Chapter 8
Evaluation Constructs and Criteria for Digital Libraries:
A Document Analysis

Soohyung Joo
University of Wisconsin – Milwaukee, USA

Iris Xie
University of Wisconsin – Milwaukee, USA

ABSTRACT

The intent of this chapter is to identify constructs and criteria for Digital Library (DL) evaluation based on document analysis. Eighty-five relevant articles and five websites were reviewed to generate the evaluation constructs and criteria. The findings consist of ten constructs, including collection, information organization, interface design, system performance, effects on users, user engagement, services, preservation, sustainability/administration, and context of use with associated criteria for each dimension. In addition, this chapter discusses challenges in DL evaluation research and practices.

INTRODUCTION

In the last decade, libraries have substantially expanded their digitization efforts and contributed a number of unique digital collections in support of scholarly research, teaching, and learning. The initial development of digital collections focused on building the technical infrastructure and es-tablishing a DL presence resulted in a number of useful digitization guidelines and best practices for creating usable and sustainable digital collections. Recommendations for undertaking periodic evaluations are also part of the guidelines for "building good digital collections" (NINCH, 2002; NISO, 2007). However, the exponential growth of DLs has not been accompanied by extensive

evaluation studies (Saracevic, 2004). According to Saracevic (1995), "Evaluation means assessing performances or value of system, process, product, or policy" (p. 138). DL evaluation is by no means a simple task, and is often overlooked in DL evaluation due to the nature of complex, multi-dimensional, and distributed systems of DLs (Bollen & Luce, 2002; Marchionini, 2000; Saracevic, 2004; Sumner & Marlino, 2004; Chowdhury, et al., 2006; Xie, 2006; Xie, 2008; Zhang, 2010). Currently, there are few guidelines for best practices of DL evaluation. Since DLs are new, complex, and multifaceted entities, researchers and practitioners in the field need a set of guidelines of what to evaluate, how to measure results, when to undertake evaluation, and how to incorporate results into the development process. Several researchers addressed the need for the multi-dimensional evaluation for the DL field as they identified the lack of a comprehensive and integrated evaluation framework as a major barrier to DL evaluation (Chowdhury, et al., 2006; Saracevic, 2000; Xie, 2006, 2008; Zhang, 2010).

This chapter adopts the definition of Digital Libraries (DLs) as: "representations of emergent and complex forms of digital information organization and design, consisting of multiple layers and building blocks, in various stages of development" (Matusiak, 2010, p. 15). DLs present a variety of resources created in the digital format as well as those converted from analog materials through digitization efforts, including print materials, manuscripts, images, audio, and video. The concepts of DLs are still evolving, correspond to a very complex notion, and cannot be captured by a simple definition (Bishop, Van House, & Buttenfield, 2003; Candela, et al., 2007; Greenstein, 2000).

As an initial step toward a comprehensive DL evaluation framework, this chapter reviewed a wide range of previous documents to identify evaluation constructs and associated criteria. Even though researchers suggested various evaluation constructs and criteria, there was little effort to analyze and integrate those criteria into a consolidated framework within the DL field. In this chapter, we identified 10 evaluation constructs and corresponding criteria based on the analysis of 85 previous documents and five websites related to DL evaluation. Furthermore, we discuss the limitations and challenges of previous evaluation efforts.

BACKGROUND

Researchers and DL practitioners recognize the importance of DL evaluation, and have proposed several models, frameworks, and various criteria for DL evaluation. Several practical tools and methods have been proposed to expand evaluation efforts, including DigiQUAL, eMetrics, EQUINOXISO/TSO-20938, and MIEL2 Project, but they are often limited to specific elements or services. For instance, DigiQUAL is proposed as a tool for assessing service quality (Kyrillidou & Giersch, 2005). Focusing more on vendor-provided data in academic libraries, the ARL's e-metrics project, COUNTER, and SUSHI protocols were designed for the purpose of assessing the outputs of DLs using the usage statistics (Pesch, 2007). In Europe, EQUINOX serves as a framework to assess performance of DLs based on the usage statistics (Brophy, 2001). Similarly, ISO/TSO-20983 and MIEL2 cover electronic resources in DL environments (Brophy & Wynne, 1997; Noh, 2010). The eVALUEd toolkit focuses on methodological issues (McNicol, 2004). Although all these efforts represent a step in the right direction, they fail to provide a comprehensive model for DL evaluation in a systematic way that addresses the needs of multiple evaluation constructs and criteria.

