Fachbeitrag

Juliane Stiller und Vivien Petras

Learning from Digital Library Evaluations: The Europeana Case

Abstract: In this paper we analyse evaluation studies of the Europeana digital library from its launch in 2009 until today. Using Saracevic's digital library evaluation framework, the studies are categorised by their constructs, contexts, criteria, and methodologies. Concentrating on studies that evaluate Europeana services or single components, we show gaps in the evaluation of certain Europeana aspects. Finally, we derive strategies for building an evaluation archive that serves as memory and supports comparisons.

Keywords: digital library, evaluation, Europeana

Aus Evaluationen digitaler Bibliotheken lernen: Das Fallbeispiel Europeana

Zusammenfassung: Im vorliegenden Artikel analysieren wir Evaluationsstudien der digitalen Bibliothek Europeana von 2009 bis heute. Unter Berücksichtigung von Saracevic' Evaluationsframework für digitale Bibliotheken werden die Studien nach ihren Konstrukten, Kontexten, Kriterien und Methodologien kategorisiert. Die Analyse konzentriert sich auf Studien, die Dienstleistungen oder einzelne Komponenten von Europeana evaluieren, und zeigt Lücken in der Evaluation bestimmter Aspekte von Europeana auf. Schließlich werden Strategien diskutiert, um ein Evaluationsarchiv zu entwickeln, welches sowohl der Langzeitarchivierung dient als auch Vergleiche von Evaluationsergebnissen unterstützt.

Schlüsselwörter: Evaluation, digitale Bibliothek, Europeana

Introduction

Europeana[1] is Europe's digital library, museum and archive. In 2009, it was launched with the mission to aggregate digital cultural heritage content from various institutions in Europe[2]. By that time, the Digital Library (DL) field had been well established with the first European Conference on Research and Advanced Technology for Digital Libraries (ECDL, now TPDL)[3] being held in Pisa, Italy, more than 10 years earlier (ECDL 1997)[4]. Today, however, Europeana has achieved exemplary status for other DLs in the area. Many more DLs such as the Digital Public Library of America (DPLA)[5] and the German Digital Library (DDB)[6] base their services on Europeana's experiences and Europeana continues its trailblazer role in metadata modelling, licensing and the aggregation of large and heterogeneous volumes of digital cultural heritage content.

Europeana is not only an access point to digital cultural heritage content, but also a platform that provides various services for different stakeholders such as cultural heritage professionals and the creative industry. The authors have accompanied Europeana's development from its first steps in the European Digital Library Network (EDLnet) to becoming the ecosystem of offerings around digital cultural heritage it is now. In this paper, we would like to investigate the evaluations of Europeana of the past decade and the changes accompanied by these studies. Applying Tefko Saracevic's framework of DL evaluation[7], we identify common evaluation objectives, the methods used and the results obtained. Based on an earlier study of a meta-analysis of 41 Europeana evaluations[8], we extracted a smaller pool of 31 papers that solely evaluate the Europeana portal and its services. Offering a closer look into the criteria and methods of these papers, we propose

1 https://www.europeana.eu/ (05.02.2018).

2 Purday, Jon. "Think Culture: Europeana.eu from Concept to Construction." *The Electronic Library* 27,6 (2009): 919–937.

3 http://www.tpdl.eu/ (05.02.2018).

4 http://www.tpdl.eu/?p=125 (05.02.2018).

5 https://dp.la/ (05.02.2018).

6 https://www.deutsche-digitale-bibliothek.de/ (05.02.2018).

7 Saracevic, Tefko. "Digital Library Evaluation: Toward an Evolution of Concepts." *Library Trends* 49,2 (2000): 350–369.

8 Petras, Vivien, Juliane Stiller. "A Decade of Evaluating Europeana – Constructs, Contexts, Methods & Criteria." In *Research and Advanced Technology for Digital Libraries. TPDL 2017*. Ed. by Jaap Kamps, Giannis Tsakonas, Yannis Manolopoulos et al. 233–245. Berlin: Springer, 2017.

the outline of an evaluation archive that can be used to store evaluations, their results and the changes initiated by them. Ideally, such a memory will avoid double work and encourage the re-use of research data.

