A Decade of Evaluating Europeana - Constructs, Contexts, Methods & Criteria

Vivien Petras[(✉)] and Juliane Stiller

Berlin School of Library and Information Science,
Humboldt-Universität zu Berlin, Dorotheenstr. 26, 10117 Berlin, Germany
{vivien.petras,juliane.stiller}@ibi.hu-berlin.de

Abstract. This meta-analysis of 41 evaluation studies of the Europeana Digital Library categorizes them by their constructs, contexts, criteria, and methodologies using Saracevic's digital library evaluation framework. The analysis shows that system-centered evaluations prevail over user-centered evaluations and evaluations from a societal or institutional perspective are missing. The study reveals, which Europeana components have received focused attention in the last decade (e.g. the metadata) and can serve as a reference for identifying gaps, selecting methodologies and re-using data for future evaluations.

Keywords: Digital library evaluation · Evaluation constructs · Europeana · Meta-analysis

1 Introduction

Almost a decade ago, Europeana, the European digital library, museum and archive was launched [43]. It was certainly not the first digital library (DL) in existence; by that time, the DL field had been well established[1]. Today, however, Europeana belongs to an elite group of DLs that has not only managed to go beyond the prototype stage of a research project, but has achieved exemplary status for other DLs. Europeana maintains a trailblazer role in metadata modelling, licensing, aggregating large and heterogeneous volumes of content and providing multilingual access to its collections - at least in its domain of cultural heritage information.

Europeana is an ecosystem of different stakeholders, collections, usage scenarios and services with the web-based portal as its primary access point to cultural heritage material. Having been part of its development from its first steps as the European Digital Library Network (EDLnet), we reflect on what progress has been achieved in this almost decade of development. Our particular lens of analysis in this paper is evaluation. Through the evaluation of a DL, we identify its important components, its strengths and its weaknesses. By applying

---

[1] Already twenty years ago, the first European Conference on Research and Advanced Technology for Digital Libraries (ECDL, now TPDL) was held in Pisa, Italy [39].

Tefko Saracevic's framework of DL evaluation [44], we use its structure-giving dimensions to inform a discussion on future developments in Europeana. Our study reviews 41 different evaluations of Europeana, from general surveys of user motivations and usability evaluations of the portal to specific evaluations of system components or the content.

The paper is structured as follows: Sect. 2 briefly reviews some of the frameworks for DL evaluation that were considered for this meta-analysis and then describes Saracevic's framework in more detail. Section 3 describes Europeana and some of the challenges for its development. Section 4 presents the analysis of Europeana evaluations with focus points that have been identified as neglected in previous evaluations. In Sect. 5, we conclude with more recommendations for future development.

2   Digital Library Evaluation Frameworks

From the mid-1990s to the early 2000s, the large NSF-funded DL initiatives dominated DL research in the US. Large evaluation initiatives were developed: the one described by Marchionini for Perseus is a quintessential example [33]. Researchers at Rutgers university, led by Saracevic [44], summarized and aggregated evaluation aspects in their evaluation frameworks [29,59,60,63]. The 5S research team at the University of Virginia [17,23] also developed models for the evaluation of DLs [24] and presented an automatic approach for the assessment of DL components [34].

Concurrently, the EU-funded DELOS Network of Excellence on Digital Libraries developed not only its reference model for DLs [6], but created an evaluation framework for it as well. A first categorization lists the three major parameters of DLs that need to be considered in evaluation: the data/collection, the technology and the users and uses [18]. The Interaction Triptych Evaluation Model developed by DELOS [19] refines and renames these components to content, system and users. It defines three axes of evaluation between these components: usability refers to the quality of interactions between the system and the users, usefulness to the relationship between content and users and performance to the relationship between system and content. It also describes criteria and methodologies to evaluate the relationships between the components of a DL. The Digital Library Evaluation Ontology DiLEO [57] formally models the strategic and procedural elements of evaluation efforts, integrating different frameworks.

The most recent evaluation framework was developed in the Multifaceted Evaluation of Digital Libraries (MEDaL) study [61], which reviewed 85 papers and 5 project websites and performed a two-round Delphi study to identify ten dimensions for DL evaluation, also describing evaluation objectives, criteria and measures.

