were also instructed to indicate at which stages of digital library development and operation that each criterion should be applied for evaluation. Seven phases were identified: planning, prototyping, building, testing, launching, operating, and upgrading. Finally, the survey solicits information in relation to purposes of digital library evaluation and the factors that hinder this process. Fig. 10.3 shows an example of the first-round survey.

The results from the first-round survey were incorporated into the design of the second round. The second-round survey concentrated on the examination of the appropriateness of measures to their corresponding criterion. For the second round, 198 measures were identified and presented to the subjects.

DIMENSION 1 - COLLECTIONS

To assess the quality and quantity of digital library collections.

Please rate the importance of the following evaluation criteria in the dimension of collections. Also, please check when to apply the criterion in evaluation

1. Audience (To assess who are the main potential users of a DL)

○ Extremely Important
○ Very Important
○ Somewhat Important
○ Neither Important nor Unimportant
○ Somewhat Unimportant
○ Very Unimportant
○ Not at all Important

1-1 When to evaluate (check multiple choices)?

Planning	Prototyping	Building	Testing	Launching	Operating	Upgrading
☐	☐	☐	☐	☐	☐	☐

FIGURE 10.3 An Example Question from the MEDaL Study About the Importance of a Digital Library Evaluation Criterion

Using a seven-point Likert scale, this study attempted to confirm the appropriateness of measurements for each criterion. Also, the subjects were given the opportunity to modify the measures provided or to suggest new measures.

Since most of the data collected through the Delphi surveys contained numerical ratings, quantitative analysis was applied. For the first-round survey, descriptive statistical analysis was performed, such as average and standard deviation, to show the importance of evaluation criteria. Based on the mean average ratings, evaluation criteria from the most important to the least were ranked. In addition, ratings of digital library evaluation criteria of stakeholders were compared in order to better understand different perspectives and needs of stakeholders. Inferential statistical tests, such as ANOVA and *t*-tests, were conducted to compare the similarities and differences among stakeholders. Similarly, descriptive statistics were used to check the appropriateness of measures to each corresponding criterion. In this chapter, only the descriptive analysis results are reported.

EVALUATION OBJECTIVES AND PHASES
EVALUATION OBJECTIVES

Identification of evaluation objectives is the first step for digital library evaluation because different purposes require different criteria and measurements. This evaluation may serve multiple purposes. Most of the research focuses on users' use of digital libraries and usability evaluation in order to identify user needs and weaknesses of digital libraries, in particular for interface enhancement. Chapters 7 and 8 offered detailed discussion related to usability and user need studies. Some evaluation studies try to compare different designs of the digital library interfaces and help select one design interface over another one (Hilary et al., 2007). Digital library evaluation is conducted for library administrations to understand users' experiences and reactions to their interactions with digital library interface (Mansor and Ripin, 2013). Digital library evaluation is also done to interpret its values based on the perceptions and uses of targeted users (Waugh et al., 2015). In addition, digital library evaluation is used to check the status of its development in a country (Alipour-Hafezi and Nick, 2015). Sometimes, digital library evaluation is performed for multiple purposes. For example, the effectiveness of its services is assessed for making budget decisions as well as to improve the quality of services (Stejskal and Hajek, 2015).

MEDaL investigates digital library evaluation purposes from the three groups of participants. Eleven types of digital library purposes were derived from the document analysis. Based on the average rating from three groups of stakeholders on a seven-point Likert scale (Fig. 10.4), the Delphi surveys reveal that the top three most important purposes are (1) understanding user needs (6.59), (2) identifying problems and weaknesses (6.54), and (3) evidence-based future planning (6.38). Interestingly, users rated "understanding user needs" the highest, 6.83, compared with the other groups (scholars: 6.61, librarians: 6.35). Both scholars (6.69) and librarians (6.48) selected "identifying problems and weaknesses" as the most important purpose of digital library. The least important purposes are (1) promotion/marketing (5.66), (2) benchmarking (5.54), and fundraising and grant writing (5.48).

EVALUATION PHASES

Digital library creation and management go through several phases: planning, prototyping, building, testing, launching, operating, and upgrading. During each phase, the focus of the assessment is different. Each phase requires its own evaluation dimensions, criteria, and methods. At some phases, multiple dimensions
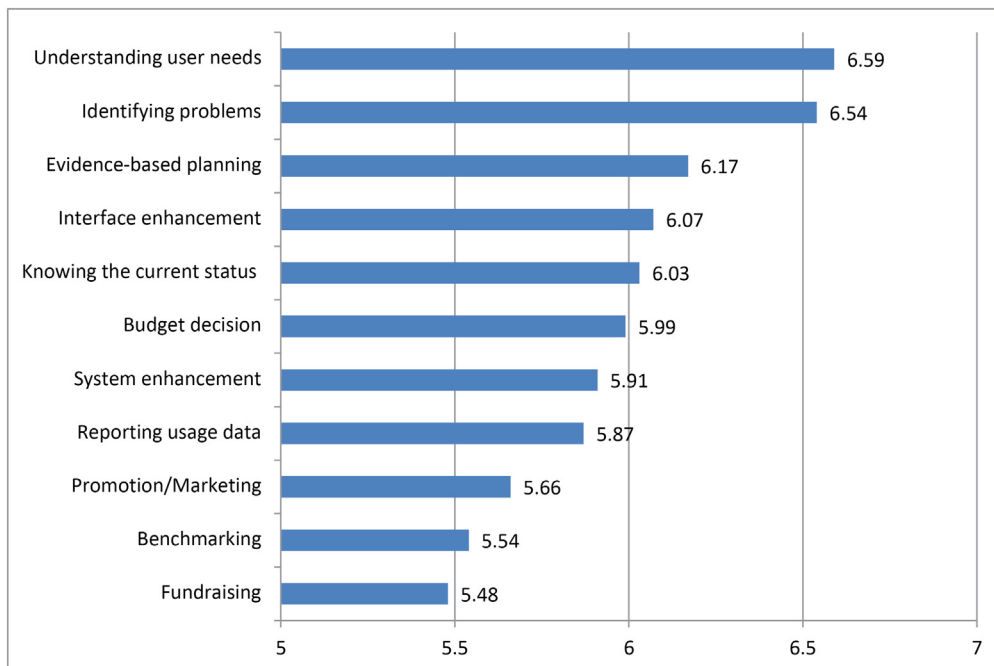
FIGURE 10.4 Average Rating of Digital Library Evaluation Purposes

of the assessment are required; at other phases, a specific evaluative dimension is needed. For example, in the prototyping phase, interface design assessment is the key; in the operating phase, most of the evaluation dimensions need to be considered. Of course, that also depends on the purpose of digital library evaluation. During the operating phase, usability evaluation places an emphasis on the interface and its corresponding criteria (Hilary et al., 2007); digital library value evaluation focuses on user engagement and the effects on user dimensions and related criteria (Waugh et al., 2015); the evaluation of the effectiveness of digital library services concentrates on the service dimension and associated criteria (Stejskal and Hajek, 2015).

Buttenfield (1999) discusses digital library evaluation, in particular usability evaluation during system design, system development, and system deployment. At different stages of digital library evaluation, researchers apply different evaluation methods. Buttenfield further specifies the methods used at different phases. For example, surveys and cognitive walkthroughs are applied at the initial design phase; online user surveys and focus groups are employed at the development phase; and transaction logs are initiated at the deployment phase.

EVALUATION DIMENSIONS AND THEIR IMPORTANCE
EVALUATION DIMENSIONS

Tsakonas and Papatheodorou (2011) define dimensions as a consideration of the scope of evaluation activities and describing evaluation purposes and outcomes. They further identify the following dimensions: effectiveness, performance measurement, service quality, outcome assessment, and technical

excellence. Dimensions are the main constructs of digital library evaluation frameworks and models (refer to the Evaluation Framework section for a discussion of frameworks with associated dimensions). Among the dimensions, Tsakonas et al. (2013) find that effectiveness, performance measurement, and technical excellence are the main research interests in digital library evaluation, which reflect the system-centered digital library evaluation approach, based on the analysis of conference presentations at the JCDL and the ECDL.

Saracevic's (2000, 2004) digital library evaluation framework specifies six dimensions: content, technology, interface, process/service, user, and context. Moreover, Zhang (2010) investigated and validated the importance of Saracevic's six dimensions from different stakeholders of digital libraries. Hu et al. (2014) further examined how the six dimensions influence users' perceptions of university digital libraries. The results yield two types of influences. On the one hand, information-providing services, information retrieval services, and individual services have a direct influence on user perception of university digital libraries. On the other hand, information organization services have an indirect influence on user perception of university digital libraries through information retrieval services and individual services. Cheng (2014) concludes that information relevance, system accessibility, and technical support are the main reasons that users continue using digital libraries.

IMPORTANCE OF DIGITAL LIBRARY EVALUATION DIMENSIONS IN THE MEDaL STUDY

Ten dimensions are identified in the MEDaL study based on document analysis, in particular, of documents regarding digital library evaluation theoretical frameworks, models, and other related literature. The MEDaL study not only discovers the key dimensions but also further reveals the importance of these dimensions from the perceptions of different digital library stakeholders. Table 10.1 presents the ten dimensions, definitions, and the average importance ratings by scholars, digital librarians, and users. Interestingly, interface design (6.38), collections (6.36), and information organization (6.31) were ranked as the most important dimensions. In contrast, administration (5.82), context (5.74), and service (5.65) were rated as the least important dimensions. Curiously, service is deemed as one of the important dimensions for library evaluation but was considered less important in the digital library environment in this study. It needs further examination.