The paper is structured as follows: section 2 presents some of the frameworks that were developed for the evaluation of digital libraries as well as Saracevic's framework that is used to categorise the evaluations in this study. Section 3 describes the Europeana digital library as well as the methodology used to assess the evaluations. In section 4, we present the results of the meta-analysis of Europeana focusing on methods and criteria used for the evaluations and identifying gaps of past evaluations. In section 5, we outline an envisioned evaluation archive for Europeana.

Digital Library Evaluations

Digital libraries became a topic of research in the mid-1990s and have remained an active research area ever since. With substantial funding initiatives in the US and Europe, a large number of digital libraries in many different cultural heritage domains were developed. Along with these projects, a theory of DLs and accompanying research developed. While Saracevic laments in his foreword to the 2016 book *Discover Digital Libraries*[9] that DL practice (the development and maintenance of a DL) and DL research still "reside in parallel universes" with few intersections, DL evaluation is certainly an area where both communities interact since an evaluation cannot take place without an actual DL to be evaluated.

The exact number of DL projects is difficult to ascertain and many of them have ceased to exist when funding or research interest ran out. The same is true for their evaluations as only few have been properly documented in the research literature. However, because of review articles that organise and synthesise evaluation approaches according to different aspects, we can get an idea of the diverse landscape of DLs.

Large evaluation frameworks, which summarised DL projects and their evaluations, were developed both in Europe and in the US. In Europe, the DELOS Network of Excellence on Digital Libraries[10] developed an evaluation framework based on its DL reference model[11]. The Interaction Triptych Evaluation Model[12] defines content, system and users as the important DL components to be evaluated. These components should be assessed on three different axes: usability (the quality of the interactions between users and the system), usefulness (the quality of the relationship between users and content) and performance (the quality of the relationship between system and content). DELOS also suggests criteria and methodologies to perform the evaluation along these axes. Blandford and colleagues[13] developed the 'PRET A Rapporter framework' to support the design of evaluation user studies. While these two frameworks provide concrete guidelines for designing an evaluation, the DiLEO DL evaluation ontology[14] is a research effort to model evaluation components from different frameworks in order to guide understanding of the whole evaluation process.

In the US, three research centres developed large-scale evaluation frameworks and initiatives. The Perseus DL, which continues its services until today, was one of the first DL projects, which made a concerted effort to evaluate its components and functionalities in a more structured way.[15] At the University of Virginia, the 5S research group[16] developed an evaluation framework based on its DL model[17] and even developed an automatic method to assess DL components[18]. The Rutgers University research group around Saracevic (next section) analysed and summarised evaluation elements in their specialized

ies. Version 0.96. DELOS Network of Excellence on Digital Libraries, 2007.

12 Fuhr, Norbert, Giannis Tsakonas, Trond Aalberg et al. "Evaluation of digital libraries." *International Journal of Digital Libraries* 8,1 (2007): 21–38.

13 Blandford, Ann, Anne Adams, Simon Attfield et al. "The PRET A Rapporter framework: Evaluating digital libraries from the perspective of information work." *Information Processing & Management* 44,1 (2008): 4–21.

14 Tsakonas, Giannis, Christos Papatheodorou. "An ontological representation of the digital library evaluation domain." *Journal of the American Society for Information Science and Technology* 62,8 (2011): 1577–1593.

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

16 5S is a framework for digital libraries (Streams, structures, spaces, scenarios, and societies). More on this can be found here: Gonçalves, M. A., E. A. Fox, L. T. Watson et al. "Streams, structures, spaces, scenarios, societies (5S): a formal model for digital libraries." *ACM Transactions on Information Systems* 22,2 (2004): 270–312.

17 Gonçalves, Marcos A., Bárbara L. Moreira, Edward A. Fox et al. "What is a good digital library? A quality model for digital libraries." *Information Processing & Management* 43,5 (2007): 1416–1437.

18 Moreira, Bárbara L., Marcos A. Gonçalves, Alberto H. F. Laender et al. "Automatic evaluation of digital libraries with 5squal." *Journal of Informetrics* 3,2 (2009): 102–123.

9 Xie, Iris, Krystyna Matusiak. *Discover Digital Libraries: Theory and Practice*. Amsterdam: Elsevier, 2016.

10 http://delosw.isti.cnr.it/ (05.02.2018).