Saracevic's Digital Library Evaluation Framework
Many evaluation frameworks, including DELOS and MEDaL, base their concepts on the dimensions provided by Tefko Saracevic's DL evaluation framework [44],

which is also used in this study. In the framework, five elements, which are needed to describe a DL evaluation, are defined:

1. **Construct:** What is evaluated? Describes the aspect, which is the focus of the evaluation, for example the metadata or the search functionality.
2. **Context:** Which perspective is used for the evaluation? Saracevic distinguishes the user-centered perspective (with social, institutional or individual levels), the interface perspective and the system-centered perspective (with engineering, process and content levels).
3. **Criteria:** Which objective is evaluated? Saracevic names library criteria such as information accuracy, information retrieval criteria such as relevance, and HCI and interface criteria such as usability.
4. **Measures:** How are the criteria evaluated? Defines the operationalization of a criterion, e.g. precision for the evaluation of relevance.
5. **Methodology:** Which approach, instrument or tool is used for data collection and analysis for the evaluation?

Saracevic applied his framework to review 80 evaluation studies of DLs [45]. He finds that the system-centered approach is used more often than the human- or usability-centered approach as the context and that the most often evaluated criteria were usability, system performance and usage. Surveys were the methodology most often used in the evaluations, followed by structured interviews, focus groups, observations and task accomplishment. Saracevic does not discuss measures, which we will also not analyze here, because they are usually very specific to a particular evaluation.

Using their DiLEO ontology, Tsakonas et al. analyzed ca. 220 evaluation studies published between 2001–2011 in the JCDL and ECDL/TPDL conferences [56]. They found that system-centered contexts are employed most often with effectiveness, performance measurement and technical excellence as the main criteria. The most frequently used methodologies were laboratory experiments and surveys.

This paper is probably closest to these analyses, but focuses on evaluation studies of just one DL - Europeana. Zooming in on a particular DL should allow to compare evaluation results, but this is not as simple, as is discussed in the conclusion to this paper.

3 A Meta-Analysis of Europeana Evaluations

3.1 Europeana

Europeana[2] is the DL for accessing Europe's cultural and scientific heritage. Originally developed as an answer to the Google Books project [43], it has evolved into a DL, which aggregates and organizes European digital heritage in its many manifestations. Europeana's slogan "we transform the world with

---

[2] http://www.europeana.eu.

culture"[3] encompasses the vision of a network of all stakeholders in the cultural heritage sector.

The development of Europeana started in 2007 with the conceptualizing work of the European Digital Library network EDLnet, which culminated in the public launch of the Europeana portal in November of 2008 [43]. The Europeana portal now provides a single access point to over 53 million cultural heritage objects from over 3200 institutions across Europe. While Europeana also makes its data available via API access points, the portal is its most visible representation.

Europeana's challenges derive from the heterogeneity and multilinguality of its content, its providers and its audiences [27,42], making data quality, data openness and value creation its biggest priorities[3]. In almost a decade, Europeana has undergone several large-scale developmental steps - from changes in the layout and design of the portal to the modelling of the content and its functionalities. The following analysis describes the evaluation efforts accompanying these changes.

3.2   Methodology

To accumulate relevant studies, we started with a list of publications created by the Europeana Task Force for Enrichment and Evaluation that aggregates evaluations in the Europeana community [28, Appendix B]. Additionally, we searched for documents in Google Scholar and Web of Science[4], which focused solely on Europeana or used Europeana in comparison for an evaluation. The date range for the selected studies is between the launch in November 2008 until early 2017. The collected sample includes some deliverables from various Europeana satellite projects named in the Task Force document. However, we did not review all 50 projects listed on Europeana websites[5] systematically to capture evaluation efforts that were not published in a journal or conference venue, assuming that the most important ones were included in the Task Force list. Based on the abstracts of the result sets, we extracted 55 papers, which we then reviewed in detail. The criteria for extraction was that the paper focused on a evaluation of Europeana. We found three different types of evaluations:

– 38 evaluations with Europeana as the object,
– 3 evaluations using Europeana data, and
– 14 meta-studies, which named Europeana as a use case.

For the detailed analysis, we looked at 41 publications, dating from 2010–2016, which conducted an evaluation with Europeana itself or the data from Europeana. In the sample, there are several publications, which use data from the same study, but describe different aspects or different results. For methodological reasons, we counted these separately in order to reflect their different *Constructs* or *Contexts* of evaluation.

