Issues for Quality Assurance of Metadata in Learning Object Repositories: The Case of Photodentro[1]

Nikos Palavitsinis[1], Elina Megalou[1]

[1]*Computer Technology Institute & Press – Diophantus (CTI), 10563, Athens, Greece*
palavitsinis@cti.gr; megalou@cti.gr

Abstract: Although digital repository projects cover numerous different domains and fields they face similar challenges. Metadata quality is one of those common challenges that defy the "raison d'être" of digital repositories. "Invisible" resources and insufficient search mechanisms are problems that could possibly be addressed through appropriate metadata. Metadata Quality Assurance (QA) mechanisms are put in place to address this problem. This paper presents such a QA approach on a learning object repository. It compares and contrasts the resulting completeness of its metadata records to the same metric of a similar repository with a similar QA approach. The aim of the paper is to gain insight in the use of specific elements in learning object repositories, aiming to serve as the basis for a user-centric, domain-specific, quality assurance process for metadata.

Keywords: learning, repository, metadata, quality, framework, element, IEEE LOM

1 Introduction

Quality problems in metadata elements in digital repositories have been apparent in studies of the last decade or so. Stvilia et al., in their study in 2004 assessed 155,000 records coming from 16 collections of academic and public librabries, museums and historical societies, showcasing lack of completeness, redundant metadata and lack of clarity. Similar cases were presented, to name a few, by Shreeves et al. (2005), Yen & Park (2006), Stvilia et al. (2007), Sanchez-

[1] In proceedings of LINQ2015: 4th International Conference on Learning Innovations and Quality, Brussels (Belgium), 11th, 12th and 13th of May, 2015

Alonso (2009) and Ochoa et al. (2011). Either through their limited use, their overuse or incorrect use metadata is one of the main areas in each digital repository project that needs to be addressed before deploying consistent search mechanisms on top of the content they host.

One of the attempts to tackle this problem has come through the involvement of domain experts of the corresponding repository discipline, in metadata design (Chen et al., 2002; Bainbridge et al., 2003; Chu et al., 2010) to allow for a better understanding and therefore use of metadata elements. Training and support of the experts with appropriate material has also been discussed extensively in relevant literature (Zhang & Dimitroff, 2005; Malaxa & Douglas, 2005; Cassella, 2010). Despite these efforts, in many cases, metadata quality remains low, in terms of the established metrics in relevant literature (Bruce & Hilman, 2004) which calls for a more focused examination of the issue. This paper presents the application of a comprehensive approach to metadata quality in a learning object repository hosting content for primary and secondary education in Greece, namely the "Photodentro LOR". First of all, our aim is to present an overview of the approach that can be deployed in other repositories. Secondly, we attempt to compare some of its results with those of the application of the same QA for a similar learning repository to deduct conclusions for the use of metadata elements.

This paper is structured as follows: Section 1 provides an introduction to the main research aims, followed by section 2 that briefly presents the background that led to this study, presenting work that are closely related to this paper. Section 3 presents the Quality Assurance methods deployed in Photodentro LOR whereas section 4 contains some preliminary quantitative results as well as the comparison with a case of a similar LOR. Section 5 draws on the main conclusions of the study, its limitations and also the road mapping of future research directions.

2 Background

Literature shows limited cases where metadata quality is addressed in the context of a repository project in a comprehensive way. The majority of studies focus on specific aspects of metadata quality, suggesting metrics of quality or measuring quality for a set of records, etc. There have been only a handful of cases where the metadata quality issue was dealt with in a more holistic way. Stvilia et al., (2004) presented a framework of metadata quality dimensions and

also used them to measure quality for almost 155,000 metadata records. Their study also offered some practical advice on how to ensure high metadata quality. Vinagre et al., (2011) presented a Library Service Quality Model, designed to evaluate digital libraries. The authors argued in favour of continuous application of their model to monitor the quality of a digital library periodically. This work showed the need for an ongoing QA process that covers all the stages in the development and operation of a digital library, a finding that was also validated from the work of Waaijers and van der Graaf (2011). Finally in similar work, Zschocke and Beniest (2011) analyzed different quality metrics for metadata and proposed a quality assurance framework that can be applied on the metadata creation process in the case of an agricultural learning repository. Similar work was also introduced in cases of educational, cultural and research repositories (Palavitsinis et al., 2014b).

The Photodentro Learning Object Repository (LOR) (http://photodentro.edu.gr/lor/) is part of the "Digital School", a large-scale program of the Greek Ministry of Education (2010-2015). Photodentro LOR is the Greek National Learning Object Repository for primary and secondary education. It aims at populating both a large but also a high quality pool of learning resources, tagged with educational metadata, open to everyone, students, teachers, parents as well as the wider public. Photodentro LOR as of January 2015 holds more than 6,500 learning objects, organized thematically based on the school curriculum. The resources are being developed by 120 project-employed, qualified teachers, in ten domain-specific workgroups, in the process of enriching Greek textbooks with digital interactive resources. Each group operated under the supervision of a coordinator, an academic with significant domain and pedagogical expertise, to ensure the quality of the learning objects. More technical details about the functionalities offered by Photodentro LOR as well as its architecture are presented in previous work (Megalou & Kaklamanis, 2014).