On the research side, many researchers have suggested a number of evaluation constructs, criteria, and indicators to create a comprehensive approach reflecting different components of DLs. The early DL research projects, funded by the National Science Foundation (NSF) as

part of Digital Libraries Initiatives I and II, laid the groundwork by producing DL prototypes and frameworks (Borgman, et al., 2000; Buttenfield, 1999; Hill, et al., 2000; Van House, et al., 1996). Although these early projects weighed much on the development of DL prototypes, several of them undertook major evaluation efforts as part of the design cycle. For instance, Hill et al. (1997) identified six criteria, including ease of use, overall appeal, usefulness, and performance, in evaluating the Alexandria Digital Library (ADL) at the University of California, Santa Barbara. Saracevic (2000, 2004) suggested a comprehensive evaluation model and identified six classes of criteria, representing: content, technology, interface, process/service, user, and context. In Europe, DELOS is one of most comprehensive and large scale DL projects, which represent joint activities aimed at integrating and coordinating the ongoing research efforts of the major European teams working in DL area. DELOS Manifesto presents a three-tier DL framework incorporating six core components such as content, functionality, quality, policy, architecture, and user (Candela, et al., 2007). DE-LOS model sets up a foundation for DL research. DELOS Network of Excellence has conducted a series of research concerning the evaluation of DLs. Fuhr et al. (2001) proposed a scheme for digital library evaluation, which contains four dimensions: data/collection, system/technology, users, and usage. Tsakonas et al. (2004) further examined the interactions of DL components and proposed the following evaluation foci: usability, usefulness, and system performance. Fuhr et al. (2007) developed a DL evaluation framework based on a DELOS model and a largescale survey of DL evaluation activities. In addition, DELOS has contributed to the large test-set-based IR research of video digital libraries in TREC. For example, the TREC video track offered an evaluation test framework for IR performance of video information in digital libraries (Smeaton & Over, 2002). Xie's (2006, 2008) evaluation framework increased the focus onto users, and proposed five

types of criteria: usability, collection quality, service quality, system performance efficiency, and user feedback solicitation. Recently, Zhang (2010) investigated the importance of multiple constructs of criteria including: content, technology, interface, service, user, and context using the empirical data from heterogeneous stakeholders.

Usability evaluation is another key area for research. Saracevic (2004) stated the usability evaluation studies play a role of connecting user-oriented and system-oriented usability studies. He suggested a list of specific usability criteria and attributes for DL evaluation, including content, process, format, and overall assessment. Dillon (1999) proposed a qualitative framework (TIME) for DL usability evaluation which covers user task (T), information model (I), manipulation facilities (M), and the ergonomic variables (E). Xie and Wolfram (2002) suggested organizational usability factors in state-level digital library. Based on an empirical study, they identified four contributing factors for organization usability in digital libraries, namely: access, promotion and training, content and format usage, and design. Ward and Hiller (2005) suggested usability evaluation criteria specific to library services—completion of the task, time and effort, and reaction to the product or service. Similarly, but more specifically, Jeng (2006) proposed a usability model for academic digital libraries employing four operational usability criteria—effectiveness, efficiency, satisfaction, and learnability. In addition, many other researchers conducted usability tests in DL environment (Eliasen, et al., 1997; Battleson, et al., 2001; Hammil, 2003; Blandford, et al., 2004; Joo & Lee, 2011).

In addition, Shim and Kantor (1999) adopted Data Envelopment Analysis (DEA) to evaluate DLs, and proposed an evaluation framework that focused on two main dimensions of effectiveness and efficiency. Missingham (2003) introduced the unique concept of a digital footprint representing the use of DLs through multidimensional measurement. Kim and Kim (2008) proposed 19

evaluation criteria tailored to digital collections, and validates them empirically using surveys from related stakeholders such as users, librarians and administrators. Noh (2010) suggested a set of evaluation indices for electronic resources in terms of electronic resource acquisition, electronic resource use, and electronic resource environment. Other researchers examined the DL evaluation criteria such as suitability, accuracy, costs, informativeness, timeliness, usefulness, and others (Kengeri, et al., 1999; Kenney, et al., 1998; Larsen, 2002). Most of the works in the research area investigate the DL frameworks and associated criteria, however they do not offer practical guidelines on how to gather data and implement evaluation.

DIGITAL LIBRARY EVALUATION CONSTRUCTS AND CRITERIA

As an effort to identify evaluation constructs and associated criteria pools for DL evaluation, the authors conducted a document analysis. Using keywords such as "digital library," "evaluation," "criteria," "assessment," and others in different combinations, the authors searched Google Scholar and EBSCOhost for relevant documents published primarily, but not exclusively, between 2000 to 2010. The relevance of a document was judged by the two following criteria: (1) whether the paper included any evaluation theories, framework, criteria, indicators, or measures; or (2) whether the paper included actual evaluation practices or pilot tests. The document analysis covers evaluation research associated with DLs and previous research related to evaluation. In this study, eighty-five relevant documents and five DL evaluation project websites (Equinox, DigiQUAL, LibQUAL+, eVALUEd, DELOS) were reviewed and analyzed.

All the evaluation constructs and criteria were collected and analyzed from the selected documents. Strauss and Corbin's (1990) open coding technique was applied to data analysis. For example, evaluation criteria for "user engagement" were identified by analyzing related literature in the perspectives of user usage and user involvement. After reviewing the constructs and criteria, the authors consolidated and organized them into ten constructs. Then, associated criteria for each construct were identified. The authors adopted all the criteria identified in the selected literature and the websites.

The document analysis yielded ten constructs and associated criteria, which include: 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 this chapter. Table 1 presents the ten constructs and associated definitions. Table 2 shows specific evaluation criteria and corresponding objectives.

