Vullo, G. (2010) A global approach to digital library evaluation towards quality interoperability. Bulletin of IEEE Technical Committee on Digital Libraries, 6(1),

Version: Published

http://eprints.gla.ac.uk/45792/

Deposited on: 11 May 2015

A Global Approach to Digital Library Evaluation

Towards Quality Interoperability

Giuseppina Vullo

Humanities Advanced Technology and Information Institute

University of Glasgow

g.vullo@hatii.arts.gla.ac.uk

Abstract

This paper describes some of the key research works related to my PhD thesis. The goal is the development of a global approach to digital library (DL) evaluation towards quality interoperability. DL evaluation has a vital role to play in building DLs, and in understanding and enhancing their role in society. Responding to two parallel research needs, the project is grouped around two tracks. Track one covers the theoretical approach, and provides an integrated evaluation model which overcomes the fragmentation of quality assessments; track two covers the experimental side, which has been undertaken through a comparative analysis of different DL evaluation methodologies, relating them to the conceptual framework. After presenting the problem dentition, current background and related work, this paper enumerates a set of research questions and hypotheses that I would like to address, and outlines the research methodology, focusing on a proposed evaluation framework and on the lessons learned from the case studies.

1. Introduction

Digital library evaluation is a growing interdisciplinary area. Researchers and practitioners have specific viewpoints of what DLs are, and they use different approaches to evaluate them. Each evaluation approach corresponds to a DL model, and there is no common agreement on its depiction [1]. Despite that, more and more efforts have been made to evaluate DLs. However, a methodology that encompasses all the approaches does not yet exist. There are two main reasons for this:

digital libraries are complex entities that need interdisciplinary approaches

digital libraries are synchronic entities: the speed of evolution of DLs coupled with their lack of historical traces makes a longitudinal analysis difficult if not impossible.

Nevertheless, DLs and DL research have reached a level of maturity such that a global approach to their evaluation is needed. It would encourage exchange of qualitative data and evaluation studies, allowing comparisons and communication between research and professional communities. Since 1999, when Christine Borgman described the gap between the perspectives of researchers and professionals [1], several initiatives have been undertaken to establish a framework for exchange between the two communities.

The reference definitions of DLs in this work are:

Waters (given in 1998; became the DLF definition in 2002):

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 [2].

DELOS (2006):

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 [3, 4].

In this work, we also keep a background concept of DLs considered as advanced extensions of traditional libraries. Using a library science approach within a continuative perspective, DLs and their role in society still follow Ranganathan's five laws :

Books are for use.

Every reader his [or her] book.

Every book its reader.

Save the time of the Reader.

The library is a growing organism.[5]

Abstracting the concepts of "book" and "reader", we think these laws are still valid and powerful; moreover, the fifth law helps us to understand how they are considered – and evaluated – holistically.

2 Background and related work

2.1 State of the art

Concepts and models for evaluating DL quality come from:

library and information science (LIS) studies (services, organisation, metadata creation)

information retrieval (IR) studies (search engines, metadata management)

human–computer interaction (HCI) studies (user interfaces) [6, 7].

Marchionini proposes 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 [8]. (For an insight into the recent state of the art, see [9]).

Reviewing the evaluation criteria identified by Lancaster [10] and by Saracevic and Kantor [11], Saracevic systematises the issue within a continuative approach, highlighting the need to focus on the DL mission and objectives [12].

Considering evaluation as the appraisal of the performance or functioning of a system, or part there of, 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) [12, p. 359].

Saracevic also indicates two evaluation levels:

user-centered level (which can be social, institutional, individual or focused on the interface)

system-centered level (which can be focused on engineering, processing or content) [12, pp. 363–364].

According to Saracevic, the issue is how to make human- and system-centered approaches work together [13].

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 [8]. Reeves, Apedoe and Woo propose some guidelines to evaluate DLs, focusing on the decision process that is behind any evaluation [14]. Chowdhury and Chowdhury highlight the need to focus on the global impact that a DL has on its users and on society in general, integrating LIS, IR and HCI criteria [15].

Through the analysis of eighty DL case studies, Saracevic outlines the approaches and methodologies that are used in practice, observing the small quantity of "real data" in comparison to the explosion of meta-literature [16]. He concludes that there is no "best" methodology: different aims can lead to different methods. He also states that DLs are still too difficult for the general public to use, although they can have a far-reaching impact on education, scholarly research publishing and society.

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 [17]

This schema has since been integrated into Saracevic's evaluation questions [16, 18].

According to the DELOS Reference Model, quality constitutes 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) [3]. A new EU-funded project, DL.org, will investigate and propose solutions for interoperability, taking into consideration core requirements for DL architecture, content, functionality, policy, quality and users identified in the DELOS Reference Model [19].

Today LIS researchers generally agree, identifying three approaches to quality evaluation:

content-based approach (digital collections, metadata)

service-based approach (DLs as organisations)

user-based approach (DLs as specific and useful digital environments).