EVALUATION CRITERIA AND THEIR IMPORTANCE
EVALUATION CRITERIA

In the area of LIS research, a number of evaluation criteria have been suggested corresponding to various dimensions of digital libraries. Fuhr et al. (2001) select the following criteria for the four key dimensions: data/collection (content, meta-content, and management), technology (user technology, information access, system structure technology, and document technology), and users/uses (user, domain, information seeking, purpose, and distribution). Saracevic (2004) identifies six classes of criteria, representing content, technology, interface, process/service, user, and context. He summarizes the digital library evaluation criteria applied by related studies consisting of:

- Usability: content (accessibly, availability, clarity, complexity, informativeness, transparency, understanding, effort to understand, adequacy, coverage, overlap, quality, accuracy, validity, reliability, and authority), process (learnability, effort/time, convenience, ease of use, lostness,

Table 10.1 MEDaL: Importance of Ten Digital Library Evaluation Dimensions		
Dimension	**Definition**	**Importance**
Interface design	Assess the usability of digital library interfaces and the extent these interfaces support users' interaction with digital libraries	6.38
Collections	Assess the quality and quantity of digital library collections	6.36
Information organization	Assess the representation, grouping, and presentation of digital information	6.31
Effects on users	Assess the impact and value of digital libraries on users' accomplishing their tasks	6.26
System and technology	Assess the efficiency, accessibility, and reliability of digital libraries as well as their retrieval performance	6.22
Preservation	Assess the extent and ways of digital library support for preservation	5.87
User engagement	Assess the extent and ways of usage of digital libraries and user involvement in their development	5.85
Services	Assess the quality and quantity of the offered digital library services	5.82
Context	Assess the extent of digital libraries fitting into, responding to, following larger context—institutional, economic, legal, social, cultural, and others	5.74
Administration	Assess administrative-related factors that affect the development of digital libraries; assess whether they can be sustained and enhanced	5.65

support, completion, interpretation difficulty, sureness in results, and error rate), format (attractiveness, sustaining efforts, consistency, representation of labels, and communicativeness of messages), overall assessment (satisfaction, success, relevance, usefulness of results, impact, value, quality of experience, barriers, irritability, preferences, and learning).

- System features consisting of technology performance (response time, processing time, speed, capacity, and load), process/algorithm performance (relevance, clustering similarity, functionality, flexibility, comparison with human performance, error rate, optimization, logical decisions, path length, clickthroughs, and retrieval time), and overall system (maintainability, scalability, interoperability, sharability, and costs).
- Usage consisting of usage patterns, use of materials, usage statistics, who uses what and when and for what reasons/decisions.

Gonçalves et al. (2007) propose a quality model for digital library evaluation, which consists of quality dimensions and associated criteria: digital object (accessibility, pertinence, preservability, relevance, similarity, significance, and timeliness), metadata specification (accuracy, completeness, and conformance), collection (completeness), catalog (completeness and consistency), repository (completeness and consistency), and services (composability, efficiency, effectiveness, extensibility, reusability, and reliability). Xie's (2006, 2008) evaluation framework posits the following criteria associated with five dimensions: collection quality (collection quality in general, scope, authority, accuracy, completeness, currency, and copyright), service quality (mission, user community, traditional library service, and unique services), system performance (system performance in general, efficiency and effectiveness,

relevance precision and recall, and usefulness), and user satisfaction (user satisfaction, user feedback, and contact information).

Zhang (2010) identifies the top digital library evaluation criteria that are agreed upon by five groups of stakeholders including administrators, developers, librarians, researchers, and users. These agreed criteria consist of content (accessibility, accuracy, and usefulness), technology (ease of use and reliability), interface (ease of use, effectiveness, and consistency), service (accessibility, integrity, reliability, responsiveness, and usefulness), user (successfulness, satisfaction, and efficiency of task completion), and context (sustainability, collaboration/sharing, and managerial support). Vullo (2010) offers some examples of digital library criteria for the four core dimensions: organization (management and policies), content (quality of data, metadata, and digital collections), service (quality of technologies and quality of design), and user (quality of interactions between users and the digital library).

In addition to comprehensive evaluation criteria, some researchers focus on the usability aspect of digital library and its associated criteria. Kengeri et al. (1999) applied ease of use ratings (easiest to read, easiest to learn, easiest to browse, easiest to search, and easiest overall), search time, and search errors as usability evaluation criteria to compare four digital libraries. Evans et al.'s (2002) usability framework proposes the following criteria: visibility of system status, match between system and the real world, user control and freedom, consistency and standards, error prevention, recognition rather than recall, flexibility and efficiency of use, aesthetic and minimalist design, recovery from errors, and help. Jeng (2005a, b) suggests an evaluation framework for usability of academic digital libraries concentrating on four criteria: effectiveness, efficiency, satisfaction (ease of use, organization of information, labeling, visual appearance, contents, and error correction), and learnability. Tsakonas and Papatheodorou (2008) select the following criteria to evaluate the usefulness (relevance, format, reliability, level, and coverage) and usability (ease of use, aesthetic, navigation, terminology, and learnability) of open access digital libraries. Similarly, Buchanan and Salako (2009) come up with the following digital library usability and usefulness criteria: effectiveness, efficiency, aesthetic appearance, terminology, navigation, learnability, relevance, reliability, and currency.

Hariri and Norouzi (2011) synthesize relevant literature on digital library user interface and usability and recommend the following 22 digital library interface evaluation criteria: navigation, searching, design, guidance, error management, presentation, learnability, user control, consistency, language, feedback, ease of use, match between system and the real world, customization, user support, user workload, interaction, compatibility, visibility of system status, user experience, flexibility, and accessibility. Following Hariri and Norouzi's research, Lai et al. (2014) rank the top five digital library evaluation criteria on user interface: ease of use, searching, language, presentation and design by applying fuzzy analytic hierarchy process (AHP) to obtain weights of evaluation criteria.

Special attention is also paid to digital reference services. Lankes et al. (2003) suggest six criteria as quality standards to assess digital reference services: courtesy, accuracy, satisfaction, repeat users, awareness, and cost. Researchers have also examined relationships among digital library evaluation criteria. According to Jeng (2005a, b), interrelated relationships are found among effectiveness, efficiency, and satisfaction.

Other aspects of digital library evaluation criteria have also been explored. For example, Lagzian et al. (2013) identify corresponding criteria for dimensions that affect motivation (e.g., top management commitment and support, middle management commitment and support, and clear digital library policies and standards), resources (e.g., content production, structural and descriptive metadata, harvesting of resources, and metadata), people (e.g., managing overall digital library, training and education, and

personnel competency), process (e.g., accuracy and reliability of service, rights statement, and digital preservation), location/network (knowledge sharing and scholarly communication), and time (provision of access and fast access to the digital library). Additional digital library evaluation criteria are also identified such as suitability, accuracy, costs, informativeness, timeliness, usefulness, use environment, and others (Kenney et al., 1998; Larsen, 2002; Kim and Kim, 2008).

IMPORTANCE OF DIGITAL LIBRARY EVALUATION CRITERIA IN THE MEDaL STUDY

The results of the Delphi surveys employed in our MEDaL study indicate the importance of digital library evaluation criteria based on the average rating from the three groups of stakeholders for each of the 10 dimensions.

Dimension 1—collections

Overall, digitization standards, authority, and cost turned out to be the three most important evaluation criteria in the dimension of collections, whereas completeness, diversity, and size were the least important. Ratings of importance vary by different groups. Authority, which is related to the reliability of collection quality, was considered the most important by users. Scholars perceived audience as the most important, whereas librarians rated digitization standards the highest. Interestingly, completeness, diversity, and size were considered less important by all three groups of subjects. Table 10.2 presents the importance of evaluation criteria in the dimension of a collection.

Table 10.2 Importance of Evaluation Criteria in the Dimension of Collections

Criterion	Definition	Importance
Digitization standards	To assess the types and methods of digitization practices conducted, as well as identify best practices to build a digital library	6.26
Authority	To assess whether the information provided by a digital library comes from trustworthy sources	6.24
Cost	To assess costs to build collections in a digital library	6.15
Item quality	To assess the quality of a digitized item provided by a digital library within its format	6.13
Format compatibility	To assess whether the format of collections in a digital library are compatible with a variety of software and systems for different purposes	6.08
Audience	To assess who are the main potential users of a digital library	6.05
Scope/Coverage	To assess the range of topics that are covered by a digital library	5.62
Contextual information	To assess what type of additional, related information, such as bibliographies, biographies, encyclopedia entries, timelines, and interpretive essays, are available to users in a digital library	5.51
Completeness	To assess whether a digital library covers all documents in each topic area	5.29
Diversity	To assess whether a digital library deals with a variety of issues in relation to a topic of interest	5.16
Size	To assess the amount of collection items provided by a digital library	5.10

Table 10.3 Importance of Evaluation Criteria in the Dimension of Information Organization		
Criterion	**Definition**	**Importance**
Appropriateness	To assess whether the organizational structure and associated categories adequately organize items in a digital library	6.21
Accessibility to metadata	To assess how easily a user can obtain the metadata information of each item	6.19
Metadata accuracy	To assess how accurately metadata elements are assigned for each item	6.16
Metadata standards	To assess whether metadata elements follow predefined standard and guides	6.04
Consistency	To assess whether metadata are consistent across collections in a digital library	5.93
Comprehensiveness	To assess whether the organization structure covers all the access points of a digital library	5.88
Depth of metadata	To assess the levels of the metadata scheme used for a digital library	5.83
Metadata interoperability	To assess whether metadata elements of a digital library are compatible to different digital libraries	5.67
Controlled vocabulary	To assess the types and numbers of controlled vocabularies used in a digital library	5.56

Dimension 2—information organization

For the dimension of information organization, appropriateness, accessibility to metadata, and metadata accuracy were ranked first, second, and third, respectively, whereas depth of metadata, metadata interoperability, and controlled vocabulary were perceived the least important. Scholars rated appropriateness the most important, whereas librarians chose accessibility to metadata. Users selected metadata accuracy, the most important with regard to information organization in digital libraries. Table 10.3 presents the importance of evaluation criteria in the dimension of information organization.