---

[3] http://strategy2020.europeana.eu/.
[4] Search terms: Europeana and (user* or evaluat* or study*) and variations.
[5] http://pro.europeana.eu/get-involved/projects/project-list.

For each of the 41 publications, we extracted information related to Sarace-vic's five elements of DL evaluation: *Construct, Context, Criteria, Measures, Methodology*. The extraction was done as close to the source as possible. Next, we followed a grounded theory approach [22] and discussed the extractions for each element to cluster the information into groups and determine categories. For example, different evaluation objects within Europeana such as metadata, enrichments or multilingual features were subsumed under the *Construct* cate-gory "Europeana component". For the *Context*, Saracevic's suggested perspec-tives were applied, dividing this element into a user-centered, system-centered and interface perspective and their subcategories (cf. Table 2).

4 Framing Europeana Evaluations

4.1 Constructs and Contexts of the Evaluations

We identified five different *Constructs* of Europeana that were evaluated (see Table 1). Most studies concentrate on the Europeana portal in general or one of its components (particularly various data quality aspects such as metadata com-pleteness or the effectiveness of automatic vocabulary enrichments). Due to the previous Europeana funding structure based on satellite projects that deliver components, a number of studies describe services, which were planned to be integrated with Europeana, but were evaluated separately from the main por-tal. Because they base their assumptions and criteria on Europeana's objectives, they were also included. Similarly, information retrieval evaluations on Euro-peana data also utilized Europeana's assumptions on user objectives and their information needs.

Table 1. Constructs used in the evaluations

Construct	Description	Number
Europeana DL	Evaluations of the portal or services overall	17
Europeana component	Studies focusing on an aspect of Europeana, e.g. metadata quality	7
External service	Evaluations of services developed for Europeana, often as part of a project, e.g. PATHS[a]	9
Algorithms	Studies using Europeana data to conduct evaluations of (search) algorithms	3
Europeana in comparison	Evaluations on DL, which compared Europeana to similar services and DLs	5

[a]http://www.paths-project.eu/.

Figure 1 shows the number of studies and their evaluation perspective, i.e. *Context*. The Venn diagram shows which *Contexts* studies have in common: 17 studies are purely system-centered, 5 purely user-centered, whereas 7 encompass all *Contexts* in the same study.
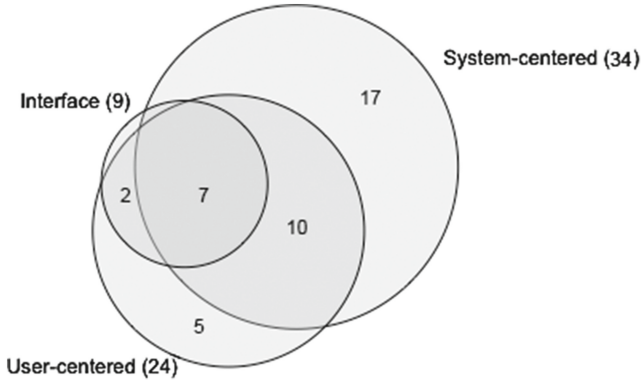
Fig. 1. Overlap of Contexts in different studies

Table 2 associates the five identified *Constructs* with their respective *Contexts*. The bold figures represent the overall number of studies for the particular *Construct* and *Context*, namely user-centered, interface or system-centered. The numbers in brackets indicate the reference. Note that the same publication can contain results that evaluate from a user-centered and a system-centered perspective, consequently, studies appear more than once in the matrix.

Equivalent to other comparative studies [45,56], system-centered evaluations are more often attempted than user-centered or interface-focused studies. It is logical that the interface is mainly evaluated when looking at the *Construct* of the portal as a whole, not when evaluating external services or components, which may not even be visible in the interface. The few algorithm studies are all system-centered, which is to be expected. Zooming in on the levels within the individual perspectives, a lack of user-centered evaluations on the social or institutional levels is apparent. Since many institutions both contribute and use Europeana, it is puzzling that we found no studies that evaluate the use of Europeana in an institutional setting. We hypothesize that the value of Europeana on a social, i.e. societal, level is "determined" through non-publication channels, for example the continued political and financial support by the EU. It is not surprising that many evaluations in our sample deal with data quality (categorized under the content *Context*) as this is an ongoing issue for Europeana. While processes are also in the focus of evaluations, we could not find any evaluations of engineering aspects. This is also not surprising as they are usually not published.