3 Metadata Quality Assurance in Photodentro

A sub-group of the teachers working in the project is responsible for the task of annotating the content produced with metadata and publishing it. To complete this task, a well-specified and standardized process is deployed, from the moment the coordinators of each of the collections assign resources to their

team until their final publication in Photodentro LOR. The resources are being annotated through the Photodentro backend, using the IEEE LOM application profile of Photodentro that is tailored to fit the specific needs of Photodentro LOR. The teams of domain experts/teachers are supported throughout the process with training courses as well as training material in the form of FAQs, guides and manuals. Once the resources are annotated, a final check is carried out by the coordinator of the collection making sure that the resource itself as well as the corresponding metadata is of the highest quality.The following table outlines the main QA methods that were deployed in Photodentro LOR, supporting the domain experts that annotated content.

Table 1: Quality Assurance Methods deployed in Photodentro

No	Date	QA Method	Description
1	3/2012	Training workshop for the metadata annotators	Introductory course to the IEEE LOM Photodentro Application Profile (AP)
2	3/2012	Training material on metadata annotation (v1.0)	Guide/handbook to support pilot metadata annotation
3	5/2012	Training material on metadata annotation (v1.2)	New version of the previous based on comments & changes in the AP
4	10/2013	Frequently Asked Questions (FAQ) on Metadata Annotation (v1.0)	Answers to common questions related to metadata annotation
5	10/2013	Metadata Quality Review	Sample of 50 metadata records reviewed by metadata experts feedback sent to teams
6	11/2013–7/2014	Training workshop for the metadata annotators	Dedicated Training workshops for the teams per subject (7 workshops).
7	1/2014	Metadata Completeness Check	Automated extraction of usage data per element and statistical analysis for 3,899 metadata records
8	2/2014	Metadata Authoring Tool Manual (v3.0)	How to use the metadata authoring tool, including application profile
9	4/2014	Metadata Completeness Check	Automated extraction of usage data per element and statistical analysis for 4,374 metadata records
10	6/2014	Training material on metadata annotation (v2.0)	New version of the previous based on comments
11	6/2014	Metadata Authoring Tool Manual (v3.1)	How to use the metadata authoring tool, including application profile
12	11/2014	Metadata Completeness Check	Automated extraction of usage data per element and statistical analysis for 5,150

			metadata records
13	11/2014	Metadata Quality Review	Sample of 299 metadata records, reviewed for mistakes by metadata expert and feedback sent to teams
14	11/2014	Frequently Asked Questions (FAQ) on Metadata Annotation (v2.0)	Answers to common questions related to metadata annotation
15	2/2015	Metadata Quality Review	Sample of 235 metadata records, reviewed for mistakes by metadata expert and feedback sent to teams

Overall, a total of fifteen (15) Quality Assurance methods were introduced through the various stages of the deployment and operation of the Photodentro LOR. The main types of QA methods included (a) workshops with the domain experts, (b) training material such as guides, FAQs and wikis, (c) reviews of metadata records and (d) usage data analyses. All of the different QA methods presented in Table 2, offer specific input to the metadata annotators to aid them in completing the metadata records appropriately. That is, completing them keeping in mind the envisaged use of the resources in Photodentro, as well as the limitations and characteristics of the application profile selected. The full scope of the QA methods deployed in Photodentro is too broad to be covered in the context of this paper. We only present an overview of them as well as a fragment of the quality results they contributed to.

4 Results

In this section, the results of the completeness check in Photodentro during November 2014 are presented and compared with a similar existing case coming from relevant literature. Our aim is to compare the use of metadata elements in Photodentro LOR, with another IEEE LOM-based learning repository and try to deduct useful conclusions. In Palavitsinis et al., (2014a) the completeness of IEEE LOM application was presented, after applying the same type of QA Methods on metadata annotation, to a group of domain experts with similar background to the one of the annotators of Photodentro. In Table 2, we can see all the common metadata elements that were adopted in both cases of learning repositories. Despite there being many more common elements, we decided to omit the ones that are (a) automatically completed and therefore are 100% complete (Format, Duration, Meta-Metadata, etc.), and (b) elements which are mandatory to complete in the metadata authoring tool, meaning that the metadata record

cannot be stored till they are completed (Title, Description, Rights, Classification, etc.). These elements were also 100% completed in both cases and therefore their comparison added nothing to the analysis.