Dimension 3—interface design

In terms of interface design, all three groups regarded search function and browsing function as the most important criteria in evaluating digital libraries. Searching and browsing are the two main approaches in the information search process, and subjects rated search and browsing functions as the top two criteria. Navigation and intuitive operation were chosen as relatively important evaluation criteria across the three groups. Visual appeal, user control, and personalized page were rated least important in the dimension of interface design. Table 10.4 presents the importance of evaluation criteria in the dimension of interface design.

Dimension 4—system and technology

In the dimension of system and technology, retrieval effectiveness, reliability, and server performance were identified in the MEDaL as the most important criteria. As digital libraries are considered as a type of information retrieval system, subjects thought retrieval effectiveness, such as precision and recall, was important in evaluating them. Simultaneously, reliability and server performance were the

Table 10.4 Importance of Evaluation Criteria in the Dimension of Interface Design		
Criterion	**Definition**	**Importance**
Search function	To assess what types of search functions are offered by a digital library and how easy it is to use them	6.54
Browsing function	To assess in what ways and to what extent the interface supports a user's ability to surf related items in a digital library	6.34
Navigation	To assess in what ways and to what extent the interface supports a user's exploration in a digital library	6.26
Intuitive operation	To assess how straightforward a digital library interface is for a user to understand its operation and how easily a user can learn to operate the interface	6.25
Search results presentation	To assess how the types of formats/options of search results are presented to users in a digital library	6.18
Consistency	To assess whether the design and layout are coherent across a digital library interface	6.12
Reliability	To assess the ability of a digital library to perform and maintain its functions under different circumstances	6.11
Help function	To assess what types of help functions are offered and how effectively they support users in their help-seeking process	5.64
Visual appeal	To assess to what extent the interface of a digital library is visually attractive to users	5.61
User control	To assess to what extent a digital library allows users to manipulate its interface	4.95
Personalized page	To assess whether a digital library offers personalized pages based on user profile	4.25

criteria that were needed to provide stable services in digital library. Less important criteria were error rate/error correction, flexibility, and linkage with other digital library. Table 10.5 presents the importance of evaluation criteria in the dimension of system and technology.

Dimension 5—effects on users
In the ratings of effects on users, research productivity and learning effects were chosen as the most important criteria. This is not surprising because this survey was conducted in an academic setting. Perceptions of digital libraries and information literacy/skill change were regarded less important. It seems that the subjects considered real effects more important than perceptions. The difficulty of conducting longitudinal studies to identify the change might have contributed to the lower rating of information literacy/skill change. Table 10.6 presents the importance of evaluation criteria in the dimension of effects on users.

Dimension 6—services
In the dimension of services, subjects chose service quality, usefulness, and user satisfaction as the three most important criteria. Service quality, usefulness, and user satisfaction are commonly applied

Table 10.5 Importance of Evaluation Criteria in the Dimension of System and Technology

Criterion	Definition	Importance
Retrieval effectiveness	To assess how effective the search algorithm is in a digital library	6.31
Reliability	To assess how stable a digital library performance is over time	6.18
Server performance	To assess the ability of a server to run a digital library	6.17
Response time	To assess how quickly a digital library responds to a user's request	6.13
Fit-to-task	To assess to what extent a digital library is adequate to perform tasks that a user requests	6.04
Connectivity	To assess how stable a digital library system is when connected to other information systems	5.99
Page loading speed	To assess how quickly a digital library presents a user-requested page	5.97
Integrated search	To assess whether a digital library offers an integrated search environment for different collections within a digital library	5.95
Error rate/ error correction	To assess the degree of errors encountered during the use of the content management system and the ability to fix the errors	5.93
Flexibility	To assess whether a digital library responds to potential internal or external changes in a timely manner	5.66
Linkage with other digital libraries	To assess the identification of and in what ways a digital library is linked to other related digital libraries	5.29

Table 10.6 Importance of Evaluation Criteria in the Dimension of Effects on Users

Criterion	Definition	Importance
Research productivity	To assess in what ways and to what extent a digital library affects a user's research outputs	5.46
Learning effects	To assess in what ways and to what extent a digital library influences a user's learning outcome	5.30
Knowledge change	To assess in what ways and to what extent a digital library influences a user's knowledge structure	5.04
Instructional efficiency	To assess in what ways and to what extent a digital library enhances a user's teaching effectiveness	4.91
Perception of digital libraries	To assess in what ways and to what extent a digital library influences a user's view of digital libraries	4.84
Information literacy/Skill change	To assess to what extenta digital library enhances a user's literacy skills	4.64

criteria for service evaluation in general. The evaluation criterion for services for users with disabilities was rated fourth. On the other hand, user education, types of unique services, and customized services were ranked as the least important criteria. It seems that subjects cared more for the quality and usefulness of the digital library services and less for the unique services offered to users. Table 10.7 presents the importance of evaluation criteria in the dimension of services.

Dimension 7—preservation

In the dimension of preservation, completeness, ability to migrate, and preservation policy were ranked first, second, and third, respectively. Institutional support, types of archiving methods, and cost per

Table 10.7 Importance of Evaluation Criteria in the Dimension of Services		
Criterion	**Definition**	**Importance**
Service quality	To assess in what ways and to what extent digital library services satisfy users' needs	6.24
Usefulness	To assess in what ways and to what extent digital library services are useful for users to achieve their tasks	6.23
User satisfaction	To assess in what ways and to what extent users are satisfied with services provided by a digital library	6.18
Types of services for users with disabilities	To assess the types of services offered to users with disabilities	6.12
Reliability	To assess how users of a digital library perceive the trustworthiness of services provided	6.01
Responsiveness	To assess the reaction time to a user's request for a digital library service	5.97
Timeliness	To assess in what ways and to what extent services are offered to users in a timely manner	5.94
Types of services	To assess the types of services provided by a digital library	5.62
Availability of digital library staff	To assess in what ways and to what extent a user can easily contact staff of a digital library for questions, feedback, and comments	5.61
Confidence	To assess in what ways and to what extent users have a positive attitude toward services offered by a digital library	5.55
Follow-up services	To assess in what ways and to what extent adequate and timely continuing services are provided to users by a digital library when necessary	5.21
FAQ/Q&A	To assess whether and how many FAQs or Q&As a digital library provides to help users in using it	5.13
User education	To assess the types of user education offered by a digital library	5.10
Types of unique services	To assess the unique types of services provided by a digital library compared to other related digital libraries	4.79
Customized services	To assess whether a digital library offers personalized services based on user profile information or user requests	4.75

Table 10.8 Importance of Evaluation Criteria in the Dimension of Preservation

Criterion	Definition	Importance
Completeness	To assess to what extent the archiving process to preserve archived materials is complete and comprehensive	6.22
Ability to migrate	To assess the ability of data migration for preservation.	6.19
Preservation policy	To assess whether a policy regarding preservation is developed and what is covered by the policy	6.00
Preservation infrastructure	To assess the types of archiving equipment and facilities a digital library uses for digital preservation	5.85
Institutional support	To assess how many staff members or resources are dedicated to the preservation tasks of a digital library	5.76
Types of archiving methods	To assess the archiving methods/approaches a digital library staff member has to apply in order to preserve information	5.73
Cost per record	To assess the average cost for archiving one record	5.44

record were less important according to the survey. Overall, librarians assigned a higher score of the importance of evaluation criteria to the dimension of preservation. This reveals that librarians emphasize preservation more than the other two groups. Table 10.8 presents the importance of evaluation criteria in the dimension of preservation.

Dimension 8—administration
Since users do not have sufficient knowledge of digital library, they were excluded from the survey in this dimension. Based on the responses from scholars and librarians, budget, planning, and staffing turned out to be the most important criteria in the dimension of administration. Management policy, fundraising/sponsor, and incentive were considered less important by the two groups. Table 10.9 presents the importance of evaluation criteria in the dimension of administration.

Dimension 9—user engagement
In the dimension of user engagement, resource use, user feedback, and site visit were the three highly rated evaluation criteria by three groups. Resource use is one of the fundamental criteria in the evaluation of libraries, and it is also perceived as an important evaluation criterion in the context of a digital library. Moreover, all three groups perceived user feedback as the important criterion to judge user engagement. On the other hand, user participation channels, user knowledge contribution, and e-commerce support were perceived as less important. Table 10.10 presents the importance of evaluation criteria in the dimension of user engagement.

Dimension 10—context
Finally, subjects selected copyright, information ethics compliance, and organizational mission as the most important evaluation criteria in this dimension. Again, the groups from the academic library setting expressed their view on the importance of information ethics compliance. On the other hand, content sharing, collaboration, and social impact were considered less important. Table 10.11 presents the importance of evaluation criteria in the dimension of context.