Interdisciplinary research is increasing. The aim of most national and international DL projects is to connect and integrate DLs with the largest range of personal activities, research communities, institutions, educational environments and societies. DLs need to identify and assess their added value in knowledge creation and reuse [8, p. 329]. However, according to Poll, the levels of usage and quality of DLs do not yet prove that users benefit from DLs [20]. This statement reveals how much DLs are still questioned. In that sense, the evaluation of DLs can also be considered as a process related to their political and social acceptance.

2.2 Integrated evaluation models

The research on integrated DL evaluation models is quite limited. Four models, from different research areas, have been taken as references at the beginning of this work. The first model [21], created in 1992, comes from information systems research but can also be applied to DLs. It is known as the D&M IS success model and was updated by the authors in 2003. The updates concerned adding a "service quality" measure as a new dimension, and the grouping all the "impact" measures into a single impact or benefit category called "net benefit" (in the original model, there were "individual impact" and "organizational impact"). The D&M IS success model identifies the interrelationships between six variable categories: "system quality", "information quality", "service quality", IS "intention to use/use", "user satisfaction", and "net benefits."

The second model was created in 2004 for the holistic evaluation of libraries services, and comes from library science studies [22]. It has an operational approach and identifies the core steps and the actors involved in the library's evaluation, taking into account the role of the administrators, put at the top of the evaluation pyramid.

The third model is the result of a DELOS supported research, and is the first generalised schema for a DL that proposes a holistic approach to its evaluation [17]. It includes the entities "users", "data/collection" (i.e. the information, the content), "system/technology" and "usage." The model identifies the communities involved in DL evaluation, within an upper research domain area. However, the schema – which includes "librarians" within the research domain - doesn't consider

the organisational level of the DL and its actors (policy makers, managers, administrators, senior librarians).

The user and digital library model [23], created in 2007, relates the DL and user points of view, focusing on their contexts (see Fig. 1).

Fig. 1. User and digital library model. © 2007 Y. Zhang & T. Saracevic

3 Research questions and hypotheses

From my literature review so far, it has become clear to me that there is a need to look at DL evaluation models from a holistic perspective and to compare them in order to offer an integrated approach and a simple framework for DL assessment. I have also realised that there is a lack of exchange between the evaluation models and their applications, of which there are still relatively few, and that qualitative studies would need to be interoperable.

I recognise the need to study the following concepts:

DL quality, considering a DL as a unique entity

DL qualitative evaluation

holistic models of DL evaluation

system vs user perspectives.

Thus I propose the following research questions:

Is a global approach to DL evaluation possible?

Who are the actors that are involved in DL evaluation?

Who needs the results from DL evaluations?

How should we make quality procedures routine?

How should we enable interoperability between qualitative DL evaluations?

In addition, I am interested in testing the following hypotheses:

DL qualitative evaluation would be more effective with a holistic approach. A common evaluation framework would avoid duplication of efforts and facilitate the exchange of materials under a comprehensive background.

Librarians, researchers and society need DL evaluation as a fundamental part of any DL project.

Qualitative DL evaluation would be more effective as a regular activity.

There is a need to make evaluation experiences comparable and interoperable.

4 Analyses and evaluations/proposed methodology

In any information system, quality can be measured from a:

system perspective (internal view) – focused on design and operation

user perspective (external view) – focused on use and value [24].

DL analyses and evaluations in this work are considered from a system perspective. The methodology I propose combines theoretical and practical comparative approaches.

4.1 A global approach to DL design/interface evaluation

A comprehensive qualitative interface evaluation of twelve DLs has been done.

I wanted to focus not just on graphic features but also on content presentation, label coherence, users' tools, transparency and personalisation services. An integrated evaluation schema was created for this purpose. It includes a description, for basic and contextual information, interface structure analysis (browse and search system, labels, information visualisation), application of quality parameters (transparency, usability, accessibility, user-centred level), qualitative evaluation, and optional notes and remarks.

The qualitative evaluations gave me an idea of which features of the interface are useful, which are not, and what improvements can be made to enhance the user experience. They also helped me to create a checklist with some best practices, and to improve my conceptual framework.

4.2 A global approach to operational DL operational

In order to apply a global approach to operational DL evaluation, I selected two assessment methodologies, both of which come from the digital preservation community: DRAMBORA [25] and InterPARES 3 [26]. The analysis of the case studies and comparisons between the methodologies are part of this work. They focus on DLs as repositories, and include a deep organisational analysis of the whole system. I applied the two methodologies to two digital repositories, and focused my attention on testing their interoperability.

5 Conceptual framework: a tentative model for a global approach to DL evaluation

While conducting heterogeneous DL evaluations and comparing them, I developed a model for a global approach to DL evaluation, as illustrated in Fig. 2.

The model includes internal and external quality views. I have highlighted in italics the starting point of an evaluation from each of the two perspectives, and I have put my evaluation experiences on the framework.

6 Limitations and risks

This study has several limitations. The main one is that it does not include user studies, despite aspiring to give a global approach to DL evaluation. Although restricting the study to one view allows an in-depth analysis of many specific facets of DL quality evaluation, future work should apply the methodology and lessons learned to user studies to test the framework. Moreover, subjectivity is a strong component in qualitative interface evaluation: an expert evaluation cannot be substituted for a user evaluation, even if the expert tries to interact with the system as a user.