External services and Europeana components are evaluated much more from the system-centered perspective than a user-centered perspective. This may point to a lack in user-centered quality assessments for these *Constructs*. However, another explanation may be that components and services can only show their value to a user if they are integrated with other functionalities and thus cannot be evaluated individually from the user-centered perspective.

Table 2. Contexts used in the evaluations (categorized by construct)

	Europeana DL	External service	Europeana component	Algorithms	Europeana in comparison
# of Studies	17	9	7	3	5
User-centered:	**15**	**4**	**1**	**0**	**4**
Social	2 [58, 62]	0	0	0	0
Institutional	0	0	0	0	0
Individual	14 [7–9, 13, 15, 16, 20, 21, 35–37, 48, 55, 58]	4 [5, 26, 46, 47]	1 [49]	0	4 [2, 14, 50, 53]
Interface	**6**	**2**	**1**	**0**	**0**
	[13, 15, 16, 35, 48, 55]	[26, 47]	[49]		
System-centered:	**11**	**8**	**7**	**3**	**5**
Engineering	0	0	0	0	0
Process	7 [8, 9, 11, 20, 25, 36, 37]	8 [3, 4, 10, 26, 28, 32, 46, 47]	5 [12, 38, 51, 52, 54]	3 [1, 40, 41]	0
Content	10 [8, 9, 11, 13, 15, 20, 35–37, 48]	4 [26, 28, 32, 46]	7 [12, 31, 38, 49, 51, 52, 54]	0	5 [2, 14, 30, 50, 53]

4.2   Criteria and Methodologies of the Evaluations

We found that almost every reported evaluation defined its own different *Criteria* (objectives of analysis). Following, we list broad *Criteria* from each *Context* whereas for each criteria example studies are given. Usability and the effectiveness of the interaction design, user behavior and algorithm performance were used as objectives more often than others, but many *Criteria* are employed in Europeana evaluations.

User-Centered Context

– impact of Europeana on society and education [58, 62]
– value of Europeana services wrt. mutuality[6], usability & reliability [35]
– value of multilingual services [20, 21, 49]
– usability & effectiveness of interaction design [5, 13, 15, 26, 46, 47]
– effectiveness of search functionalities [7]
– usage patterns and criteria of Human Computer Interactions [8, 9, 20, 36, 37], such as task completion and time performance
– behaviour of particular user groups [16]
– value of engagement and access features [50, 53].

---

[6] The study defines this term as a criteria to act as multiplier for member institutions.

System-Centered Context

- metadata quality [12,30,31]
- impact of semantic enrichments [38,52,54] and components of workflow [51]
- performance of item similarity algorithms [3,4,10,25,26]
- content characteristics compared to other DLs [2,14]
- usage of particular content [8,9,36,37]
- accessibility of content [2]
- information retrieval criteria, e.g. precision of search results [1,11,40,41,46]
- performance of enrichment tools [32].

Interface Context

- user foci in interface [13,55] and look and feel [48].

Table 3 categorizes the *Methodologies* used in the different evaluations. Most studies first defined their own *Criteria* and then assessed the quality of a component or service based on these *Criteria*. These are usually expert assessments that are performed on quantitative or qualitative data. A gold standard is commonly used for algorithm testing or the quality of a service where a particular outcome can be expected, for example, automatic enrichment of keywords in the metadata. Criteria-based, gold standard-based evaluations often assess the DL from the system-centered perspective. Logfile analyses are used for both system- and user-centered evaluations. User-centered methodologies such as usability studies usually take more effort and occur not as often.

Table 3. Methodologies used in the evaluations (multiple possible)

Method	Description	Number
Criteria-based	Certain criteria were determined to assess a service or algorithm	16
Gold standard-based	Use of a manually created gold standard to assess performance	9
Logfile analysis	Uses an automatically created logfile of user interactions	8
Usability study	Several methods to assess usability of a service, e.g. user studies, interviews, surveys	7
Impact study	Expert assessment of the overall value of a service	2

While it may be challenging to identify *Criteria* that have not been evaluated for Europeana, we argue that *Criteria* and *Methodologies* to rate the overall impact and added value of Europeana are still missing. Europeana has recognized this gap and is about to implement an Impact Framework.[7] Evaluations targeting aspects such as user satisfaction or performance of algorithms refer to established methods and criteria. Defining success criteria specifically for cultural heritage DLs seems to be a gap that should receive more attention in future evaluations.