Table 10.9 Importance of Evaluation Criteria in the Dimension of Administration

Criterion	Definition	Importance
Budget	To assess the amount and arrangement of monetary resources to efficiently manage a digital library	6.18
Planning	To assess whether strategic plans are established to create, manage, maintain, and enhance a digital library	6.10
Staffing	To assess the quantity and arrangement of human resources to efficiently manage a digital library	6.02
Staff training	To assess the types, frequency, and efficiency of training programs offered to digital library staff	5.79
Marketing	To assess the publicity efforts of a digital library to attract potential users and inform related communities	5.67
Regular assessment	To assess whether regular, continuous evaluation is performed to maintain and enhance a digital library	5.59
Management policy	To assess whether a well-defined policy for administration is offered by a digital library	5.40
Fundraising/sponsor	To assess the effort of fundraising to support a digital library financially	5.28
Incentive	To assess the types of incentives provided to digital library staff	4.32

Table 10.10 Importance of Evaluation Criteria in the Dimension of User Engagement

Criterion	Definition	Importance
Resource use	To assess in what ways and to what extent users use resources in a digital library	6.04
User feedback	To assess the types of user comments and suggestions received by a digital library and in what ways and to what extent these comments and suggestions are incorporated into the enhancement of the digital library	5.97
Site visit	To assess how frequently users visit a digital library web site and the duration for each visit	5.74
Integration with external applications	To assess in what ways users can export digital objects and integrate with external applications, such as slide presentation software	5.51
Help feature use	To assess which help features are offered to users, how frequently, and in what context users try to use help-related features in a digital library	5.43
User participation channels	To assess the types of channels available to users to communicate with the staff of a digital library	5.39
User knowledge contribution	To assess the ways users can contribute to digital library content and organization through tagging, commenting, and adding their own objects	5.26
E-commerce support	To assess the capabilities of ordering digital objects online	4.61

Table 10.11 Importance of Evaluation Criteria in the Dimension of Context

Criterion	Definition	Importance
Copyright	To assess whether a digital library identifies and conforms to copyright issues	6.26
Information ethics compliance	To assess whether a digital library identifies and conforms to ethical issues related to its creation and use	6.23
Organizational mission	To assess in what ways and to what extent digital library creation and use conform to organizational objectives	5.79
Targeted user community	To assess in what ways and to what extent a digital library engages in targeted user groups	5.77
Content sharing	To assess in what ways and to what extent stakeholders of a digital library are willing to share their content	5.69
Collaboration	To assess in what ways and to what extent stakeholders of a digital library work together; to assess in what ways and to what extent stakeholders of a digital library cooperate with stakeholders of another one	5.30
Social impact	To assess in what ways and to what extent the use of a digital library influences society	4.97

EVALUATION MEASUREMENTS
DIGITAL LIBRARY EVALUATION MEASUREMENTS

In order to effectively perform digital library evaluations, not only the criteria but also measurements are essential. Compared to the evaluation, it is more difficult to validate its measurements. Marchionini (2000) makes two suggestions for digital library evaluation measurements. First, it is important to integrate as many specific measures instead of depending on one single measure. Second, it is vital to integrate statistical data and qualitative data. Albertson (2015) also stresses the importance of having both quantitative and qualitative measures for user-centered visual digital library evaluation. Just as Heradio et al. (2012) states, "there is an increasing trend to blend quantitative and qualitative data within a study to provide a broader and deeper perspective" (p. 277).

Researchers have proposed different types of measurements for either overall digital library evaluation or its specific aspects. Gonçalves et al. (2007) suggest mostly quantitative measurements for the quality evaluation of digital libraries—for example, accurate attributes, number of attributes in the record for accuracy of and missing attributes, schema size for completeness of and conformant attributes, and schema size for conformance of metadata specification. Heradio et al. (2012) review the quality digital library evaluation from users' perceptions, in particular related to utility, usability, and their costs. Specifically, they introduce two alternative measurements: Likert scales and fuzzy linguistic information. They conclude that Likert scales have the advantage of measuring user's opinion distribution although "the assumption of interval for Likert data in the digital library context has to be justified" (p. 280) whereas fuzzy linguistic modeling generates better results. Buchanan and Salako (2009) recommend measures for each of the digital library usability and usefulness criteria. For example, tasks completed for effectiveness, time to complete for efficiency, attractiveness for aesthetic,

appropriateness for appearance, comprehension and consistency for terminology, steps to complete for navigation, repetition of failed commands for learnability, relevant results and utility for relevance, credibility for reliability, and creation date and last citation for currency.

Electronic resource evaluation accounts for a large part of digital library evaluation. Noh (2010) identifies multiple sectors and their corresponding evaluation indices for electronic resource development and uses. Based on three rounds of the Delphi survey, Noh presents evaluation measurements for evaluation criteria in relation to e-resource acquisition, e-resource use, and environment for e-resource use. For example, annual number of sessions to web database (DB) per service recipient, and the annual number of DB hits per service recipient, and annual number of Web DB downloads per service recipient are the measurements of use of Web DB. Blixrud (2002, 2003) and Shim (2002) report the project conducted by members of the Association of Research Libraries (ARL). They developed measures for the assessment of electronic resources in terms of resources, expenditures, and usage. Among the measures, library digitization activities are assessed by size of library digital collection, use of library digital collection, and cost of digital collection construction and management.

Evaluation of services offered by digital libraries is an area of keen interest to researchers and practitioners. Brophy (2001) summarizes digital library performance indicators based on the opinions from the professional community as part of the EQUINOX project: percentage of the population reached; number of sessions on each digital library service per member of the target population; number of remote sessions on these services per member of the population to be served; number of documents and entries viewed per session for each service; cost per session for each service; cost per document or entry viewed for each service; percentage of information requests submitted electronically; library computer workstation use rate; number of library computer workstation hours available per member of the population to be served; rejected sessions as a percentage of total attempted sessions; percentage of total acquisitions of expenditure spent on acquisition of services; number of attendees at formal service training sessions per member of the population to be served; number of library staff developing, managing, and providing digital libraries and user training as a percentage of total library staff; and user satisfaction with digital library services. Lankes et al. (2003) propose five types of performance measures to assess digital reference services: descriptive statistics and measures, log analysis, user satisfaction measures, cost, and staff time expended.

- *Descriptive measures:* the number of digital reference questions received, number of digital reference responses, number of digital reference answers, total reference activities, percentage of digital reference questions to total reference questions, digital reference fill rate, digital reference completion rate, number of unanswered digital reference questions, types of digital reference questions received, saturation rate, sources used per question, and repeat users.
- *Examples of log measures:* the number of digital reference sessions, usage of digital reference services by day of the week.
- *User measures:* awareness of service, accessibility of service, expectation for service, etc.
- *Cost measures:* cost of digital reference service, cost of digital reference service as a percentage of total reference budget, etc.
- *Staff measures:* percentage of staff time spent overseeing technology and assisting users with technology.

Even though specific measures have been suggested for different aspects of digital library evaluation, there is a lack of measure for all the dimensions. Moreover, there is no systematic analysis and

discussion about measures for each criterion under each dimension. Our MEDaLstudy presented, as follows, fills in the gap.

APPROPRIATENESS OF MEASUREMENTS IN THE MEDaL STUDY

Measures are the foci of the second part of the MEDaL study presented in this chapter. In the second round, we suggested specific measures and operational definitions for evaluation criteria. In this round, the user group was excluded from the survey because they do not have in-depth knowledge of measurements. Fifty-five subjects out of 61 participated in the second-round survey, which is a 90.1% participation rate. They were asked to rate the appropriateness of measures based on the seven-item Likert scale (7 as the most appropriate and 1 as not at all appropriate). This section summarizes the survey results. Different measurements for each criterion are listed based on their rated appropriateness from high to low in the following tables.

Dimension 1—collections

Twenty-one measures in this dimension were suggested in this study, and scholar and librarian subjects were instructed to rate the appropriateness of each measure to its corresponding criterion. Subjects gave a rating over 6 for compliance with digitization standards (6.327), quality specification (6.164), and presence of resource reference information (6.036). However, level of domain knowledge (4.985), presence of diverse perspectives (4.945), and potential user demographic data (4.873) were rated the least appropriate to explain their criteria, respectively. Table 10.12 presents the appropriateness of evaluation measurements in the dimension of collections.

Dimension 2—information organization

In the second dimension—information organization—total 17 measures for 9 criteria were presented to the subjects. All measures were rated over 5 in terms of their appropriateness. There were three measures that are rated over 6: compliance to the metadata standards (6.164), compliance to interoperability standards (6.109), and incorrect data value (6.091). The least three appropriate measures were metadata elements used (5.127), subject analysis (5.036), and depth of description (5.036). Table 10.13 presents the appropriateness of evaluation measures in the dimension of information organization.

Dimension 3—interface design

In the interface design dimension, 32 measures were identified for 10 criteria. Eight measures were rated over 6 for their appropriateness, which indicates at least "very appropriate." These 8 measures are as follows: search function usefulness (6.327), types of search features (6.309), search function ease of use (6.291), overall ease of use (6.200), design consistency (6.164), browsing function usefulness (6.036), navigation usefulness (6.036), and browsing function ease of use (6.018). However, 2 measures suggested for the criterion of user participatory design were perceived less appropriate: extent of user participation (5.055) and types of user participation in interface design (5.018). Also, use of help features was considered least appropriate among 32 measures in this dimension. Table 10.14 presents the appropriateness of evaluation measures in the dimension of interface design.

Dimension 4—system and technology

In Dimension 4, 21 measures for 15 criteria were shown to the subjects. These measures are mostly adopted from evaluation research on information retrieval evaluation or system evaluation. Six

Table 10.12 Appropriateness of Evaluation Measures in the Dimension of Collections

Criterion	Measure and Definition
Digitization standards	*Compliance with digitization standards:* Whether a digital library adheres to the established digitization standards
Authority	*Presence of resource reference information:* Whether resource reference information for each item is available
Item quality	*Quality specification:* Technical specification for creating digitized objects
Cost	*Cost of metadata:* Average cost for creating metadata per record
	Cost for building a digital collection: Average cost for building a collection
	Cost of conversion: Average cost for converting to a digitized item
Format compatibility	*Types of access files:* Types of access files used in the collection
	Data type: Types of data used in the collection
Audience	*User demographic data:* Whether user information data are collected
	Potential user demographic data: Types of potential users and their demographic characteristics
Scope/Coverage	*Time span of coverage:* Time period covered in the collections
	Subject coverage: Number of topics in a digital library
Contextual information	*Presence of contextual information for collection:* Whether there are secondary resources for digital collections to provide contextual information
Completeness	*Item size in specific topic:* Number of items per topic
Diversity	*Presence of diverse perspectives:* Whether a digital library contains diverse perspectives on a topic
Size	*Collection size:* Number of digitized objects
Collection development policy	*Presence of collection development policy:* Whether a digital library has a documented policy about collection development
	Components of policy: Types of components of collection development policy
Currency	*Currency of collections:* Proportions of newly archived collections in recent years
Re-use	*Re-use of digital objects:* Types of digital objects that can be reused
Collection developer knowledge	*Level of domain knowledge:* To what extent collection developers have domain knowledge

measures were perceived "very appropriate," which scored over 6 on average on a seven-point scale: precision (6.273), system failure (6.218), response time to search results (6.200), recall (6.055), system response time (6.036), and search across collections (6.000). On the contrary, replacement and update of equipment (5.109), use of emerging technologies (5.018), and presence of open source (4.909) were rated relatively lower in this dimension. Table 10.15 presents the appropriateness of evaluation measures in the dimension of system and technology.