CHALLENGES IN DIGITAL LIBRARY EVALUATION

Many researchers and practitioners have expanded their efforts in digital library evaluation in the recent decade. However, there are still many challenges and problems in DL evaluation. This chapter discusses some key challenges that hinder the research and practices of evaluation in the DL community.

As discussed in Background, there are few evaluation standards and comprehensive frameworks that are directly applicable to DL fields, despite scholarly and practical efforts in the DL area. Early DL projects funded by the National Science Foundation (NSF) primarily focus on the design of DL prototypes, and do not offer direct insights and implications for DL evaluation. Several models created for DL evaluation have not been recognized and adopted by researchers and practitioners as standards. The absence of stan-

Table 1. DL evaluation constructs and definitions

Construct	Definition
Collection	Assess the quality and quantity of DL collections.
Information Organization	Assess the representation, grouping, and presentation of digital information.
Interface design	Assess the usability of DL interfaces, and the extent these interfaces support users' interaction with DLs.
System performance	Assess the efficiency, accessibility, and reliability of DLs, as well as their retrieval performance.
Effects on users	Assess the impact and value of DLs on users' accomplishment of their tasks.
Services	Assess the quality and quantity of the offered DL services.
Preservation	Assess the extent and ways of DL support for preservation.
Administration/ Sustainability	Assess administrative-related factors that affect the development of DLs; Assess whether DLs can be sustained and enhanced.
User engagement	Assess the extent and ways of usage of DLs, and the extent and ways of user involvement in the DL development.
Context	Assess the extent of DLs fitting into, responding to, following larger context—institutional, economic, legal, social, cultural—and others (Saracevic, 2004).

dards is a major obstacle to building an evaluation framework and tools in the DL area. At the same time, most of current evaluation practices are prone to focus narrowly on specific systems or particular aspects or services (e.g. DigiQUAL for service quality; Equinox for usage). A one-time evaluation of a single portion of the system, while convenient, might leave out other significant portions of the system that could lend other crucial viewpoints for decision-making (Nicholson, 2004). Several researchers note that holistic evaluation studies are conspicuously absent from the DL field, and discuss the potential benefits of comprehensive approaches (Chowdhury, et al., 2006; Xie, 2006, 2008; Zhang, 2010). Moreover, the validation of an evaluation model remains one of the most challenging tasks.

There is a gap between research and practice. As discussed above, researchers have suggested various models and criteria, but their works remain mostly conceptual or theoretical. Most of the theoretical research investigates the importance of criteria, but fails to offer detailed, practical guidelines on how to gather data, implement the analysis, and use the results of evaluation. In addition, they have not sufficiently suggested practi-

cal criteria and methodology, such as providing measureable criteria, methods to collect the data, and how to incorporate the evaluation results into the development of DLs. As a result, practitioners have adopted few of those theoretical or conceptual models into the DL fields. Researchers will need to deal with practical matters in detail to fill the gap and to create more practical evaluation tools.

Less effort was made to evaluate contextual elements surrounding digital libraries. The majority of the studies in DL evaluation adopted traditional IR criteria at a restricted level, such as precision/ recall and error rate, usability criteria, or service quality criteria. Very few address how well a DL fits into or even improves people's daily work and life, via contextual criteria at higher levels, such as social, legal, and cultural aspects (Zhang, 2010). Context is another key aspect of DL evaluation, distinguished from predominant system- or user-centered evaluation framework (Adams & Bladford, 2001; Bollen & Luce, 2002). Saracevic (2004) is one of few researchers who emphasizes context criteria and suggests related measures (e.g. institutional fit, sustainability, productivity of and impact on community, members, etc.) in evaluating DLs. However, there are few practical

Table 2. DL evaluation criteria and objectives

Construct	Criterion	Objective
Collection	audience	To assess who are the main potential users of a DL.
	authority	To assess whether information provided by a DL comes from trustworthy sources.
	completeness	To assess whether a DL covers all documents in each topic area.
	cost	To assess costs to build collections in a DL.
	digitizing practices	To assess the types and methods of digitization practices conducted, as well as identifying best practices to build a DL.
	diversity	To assess whether a DL deals with a variety of issues in relation to a topic of interest.
	format flexibility	To assess whether the format of collections in a DL are compatible with a variety of software and systems for different purposes.
	quality of format	To assess the quality of a digitized item provided by a DL within its format (e.g. resolution, scanned image quality, recorded audio quality, etc.).
	size	To assess the amount of collection items provided by a DL.
	scope/coverage	To assess the range of topics that is covered by a DL.
	sufficiency	To assess whether a DL provides enough documents to satisfy a user's need.
	types of format	To assess what kinds of document formats are available in a DL.
Information Organization	accessibility to metadata	To assess how easily a user can obtain the metadata information of each item.
	appropriateness	To assess whether the organizational structure and associated categories adequately organize items in a DL.
	completeness/ comprehensiveness	To assess whether the organization structure covers all the access points of a DL.
	conformance	To assess whether metadata elements follow predefined standard and guides.
	consistency	To assess whether metadata is consistent across collections in a DL.
	controlled vocabulary	To assess the type and amount of controlled vocabularies used in a DL.
	interoperability	To assess whether metadata elements of a DL are compatible to different DLs.
	metadata accuracy	To assess how accurately metadata elements are assigned for each item.
	metadata authority	To assess whether credible metadata elements are selected for items in a DL.
	metadata elements	To assess how many metadata elements are employed to represent items in a DL.
	quality of metadata scheme	To assess the quality of metadata system selected for a DL.
Interface design	browsing function	To assess in what ways and to what extent the interface supports a user's ability to surf related items in a DL.
	consistency	To assess whether the design and layout are coherent across a DL interface.
	emotional appeal	To assess to what extent the interface of a DL is effectively attractive to users.
	help function	To assess what types of help functions are offered and how effectively they support users in their help-seeking process.
	intuitive operation	To assess how straightforward a DL interface is for a user to understand its operation, and how easily a user can learn to operate the interface.
	navigation	To assess in what ways and to what extent the interface supports a user's exploration in a DL.
	personalized page	To assess whether a DL offers personalized pages based on user profile.
	reliability	To assess the ability of a DL to perform and maintain its functions under different circumstances.