11 Candela, Leonardo, Donatella Castelli, Nicola Ferro et al. *The DELOS digital library reference model: Foundations for Digital Librar*

frameworks[19], which are usually based on Saracevic's. In their recent book, Xie and Matusiak describe their Multi-faceted Evaluation of Digital Libraries (MEDaL)[20] review study, which included a literature analysis of 85 papers and a Delphi study to describe DL evaluation dimensions, objectives, criteria and measures also based on Saracevic.

Saracevic's Evaluation Framework

In this study, we used the evaluation framework of Saracevic[21], which appears to be one of the more widely adopted frameworks in evaluation research. For example, the elements introduced in this framework were adapted to others in the domain such as DELOS and MEDaL. Saracevic introduces five elements that frame a DL evaluation, namely *Construct*, *Context*, *Criteria*, *Measures* and *Methodology*. These elements are the components of each evaluation and are described as follows:

1. *Construct* describes the object of the evaluation: What is evaluated? Which aspect is at the centre of the evaluation?
2. The *Context* determines the perspective that is used for the evaluation. Saracevic distinguishes the user-centred perspective (with social, institutional or individual levels), the interface perspective and the system-centred perspective (with engineering, process and content levels).
3. The *Criteria* element describes, which objectives are evaluated. Saracevic names library criteria such as information accuracy, information retrieval criteria such as relevance, human-computer-interaction and interface criteria such as usability.
4. The *Measures* element determines, how the criteria are evaluated.
5. The *Methodology* describes the approach, process or tool that is used for the evaluation.

In 2004, Saracevic reviewed 80 DL evaluation studies using his framework.[22] Results indicate that for the component *Context*, the system-centred perspective is taken more often than the human- or usability-centred one. The prevailing criteria used in the evaluations are usability, system performance and usage. For methodologies, Saracevic finds surveys most often applied to answer research questions. Structured interviews, focus groups, observations and task accomplishments are further methods identified in the evaluations. Saracevic does not name measures with which the criteria are evaluated. We also found them to be very tailored to particular research questions.

Tsakonas et al. analyse about 220 evaluation studies using their DiLEO ontology. They focus on evaluations published between 2001 and 2011 in the JCDL and ECDL/TPDL conferences.[23] Similar to Saracevic's findings, system-centred contexts are prevailing, employing mainly criteria such as effectiveness, performance measurement and technical excellence. They identify laboratory experiments and surveys as the methodologies primarily used.

The analysis in this paper and the study it is based upon are both similar to the studies mentioned above, but focus on evaluation studies of just one use case – Europeana. We detail the evaluations of one particular DL to underline the necessity of tracking evaluations over time and recording the improvements driven by the studies. Therefore, this paper concentrates on the evaluation of the Europeana portal, the learnings and outcomes of these evaluations and how an evaluation framework can serve for better re-use of processes.

A Meta-analysis of Europeana Evaluations

The portal Europeana[24] offers a single access point to Europe's digital cultural and scientific heritage aggregated from libraries, archives, museums and audio-visual archives. Currently, over 55 million objects can be accessed via Europeana provided by over 3,200 memory institu-

19 See, for example, Xie, Hong. "Evaluation of digital libraries: Criteria and problems from users' perspectives." *Library and Information Science Research* 28,3 (2006): 433–452; Zhang, Ying. "Developing a holistic model for digital library evaluation." *Journal of the American Society for Information Science and Technology* 61,1 (2010): 88–110; Joo, Soohyung, Iris Xie. "Evaluation constructs and criteria for digital libraries: A document analysis." In *Recent Developments in the Design, Construction, and Evaluation of Digital Libraries: Case Studies*. Ed. by Colleen Cool, Kwong Bor Ng. 126–140. Hershey: IGI Global, 2013.
20 Xie, Iris., Krystyna Matusiak. *Discover Digital Libraries*. 281–318. Oxford: Elsevier, 2016.
21 Saracevic 2000.

22 Saracevic, Tefko. "Evaluation of digital libraries: An overview." In *Notes of the DELOS WP7 workshop on the evaluation of Digital Libraries*. Ed. by Maristella Agosti, Norbert Fuhr. 13–30. Padua, 2004.
23 Tsakonas, Giannis, Angelos Mitrelis, Leonidas Papachristopoulos et al. "An exploration of the digital library evaluation literature based on an ontological representation." *Journal of the Association for Information Science and Technology* 64,9 (2013): 1914–1926.
24 http://www.europeana.eu (05.02.2018).

tions. Europeana does not only provide search and browsing functionalities for the aggregated metadata objects, but also maintains API access points for integrating Europeana's data into other contexts. Due to its scale and the heterogeneity of its data in terms of formats, media types and languages, Europeana assumes a trailblazing role in maintaining and offering digital cultural heritage material. Over the past decade, Europeana was the object of interest for many evaluations.