Table 10.13 Appropriateness of Evaluation Measures in the Dimension of Information Organization

Criterion	Measure and Definition
Appropriateness	*Domain appropriateness:* Metadata appropriateness as judged by domain experts
	User-perceived appropriateness: Metadata appropriateness as judged by users
Accessibility of metadata	*Ease of access to metadata:* Users' perceived accessibility of metadata
Metadata accuracy	*Incorrect data value:* Percentage of incorrect data values
	Inaccurate data entry: Percentage of inaccurate data entry
Metadata standards	*Compliance to the metadata standards:* Proportion of accurately mapped elements compared to the selected metadata schema
	Types of metadata standards: Whether a digital library adheres to the selected metadata standard
Consistency	*Metadata element consistency:* The extent to which the selected metadata elements are used for data input across collections in a digital library
	Metadata schema consistency: The extent to which the selected metadata schema is used for data input across collections in a digital library
Comprehensiveness	*Completed metadata:* Average number of metadata fields populated per record
	Metadata elements used: Percentage of metadata elements used compared to the selected metadata schema
Depth of metadata	*Subject analysis:* Average number of subject terms per record
	Depth of description: Length of item description
Metadata interoperability	*Compliance to interoperability standards:* Whether a digital library complies with interoperability standards
Controlled vocabulary	*Presence of controlled vocabularies:* Whether a digital library uses controlled vocabularies in organizing objects
	Ease of access to controlled vocabularies: Users' perceived access to controlled vocabularies
	Presence of controlled vocabularies: Whether a digital library offers controlled vocabularies

Dimension 5—effects on users

In the effects on users dimension, 11 measures were identified for 5 evaluation criteria. In this dimension, no measure scored over 6. This dimension contains many criteria that are related to change, and it is hard to measure them by a simple study. The top three ranked measures in terms of appropriateness are willingness to continue use of the digital library (5.982), in teaching (5.691), and for research (5.636). On the contrary, two measures were rated less than 5: attitude change after digital library uses (4.982) and perceived information literacy/skill (4.982). Table 10.16 presents the appropriateness of evaluation measures in the dimension of effects on users.

Table 10.14 Appropriateness of Evaluation Measures in the Dimension of Interface Design

Criterion	Measure and Definition
Search function	*Usefulness:* Users' perceived usefulness of the search function
	Types of search features: Types of search features available
	Ease of use: Users' perceived ease of use
	Use of search features: Average frequency of and time spent on search feature use in a session
Browsing function	*Usefulness:* Users' perceived usefulness of the browsing function
	Ease of use: Users' perceived ease of use
	Browsing access points and paths: Types of browsing access points and paths available
	Organization of browsing structure: Experts' assessment of logic and quality of the browsing structure
	Use of browsing features: Average frequency of and time spent on browsing feature use in a session
Navigation	*Usefulness:* Users' perceived usefulness of the navigation features
	Ease of use: Users' perceived ease of use
	Navigation features: Types of navigation features available
	Use of navigation features: Average frequency of and time spent using the navigation features in a session
Intuitive operation	*Overall ease of use:* Users' perceived ease of use to operate the interface of a digital library
Consistency	*Design consistency:* Consistency in fonts, layout, menus, colors, etc
	Consistency from user perspective: Users' perceived consistency of the interface
Help function	*Ease of use:* Users' perceived ease of use of the help function
	Usefulness: Users' perceived usefulness of the help function
	Help use situations: Under what situations users use help features
	Types of help features: Types of help features available
	Use of help features: Average frequency of and time spent on help feature use in a session
Visual appeal	*Visual aesthetics:* Users' perception of the interface aesthetics
User control	*Usefulness:* Users' perceived usefulness of user control features
	Ease of use: Users' perceived ease of use
	Types of user control features: Types of user control features available
	Use of user control feature: Average frequency of and time spent on using each type of user control feature in a session
Personalization feature	*Ease of use:* Users' perceived ease of use
	Usefulness: Users' perceived usefulness of personalization features
	Types of personalization features: Types of personalization features provided
	Use of personalization features: Average frequency of and time spent on personalization feature use in a session
User participatory design	*Extent of user participation:* To what extent users participate in interface design
	Types of user participation in interface design: Types of user participation in the process of interface design

Table 10.15 Appropriateness of Evaluation Measures in the Dimension of System and Technology

Criterion	Measure and Definition
Retrieval effectiveness	*Precision:* Precision = Number of relevant items retrieved/number of retrieved items
	Recall: Recall = Number of relevant items retrieved/number of relevant items in the digital library collection
	Aspectual recall: Ratio of aspects of the search topic identified in the documents saved by the subject to the total number of aspects of the topic
Retrieval efficiency	*Response time to search results:* Response time to present search results after a search request is submitted
Server performance	*Bandwidth:* Bandwidth speed
	Traffic: The volume of total traffic accessing a digital library site
Reliability	*System failure:* Number of system failures occurring in a specific period of time
Response time	*System response time:* End-to-end response time after a page request is made
Fit to task	*Perceived fit to task:* The extent to which a user perceives the appropriateness of a digital library to carry out his/her search task
Speed of page loading	*Page loading speed:* Average downloading speed per page
System connectivity	*Compatibility with other types of systems:* Technical ability to connect to other types of systems
	Ease of connection: Experts' assessment of ease of connection to other systems
Error rate and correction	*Error rate:* Ratio of number of error occurrences over number of page attempts
	Error correction rate: Ratio of corrected errors compared with errors encountered
Integrated search	*Search across collections:* Whether a digital library provides an integrated search function across multiple collections
Customizability	*Capability to adopt system features:* To what extent a digital library is able to add or customize new features
	Availability of API: Whether a digital library provides APIs to developers or users
Open source	*Presence of open source:* Whether a digital library provides a platform for open source
Emerging technologies	*Use of emerging technologies:* Types of emerging technologies incorporated in a digital library
Lifecycle of equipment	*Replacement and update of equipment:* Frequency of replacement and upgrade of equipment
Technical support	*Technical support from the IT team:* Types of technical support from institution's IT team

Table 10.16 Appropriateness of Evaluation Measures in the Dimension of Effects on Users	
Criterion	**Measure and Definition**
Research productivity	*Digital library uses for research:* Frequency of digital library use for research purposes
	Effects of digital library uses on research: In what ways a digital library enhances a user's research productivity
	Research productivity change: To what extent a digital library enhances a user's research productivity
Domain knowledge change	*Perceived domain knowledge change:* Perceived increase of domain knowledge after using a digital library
	Domain knowledge change after digital library uses: Domain knowledge change between the pretest and posttest after digital library use study
Instructional effectiveness	*Digital library uses in teaching:* Frequency of digital library uses in teaching
	Effects of digital library uses on teaching effectiveness: In what ways a digital library improves users' teaching effectiveness
Perception of digital libraries	*Attitude change after digital library uses:* To what extent users change attitude toward digital libraries in general
	Willingness to continue use of a digital library: Perceived willingness to continue use of a digital library
Information literacy/ skill change	*Perceived information literacy/skill change:* Perceived improvement of information literacy skill after using a digital library
	Change of information literacy/skill: Change between the pretest and posttest after using a digital library

Dimension 6—services

In the services dimension, 26 measures were identified for 10 criteria. There were two measures rated over 6: overall usefulness (6.145) and overall satisfaction (6.091). These two measures are frequently used in service evaluation, and they were also selected as highly appropriate measures in the context of digital libraries. However, ratings for types of services—uniqueness (4.945), number of reference services provided (4.909), and staff accessible hours (4.873) were comparatively lower. Table 10.17 presents the appropriateness of evaluation measures in the dimension of services.

Dimension 7—preservation

In the preservation dimension, 11 measures were identified for 6 criteria. Among them, exporting capability (6.309), presence of preservation policy (6.164), and migratable data type (6.036) were rated highly appropriate. On the other hand, components of preservation policy (5.618), refresh frequency (5.473), and preservation cost per record (5.400) were rated less appropriate to account for associated evaluation criteria. Table 10.18 presents the appropriateness of evaluation measures in the dimension of preservation.