Table 2. Continued

Construct	Criterion	Objective
Interface design	search function	To assess what types of search functions are offered by a DL and how easy to use them.
	search results presentation	To assess the types of formats/options of search results are presented to users in a DL.
	usability	To assess the efficiency, effectiveness, and satisfactory use of a DL's interface (ISO, 1997).
	user control	To assess to what extent the DL allows users to manipulate its interface.
	view option	To assess whether and to what extent a user can customize viewing options in a DL.
	visual appeal	To assess to what extent the interface of a DL is visually attractive to users.
System performance	connectivity	To assess how stable a DL system can be connected to other information systems.
	error rate/error correction	To assess the degree of errors encountered during system operation and the ability of a DL to fix errors.
	fit-to-task	To assess to what extent a DL is adequate to perform tasks that a user requests.
	flexibility	To assess whether a DL responds to potential internal or external changes in a timely manner.
	integrated search	To assess whether a DL offers an integrated search environment for different collections within a DL.
	linkage with other DLs	To assess the identification of and in what ways a DL is linked to other related DLs.
	reliability	To assess how stable a DL's performance is over time.
	response time	To assess how quickly a DL responds to a user's request.
	retrieval effectiveness	To assess how effective the search algorithm is in a DL (e.g. recall and precision).
	server performance	To assess the ability of a server to run a DL.
	speed of page loading	To assess how quickly a DL presents a user-requested page.
	storage	To assess the size of a DL's storage capacity.
Effects on users	behavior change	To assess in what ways and to what extent a DL influences a user's behaviors.
	information literacy	To assess in what ways and to what extent a DL enhances a user's ability to find useful information for his/her tasks.
	instructional efficiency	To assess in what ways and to what extent a DL enhances user's teaching effectiveness.
	knowledge change	To assess in what ways and to what extent a DL influences a user's knowledge structure.
	learning effects	To assess in what ways and to what extent a DL influences a user's learning outcome.
	research productivity	To assess in what ways and to what extent a DL affects a user's research outputs (e.g. publications, grants, etc.).
Services	accessibility to managerial staff	To assess in what ways and to what extent a user can easily contact staff of a DL for questions, feedback, and comments.
	confidence	To assess in what ways and to what extent users have a positive attitude toward services offered by a DL.
	customized services	To assess whether a DL offers personalized services based on user profile information or user requests.
	FAQ/Q&A	To assess whether and how many FAQs or Q&A a DL provides to help users in using the DL.
	follow-up services	To assess in what ways and to what extent adequate and timely continuing services are provided to users by a DL when necessary.
	reliability	To assess how users of a DL perceive the trustworthiness of service provided.
	responsiveness	To assess the reaction time to a user's request for a DL service.

Table 2. Continued

Construct	Criterion	Objective
Services	service quality	To assess in what ways and to what extent DL services satisfy users' needs
	timeliness	To assess in what ways and to what extent services are offered to users in a timely manner.
	types of services	To assess the types of services provided by a DL.
	types of services for users with disabilities	To assess the types of services offered to users with disabilities.
	types of unique services	To assess the unique types of services provided by a DL compared to other related DLs.
	usefulness	To assess in what ways and to what extent DL services are useful for users to achieve their tasks.
	user education	To assess the types of user education offered by a DL.
	user satisfaction	To assess in what ways and to what extent users are satisfied with services provided by a DL.
Preservation	ability to migrate	To assess the ability of data migration for preservation.
	completeness	To assess to what extent the archiving process to preserve archived materials is complete and comprehensive.
	cost per record	To assess the average cost for archiving one record.
	data processing	To assess in what ways and to what extent data are processed for document preservation.
	incentives for archives	To assess the types and amount of incentives for performing archiving tasks.
	preservation policy	To assess whether a policy regarding preservation is developed and what is covered in the policy.
	staff dedicated to preservation	To assess how many staff members are dedicated to preservation task of a DL.
	types of archiving equipments/facilities	To assess the types of archiving equipments and facilities a DL uses for digital preservation.
	types of archiving methods	To assess the archiving methods/approaches a DL staff member has to apply in order to preserve information.
Administration /Sustainability	budget	To assess the amount and arrangement of monetary resources to efficiently manage a DL.
	fundraising/sponsor	To assess the effort of fundraising to support a DL financially.
	incentive	To assess the types of incentives provided to DL staff.
	management policy	To assess whether a well-defined policy for administration is offered by a DL.
	marketing	To assess the publicity efforts of a DL to attract potential users and inform related communities.
	planning	To assess whether strategic plans are established to create, manage, maintain, and enhance a DL.
	regular assessment	To assess whether regular, continuous evaluation is performed to maintain and enhance a DL.
	staffing	To assess the quantity and arrangement of human resources to efficiently manage a DL.
	staff training	To assess the types of, frequency of, and efficiency of training programs offered to DL staff.
User engagement	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 DL.
	resource use	To assess in what ways and to what extent users uses resources in a DL (item view, download, etc.)