There have been several works that analyse DLs' quality, but during my literature review I realised that quality interoperability is used very loosely. One of the challenges that I face is defining

precisely what quality interoperability means within DLs and how it can be improved. I would like to discuss this definition and understanding further in the light of my proposed research.

Another limitation of my approach is the number of methodologies I have tested. I would like to use more methodologies to assess the same DLs. A holistic evaluation of DLs is difficult, and it is more complex in reality than in a model.

I would like to get some suggestions and insights into how to design a global qualitative evaluation framework.

7 Conclusion

There is no common agreement on how to evaluate DLs. DL evaluation as a field is still in its formative years and is generally unfunded; as yet, no general models have been embraced. Only a small fraction of all the works on DL are devoted to evaluation [23]. However, several assessment methodologies have been built, and the interdisciplinary research on the field is growing. A general global approach is still lacking.

I believe my proposed research will help people to understand the need for a global approach to DL evaluation and for a quality framework which allows DLs to communicate with each other and to exchange their evaluation experiences. In addition to this, I hope to contribute by developing a methodology to promote qualitative DL evaluation environments and exchanges.

Fig. 2. A global approach to DL evaluation

References

[1]   C.L. Borgman, "What are digital libraries? Competing visions", Information Processing and Management, 35(3), 1999, 227–243.

[2]   D.J. Waters, "What are digital libraries?" CLIR Issues, 4, 1998, http://www.clir.org/pubs/issues/issues04.html#dlf

[3]   L. Candela L. et al., The Digital Library Manifesto, DELOS, 2006, http://146.48.87.21:80/OLP/UI/1.0/Disseminate/1243690447v496IUssFV/

a221243690447dMYq1WeE

[4]   L. Candela L. et al., "Setting the foundation of digital libraries", D-Lib Magazine, 13, 2007, http://www.dlib.org/dlib/march07/castelli/03castelli.html

[5]   S. Ranganathan, The Five Laws of Library Science, Edward Goldston, 1931.

[6]   L.T. Su, "Evaluation measures for interactive information retrieval", Information Processing and Management,28(4), 1992, 503–516.

[7]   B. Shneiderman, Designing the User Interface: Strategies for Effective Human–Computer Interaction, 3rd edn. Addison-Wesley, 1998.

[8]   G. Marchionini, "Evaluating digital libraries: a longitudinal and multifaceted view", Library Trends, 49(2), 2000, 304–333.

[9]   H.I. Xie, "Users' evaluation of digital libraries (DLs): their uses, their criteria, and their assessment", Information Processing and Management, 44(4), 2008, 1348–1351.

[10]   F.W. Lancaster, If You Want to Evaluate Your Library, 2nd edn. Morgan Kaufman, 1993.

[11]   T. Saracevic, P.B. Kantor, "Studying the value of library and information services. Part II. Methodology and taxonomy", J. American Society for Information Science, 48(6), 1997, 543–563.

[12]   T. Saracevic, "Digital library evaluation: toward an evolution of concepts", Library Trends, 49(3), 2000, 350–369.

[13]   T. Saracevic, "Information science", J. American Society for Information Science, 50(12), 1999, 1051–1063.

[14]   T.C. Reeves, X. Apedoe, Y.H. Woo, Evaluating Digital Libraries: A User-Friendly Guide, National Science Digital Library, 2003.

[15]   G.G. Chowdhury, S. Chowdhury, Introduction to Digital Libraries, Facet Publishing, 2003.

[16]   T. Saracevic, Evaluation of digital libraries: an overview, paper presented at DELOS Workshop on the Evaluation of Digital Libraries (2004), http://dlib.ionio.gr/wp7/WS2004_Saracevic.pdf

[17]   N. Fuhr et al., "Digital libraries: a generic classification and evaluation scheme", ECDL 2001, LNCS 2163, Springer, 2001, 187–199.

[18]   N. Fuhr et al., "Evaluation of digital libraries", Int. J. Digital Libraries, 8(1), 2007, 21–38.

[19]   DL.org, http://www.dlorg.eu/

[20]   R. Poll, "Measuring the impact of the new library services", Proc. World Library and Information Congress (IFLA 2005), 2005, http://archive.ifla.org/IV/ifla71/papers/081e-Poll.pdf

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

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

[23]   Y. Zhang, T. Saracevic, "Criteria in evaluation of use & usability in digital libraries", LIDA2007, 2007, http://www.ffos.hr/lida/datoteke/LIDA2007-Zhang_Saracevic.ppt

[24]   Y. Wand, R.Y. Wang, "Anchoring data quality dimensions in ontological foundations", Communications of the ACM, 39(11), 1996, 86–95.

[25]   Digital Curation Centre (DCC), DigitalPreservationEurope (DPE), "Digital Repository Audit Method Based on Risk Assessment", v. 1.0, 2007, http://www.repositoryaudit.eu/

[26]   InterPARES 3 Project, http://www.interpares.org/ip3/ip3_index.cfm