Methodology

In our previous study[25], we accumulate relevant studies based on an existing list created by the Europeana Task Force for Enrichment and Evaluation[26] as well as searches for documents in Google Scholar and Web of Science[27]. The resulting list of 55 papers was reviewed in detail to extract 41 papers that reflect studies with Europeana as research object or used Europeana data to conduct evaluations. Looking for information on Saracevic's five elements, *Construct, Context, Criteria, Measures* and *Methodology*, each of the 41 publications was analysed and relevant information extracted. We followed a grounded theory approach to form categories from the emerging data through discussion and a constant switch between the publication and the categories[28]. For the *Construct* category, the analysis resulted in five groups: Europeana, Europeana Component, Europeana in Comparison, Europeana Service, and Europeana Data.

For this analysis, we only look at studies that focus their efforts on Europeana. We reduce the pool of articles to the ones that have one of the following *Constructs:* Europeana, Europeana Component or Europeana in Comparison. The category Europeana Services aggregates evaluations where developments within Europeana's satellite projects were assessed. As these services often did not make it into the Europeana production system, we exclude these evaluations from this analysis. Similarly, the category Europeana Data focuses on evaluations that solely use data of Europeana to evaluate algorithms or retrieval performance. As these evaluations were not used to assess and improve the portal, we omit them, too. Three documents are added to the pool, as these are the

---

25 Petras/Stiller 2017.
26 https://pro.europeana.eu/project/evaluation-and-enrichments (05.02.2018).
27 Search terms: Europeana and (user* or evaluat* or study*) and variations.
28 Glaser, Barney G., Anselm Strauss. *The discovery of grounded theory: strategies for qualitative research.* Chicago: Aldine, 1967.

most recent evaluations from 2017. In total, we assess 31 papers in detail using Saracevic's framework, defining the used criteria and describing the methods in more detail.

Results from the Meta-analysis

In this section, we present the results of assessing the evaluations with Saracevic's framework, looking into the process and categorising the methods in more detail. Most of the evaluations (17 in total) focus on evaluating Europeana and its services. Nine evaluations specifically choose a component of Europeana and assess it, whereas five studies evaluate Europeana in comparison to other large-scale digital libraries.

Constructs and Context

We further analysed, which perspective the evaluations adopted: a user-centred, system-centred or interface perspective and their corresponding levels. Table 1 lists the number of evaluations for each perspective as well as the overlap with the *Construct* of the evaluation.

Tab. 1: Number of studies in relation to their Construct and their Context. Note that studies can focus on more than one perspective.

	Europeana Digital Library	Europeana Component	Europeana in Comparison
Number of Studies	17	9	5
User-centred	16	2	4
Social	2	0	0
Institutional	0	0	0
Individual	15	2	4
Interface	6	1	0
System-centred	11	9	5
Engineering	0	0	0
Process	6	5	0
Content	11	9	5

As the results show, system-centred evaluations are prevailing, although the focus is clearly on evaluating the content or the technical processes. We did not find any evaluations that focus on the engineering perspective. With regard to the user-centred perspective, there is a clear majority of evaluations that target the individual perspective meaning that either users were directly involved or experts evaluated the system from the users' point of view. Evaluations with a focus on the institutional per-

spective were not found. The societal perspective is only taken into account in two studies that we subsume under the category 'Impact studies'.

Methods and Criteria

To re-use the results of evaluations or learn from their outcomes, it is essential to identify the applied methods as well as the criteria used for the assessment. The methodologies used were manifold and we subsume them under the four categories presented in table 2. Most of the evaluations determine custom criteria for their assessment; these types of evaluations were aggregated under the category 'criteria-based study'.

Tab. 2: Methods used in the evaluations and the number of studies applying them.