Dimension 8—administration

Regarding the administration dimension, 25 measures were identified for 11 evaluation criteria. Among them, presence of copyright policy (6.309), presence of sustainability plans (6.091), and presence of strategic plans (5.782) were determined to be the three most appropriate measures for their associated

Table 10.17 Appropriateness of Evaluation Measures in the Dimension of Services

Criterion	Measure and Definition
Overall usefulness	*Perceived overall usefulness:* Users' perceived usefulness of overall services provided in a digital library
	Ways of usefulness: In what ways digital library services are useful to users
Overall satisfaction	*Perceived overall satisfaction:* Users' perceived satisfaction to overall services provided in a digital library
Services for users with disabilities	*Types of services:* Types of services for people with disabilities offered in a digital library
	Usefulness: Disabled users' perceived usefulness
	Ease of use: Disabled users' perceived ease of use of digital library services
	Frequency of service uses by people with disabilities: Frequency of each type of service used by people with disabilities in a specific time period
Overall reliability	*Reliability of services:* Users' perceived reliability
Overall responsiveness	*Perceived responsiveness:* Service responsiveness rated by users
Update	*Update frequency:* Frequency of and types of services updated in a specific time period
Types of services	*Types of user services:* Types of user services offered in a digital library
	Usefulness: Users' perceived usefulness for each type of service
	Ease of use: Users' perceived ease of use for each type of service
	Uniqueness: Number of unique services offered in a digital library compared to library services or other digital library services
	Frequency of service uses: Frequency of each type of service used in a specific time period
Accessibility to managerial staff	*Perceived availability of staff:* User perception of the availability of digital library management staff
	Staff accessible hours: Number of hours users can access digital library management staff
Reference services	*Types of reference services:* Types of reference services offered in a digital library
	Usefulness: Users' perceived usefulness of reference services offered in a digital library
	Ease of use: Users' perceived ease of use of reference services
	Number of reference services provided: Number of times/instances that reference services are provided to users in a specific period of time
	Response time to digital library reference requests: Average response time to a reference request regarding digital library resources
Customized services	*Types of customized services:* Types of customized services offered by a digital library
	Usefulness: Users' perceived usefulness of customized services
	Ease of use: Users' perceived ease of use of customized services
	Use of customized services: Frequency of each type of customized service used in a specific period of time

Table 10.18 Appropriateness of Evaluation Measures in the Dimension of Preservation	
Criterion	**Measure and Definition**
Ability to migrate	*Migratable data type:* Types of data that can be migrated to a digital library
	Exporting capability: Whether a digital library has a function to export data in different formats for preservation
Preservation policy	*Presence of preservation policy:* Whether a digital library has a documented policy regarding preservation practices
	Components of preservation policy: Types of components in the preservation policy
	Strategies of preservation: Types of strategies presented in the preservation policy
Preservation infrastructure	*Types of preservation tools:* Types of preservation tools offered
Institutional support	*Types of support:* Types of support offered by the institution
	Level of support: The extent of support offered from the institution
Cost per record	*Preservation cost per record:* Average cost for preserving a record
Ability to refresh	*Refreshable data type:* Types of data that can be refreshed
	Refresh frequency: How frequently data are refreshed

criteria. On the contrary, the three least appropriate measures were as follows: frequency of marketing/promotion activities (4.982), student hours on digital library (4.800), and number/amount of grant/fundraising attempts (4.673), each of which was given less than five points in the survey. Table 10.19 presents the appropriateness of evaluation measures in the dimension of administration.

Dimension 9—user engagement
In the user engagement dimension, 18 measures were identified for 7 criteria. In particular, 5 measures were suggested for the criterion of site visits, such as frequency, session length, and unique visits. In this dimension, most of the measures were related to measuring resource usage in digital library. As to the most appropriate measures, item viewed (6.000), item downloading (6.000), and frequency of site visits (5.982) were highly rated. Also, the top seven most appropriate measures are related to resource uses in digital library. On the contrary, quantity of user feedback (5.473), ease of purchasing (5.091), and number of orders (4.909) were rated less appropriate. Table 10.20 presents the appropriateness of evaluation measures in the dimension of user engagement.

Dimension 10—context
Finally, 16 measures were suggested for 9 criteria in the context dimension. Interestingly, there was no measure rated over 6 on average, and overall the ratings were relatively lower compared to measures in other dimensions. This implies that they are more difficult to measure. Most of the measures in this dimension are qualitative, which are more prone to subjectivity. The three most appropriate measures were deemed to be user community engagement (5.727), organizational mission—ways of support (5.691), and level of user community engagement (5.691). On the contrary, types of collaborations (5.182), number of collaborations (5.127), and components of guidelines for ethics (5.073) were rated least appropriate. Table 10.21 presents the appropriateness of evaluation measures in the dimension of context.

Table 10.19 Appropriateness of Evaluation Measures in the Dimension of Administration

Criterion	Measure and Definition
Copyright	*Presence of copyright policy:* Whether a digital library has a documented policy regarding copyright management
	Components of copyright policy: Types of components in the copyright policy
Budget	*Budget amount:* Total amount of budget for a digital library
	Proportion of digital library budget: Proportion of digital library budget over total library budget
	Distribution of digital library budget: Distribution of budget for different components of a digital library
Planning	*Presence of strategic plans:* Whether there are documented strategic plans for a digital library
	Components of plans: Components of strategic plans
Staffing	*Number of staff:* Number of staff dedicated to a digital library
	Staff hours on digital library: Number of professional staff hours dedicated to a digital library
	Student hours on digital library: Number of student worker hours dedicated to a digital library
Staff training	*Types of training for digital library staff:* Types of training offered to digital library staff for a specific period of time
	Resources: Amount of resources for training allocated to a digital library
Marketing/ Promotion	*Marketing/promotion methods:* Types of marketing/promotion methods used for promoting a digital library
	Frequency of marketing/promotion activities: Frequency of each type of marketing activity taking place for a digital library in a specific period of time
	Recognition of digital libraries: Number of people aware of a digital library based on survey
Assessment	*Frequency of assessment:* Frequency of digital library assessment in a specific period of time
	Dimensions of assessment: Dimensions of a digital library included in the assessment
Management policy	*Presence of management policy*: Whether there is a documented management policy on a digital library
	Components of management policy: Types of components in the management policy related to a digital library
Grant/Fundraising	*Number/Amount of grant/fundraising:* Total number/amount of grants/fundraising for digital libraries in a specific period of time
	Number/Amount of grant/fundraising attempts: Number/amount of grants/fundraising attempted on digital libraries in a specific period of time
	Number/Amount of grant/fundraising received: Number/amount of grants/fundraising received in a specific period of time
Cost effectiveness	*Cost effectiveness:* The ratio of total digital library uses to the cost
Sustainability plan	*Presence of sustainability plans:* Whether there are documented sustainability plans for a digital library
	Components of plans: Components of sustainability plans for a digital library

Table 10.20 Appropriateness of Evaluation Measures in the Dimension of User Engagement	
Criterion	**Measure and Definition**
Digital object use	*Items viewed:* Number of items viewed
	Time spent on an item: Average time spent on viewing an individual item
	Item downloading: Number of items downloaded.
User feedback	*User feedback channels:* Types of user feedback channels offered in a digital library
	Quantity of user feedback: Amount of user feedback submitted for a digital library
Site visit	*Frequency of site visits:* Number of site visits within a specific period of time
	Session length: Average time spent on a digital library from the beginning to the end of a session
	Frequency of page visits: Number of page visits within a specific period of time
	Unique site visits: Number of unique (site) visits within a specific period of time
	Unique page visits: Number of unique (page) visits within a specific period of time
Integration with external applications	*Compatibility with external applications:* Types of external applications that are compatible with a digital library
	Ease of integration: Degree of ease to integrate a digital library to external application as assessed by an expert
User/community knowledge contribution	*Types of user/community knowledge contributions:* Types of user/community knowledge contribution channels available in a digital library
E-commerce support	*Number of orders:* Number of orders placed for items in a digital library within a specific period of time
	Ease of purchasing: Users' perceived ease of purchasing digital objects
Search pattern	*Queries entered:* Number of queries entered per session
	Categories viewed: Number of category pages viewed per session
	Search result evaluation: Number of search result pages viewed per session

FACTORS HINDERING DIGITAL LIBRARY EVALUATION

Although there are more studies on digital library evaluation criteria and measures, fewer studies concentrate on factors. Multiple factors affect digital library evaluation. Our study explored what factors negatively influence digital library evaluation research and practices. Twelve factors hindering the evaluation of digital libraries were specified. Among them, the three most influential factors were as follows: limited evaluation tools directly applicable to practices (5.82), insufficient experience in evaluation (5.76), and limited awareness of the importance of digital library evaluation (5.59). It seems that the lack of evaluation tools and experience contributed the most to the impediment of digital library evaluation. On the contrary, the three least influential factors were selected as lack of user participation (5.21), limited application of evaluation results (5.18), and lack of incentive for evaluation (5.16). Fig. 10.5 presents the hindering factors affecting digital library evaluation.

Table 10.21 Appropriateness of Evaluation Measures in the Dimension of Context	
Criterion	**Measure and Definition**
Information ethics	*Presence of ethics guidelines:* Whether a digital library has guidelines for ethical issues
	Components of guidelines for ethics: Types of components in the guidelines for ethics
Organizational mission	*Conformity to organizational mission:* The extent to which a digital library conforms to organizational mission
	Ways of support: In what ways a digital library supports the organizational mission
Targeted user community	*User community engagement:* Types of user community engagements (e.g., outreach, collaboration, participation in collection development, etc.)
	Level of user community engagement: To what extent a digital library engages in user communities
Content sharing	*Types of content sharing:* Types of digital library content sharing partners
	Types of resources shared: Types of digital library items shared with partners
	Number of items shared: Number of items shared with partners
Collaboration	*Types of collaboration:* Types of digital library collaboration partners or stakeholders
	Number of collaborations: Number of digital library collaborations in a specific period of time
Social impact	*Types of social impacts:* Types of social impacts of a digital library on community and society
	Level of social impact: To what extent a digital library influences society
Knowledge change in communities	*Ways of knowledge change:* In what ways a digital library supports knowledge change in communities or societies
Multilingual access	*Types of languages:* Types of languages supported by a digital library
Multicultural audiences	*Types of multicultural audiences:* To what extent a digital library engages in multicultural audiences

EVALUATION CHALLENGES

Even though researchers and practitioners realize the importance of digital library evaluation, there are still challenges and problems for performing successful evaluation. First, definitions of digital library evaluation criteria need clarification and to be agreed upon. Heradio et al. (2012) point out that the lack of a standard definition for usability and usefulness is one of the main digital library evaluation challenges from user perspectives. Researchers and practitioners use different terms for the same concept, or they use the same terms but with differing meanings. It is critical to develop standard definitions of digital library evaluation criteria that are agreed upon and adopted as a standard by researchers and practitioners.