---

[7] More information can be found here: https://impkt.tools/.

5   Conclusions

The meta-analysis showed that Europeana has been evaluated from many perspectives with a vast number of *Criteria* and many different *Methodologies*. The categorization of evaluation *Constructs* and *Contexts* showed that societal and institutional perspectives appear less often than system-centered quantitative approaches and that only a marginal number of studies tried to assess the impact of Europeana on different stakeholders. Concluding, we argue for more standardization with regard to evaluations of large-scale DLs and how the evaluation process can be improved.

Establish an Evaluation Archive
Our analysis showed that evaluations usually evolve individually and rarely refer to previous results or to similar efforts in the Europeana ecosystem. Logging data and other evaluation outcomes are rarely re-used, except within the same research groups. The lack of published documentation and coordination between evaluations calls for a more concerted effort. Europeana has been the target of over 50 satellite projects, which probably included more unpublished evaluation efforts. Europeana as an ecosystem needs an evaluation archive, which helps to build a common memory in the community and promotes learning from past results.

Track Improvements Over Time
We found that Europeana development has progressed due to evaluations, although implementations cannot be traced back to certain evaluation results. In general, more coordination and documentation is needed to learn from previous experiences and also track improvements over time. Europeana has recognized the importance of evaluation by integrating permanent activities into their ecosystem, for example by integrating a logging framework in the portal to understand user interactions better. The Europeana Statistics Dashboard[8] provides current interaction log- and content-based statistics for providers and users, while the Europeana Data Quality Committee[9] develops standards to continuously improve the data quality. System development should be traced alongside evaluation efforts.

Standardize Evaluations
The heterogeneity of methodologies and criteria used makes it hard to draw concrete solutions for Europeana development in general and aggravates the reuse of the data. Here, a more standardized format is desirable that allows to compare results over time. However, it is important to note that the evaluations, which created gold standards from Europeana data (for example [40]) or developed framework or experimental set-ups for evaluation (for example [32,52]) can be re-used in other evaluation efforts. This means that Europeana evaluations have pushed the envelope further and also contributed to DL evaluation research in general.

---

[8] http://statistics.europeana.eu/.
[9] http://pro.europeana.eu/page/data-quality-committee.

With evaluation as an integral part of development in Europeana, these efforts can hopefully be better organized so that new services and partners can learn from the large experience that has already been accumulated.

References

1. Akasereh, M., Malak, P., Pawłowski, A.: Evaluation of IR strategies for Polish. In: Przepiórkowski, A., Ogrodniczuk, M. (eds.) NLP 2014. LNCS, vol. 8686, pp. 384–391. Springer, Cham (2014). doi:10.1007/978-3-319-10888-9_38

2. van den Akker, C., van Nuland, A., van der Meij, L., van Erp, M., Legêne, S., Aroyo, L., Schreiber, G.: From information delivery to interpretation support: evaluating cultural heritage access on the web. In: WebSci 2013, pp. 431–440. ACM (2013)

3. Aletras, N., Stevenson, M.: Computing similarity between cultural heritage items using multimodal features. In: LaTeCH 2012, pp. 85–93. ACL (2012)

4. Aletras, N., Stevenson, M., Clough, P.: Computing similarity between items in a digital library of cultural heritage. J. Comp. Cult. Heri. **5**(4), 16:1–16:19 (2012)

5. Aloia, N., Concordia, C., Gerwen, A.M., Hansen, P., Kuwahara, M., Ly, A.T., Meghini, C., Spyratos, N., Sugibuchi, T., Tanaka, Y., Yang, J., Zeni, N.: Design, implementation and evaluation of a user generated content service for Europeana. In: Gradmann, S., Borri, F., Meghini, C., Schuldt, H. (eds.) TPDL 2011. LNCS, vol. 6966, pp. 477–482. Springer, Heidelberg (2011). doi:10.1007/978-3-642-24469-8_54

6. Candela, L., Castelli, D., Ferro, et al.: The DELOS digital library reference model. Version 0.96. Technical report, DELOS Network of Excellence on Digital Libraries (2007)