Table 2. Continued

Construct	Criterion	Objective
User engagement	site visit	To assess how frequently users visit a DL website and the duration for each visit.
	user feedback	To assess the types of user comments and suggestions received by a DL, and in what ways and to what extent these comments and suggestions are incorporated into the enhancement of the DL.
	user participation channels	To assess the types of channels available to users to communicate with the staff of a DL.
Context	collaboration	To assess in what ways and to what extent stakeholders of a DL work together; To assess in what ways and to what extent stakeholders of a DL cooperate with stakeholders in another DL.
	content sharing	To assess in what ways and to what extent stakeholders of a DL are willing to share their content.
	copyrights	To assess whether a DL identifies and conforms to copyright issues.
	information ethics compliance	To assess whether a DL identifies and conforms to ethical issues related to DL creation and use.
	organizational mission	To assess in what ways and to what extent DL creation and use conform to organizational objectives.
	social impact	To assess in what ways and to what extent the use of a DL influences the society.
	targeted user community	To assess in what ways and to what extent a DL engages in targeted user groups.

attempts to evaluate the contextual elements in DL settings. Evaluating context, such as social, legal, and community is an unsolved problem due to the difficulty in measurement.

Few studies or evaluation practices attempted to employ multiple methods and data collection methods to assess components of DLs in an evaluation framework. There are few DL evaluation studies that applied multiple methods and data collection techniques at the same time. Moreover, quantitative data from surveys, transaction logs, and usage statistics, were dominantly used for a summative evaluation of DL performance or service quality. However, these quantitative data are less helpful in giving timely feedbacks about potential problems for DL design and implementation (Kaplan & Shaw, 2004). Qualitative methods, by providing findings that connect more directly with these individuals' perspectives, can increase the credibility and usefulness of evaluations (Patton, 2001).

CONCLUSION

In this chapter, we identified constructs and associated criteria for DL evaluation. Eighty-five documents and five websites in relation to DL were collected and analyzed to come up with ten constructs and associated DL evaluation criteria. This chapter presents multiple constructs of DL evaluation ranging from collection, information organization, interface design, system performance, and effects on users, services, preservation, sustainability/administration, user engagement, and context. In addition, corresponding criteria were also identified. We also discussed challenges in DL evaluation research and practices including lack of DL evaluation standards, few comprehensive evaluation models, the gap between research and practice, limited context evaluation, and few studies applying multiple research methods, in particular qualitative approaches, to their DL evaluations. These challenges call for the need

for researchers to develop DL evaluation frameworks, going beyond constructs and criteria, and further exploring associated measures and data collection methods.

REFERENCES

Adams, A., & Blandford, A. (2001). Digital libraries in a clinical setting: friend or foe? Research and advanced technology for digital libraries. In *Proceedings of 5th European Conference Digital Library, ECDL 2001*. Darmstadt, Germany: ECDL.

Battleson, B., Booth, A., & Weintrop, J. (2001). Usability testing of an academic library web site: A case study. *Journal of Academic Librarianship, 27*(3), 188–198. doi:10.1016/S0099-1333(01)00180-X

Bishop, A. P. (1995). Working towards an understanding of digital library use: A report on the user research efforts of the NSF/ARPA/NASA DLI projects. *D-Lib Magazine*. Retrieved from http://www.dlib.org/dlib/october95/10bishop.html

Bishop, A. P. Van House, N. A., & Buttenfield, B. P. (2003). *Digital library use: Social practice in design and evaluation*. Cambridge, MA: MIT Press.

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, (pp. 27-36). ACM/IEEE.

Bollen, J., & Luce, R. (2002). Evaluation of digital library impact and user communities by analysis of usage patterns. *D-Lib Magazine, 8*(6). Retrieved from http://www.dlib.org/dlib/june02/bollen/06bollen.html doi:10.1045/june2002-bollen

Bollen, J., & Luce, R. (2002). Evaluation of digital library impact and user communities by analysis of usage patterns. *D-Lib Magazine, 8*(6). Retrieved from http://www.dlib.org/dlib/june02/bollen/06bollen.html doi:10.1045/june2002-bollen

Borgman, C. L., Gilliland-Swetland, A. J., Leazer, G. H., Mayer, R., Gwynn, D., Gazan, R., & Mautone, P. (2000). Evaluating digital libraries for teaching and learning in undergraduate education: A case study of the Alexandria digital earth prototype (ADEPT). *Library Trends, 49*(2), 228–250.