Second, digital library evaluation is complex because digital libraries mean different things to different groups (Van House, 2003; Zhang, 2010). Scholars, librarians, and users are the three main stakeholders of digital libraries. However, they do have differing opinions regarding the important criteria for digital library evaluation because they play different roles in its research, creation, management, and use. Although researchers and users consider the ideal situations in digital library evaluation, librarians
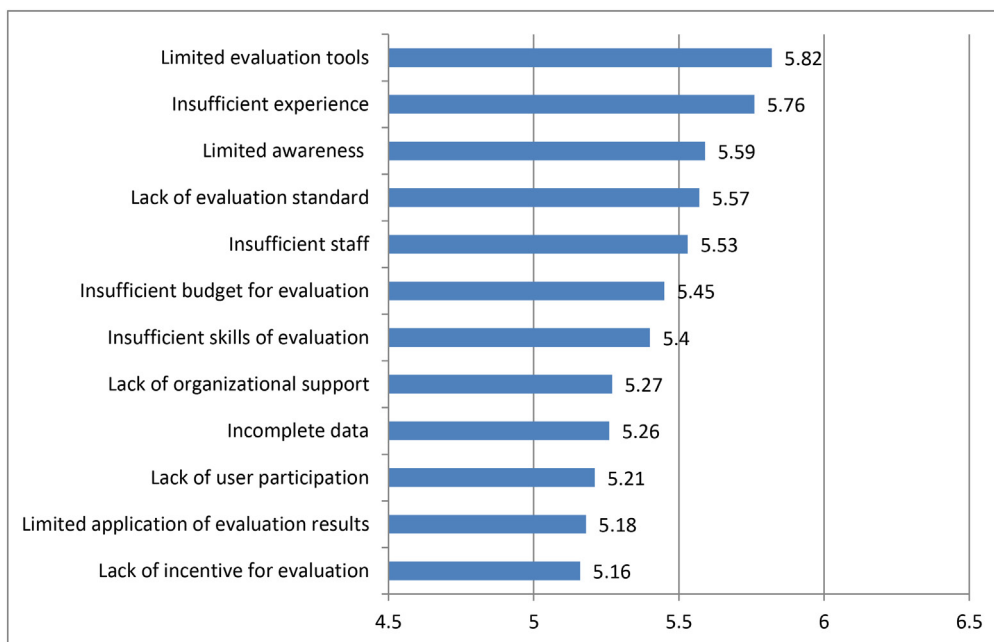
FIGURE 10.5 Hindering Factors of Digital Library Evaluation

have to consider the practical side of its management. It is beneficial to engage scholars, librarians, and users in digital library evaluation in order to have a more comprehensive picture.

Third, evaluation objectives or purposes are the leading forces for digital library evaluation. It is a challenge to match digital library evaluation criteria and measurements with diverse evaluation objectives or purposes. Future research needs to identify the relationships between evaluation objectives and associated criteria and measurements and further offer a set of specific evaluation criteria with associated measures for each specific evaluation objective so practitioners and researchers can mix and match different digital library criteria and measurements for their specific objectives.

Fourth, challenges related to measurements mainly focus on the identification of types of data used for measurements. There are disagreements on the acceptance of different types of measurements, such as the interval status for Likert scales and Likert scales versus fuzzy linguistic modeling (Heradio et al., 2012). It requires further research to identify the most appropriate measurements for different types of evaluation criteria.

Fifth, existing research has suggested that both qualitative and quantitative methods, as well as multiple data collection methods and multiple measurements, are needed for digital library evaluation. As such, the challenge is to find a research team or person who can integrate all the research methods into the assessment of digital libraries. Digital library evaluation requires the involvement of multiple personnel who have expertise in different data collection and data analysis methods.

Sixth, in order to create a comprehensive digital library evaluation framework and associated criteria and measurement, it is beneficial to create a community for evaluation research and to provide data repositories to share research findings and reach consensus (Fuhr et al., 2007). The challenge is how to

build this type of community and repositories, in particular internationally. The other related issues are data reuse and data protection.

Seventh, this challenge is related to how to evaluate digital libraries as emergent systems. Digital libraries are dynamic and complex systems (Marchionini, 2000; Saracevic, 2004). Can we evaluate digital libraries while taking into consideration their dynamic changes librarys and all the factors affecting them? Longitudinality and flexibility in evaluation might help overcome this challenge (Marchionini, 2000).

Eighth, the final challenge is related to the diverse types of digital libraries. Digital libraries can be classified based on content, audience, formats, and sponsors, as well as languages. The challenges for building these digital libraries vary, and so do the challenges for evaluating them. For example, Diekema (2012) enumerates the challenges that multilingual digital libraries encounter: cross language barrier, data management, representation in relation to standardization of encoding schemes, development considering cultural differences, and interoperability. Albertson (2015) emphasizes the uniqueness of evaluating visual digital libraries, which demands more user–system interaction. These challenges have to be considered for digital library evaluation. Different types of digital libraries may bring distinctive challenges for digital library evaluation.

REFERENCES

Adams, A., Blandford, A., Lunt, P., 2005. Social empowerment and exclusion: a case study on digital libraries. ACM Trans. Comput.-Hum. Interact. (TOCHI) 12 (2), 174–200.

Albertson, D., 2015. Synthesizing visual digital library research to formulate a user-centered evaluation framework. New Lib. World 116 (3/4), 122–135.

Alipour-Hafezi, M., Amanollahi Nick, H., 2015. Evaluation of digital libraries of Iranian research institutions based on the DigiQUAL protocol. Electron. Lib. 33 (4), 824–841.

Bishop, A.P., Mehra, B., Bazzell, I., Smith, C., 2003. Participatory action research and digital libraries: reframing evaluation. In: Bishop, A.P., Van House, N.A., Buttenfield, B.P. (Eds.), Digital Library Use: Social Practice in Design and Evaluation. The MIT Press, Cambridge, MA, pp. 161–189.

Blandford, A., Keith, S., Connell, I., Edwards, H., 2004. Analytical usability evaluation for digital libraries: a case study. In: Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries. ACM, New York, NY, pp. 27–36.

Blixrud, J.C., 2002. Issues in research library measurement. J. Lib. Admin. 35 (4), 3–5.

Blixrud, J.C., 2003. Measures for electronic use: the ARL E-Metrics project. Statistics in Practice—Measuring and Managing, 73–84. Available from: http://www.lboro.ac.uk/microsites/infosci/lisu/downloads/statsinpractice-pdfs/blixrud.pdf.

Bollen, J., Luce, R., 2002. Evaluation of digital library impact and user communities by analysis of usage patterns. D-Lib Mag. 8 (6), 1–13.

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). Research and Advanced Technology for Digital Libraries (EBT). Springer, Berlin, Heidelberg, pp. 390–401.

Brophy, P., 2001. Electronic Library Performance Indicators: the EQUINOX project. Serials 14 (1), 5–9.

Buchanan, S., Salako, A., 2009. Evaluating the usability and usefulness of a digital library. Lib. Rev. 58 (9), 638–651.

Buttenfield, B., 1999. Usability evaluation of digital libraries. Sci. Technol. Lib. 17 (3–4), 39–59.

Candela, L., Castelli, D., Pagano, P., Thanos, C., Ioannidis, Y., Koutrika, G., Ross, S., Schek, H., Schuldt, H., 2007. Setting the foundations of digital libraries: the DELOS Manifesto. D-Lib Mag. 13 (3/4.), Available from: http://www.dlib.org/dlib/march07/castelli/03castelli.html.

Carr, R., 2006. What users want: an academic "hybrid" library perspective. Ariadne 46. Available from: http://www.ariadne.ac.uk/issue46/carr/.

Chowdhury, G., Chowdhury, S., 2003. Introduction to Digital Libraries. Facet, London.

Chowdhury, G., Poulter, A., McMenemy, D., 2006. Public library 2.0: towards a new mission for public libraries as a "network of community knowledge." Online Inform. Rev. 30 (4), 454–460.

Choudhury, S., Hobbs, B., Lorie, M., Flores, N., 2002. A framework for evaluating digital library services. D-Lib Mag. 8 (7/8), 1082–9873.

Crabtree, A., Twidale, M.B., O'Brien, J., Nichols, D.M., 1997. Talking in the library: implications for the design of digital libraries. In: Proceedings of the Second ACM International Conference on Digital Libraries. ACM, New York, NY, pp. 221–228.

Diekema, A.R., 2012. Multilinguality in the digital library: a review. The Electron. Lib. 30 (2), 165–181.

Dillon, A., 1999. TIME—A multi-level framework for the design and evaluation of digital libraries. Int. J. Digit. Lib. 2 (2/3), 170–177.

Evans, J., O'Dwyer, A., Schneider, S., 2002. Usability evaluation in the context of digital video archives. In: Fourth DELOS Workshop. Evaluation of Digital Libraries: Testbeds, Measurements, and Metrics, pp. 79–86. Available from: http://wwwold.sztaki.hu/conferences/deval/presentations/DELOSWorkshop4OnEval_report.pdf#page=79.

Ferreira, S.M., Pithan, D.N., 2005. Usability of digital libraries: a study based on the areas of information science and human-computer-interaction. OCLC Syst. Serv. 21 (4), 311–323.