Method	Description	Number of studies
Criteria-based study	Certain criteria were determined to assess a service or algorithm.	18
Log file analysis	Evaluation uses an automatically created log file of user interactions.	7
Usability study	Evaluation incorporates several methods to assess usability of a service, e.g. user studies, interviews, surveys, etc.	4
Impact Study	Study uses an expert assessment of the overall value of a service within one or more specific areas.	2

For further detailing the objectives of the evaluations, we specifically look at the criteria assessed and try to identify patterns in the criteria used. Although, many studies do not specifically name the criteria they applied to evaluate a given *Construct*, we were able to extract the objectives described in table 3.

Tab. 3: Criteria applied in evaluation studies and the number of studies applying them (more than one criterium per study is possible).

Criteria	Description	Number of studies
Accessibility	This criterion covers the forms of access used to navigate the content of Europeana.	7
Coverage	This focuses on the content of Europeana and covers linguistic, thematic or geographic coverage; also different media types are covered here, as well as the size of the collection.	9
Data quality	It describes all criteria used to identify the quality of data and metadata.	10
Error rate	It covers the quantification of errors of workflows or in data quality.	2
Impact criteria	It covers all criteria that express change on users or society.	2
Performance evaluation	Criteria that measures how easy workflows, algorithms are executed. It is mainly system-focused.	7
Usability	It covers established criteria in usability research such as efficiency and effectiveness as well as ease of use and related criteria.	4
Usage statistics and patterns	It covers user behaviour such as paths through the system, items clicked or viewed.	9
User satisfaction	These are subjective criteria of users' perception of the system.	2

This classification gives a concise picture of the nature of the evaluation and could be helpful in guiding future evaluations.

Furthermore, we identify only five evaluations involving users, the others are expert evaluations. We consider 12 studies to be quantitative, whereas 19 studies used a qualitative approach.

To allow other researchers interested in past evaluations of Europeana to get an overview of the assessments, we aggregated all information on the studies in table 4.

Tab. 4: Summary of the meta-analysis of the 31 evaluations studied.[29]

Paper	Year	Construct	Method	Criteria	Perspective	Type	User-centred	Interface	System-Centred
Ceccarelli et al.[30]	2011	Europeana	log file analysis	usage statistics & patterns	expert	quantitative	x		
CIBER Research Ltd[31]	2013	Europeana	log file analysis	usage statistics & patterns, performance evaluation	expert	quantitative	x		x
Clark et al.[32]	2011	Europeana	log file analysis	usage statistics & patterns, performance evaluation	expert	quantitative	x		x
Dangerfield et al.[33]	2015	Europeana component	criteria-based evaluation	data quality	expert	qualitative			x
Dani et. al[34]	2015	Europeana	usability study	usability	user	qualitative	x	x	x
Dickel[35]	2015	Europeana in comparison	criteria-based evaluation	coverage, data quality, usability	expert	qualitative	x		x
Dobreva et al.[36]	2010	Europeana	usability study	user satisfaction, coverage, usability	user	qualitative	x	x	x
Dobreva et al.[37]	2010	Europeana	usability study	usability	user	qualitative	x	x	
Gäde[38]	2014	Europeana	log file analysis	usage statistics & patterns, coverage, accessibility	expert	quantitative	x		x
Gäde et al.[39]	2014	Europeana	log file analysis	usage statistics & patterns	expert	quantitative	x		

29 The table is sorted by surname of the first author.

30 Ceccarelli, Diego, Sergiu Gordea, Claudio Lucchese et al. "Improving europeana search experience using query logs." In *International Conference on Theory and Practice of Digital Libraries*. Ed. by Stefan Gradmann, Francesca Borri, Carlo Meghini. 384–395. Berlin: Springer, 2011.

31 CIBER Research Ltd. Europeana *2012–2013: usage and performance update*. CIBER Research Ltd. 2013. http://ciber-research.eu/download/20130623-Europeana_2013_usage_and_performance_update.pdf (05.02.2018).

32 Clark, David, David Nicholas, Ian Rowlands. *Report on best-practice and how users are using the Europeana service* [D 3.1.3], 2011.

33 Dangerfield, Marie-Claire, Lisette Kalshoven. *Report and Recommendations from the Task Force on Metadata Quality*. 2015.