Brophy, P. (2001). Electronic library performance indicators: The EQUINOX project. *Serials, 14*(1), 5–9.

Brophy, P., & Wynne, P. M. (1997). *Management information systems and performance measurement for the electronic library: eLib supporting study*. Lancashire, UK: University of Central Lancashire, Centre for Research in Library and Information Management.

Buttenfield, B. (1999). Usability evaluation of digital libraries. *Science & Technology Libraries, 17*(3-4), 39–59. doi:10.1300/J122v17n03_04

Candela, L., Akal, F., Avancini, H., Castelli, D., Fusco, L., & Guidetti, V. (2007). DILIGENT: Integrating digital library and grid technologies for a new earth observation research infrastructure. *International Journal on Digital Libraries, 7*, 59–80. doi:10.1007/s00799-007-0023-8

Chowdhury, S., Landoni, M., & Gibb, F. (2006). Usability and impact of digital libraries: A review. *Online Information Review, 30*(6), 656–680. doi:10.1108/14684520610716153

Dillon, A. (1999). TIME - A multi-leveled framework for evaluating and designing digital libraries. *International Journal on Digital Libraries, 2*(2/3), 170–177. doi:10.1007/s007990050045

Eliasen, K., McKinstry, J., & Fraser, B. M. (1997). Navigating online menus: A quantitative experiment. *College & Research Libraries, 58*, 509–516.

Fuhr, N., Hansen, P., Mabe, M., Micsik, A., & Sølvberg, I. (2001). Digital libraries: A generic classification and evaluation scheme. *Lecture Notes in Computer Science, 2163*, 187–199. doi:10.1007/3-540-44796-2_17

Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., & Kapidakis, S. (2007). Evaluation of digital libraries. *International Journal on Digital Libraries, 8*, 21–38. doi:10.1007/s00799-007-0011-z

Greenstein, D. (2000). Digital libraries and their challenges. *Library Trends, 49*(2), 290–303.

Hammill, S. (2003). Usability testing at Florida international university libraries: What we learned. *Electronic Journal of Academic and Special Librarianship, 4*(1).

Hill, L. L., Dolin, R., Frew, J., Kemp, R. B., Larsgaard, M., & Montello, D. R. Simpson, J. (1997). User evaluation: Summary of the methodologies and results for the Alexandria digital library, University of California at Santa Barbara. In *Proceedings of 60th ASIST Annual Meeting*, (pp. 225-243, 369). Medford, NJ: Information Today.

International Standard Organization. (1997). *ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs), Part 11: Guidance on usability specification and measures. Technical Report*. Geneva, Switzerland: International Standard Organization.

Jeng, J. (2006). *Usability of the digital library: An evaluation model*. (Unpublished Doctoral Dissertation). Rutgers University. Hoboken, NJ.

Joo, S., & Lee, J. (2011). Measuring the usability of academic digital libraries: Instrument development and validation. *The Electronic Library, 29*(4), 523–537. doi:10.1108/02640471111156777

Kaplan, B., & Shaw, N. T. (2004). People, organizational, and social issues: Future directions in evaluation research. *Methods of Information in Medicine, 43*, 215–231.

Kengeri, R., Seals, C. D., Reddy, H. P., Harley, H. D., & Fox, E. A. (1999). Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, and NDLTD. *International Journal on Digital Libraries, 2*(2/3), 157–169. doi:10.1007/s007990050044

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

Kim, Y. H., & Kim, H. H. (2008). Development and validation of evaluation indicators for a consortium of institutional repositories: A case study of collection. *Journal of the American Society for Information Science and Technology, 59*(8), 1282–1294. doi:10.1002/asi.20818

Kyrillidou, M., & Giersch, S. (2005). Developing the DigiQUAL protocol for digital library evaluation. In M. Marlino, T. Summer, & F. Shipman (Eds.), *Proceedings of the Fifth ACM/IEEE-CS Joint Conference on Digital Libraries,* (pp. 172-173). New York, NY: ACM Press.

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*. Retrieved from http://www.dlib.org/metrics/public/papers/The_Dlib_Test_Suite_and_Metrics.pdf

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

Matusiak, K. K. (2010). Use of digital resources in an academic environment: A qualitative study of students' perceptions, experiences, and digital literacy skills. (Unpublished Dissertation). University of Wisconsin – Milwaukee. Milwaukee, WI.

McNicol, S. (2004). The eVALUEd toolkit: A framework for the qualitative evaluation of electronic information services. *Vine, 34*(4), 172–175. doi:10.1108/03055720410570957

Missingham, R. (2003). *What makes libraries relevant in the 21st century? Measuring digital collections from three perspectives. information online conference.* Retrieved from http://www.nla.gov.au/nla/staffpaper/2003/missingham2.html

Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *The Journal of Documentation, 60*(2), 164–182. doi:10.1108/00220410410522043

NINCH. (2002). *The NINCH guide to good practice in the digital representation and management of cultural heritage materials.* Retrieved from http://www.nyu.edu/its/humanities/ninchguide/

NISO Framework Advisory Group. (2007). *A framework of guidance for building good digital collections* (3rd Ed.). Retrieved 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. *Journal of Academic Librarianship, 36*(1), 41–52. doi:10.1016/j.acalib.2009.11.005

Patton, M. Q. (2001). *Qualitative research and evaluation methods.* Thousand Oaks, CA: Sage Publications.

