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 establishing 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, EQUINOX ISO/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