Fuhr, N., Hansen, P., Mabe, M., Micsik, A., Sølvberg, I., 2001. Digital libraries: a generic classification and evaluation scheme. Research, Advanced Technology for Digital Libraries, 5th European Conference. Springer, Berlin, Heidelberg, pp. 187–199.

Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., Sølvberg, I., 2007. Evaluation of digital libraries. Int. J. Digit. Lib. 8 (1), 21–38.

Gonçalves, M.A., Moreira, B.L., Fox, E.A., Watson, L.T., 2007. What is a good digital library? A quality model for digital libraries. Inform. Process. Manag. 43 (5), 1416–1437.

Gonçalves, M.A., Fox, E.A., Watson, L.T., Kipp, N., 2004. Streams, structures, spaces, scenarios, societies (5S): a formal model for digital libraries. ACM Trans. Inform. Syst. 22 (2), 270–312.

Hariri, N., Norouzi, Y., 2011. Determining evaluation criteria for digital libraries' user interface: a review. The Electron. Lib. 29 (5), 698–722.

Heradio, R., Fernández-Amorós, D., Cabrerizo, F.J., Herrera-Viedma, E., 2012. A review of quality evaluation of digital libraries based on users' perceptions. J. Inform. Sci. 38 (3), 269–283.

Hilary, B.H., Druin, A., Bederson, B.B., 2007. Supporting elementary-age children's searching and browsing: design and evaluation using the International Children's Digital Library. J. Am. Soc. Inform. Sci. Technol. 58 (11), 1618. Available from: https://ezproxy.lib.uwm.edu/login?url=http://search.proquest.com/docview/231417597?accountid=15078.

Hu, C., Hu, Y., Yan, W., 2014. An empirical study of factors influencing user perception of university digital libraries in China. Lib. Inform. Sci. Res. 36 (3/4), 225–233.

Jeng, J., 2005a. Usability assessment of academic digital libraries: effectiveness, efficiency, satisfaction, and learnability. Libri 55 (2–3), 96–121.

Jeng, J., 2005b. What is usability in the context of the digital library and how can it be measured. Inform. Technol. Lib. 24 (2), 47–56.

Joo, S., Xie, I., 2013. Evaluation constructs and criteria for digital libraries: a document analysis. In: Cool, C., Ng, K.B. (Eds.), Recent Developments in the Design, Construction and Evaluation of Digital Libraries. IGI Global, Hershey, PA, pp. 126–140.

Kani-Zabihi, E., Ghinea, G., Chen, S.Y., 2006. Digital libraries: what do users want? Online Inform. Rev. 30 (4), 396–412.

Kengeri, R., Seals, C.D., Harley, H.D., Reddy, H.P., Fox, E.A., 1999. Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD. Int. J. Digit. Lib. 2 (2–3), 157–169.

Kenney, A.R., Sharpe, L.H., Berger, B., 1998. Illustrated book study: digital conversion requirements of printed illustration. In: Nikolaou, C., Stephanidis, C. (Eds.), Research and Advanced Technology for Digital Libraries: Proceedings of the Second European Conference. Springer, Berlin, pp. 279–293.

Khoo, M., Kusunoki, D., MacDonald, C., 2012. Finding problems: when digital library users act as usability evaluators. In: System Science (HICSS), 2012 45th Hawaii International Conference, pp. 1615–1624.

Kim, Y.H., Kim, H.H., 2008. Development and validation of evaluation indicators for a consortium of institutional repositories: a case study of dCollection. J. Am. Soc. Inform. Sci. Technol. 59 (8), 1282–1294.

Kyrillidou, M., Cook, C., Lincoln, Y., 2009. Digital library service quality: what does it look like? In: Tsakonas, G., Papatheodorou, C. (Eds.), Evaluation of Digital Libraries: An Insight into Useful Applications and Methods. Chandos Publishing, Oxford, pp. 187–214.

Kyrillidou, M., Giersch, S., 2005. Developing the DigiQUAL protocol for digital library evaluation. In: Digital Libraries, 2005. Proceedings of the Fifth ACM/IEEE-CS Joint Conference on JCDL'05, pp. 172–173. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.107.9245&rep=rep1&type=pdf.

Lagzian, F., Abrizah, A., Wee, M.C., 2013. An identification of a model for digital library critical success factors. The Electron. Lib. 31 (1), 5–23.

Lai, C.F., Chiu, P.S., Huang, Y.M., Chen, T.S., Huang, T.C., 2014. An evaluation model for digital libraries' user interfaces using fuzzy AHP. Electr. Lib. 32 (1), 83–95.

Lankes, R.D., Gross, M., McClure, C., 2003. Cost, statistics, measures, and standards for digital reference services: a preliminary view. Lib. Trend 51 (3), 401–413.

Larsen, L., 2002. The DLib Test Suite and Metrics Working Group: Harvesting the Experience from the Digital Library Initiative. D-Lib Working Group on Digital Library Metrics Website. Available from: http://www.dlib.org/metrics/public/papers/The_Dlib_Test_Suite_and_Metrics.pdf.

Mansor, Y., Ripin, F.M., 2013. Usability evaluation of online digital manuscript interface. Lib. Phil. Prac. 986, 1–12.

Marchionini, G., 2000. Evaluating digital libraries: a longitudinal and multifaceted view. Lib. Trends 49 (2), 304–333.

Marchionini, G., Plaisant, C., Komlodi, A., 2003. The people in digital libraries: multifaceted approaches to assessing needs and impact. In: Bishop, A.P., Van House, N.A., Buttenfield, B.P. (Eds.), Digital Library Use: Social Practice in Design and Evaluation. The MIT Press, Cambridge, MA, pp. 119–160.

Meyyappan, N., Foo, S., Chowdhury, G.G., 2004. Design and evaluation of a task-based digital library for the academic community. J. Document. 60 (4), 449–475.

Nicholson, S., 2004. A conceptual framework for the holistic measurement and cumulative evaluation of library services. J. Document. 60 (2), 164–182. Available from: http://arizona.openrepository.com/arizona/bitstream/10150/106172/1/holistic.pdf.

The National Initiative for a Networked Cultural Heritage (NINCH), Humanities Advanced Technology and Information Institute (HATII) (2002). The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials. Available from: http://www.nyu.edu/its/humanities/ninchguide/.

NISO, 2007. A Framework of Guidance for Building Good Digital Collections, 3rd ed. Available from: http://www.niso.org/publications/rp/framework3.pdf.

Noh, Y., 2010. A study on developing evaluation criteria for electronic resources in evaluation indicators of libraries. J. Acad. Librar. 36 (1), 41–52.

Park, S., 2000. Usability, user preferences, effectiveness, and user behaviors when searching individual and integrated full-text databases: implications for digital libraries. J. Am. Soc. Inform. Sci. 51 (5), 456–468.

Saracevic, T., 2000. Digital library evaluation: toward evolution of concepts. Lib. Trends 49 (2), 350–369.

Saracevic, T., 2004. Evaluation of digital libraries: an overview. In: Notes of the DELOS WP7 Workshop on the Evaluation of Digital Libraries, pp. 13–30. Available from: http://comminfo.rutgers.edu/~tefko/DL_evaluation_Delos.pdf.

Shim, W., 2002. Measuring services, resources, users and use in the networked environment. J. Lib. Admin. 35 (4), 71–84.

Strauss, A.L., Corbin, J., 1990. Basics of Qualitative Research. Sage Publications, Newbury Park, CA.

Stejskal, J., Hajek, P., 2015. Effectiveness of digital library services as a basis for decision-making in public organizations. Lib. Inform. Sci. Res. 37 (4), 346–352.

Tsakonas, G., Kapidakis, S., Papatheodorou, C., 2004. Evaluation of user interaction in digital libraries. In: Notes of the DELOS WP7 Workshop on the Evaluation of Digital Libraries, pp. 45–60. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.9642&rep=rep1&type=pdf.

Tsakonas, G., Mitrelis, A., Papachristopoulos, L., Papatheodorou, C., 2013. An exploration of the digital library evaluation literature based on an ontological representation. J. Am. Soc. Inform. Sci. Technol. 64 (9), 1914–1926.

Tsakonas, G., Papatheodorou, C., 2008. Exploring usefulness and usability in the evaluation of open acess digital libraries. Inform. Process. Manag. 44 (3), 1234–1250.

Tsakonas, G., Papatheodorou, C., 2011. An ontological representation of the digital library evaluation domain. J. Am. Soc. Inform. Sci. Technol. 62 (8), 1577–1593.

Van House, N.A., 2003. Digital libraries and collaborative knowledge construction. In: Bishop, A.P., Van House, N.A., Buttenfield, B.P. (Eds.), Digital Library Use: Social Practice in Design and Evaluation. The MIT Press, Cambridge, MA, pp. 271–295.

Van House, N., Weil, B., McClure, C., 1990. Measuring Academic Library Performance: A Practical Approach. American Library Association, Chicago.

Vullo, G., 2010. A global approach to digital library evaluation. Liber Quart. 20 (2), 169–178.

Waugh, L., Hamner, J., Klein, J., Brannon, S., 2015. Evaluating the University of North Texas' Digital Collections and Institutional Repository: an exploratory assessment of stakeholder perceptions and use. J. Acad. Librar. 41 (6), 744–750.

Xie, H., 2006. Evaluation of digital libraries: criteria and problems from users' perspectives. Lib. Inform. Sci. Res. 28 (3), 433–452.

Xie, H.I., 2008. Users' evaluation of digital libraries (DLs): their uses, their criteria, and their assessment. Inform. Process. Manag. 44 (3), 1346–1373.

Cheng, Y.M., 2014. Why do users intend to continue using the digital library? An integrated perspective. Aslib J. Inform. Manag. 66 (6), 640–662.

Zhang, Y., 2010. Developing a holistic model for digital library evaluation. J. Am. Soc. Inform. Sci. Technol. 61 (1), 88–110.