34 Dani, Anxhela, Chrysanthi Chatzopoulou, Rania Siatri et al. "Digital libraries evaluation: Measuring europeana's usability." In *Metadata and Semantics Research: 9th Research Conference, MTSR 2015, Manchester, UK, September 9–11, 2015*. Ed. by Emmanouel Garoufallou, Richard J. Hartley, Panorea Gaitanou. 225–236. Berlin: Springer, 2015.

35 Dickel, Julia. "Digitale Bibliotheken im Vergleich: Europeana & WDL." *Perspektive Bibliothek* 4,1 (2015): 45–67.

36 Dobreva, Milena, Sudatta Chowdhury. "A User-Centric Evaluation of the Europeana Digital Library." In *The Role of Digital Libraries in a Time of Global Change: 12th International Conference on Asia-Pacific Digital Libraries, ICADL 2010, Gold Coast, Australia, June 21–25, 2010. Proceedings*. Ed. by Gobinda Chowdhury, Chris Koo, Jane Hunter. 148–157. Berlin: Springer, 2010.

37 Dobreva, Milena, Emma McCulloch, Duncan Birrell et al. "Digital Natives and Specialised Digital Libraries: A Study of Europeana Users." In *Technological Convergence and Social Networks in Information Management*. Vol. 96. Ed. by Serap Kurbanoğlu, Umut Al, Phyllis Lepon Erdoğan et al. 45–60. Berlin: Springer, 2010.

38 Gäde, Maria. *Country and language level differences in multilingual digital libraries (Dissertation)*. Berlin: Humboldt-Universität zu Berlin, Philosophische Fakultät I, 2014.

39 Gäde, Maria, Vivien Petras. "Multilingual Interface Preferences." In *Proceedings of the 5th Information Interaction in Context Symposium*. 231–234. New York: ACM, 2014.

Tab. 4 (fortgesetzt)

Paper	Year	Construct	Method	Criteria	Perspective	Type	User-centred	Interface	System-Centred
Kapidakis[40]	2012	Europeana in comparison	criteria-based evaluation	data quality	expert	quantitative			x
Király[41]	2015	Europeana component	criteria-based evaluation	data quality	expert	quantitative			x
Navarrete[42]	2016	Europeana	criteria-based evaluation	coverage, usage statistics and patterns	expert	qualitative	x	x	x
Nicholas et al.[43]	2013	Europeana	log file analysis	usage statistics & patterns, coverage, accessibility	expert	quantitative	x		x
Nicholas et al.[44]	2013	Europeana	log file analysis	usage statistics & patterns, coverage, accessibility	expert	quantitative	x		x
Olensky et al.[45]	2012	Europeana component	criteria-based evaluation	performance evaluation, data quality, error rate	expert	qualitative			x
Schweibenz[46]	2010	Europeana	usability study	user satisfaction	user	qualitative	x	x	x
Şencan[47]	2013	Europeana	criteria-based evaluation	performance evaluation	expert	qualitative			x
Stiller et al.[48]	2014	Europeana component	criteria-based evaluation	performance evaluation, coverage, error rate	expert	qualitative			x
Stiller[49]	2014	Europeana in comparison	criteria-based evaluation	accessibility	expert	qualitative	x		x

40 Kapidakis, Sarantos. "Comparing metadata quality in the Europeana context." In *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*. 25. New York: ACM, 2012.

41 Király, Péter. *A Metadata Quality Assurance Framework*. Göttingen: Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, 2015.

42 Navarrete, Trilce. *Europeana as online cultural information service: Study Report*. 2016.

43 Nicholas, David, David Clark. "The second digital transition: to the mobile space – an analysis of Europeana." *Learned Publishing* 26,4 (2013): 240–252.

44 Nicholas, David, David Clark, Ian Rowlands et al. "Information on the Go: A Case Study of Europeana Mobile Users." *Journal of the American Society for Information Science and Technology* 64,7 (2013): 1311–1322.

45 Olensky, Marlies, Juliane Stiller, Evelyn Dröge. "Poisonous India or the importance of a semantic and multilingual enrichment strategy." In *Metadata and Semantics Research: 6th Research Conference,*

MTSR 2012, Cadiz, Spain, November 28–30, 2012. Ed. by Juan Manuel Dodero, Manuel Palomo-Duarte, Pythagoras Karampiperis. 252–263. Berlin: Springer, 2012.

46 Schweibenz, Werner. „Eine erste Evaluation der Europeana: Wie Benutzer das Look & Feel des Prototypen der Europäischen Digitalen Bibliothek beurteilen." *Information – Wissenschaft und Praxis* 61,5 (2010): 277–284.