7. Ceccarelli, D., Gordea, S., Lucchese, C., Nardini, F.M., Tolomei, G.: Improving Europeana search experience using query logs. In: Gradmann, S., Borri, F., Meghini, C., Schuldt, H. (eds.) TPDL 2011. LNCS, vol. 6966, pp. 384–395. Springer, Heidelberg (2011). doi:10.1007/978-3-642-24469-8_39

8. CIBER Research Ltd.: Europeana 2012–2013: usage and performance update. Technical report, CIBER Research Ltd. (2013). http://bit.ly/2rgWjrl

9. Clark, D., Nicholas, I., Rowlands, I.: D3.1.3 Report on best-practice and how users are using the Europeana service. Technical report, D 3.1.3, EuropeanaConnect (2011)

10. Clough, P., Otegi, A., Agirre, E., Hall, M.: Implementing recommendations in the PATHS system. In: Bolikowski, Ł., Casarosa, V., Goodale, P., Houssos, N., Manghi, P., Schirrwagen, J. (eds.) TPDL 2013. CCIS, vol. 416, pp. 169–173. Springer, Cham (2014). doi:10.1007/978-3-319-08425-1_17

11. Şencan, I.: The performance evaluation of the information retrieval system of the Europeana website. In: IMCW 2013, p. 59 (2013)

12. Dangerfield, M.C., Kalshoven, L.: Report and recommendations from the task force on metadata quality. Technical report, Europeana (2015). http://bit.ly/2rM113M

13. Dani, A., Chatzopoulou, C., Siatri, R., Mystakopoulos, F., Antonopoulou, S., Katrinaki, E., Garoufallou, E.: Digital libraries evaluation: measuring Europeana's usability. In: Garoufallou, E., Hartley, R.J., Gaitanou, P. (eds.) MTSR 2015. CCIS, vol. 544, pp. 225–236. Springer, Cham (2015). doi:10.1007/978-3-319-24129-6_20

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

15. Dobreva, M., Chowdhury, S.: A user-centric evaluation of the Europeana digital library. In: Chowdhury, G., Koo, C., Hunter, J. (eds.) ICADL 2010. LNCS, vol. 6102, pp. 148–157. Springer, Heidelberg (2010). doi:10.1007/978-3-642-13654-2_19

16. Dobreva, M., McCulloch, E., Birrell, D., Ünal, Y., Feliciati, P.: Digital natives and specialised digital libraries: a study of Europeana users. In: Kurbanoğlu, S., Al, U., Lepon Erdoğan, P., Tonta, Y., Uçak, N. (eds.) IMCW 2010. CCIS, vol. 96, pp. 45–60. Springer, Heidelberg (2010). doi:10.1007/978-3-642-16032-5_5

17. Fox, E.A., Gonçalves, M.A., Shen, R.: Theoretical foundations for digital libraries: the 5s (societies, scenarios, spaces, structures, streams) approach. Synth. Lect. Inform. Concepts, Retrieval, Serv. **4**(2), 1–180 (2012)

18. Fuhr, N., Hansen, P., Mabe, M., Micsik, A., Sølvberg, I.: Digital libraries: a generic classification and evaluation scheme. In: Constantopoulos, P., Sølvberg, I.T. (eds.) ECDL 2001. LNCS, vol. 2163, pp. 187–199. Springer, Heidelberg (2001). doi:10.1007/3-540-44796-2_17

19. Fuhr, N., Tsakonas, G., Aalberg, T., et al.: Evaluation of digital libraries. IJDL **8**(1), 21–38 (2007)

20. Gäde, M.: Country and language level differences in multilingual digital libraries. Dissertation, Humboldt-Universität zu Berlin, April 2014

21. Gäde, M., Petras, V.: Multilingual interface preferences. In: IIiX 2014, pp. 231–234. ACM (2014)

22. Glaser, B.G., Strauss, A.: The discovery of ground theory (1967)

23. Gonçalves, M.A., Fox, E.A., Watson, L.T., Kipp, N.A.: Streams, structures, spaces, scenarios, societies (5s): a formal model for digital libraries. ACM TOIS **22**(2), 270–312 (2004)

24. Gonçalves, M.A., Moreira, B.L., Fox, E.A., Watson, L.T.: "What is a good digital library?" - a quality model for digital libraries. IPM **43**(5), 1416–1437 (2007)

