A Global Approach to Digital Library Evaluation

Giuseppina Vullo

Humanities Advanced Technology and Information Institute (HATII),
University of Glasgow,
g.vullo@hatii.arts.gla.ac.uk

Abstract

This paper describes the key advances on digital library evaluation research. The paper provides a comparison of the existing models, the current research questions in this area, an integrated LIS-oriented evaluation framework, and a selection of international projects.

Key Words: digital libraries; digital library evaluation; digital library evaluation models; quality

Introduction

More and more efforts have been made in the last decade to evaluate digital libraries and to build global evaluation models, even if an accepted methodology that encompasses all the approaches does not exist. Research and professional communities have specific viewpoints on what digital libraries are, and they use different approaches to evaluate them. Evaluating digital libraries is a challenging activity, as digital libraries are complex, dynamic and synchronic entities which need flexible approaches.

Since 1999, when Christine Borgman described the gap between the perspectives of digital library researchers and professionals (Borgman, 1999), several initiatives have been undertaken to establish a framework for exchange

between the two communities. The reference definitions of 'digital library' in this work are:

1. 'Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities' (Waters, 1998), which has been adopted by the Digital Libraries Federation in 2002.
2. 'A possibly virtual organization that comprehensively collects, manages, and preserves for the long term rich digital content, and offers to its user communities specialized functionality on that content, of measurable quality and according to codified policies', formulated within the DELOS project (Candela *et al*., 2008).

We will also keep a background library science approach taking into account Ranganathan's five laws (Ranganathan, 1931):

1. Books are for use.
2. Every reader his [or her] book.
3. Every book its reader.
4. Save the time of the Reader.
5. The library is a growing organism.

State-of-the-art

Digital library (subsequently DL) evaluation has been investigated since the end of the 1990s, when Saracevic and Kantor (Saracevic & Kantor, 1997) reviewed the traditional libraries' evaluation criteria identified by Lancaster (Lancaster, 1993), and Saracevic (Saracevic, 2000) systematised the issue within a continuative approach, highlighting the need to focus on the DL mission and objectives. According to Saracevic, considering evaluation as the appraisal of the performance or functioning of a system, or part thereof, in relation to some objective(s), the performance can be evaluated as to: effectiveness (how well does a system do what it was designed for?)

 – efficiency (at what cost, in terms of money or time?)
 – a combination of these two (i.e. cost-effectiveness) (Saracevic, 2000, p. 359).

Saracevic also indicated two evaluation levels, which is challenging to integrate:

– user-centred level (which can be social, institutional, individual or focused on the interface)
– system-centred level (which can be focused on engineering, processing or content) (Saracevic, 2000, p. 363–364).

In the same year, Marchionini (Marchionini, 2000) proposed the application of the same techniques and indicators used for traditional libraries, such as circulation, creation and growth of collections, user data, user satisfaction, and financial stability indicators. According to Marchionini, a DL evaluation can have different aims, from the understanding of basic phenomena (e.g. the users' behaviour towards IR tools) to the effective evaluation of a specific object. He presented the results of a longitudinal analysis of the Perseus DL, which lasted more than ten years (Marchionini, 2000). Among the evaluation corollaries of that study, he stated that successful DLs should have:

- clear missions
- strong leadership and a strong talent pool
- good technical vision and decisions
- quality content and data management
- giving users multiple access alternatives
- ongoing evaluation effort (Marchionini, 2000).

Successively, some guidelines to evaluate DLs were proposed (Reeves, Apedoe & Woo, 2003), focusing on the decision process that is behind any evaluation; and the need to focus on the 'global' impact that a DL has on its users and on society was highlighted (Chowdhury & Chowdhury, 2003) by integrating LIS, IR and HCI criteria. Through the analysis of eighty DL case studies, Saracevic observed the small quantity of 'real data' in comparison to the explosion of meta-literature (Saracevic, 2004), concluding that there is no 'best' methodology: different aims can lead to heterogeneous methods.

The development of an evaluation model has been carried forward by the DELOS project; its evaluation schema initially had three dimensions:

– data/collection
– system/technology
– users/uses (Fuhr *et al.*, 2001).

This schema was then integrated into Saracevic's four evaluation categories (Saracevic, 2004; Fuhr *et al.*, 2007).