47 Şencan, İpek. "The Performance Evaluation of the Information Retrieval System of the Europeana Website." In *Beyond the Cloud: Information...Innovation...Collaboration...* 59–60. 2013.

48 Stiller, Juliane, Marlies Olensky, Vivien Petras. "A Framework for the Evaluation of Automatic Metadata Enrichments." In *Metadata and Semantics Research: 8th Research Conference, MTSR 2014, Karlsruhe, Germany, November 27–29, 2014. Proceedings*. Ed. by Sissi Closs, Rudi Studer, Emmanouel Garoufallou et al. 238–249. Berlin: Springer, 2014.

49 Stiller, Juliane. *From curation to collaboration: a framework for interactions in cultural heritage information systems*. Berlin: Humboldt-Universität zu Berlin, Philosophische Fakultät I, 2014.

Tab. 4 (fortgesetzt)

Paper	Year	Construct	Method	Criteria	Perspective	Type	User-centred	Interface	System-Centred
Stiller et al.[50]	2015	Europeana in comparison	criteria-based evaluation	accessibility	expert	qualitative	x		x
Stiller et al.[51]	2013	Europeana component	criteria-based evaluation	accessibility	expert	qualitative	x	x	x
Stiller et al.[52]	2014	Europeana component	criteria-based evaluation	data quality, performance evaluation	expert	qualitative			x
Stiller et al.[53]	2014	Europeana component	criteria-based evaluation	data quality, performance evaluation	expert	qualitative			x
Sykes et al.[54]	2010	Europeana	criteria-based evaluation	usage statistics and patterns	user	qualitative	x	x	
Valtysson et al.[55]	2012	Europeana	impact study	impact criteria	expert	qualitative	x		
Van den Akker et al.[56]	2013	Europeana in comparison	criteria-based evaluation	accessibility	expert	qualitative	x		x
Yankova et al.[57]	2015	Europeana	impact study	impact criteria	expert	qualitative	x		
Charles et al.[58]	2017	Europeana component	criteria-based evaluation	coverage, data quality	expert	quantitative	x		x
Stiller et al.[59]	2017	Europeana component	criteria-based evaluation	coverage, data quality	expert	quantitative			x
Gaona-Garcia et al.[60]	2017	Europeana	criteria-based evaluation	data quality	expert	quantitative	x		x

50 Stiller, Juliane, Vivien Petras. "A framework for classifying and comparing interactions in cultural heritage information systems." In *Cultural Heritage Information: Access and management*. Ed. by Ian Ruthven, Gobinda Chowdhury. 153–176. London: Facet Publishing, 2015.

51 Stiller, Juliane, Maria Gäde, Vivien Petras. "Multilingual access to digital libraries: The Europeana use case." *Information – Wissenschaft und Praxis*. 64,2–3 (2013): 86–95.

52 Stiller, Juliane, Antoine Isaac, Vivien Petras. *EuropeanaTech Task Force on a Multilingual and Semantic Enrichment Strategy: final report*. 2014.

53 Stiller, Juliane, Vivien Petras, Maria Gäde et al. "Automatic Enrichments with Controlled Vocabularies in Europeana: Challenges and Consequences." In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*. Ed. by Marinos Ioannides, Nadia Magnenat-Thalmann, Eleanor Fink et al. 238–247. Berlin: Springer, 2014.

54 Sykes, Jonathan, Milena Dobreva, Duncan Birrell et al. "A new focus on end users: Eye-tracking analysis for digital libraries." In *Research and Advanced Technology for Digital Libraries: 14th European Conference, ECDL 2010, Glasgow, UK, September 6–10, 2010*. Ed. by Mounia Lalmas, Joemon Jose, Andreas Rauber et al. 510–513. Berlin: Springer, 2010.

55 Valtysson, Bjarki. "Europeana – The digital construction of Europe's collective memory." *Information, Communication & Society* 15,2 (2012): 151–170.

56 van den Akker, Chiel, Ardjan van Nuland, Lourens van der Meij et al. "From Information Delivery to Interpretation Support: Evaluating Cultural Heritage Access on the Web." In *Proceedings of the 5th Annual ACM Web Science Conference*. 431–440. New York: ACM, 2013.