25. Gonzalez-Agirre, A., Rigau, G., Agirre, E., Aletras, N., Stevenson, M.: Why are these similar? Investigating item similarity types in a large digital library. JASIST **67**(7), 1624–1638 (2016)

26. Griffiths, J., et al.: D5.2 evaluation of the second PATHS prototype. Technical report (2014). http://bit.ly/2rh3hwJ

27. Hill, T., Charles, V., Stiller, J., Isaac, A.: Searching for inspiration: user needs and search architecture in Europeana collections. In: ASIS&T, pp. 1–7. Wiley (2016)

28. Isaac, A., Manguinhas, H., Stiller, J., Charles, V.: Final report on evaluation and enrichment. Technical report, Europeana (2015). http://bit.ly/2rcKuHr

29. Joo, S., Xie, I.: Evaluation constructs and criteria for digital libraries: a document analysis. In: Recent Developments in the Design, Construction, and Evaluation of Digital Libraries: Case Studies, pp. 126–140. IGI Global (2013)

30. Kapidakis, S.: Comparing metadata quality in the Europeana context. In: PETRA 2012, p. 25:1–25:8. ACM (2012)

31. Király, P.: A metadata quality assurance framework. Technical report, Europeana (2015). https://pkiraly.github.io/metadata-quality-project-plan.pdf

32. Manguinhas, H., Freire, N., Isaac, A., Stiller, J., Charles, V., Soroa, A., Simon, R., Alexiev, V.: Exploring comparative evaluation of semantic enrichment tools for cultural heritage metadata. In: Fuhr, N., Kovács, L., Risse, T., Nejdl, W. (eds.) TPDL 2016. LNCS, vol. 9819, pp. 266–278. Springer, Cham (2016). doi:10.1007/978-3-319-43997-6_21

33. Marchionini, G.: Evaluating digital libraries: a longitudinal and multifaceted view. Libr. Trends **49**(2), 304 (2000)

34. Moreira, B.L., Gonçalves, M.A., Laender, A.H., Fox, E.A.: Automatic evaluation of digital libraries with 5squal. J. Informetrics **3**(2), 102–123 (2009)

35. Navarrete, T.: Europeana as online cultural information service. Study Report (2016). http://bit.ly/2rcMsaT

36. Nicholas, D., Clark, D.: The second digital transition: to the mobile space - an analysis of Europeana. Learn. Publishing **26**(4), 240–252 (2013)

37. Nicholas, D., Clark, D., Rowlands, I., Jamali, H.R.: Information on the Go: a case study of Europeana mobile users. JASIST **64**(7), 1311–1322 (2013)

38. Olensky, M., Stiller, J., Dröge, E.: Poisonous India or the importance of a semantic and multilingual enrichment strategy. In: Dodero, J.M., Palomo-Duarte, M., Karampiperis, P. (eds.) MTSR 2012. CCIS, vol. 343, pp. 252–263. Springer, Heidelberg (2012). doi:10.1007/978-3-642-35233-1_25

39. Peters, C., Thanos, C. (eds.): ECDL 1997. LNCS, vol. 1324. Springer, Heidelberg (1997)

40. Petras, V., Bogers, T., Toms, E., Hall, M., Savoy, J., Malak, P., Pawłowski, A., Ferro, N., Masiero, I.: Cultural heritage in CLEF (CHiC) 2013. In: Forner, P., Müller, H., Paredes, R., Rosso, P., Stein, B. (eds.) CLEF 2013. LNCS, vol. 8138, pp. 192–211. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40802-1_23

41. Petras, V., Ferro, N., Gäde, M., Isaac, A., Kleineberg, M., Masiero, I., Nicchio, M., Stiller, J.: Cultural Heritage in CLEF (CHiC) Overview 2012. In: CLEF (2012)

42. Petras, V., Hill, T., Stiller, J., Gäde, M.: Europeana - a search engine for digitised cultural heritage material. Datenbank-Spektrum **17**(1), 41–46 (2017)

43. Purday, J.: Think culture: Europeana.eu from concept to construction. Electron. Libr. **27**(6), 919–937 (2009)

44. Saracevic, T.: Digital library evaluation: toward an evolution of concepts. Libr. Trends **49**(2), 350–369 (2000)

45. Saracevic, T.: Evaluation of digital libraries: an overview. In: Notes of the DELOS WP7 Workshop on the Evaluation of Digital Libraries, Padua, Italy (2004)