A quality model for digital libraries was elaborated in 2007 within the 5S (Streams, Structures, Spaces, Scenarios, and Societies) theoretical framework (Gonçalves *et al.*, 2004; Gonçalves *et al.*, 2007); the model was addressed to digital library managers, designers and system developers, and defined a number of dimensions and metrics which were illustrated with real case studies.

Within the DELOS Digital Library Reference Model (Candela *et al.*, 2008), quality facets and parameters have been investigated to model the Quality domain. Quality is defined as 'the parameters that can be used to characterise and evaluate the content and behaviour of a DL. Some of these parameters are objective in nature and can be measured automatically, whereas others are inherently subjective and can only be measured through user evaluation (e.g. in focus groups' (Candela *et al.*, 2008, p. 20).

The ongoing EU-funded project <u>DL.org</u> is currently investigating and identifying solutions for interoperability, according to the six domains of the DELOS Reference Model (Architecture, Content, Functionality, Policy, Quality and Users) (Candela *et al.*, 2008), and its <u>Quality Working Group</u> aims to continuing the research on quality parameters and dimensions developed within DELOS.

Research advances on digital library evaluation and quality are especially needed considering the amount of national and international collaborative projects aiming the interoperation between diverse DLs, and their connection with individuals, groups, institutions, and societies; they can also have a crucial role within DL projects political and social acceptance.

Digital Library Evaluation Models

Concepts and models for evaluating digital libraries come mainly from three research areas: library and information science (LIS) studies, computer science studies, and human-computer interaction (HCI) studies.

They can adopt the following types of approach:

- content-based approach (DLs as collections of data and metadata);
- technical-based approach (DLs as software systems);

–   service-based approach (DLs as organisations providing a set of intangible goods, i.e. benefits);
–   user-based approach (DLs as personal and social environments).

Considering the dynamic nature of digital libraries and the spread of projects dedicated to them, the research on comprehensive evaluation models is quite limited. In this study we present a selection of models that not always come from the DL field; however, they are considered relevant for building global DL evaluation frameworks.

The first comprehensive and multidimensional model comes from information systems research. It is known as the D&M IS success model (DeLone & McLean, 1992) and was updated in 2003 (DeLone & McLean, 2003), when the authors decided to revise it considering the advent and explosive growth of e-Commerce. The updates concerned the adding of a 'service quality' measure as a new dimension, and the grouping of all the 'impact' measures into a single impact or benefit category called 'net benefit' (in the original model, there were 'individual impact' and 'organisational impact'). The D&M IS success model identifies the interrelationships between six variable categories involved in the 'success' of information systems: 'information quality', 'system quality', 'service quality', 'intention to use/use', 'user satisfaction', 'net benefits'.

The three types of quality (which — in the DL world — correspond to content quality, DL system quality and DL services quality) concur to the quantitative and qualitative interactions with the information system, respectively 'intention to use/use' and 'user satisfaction'. The final entity of the model is 'net benefits', which within the DL world could become 'social benefits', either to the individuals or to groups and communities. This model does not offer any specific quality parameter or metrics: on the contrary, it aims to be simple and as general and applicable as possible.

The second model comes for the Library Science field and was created in 2004 for the holistic evaluation of traditional library services (Nicholson, 2004). The model is a pyramid describing the evaluation workflow; it adopts an operational approach, and identifies core steps and actors involved in the evaluation process. Nicholson's model is relevant not only because it is the first one that aims to consider libraries' evaluation holistically, but mostly because it takes into account the role of the administrators — who are also the decision-makers — at the head of the evaluation pyramid. At the basis of Nicholson's

pyramid there is the measurement matrix, i.e. the measurements from different topics and perspectives; the upper level is constituted by the evaluation criteria; the highest level corresponds to the evaluation viewpoints which are classified hierarchically (from the lowest to the highest) as evaluations by 'users', by 'library personnel', and by 'decision-makers'. Once the decisions have been made, the evaluation cycle moves from the pyramid's top to its bottom, with the 'changes implemented by the library personnel', 'users impacted by changes', 'evaluation criteria selected to measure impact on system' and, at the basis the pyramid again, the 'measurements from different topics and perspectives selected', where the cycle starts again (Nicholson, 2004). Nicholson's focus is the organisational context of evaluation, and it does not explain how the different viewpoints and measurements can be combined or integrated, nor the quality parameter and metrics involved.