57 Yankova, Ivanka, Evgeni Velev, Kamelia Nusheva et al. "The European Digital Library-Factor for Long-life learning in Arts and Cultural Studies." *Qualitative & Quantitative Methods in Libraries* 4,4 (2015): 965–971.

58 Charles, Valentine, Juliane Stiller, Péter Király et al. "Data Quality Assessment in Europeana: Metrics for Multilinguality." In *Workshop on (meta-)data quality at 21st International Conference on Theory and Practice of Digital Libraries*. 2017.

59 Stiller, Juliane, Péter Király. "Multilinguality of Metadata. Measuring the Multilingual Degree of Europeana's Metadata." In *ISI 2017 – 15th International Symposium of Information Science*. 164–176. 2017.

60 Gaona-Garcia, Paulo Alonso, Salvador Sanchez-Alonso et al. "Visual analytics of Europeana digital library for reuse in learning environments: A premier systematic study." *Online Information Review* 41,6 (2017): 840–859.

The framework developed by Saracevic allows us to identify gaps in evaluation and perspectives that might have not been taken into account in the assessments. Given the pool of 31 studies we looked at, we could identify the following gaps in evaluations – or the documentation thereof – applied to Europeana so are:

– The user-centred perspective is often chosen in the assessments and the individual perspective is presented in the majority of studies, showing a lack of other perspectives.
– There are no studies looking at the institutional context of Europeana, and only two looking at the societal perspective and the value of the service for the society.
– Evaluation criteria are not described well enough to allow a repetition of the evaluation on updated Europeana components or the re-use of the evaluation results for comparisons.
– Re-use of results or data for evaluation only happens within the same research group, probably also for lack of documentation.
– The methodologies applied are often insufficiently described, making it difficult to compare evaluations even by methodology.

A Call for an Evaluation Archive

This analysis of over 10 years of published evaluation studies of Europeana provides an overview over the variety of evaluated components, applied criteria, methods and perspectives. At the same time, the lack of detail in these studies frustrates as evaluation results are difficult to compare and evaluations can hardly be repeated with the same experimental design. But isn't it only the continuous cycle of evaluations, which will show progress in comparison to previous versions of a component or service? And isn't it only the standardised experimental designs, methods and measures that allow a validated comparison between different DL versions or even different DLs? This initiative to categorise evaluations for Europeana has demonstrated that a more considerable effort should be invested into an evaluation archive, which would allow progress to be tracked. This is not only true for Europeana, but for all large-scale DL projects where an institutional memory needs to survive past individuals. An even grander vision would document evaluation studies across DL in order to drive the standardisation of this research area. First efforts are underway: for example, the RePast Repository of Assigned Search Tasks[61], the DIRECT portal for information retrieval evaluation campaign data[62] or a planned workshop on the re-use of interactive information retrieval resources[63]. However, as cross-organisational initiatives take time, we recommend the development of an evaluation archive even for an individual DL. The elements from Saracevic's framework can serve as a first structural organisation for such an envisioned repository, however, we learned that each element needs to be documented in more detail in order to allow for comparisons over time and over evaluation component. Such a repository would also aggregate the research data and relevant analysis results in one place, providing all the necessary ingredients to support both the comparison between studies, but also the design and planning for new evaluations as parallel work can be avoided and earlier mistakes rectified. An evaluation archive serves all stakeholders of a DL, not just the developers, as developments and progress can be tracked over time and gaps in the evaluation coverage for a DL can be identified. Now that the data modelling and information architecture of DLs have become standardised, it is high time that evaluation efforts follow suit.

Autoreninformationen

Juliane Stiller
Berlin School of Library and Information Science
Humboldt-Universität zu Berlin
Dorotheenstr. 26
10117 Berlin
juliane.stiller@ibi.hu-berlin.de
http://orcid.org/0000-0001-8184-6187

Vivien Petras
Berlin School of Library and Information Science
Humboldt-Universität zu Berlin
Dorotheenstr. 26
10117 Berlin
vivien.petras@ibi.hu-berlin.de
http://orcid.org/0000-0002-8113-1509

---

61 https://ils.unc.edu/searchtasks/index.php (05.02.2018).
62 http://direct.dei.unipd.it/ (05.02.2018).
63 http://biirrr2018.aau.dk (05.02.2018).