Pesch, O. (2007). Usage statistics: About COUNTER and SUSHI. *Information Services & Use, 27*(4), 207–213.

Saracevic, T. (1995). *Evaluation of evaluation in information retrieval.* Paper Presented at the SIGIR 1995. Seattle, CA.

Saracevic, T. (2004). *Evaluation of digital libraries: An overview.* Paper presented at the DELOS Workshop on the Evaluation of Digital Libraries. Retrieved from http://www.scils.rutgers.edu/~tefko/DL_evaluation_Delos.pdf

Saracevic, T., & Covi, L. (2000). Challenges for digital library evaluation. In D. H. Kraft (Ed.), *ASIS 2000: Proceedings of the 63rd ASIS Annual Meeting,* (Vol. 37, pp. 341-350). Medford, NJ: Information Today.

Shim, W., & Kantor, P. (1999). Evaluation of digital libraries: A DEA approach. In *Proceedings of the 62nd ASIST Annual Meating,* (pp. 605-615). ASIST.

Smeaton, A., & Over, A. (2002). The TREC-2002 video track report. In E. M. Voorhees & D. K. Harman (Eds.), *The Eleventh Text REtrieval Conference* (TREC 2002). TREC.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage Publications.

Sumner, T., & Marlino, M. (2004). Digital libraries and educational practice: A case for new models. In *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries (JCDL 2004),* (pp. 170-178). ACM/IEEE.

Tsakonas, G., Kapidakis, S., & Papatheodorou, C. (2004). Evaluation of user interaction in digital libraries. In M. Agosti & N. Fuhr (Eds.), *Notes of the DELOS WP7 Workshop on the Evaluation of Digital Libraries.* Padua, Italy: Academic Press.

Van House. N., Weil, B., & McClure, C. (1990). *Measuring academic library performance: A practical approach.* Chicago, IL: American Library Association.

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

Ward, J. L., & Hiller, S. (2005). Usability testing, interface design, and portals. *Journal of Library Administration, 43*(1), 155–171. doi:10.1300/J111v43n01_10

Xie, H. (2008). Users' evaluation of digital libraries (DLs): Their uses, their criteria, and their assessment. *Information Processing & Management, 44*(3), 1346–1373. doi:10.1016/j.ipm.2007.10.003

Xie, I. (2006). Evaluation of digital libraries: Criteria and problems from users' perspectives. *Library & Information Science Research, 28*(3), 433–452. doi:10.1016/j.lisr.2006.06.002

Xie, I., & Wolfram, D. (2002). State digital library usability: Contributing organizational factors. *Journal of the American Society for Information Science and Technology, 53*(13), 1085–1097. doi:10.1002/asi.10152

Zhang, Y. (2010). Developing a holistic model for digital library evaluation. *Journal of the American Society for Information Science and Technology, 61*(1), 88–110. doi:10.1002/asi.21220

ADDITIONAL READING

Baker, S., & Lancaster, F. W. (1991). *The measurement and evaluation of library services* (2nd ed.). Arlington, VA: Information Resources Press. doi:10.2307/40323203

Bearman, D. (2007). Digital libraries. *Annual Review of Information Science & Technology, 41*, 223–263. doi:10.1002/aris.2007.1440410112

Bertot, J. C., Snead, J. T., Jaeger, P. T., & McClure, C. R. (2006). Functionality, usability, and accessibility: Iterative user-centered evaluation strategies for digital libraries. *Performance Measurement and Metrics, 7*(1), 17–28. doi:10.1108/14678040610654828

Bishop, A. P. (1995). Working towards an understanding of digital library use: A report on the user research efforts of the NSF/ARPA/NASA DLI Projects. *D-Lib Magazine.* Retrieved from http://www.dlib.org/dlib/october95/10bishop.html

Bollen, J., & Luce, R. (2002). Evaluation of digital library impact and user communities by analysis of usage patterns. *D-Lib Magazine, 8*(6). Retrieved from http://www.dlib.org/dlib/june02/bollen/06bollen.html doi:10.1045/june2002-bollen

Borgman, C. L., Gilliland-Swetland, A. J., Leazer, G. H., Mayer, R., Gwynn, D., Gazan, R., & Mautone, P. (2000). Evaluating digital libraries for teaching and learning in undergraduate education: A case study of the Alexandria digital earth prototype (ADEPT). *Library Trends, 49*(2), 228–250.

Borlund, P. (2003). The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research, 8*(3). Retrieved from http://informationr.net/ir/8-3/paper152.html

Brantley, P. (2008). Roles and expectations for DIGITAL libraries. *EDUCAUSE Review, 43*(2), 30–38.

Brophy, P. (2006). *Measuring library performance: Principles and techniques.* London, UK: Facet Publishing.