46. Schindler, A., Gordea, S., Biessum, H.: The Europeana sounds music information retrieval pilot. In: Ioannides, M., Fink, E., Moropoulou, A., Hagedorn-Saupe, M., Fresa, A., Liestøl, G., Rajcic, V., Grussenmeyer, P. (eds.) EuroMed 2016. LNCS, vol. 10059, pp. 109–117. Springer, Cham (2016). doi:10.1007/978-3-319-48974-2_13

47. Schlötterer, J., Seifert, C., Wagner, L., Granitzer, M.: A game with a purpose to access europe's cultural treasure. ECIR **2015**, 13–18 (2015)

48. Schweibenz, W.: Eine erste evaluation der Europeana: Wie Benutzer das look & feel des prototypen der Europäischen digitalen bibliothek beurteilen. Inf. Wissenschaft und Praxis **61**(5), 277–284 (2010)

49. Stiller, J., Gäde, M., Petras, V.: Multilingual access to digital libraries: the Europeana use case. Inf.-Wissenschaft und Praxis **64**, 86–95 (2013)

50. Stiller, J.: From curation to collaboration. Dissertation, Humboldt-Universität zu Berlin, Philosophische Fakultät I (2014)

51. Stiller, J., Isaac, A., Petras, V.: EuropeanaTech task force on a multilingual and semantic enrichment strategy. Technical report, Europeana (2014). http://bit.ly/2soSHs7

52. Stiller, J., Olensky, M., Petras, V.: A framework for the evaluation of automatic metadata enrichments. In: Closs, S., Studer, R., Garoufallou, E., Sicilia, M.-A. (eds.) MTSR 2014. CCIS, vol. 478, pp. 238–249. Springer, Cham (2014). doi:10.1007/978-3-319-13674-5_23

53. Stiller, J., Petras, V.: A framework for classifying and comparing interactions in cultural heritage information systems. In: Cultural Heritage Information: Access and management. Facet Publishing (2015)

54. Stiller, J., Petras, V., Gäde, M., Isaac, A.: Automatic enrichments with controlled vocabularies in Europeana: challenges and consequences. In: Ioannides, M., Magnenat-Thalmann, N., Fink, E., Žarnić, R., Yen, A.-Y., Quak, E. (eds.) EuroMed 2014. LNCS, vol. 8740, pp. 238–247. Springer, Cham (2014). doi:10.1007/978-3-319-13695-0_23

55. Sykes, J., Dobreva, M., Birrell, D., McCulloch, E., Ruthven, I., Ünal, Y., Feliciati, P.: A new focus on end users: eye-tracking analysis for digital libraries. In: Lalmas, M., Jose, J., Rauber, A., Sebastiani, F., Frommholz, I. (eds.) ECDL 2010. LNCS, vol. 6273, pp. 510–513. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15464-5_69

56. Tsakonas, G., Mitrelis, A., Papachristopoulos, L., Papatheodorou, C.: An exploration of the digital library evaluation literature based on an ontological representation. JASIST **64**(9), 1914–1926 (2013)

57. Tsakonas, G., Papatheodorou, C.: An ontological representation of the digital library evaluation domain. JASIST **62**(8), 1577–1593 (2011)

58. Valtysson, B.: Europeana - the digital construction of Europe's collective memory. Inf. Commun. Soc. **15**(2), 151–170 (2012)

59. Xie, H.: Evaluation of digital libraries: criteria and problems from users' perspectives. Libr. Inf. Sci. Res. **28**(3), 433–452 (2006)

60. Xie, H.I.: Users' evaluation of digital libraries (DLs): their uses, their criteria, and their assessment. IPM **44**(3), 1346–1373 (2008)

61. Xie, I., Matusiak, K.K.: Chapter 10 - evaluation of digital libraries. In: Discover Digital Libraries, pp. 281–318. Elsevier, Oxford (2016)

62. Yankova, I., Velev, E., Nusheva, K., Sonia, S.: The European digital library-factor for long-life learning in arts and cultural studies. Qual. Quant. Methods Libr. **4**(4), 965–971 (2015)

63. Zhang, Y.: Developing a holistic model for digital library evaluation. JASIST **61**(1), 88–110 (2010)