The third model — known as 'a generalised schema for a digital library' — is the result of research developed within the EU-funded DELOS project. It constitutes the first holistic model specifically created for DL evaluation from the research community (Fuhr *et al.*, 2001). The model describes the DL domain and its three core entities — 'system/technology', 'data/collection', and 'users' — all directing to a fourth entity called 'usage'. The DL domain is over-arched by a larger circle called 'research domain'. The research domain identifies the research areas involved in the four entities of the DL domain as follows:

- system/technology: system and technology researchers;
- data/collection: librarians, LIS researchers;
- users: publishers, sociology of science, communication researchers;
- usage: HCI, librarians, systems researchers (Fuhr *et al.*, 2001).

This model effectively illustrates the heterogeneity of research fields involved in DLs. However — excluding policy makers, managers, senior librarians and administrators — does not take into account the organisational context of the DL.

The fourth and most recent model, grounded on Saracevic's evaluation dimensions, identifies the most important evaluation criteria according to the different stakeholders of digital libraries, within a holistic perspective ('Holistic DL evaluation model') (Zhang, 2010). Zhang identified the most relevant digital library evaluation criteria among five groups of survey participants (administrators, developers, librarians, researchers, users), according to Saracevic's six dimensions (content, technology, interface, service, user, context).

While conducting a comparative study on digital library evaluation models, I developed a LIS-oriented framework, as illustrated in Figure 1. The model includes both the user and system perspective. The arrows indicate the routes corresponding to the two perspectives, i.e. the DL core directions, respectively 'use' and 'mission'.

The core entities of the DL are organisation, content, services and users. According to these four entities the evaluation can focus on organisational aspects (such as management and policies), content aspects (quality of data, metadata, digital collections and even digital libraries), service aspects (quality of technologies or quality of design), or users' aspects (quality of interactions between users and the DL).

Fig. 1: A LIS-oriented framework for DL evaluation.

Developing metrics: EQUINOX, the DLib Test Suite and DigiQUAL

Among the first European projects on digital library evaluation, EQUINOX — Library Performance Measurement and Quality Management System, run from 1998 to 2000 — was the first one within which a set of specific performance indicators were developed. The intention of this set of indicators was to enhance the indicators for traditional library services presented in ISO 11620: Library Performance Indicators. The project delivered fourteen performance indicators, a consolidated list of datasets, a glossary and a list of collection methodologies which are still publicly available on the project website.

During the same years, the D-Lib Working Group on <u>Digital Library Metrics</u> ran 'The DLib Test Suite' (Larsen, 2002), which was an 'early attempt at organizing a rigorous and well supported test bed to enable comparative evaluation of digital library technologies and capabilities' (Larsen, 2002), which was sponsored by the DLib Forum. The DLib Working Group elaborated a standard set of data for quantitative and comparative research, to evaluate and compare the effectiveness of digital libraries and component technologies in a distributed environment (Larsen, 2002).

Regardless of the efforts of EQUINOX and the DLib Metrics Working Group, there is a lack of common strategies for digital libraries evaluation. Within the <u>ARL</u> (Association of Research Libraries) an operational evaluation protocol has been developed, called <u>DigiQUAL</u>. DigiQUAL aims to provide a standard methodology to measure DL service quality and is grounded on the <u>LibQUAL</u> protocol, which is used to measure service quality by traditional libraries. DigiQUAL has identified more than one hundred eighty items around twelve themes related to digital library service quality (Lincoln, Cook & Kyrillidou, 2004).

Conclusions

In this paper we indicated interdisciplinary routes towards a global approach to DL evaluation, describing the state-of-the art, proposing models from different research fields, presenting a LIS-oriented framework and selecting the most relevant projects in this area. There is no common agreement on how to evaluate DLs and evaluation activities are still low-prioritised issues within the DL field. Several assessment methodologies have been built and the interdisciplinary research is growing, while a broadly accepted model is still lacking.

References

Borgman, C.L. (1999): 'What are digital libraries? Competing visions', *Information Processing and Management*, 35(3), 227–243.