Brophy, P., & Wynne, P. M. (1997). *Management information systems and performance measurement for the electronic library (MIEL2)*. Retrieved from http://www.ukoln.ac.uk/services/elib/papers/supporting/pdf/mis.pdf

Buchanan, S. (2009). Evaluating the usability and usefulness of a digital library. *Library Review*, *58*(9), 638–651. doi:10.1108/00242530910997928

Candela, L. (2007). Setting the foundations of digital libraries: The DELOS manifesto. *D-Lib Magazine*, *13*(3/4). Retrieved from http://www.dlib.org/dlib/march07/castelli/03castelli.html

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

Chowdhury, S., Landoni, M., & Gibb, F. (2006). Usability and impact of digital libraries: A review. *Online Information Review*, *30*(6), 656–680. doi:10.1108/14684520610716153

Comeaux, D. J. (2008). Usability studies and user-centered design in digital libraries. *Journal of Web Librarianship*, *2*(2), 457–475. doi:10.1080/19322900802190696

Cook, C., & Heath, F. (2001). Users' perceptions of library service quality: A LibQUAL+ qualitative study. *Library Trends*, *49*(4), 548–584.

Cullen, R. (2001). Perspectives on user satisfaction survey. *Library Trends*, *49*(4), 662–686.

Foo, S., Goh, D., & Na, J.-C. (2009). *Handbook of research on digital libraries: Design, development, and impact*. Hersey, PA: IGI Global.

Fox, E., & Marchionini, G. (2001). Digital libraries. *Communications of the ACM*, *44*(5), 31–32. doi:10.1145/374308.374329

Jeng, J. (2005). Usability assessment of academic digital libraries: Effectiveness, efficiency, satisfaction, and learnability. *Libri*, *55*, 96–121. doi:10.1515/LIBR.2005.96

Kani-Zabihi, E., Ghinea, G., & Chen, S. Y. (2006). Digital libraries: What do users want? *Online Information Review*, *30*(4), 396–412. doi:10.1108/14684520610686292

Kresh, D. (2007). *The whole digital library handbook*. Chicago, IL: American Library Association.

Landon, G. V. (2009). Toward digitizing all forms of documentation. *D-Lib Magazine*, *15*(3). Retrieved from http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/march09/landon/03landon.html doi:10.1045/march2009-landon

Lesk, M. (2004). *Understanding digital libraries*. Amsterdam, The Netherlands: Morgan-Kaufmann.

Long, H., Lage, K., & Cronin, C. (2005). The flight plan of a digital initiatives project, part 2: Usability testing in the context of user-centered design. *OCLC Systems & Services*, *21*(4), 324–345. doi:10.1108/10650750510631703

Morse, E. L. (2002). Evaluation methodologies for information management systems. *D-Lib Magazines*, *8*(9). Retrieved from http://www.dlib.org/dlib/september02/morse/09morse.html

Nielson, J. (1993). *Usability engineering*. Cambridge, UK: Academic Press.

Norlin, E., & Winters, C. M. (2002). *Usability testing for library web sites*. Chicago, IL: ALA.

Reeve, T., Apedoe, X., & Woo, Y. H. (2003). Evaluating digital libraries: A user-friendly guide. *Digital Library of Information Science & Technology*. Retrieved from http://www.dpc.ucar.edu/projects/evalbook/EvaluatingDigitalLibraries.pdf

Ross, S. (2007). *Digital preservation, archival science and methodological foundations for digital libraries*. Keynote address at the 11th European Conference on Digital Libraries (ECDL). Budapest, Hungary. Retrieved from http://www.ecdl2007.org/Keynote_ECDL2007_SROSS.pdf

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage Publications.

Tsakonas, G., & Papatheodorou, C. (2009). *Evaluation of digital libraries: An insight into useful applications and methods*. Oxford, UK: Chandos Publishing.

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

Witten, I. H., Bainbridge, D., & Nichols, D. M. (2010). *How to build a digital library*. Amsterdam, The Netherlands: Morgan Kaufmann.

Xie, H., & Wolfram, D. (2002). State digital library usability: Contributing organizational factors. *Journal of the American Society for Information Science and Technology, 53*, 1085–1097. doi:10.1002/asi.10152

Xie, I. (2008). *Interactive information retrieval in digital environments*. Hershey, PA: IGI Global. doi:10.4018/978-1-59904-240-4

KEY TERMS AND DEFINITIONS

Digital Libraries: "Digital Libraries (DLs) are defined as representations of emergent and complex forms of digital information organization and design, consisting of multiple layers and building blocks, in various stages of development. DLs present a variety of resources created in the digital format as well as those converted from analog materials through digitization efforts, including print materials, manuscripts, images, audio, and video" (Matusiak, 2010, p. 15).

Digital Library Evaluation Framework: Digital library evaluation framework refers to a comprehensive evaluation structure that includes evaluation constructs, criteria, and measures to assess multiple aspects of digital libraries.

Evaluation: "Evaluation means assessing performances or value of system, process, product, or policy" (Saracevic, 1995, p. 138).

Evaluation Construct: Evaluation construct refers to a major concept or a dimension of digital libraries identified in the evaluation framework.

Evaluation Criterion: Evaluation criterion refers to a specific standard or benchmark, associated with an evaluation construct of digital libraries, on which a judgment or decision is based.