Candela, L. *et al.* (2007): 'Setting the foundation of digital libraries', *D-Lib Magazine*, 13, 2007, <<u>http://www.dlib.org/dlib/march07/castelli/03castelli.html</u>>.

Candela, L. *et al.* (2008): *The DELOS Digital Library Reference Model. Foundations for Digital Libraries*, Version 0.98, Project no. 507618, DELOS, <<u>http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf</u>>.

Chowdhury, G.G. and S. Chowdhury (2003): *Introduction to Digital Libraries*. Facet Publishing, London.

DeLone, W.H. and E.R. McLean (2003): 'The DeLone and McLean model of information systems success: a ten-year update', *J. Management Information Systems*, 19(4), 9–30.

Fuhr, N. *et al.* (2001): 'Digital libraries: a generic classification and evaluation scheme', Proceedings of *ECDL 2001*. LNCS, 2163, 187–199. Springer, Heidelberg.

Fuhr, N. *et al.* (2007): 'Evaluation of digital libraries', *Int. J. Digital Libraries*, 8(1), 21–38.

Gonçalves, M.A. *et al.* (2004): 'Streams, Structures, Spaces, Scenarios, Societies (5s): A formal model for digital libraries', *ACM Trans. Inf. Syst.*, 22(2), 270–312.

Gonçalves, M.A. *et al.* (2007): '"What is a good digital library?" A quality model for digital libraries', *Information Processing and Management*, 43(5), 1416–1437.

Lancaster, F.W. (1993): *If you want to evaluate your library*, 2nd edn. Morgan Kaufman, San Francisco.

Larsen, R.L. (2002): *The DLib Test Suite and Metrics Working Group. Harvesting the Experience from the Digital Library Initiative*, University of Maryland, publicly available at <http://www.dlib.org/metrics/public/papers/The_Dlib_Test_Suite_and_Metrics.pdf>.

Lincoln, Y.S., C. Cook and M. Kyrillidou (2004): 'Evaluating the NSF National Science Digital Library Collections: Categories and Themes from MERLOT and DLESE', paper presented at the MERLOT International Conference, Costa Mesa, California.

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

Nicholson, S. (2004): 'A conceptual framework for the holistic measurement and cumulative evaluation of library services', *J. Documentation*, 60(2), 164–182, <http://www.bibliomining.com/nicholson/holisticfinal.html>.

Ranganathan, S. (1931): *The Five Laws of Library Science*. Edward Goldston, London.

Reeves, T.C. X. Apedoe and Y.H. Woo (2003): *Evaluating Digital Libraries: A User-Friendly Guide*. National Science Digital Library, University of Georgia, Athens, GA.

Saracevic, T. and P.B. Kantor (1997): 'Studying the value of library and information services. Part II. Methodology and taxonomy', *J. American Society for Information Science*, 48(6), 543–563.

Saracevic, T. (2000): 'Digital library evaluation: toward an evolution of concepts', *Library Trends*, 49(3), 350–369.

Saracevic, T. (2004): 'Evaluation of digital libraries: an overview', paper presented at the DELOS Workshop on the Evaluation of Digital Libraries, <http://dlib.ionio.gr/wp7/WS2004_Saracevic.pdf>.

Su, L.T. (1992): 'Evaluation measures for interactive information retrieval', *Information Processing and Management*, 28(4), 503–516.

Waters, D.J. (1998): 'What are digital libraries?', *CLIR Issues*, 4, 1998, <http://www.clir.org/pubs/issues/issues04.html#dlf>.

Zhang, Y. (2010): 'Developing a holistic model for digital library evaluation', *J. American Society for Information Science*, 61(1), 88–110.

Websites Referred to in the Text

ARL, Association of Research Libraries, http://www.arl.org/

Delos project, http://www.delos.info/

DigiQUAL, http://www.digiqual.org/digiqual/index.cfm

D-Lib Working Group on Digital Library Metrics, http://www.dlib.org/metrics/public/metrics-home.html

DL.org, http://www.dlorg.eu/; Quality Working Group, https://workinggroups.wiki.dlorg.eu/index.php/Quality_Working_Group

EQUINOX, http://equinox.dcu.ie, and http://equinox.dcu.ie/reports/pilist.html

LibQUAL, http://www.libqual.org/home