Table 2: Comparison between completeness measurement in Organic.Edunet and Photodentro

	Organic.Edunet	Obligation in AP	Photodentro	Obligation in AP	Difference
1. General					
1.3 Language	99.9%	M	99.7%	R	-0.2%
1.5 Keyword	99.9%	R	100%	R	0.1%
1.6 Coverage	82.6%	R	16.7%	R	-65.9%
2. LifeCycle					
2.1 Version	18.2%	O	22.3%	O	4.1%
2.2 Status	39.7%	O	28.2%	O	-11.5%
2.3.1 Contribute Role	75%	R	100%	R	25%
2.3.2 Contribute Entity	75.8%	R	100%	R	24.2%
2.3.3 Contribute Date	62.8%	R	72.7%	O	9.9%
4. Technical					
4.4.1 Requirement	6.7%	O	97.8%	R	93.1%
5. Educational					
5.1 Interactivity Type	36.9%	O	30,8%	O	-6.1%
5.2 Learning Resource Type	3.1%	R	98.6%	R	95.5%
5.5 Intended End User Role	82.4%	R	36.9%	O	-45.5%
5.6 Context	81%	R	100%	R	19%
5.7 Typical Age Range	63.9%	R	100%	R	36.1%
5.9 Typical Learning Time	0.4%	O	11.6%	O	11.2%
5.10 Description	14.7%	R	20.2%	O	5.5%
5.11 Language	52.3%	O	31%	O	-21.3%

Overall, Photodentro LOR has an average metadata completeness that is 10% higher than Organic.Edunet. The average completeness for all the recommended elements in Organic.Edunet is 64.1% whereas in Photodentro it is 90.3% which is really high for this kind of element obligation. The percentage for the optional elements is 25.7% and 31.7% respectively, which shows a convergence in the behavior of the metadata annotators, regarding the optional elements that are common in both application profiles.

5 Conclusions

The present study outlined an overview of the main QA methods that were deployed in the Photodentro Learning Object Repository to support metadata annotation from domain experts. It continued by comparing outcomes of the combined QA methods in terms of metadata completeness with outcomes of similar processes applied in a similar learning repository, in terms of size, type of material and domain experts involved. This first discussion on similarities and differences shows that optional elements are treated similarly across the two projects whereas recommended elements were completed more in Photodentro. Fields that did not seem to be affected by their obligation and sustain similar completeness percentages were: Language, Keyword, Version, Status, Date, Interactivity Type and Educational Description.

The main limitation of this paper lies in the fact that it does not present in full length the two approaches followed in each project and therefore it does not establish a firm basis for a true comparison of the resulting completeness or other quality metrics for metadata. Nevertheless, this paper serves as an introduction to the work that will follow, and therefore all the details about QA methods were kept to a minimum. Overall, the authors feel that the outcomes offered within this work will serve as a starting point for a discussion on metadata quality in all phases of a LOR, including the post-funding phase. Future research will attempt to examine metadata elements closer, suggesting mechanisms to lower the costs and effort associated with QA methods, as these were described in the respective chapter of this paper. It would also be interesting to look at different metadata quality metrics, such as appropriateness, consistency, correctness, etc., across repository projects with similar metadata QA methods and similar content/collections. Through results of such an analysis, interesting research directions could evolve for metadata training, metadata authoring and repository management in general.

6 References

Bainbridge, D., Thompson, J. and Witten, I.H. (2003), "Assembling and enriching digital library collections", Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 323-34.

Bruce, T. R., & Hillmann, D. I. (2004). The continuum of metadata quality: defining, expressing, exploiting.

Cassella, M. (2010). Institutional repositories: an internal and external perspective on the value of IRs for researchers' communities. Liber quarterly, 20(2), 210-225.

Chen, C. C., Chen, H. H., Chen, K. H., & Hsiang, J. (2002). The design of metadata for the Digital Museum Initiative in Taiwan. Online Information Review,26(5), 295-306.

Chu, H. C., Hwang, G. J., & Tseng, J. C. (2010). An innovative approach for developing and employing electronic libraries to support context-aware ubiquitous learning. Electronic Library, The, 28(6), 873-890.

Malaxa, V., & Douglas, I. (2005). A framework for metadata creation tools. Interdisciplinary Journal of E-Learning and Learning Objects, 1(1), 151-162.

Megalou, E., & Kaklamanis, C. (2014). PHOTODENTRO LOR, THE GREEK NATIONAL LEARNING OBJECT REPOSITORY. *INTED2014 Proceedings*, 309-319.

Ochoa, X., Klerkx, J., Vandeputte, B., & Duval E. (2011). On the use of Learning Object Metadata: the GLOBE experience. Lecture Notes in Computer Science, 6964, 271-284.

Palavitsinis, N., Manouselis, N., & Sanchez-Alonso, S. (2014a). Metadata quality in learning object repositories: a case study. *The Electronic Library*, 32(1), 62-82.

Palavitsinis, N., Manouselis, N., & Sanchez-Alonso, S. (2014b). Metadata quality in digital repositories: Empirical results from the cross-domain transfer of a quality assurance process. *Journal of the Association for Information Science and Technology*, 65(6), 1202-1216.

Sánchez-Alonso, S. (2009). Enhancing availability of learning resources on organic agriculture and agroecology. The Electronic Library 27(5): 792-813.

Stvilia, B., Gasser, L., Twidale, M., Shreeves, S., & Cole, T. (2004). Metadata quality for federated collections. In Proceedings of ICIQ04 - 9th International Conference on Information Quality, (pp. 111-125). Boston, MA.

Stvilia, B., Gasser, L., Twidale, M. B., & Smith, L. C. (2007). A framework for information quality assessment. Journal of the American Society for Information Science and Technology, 58(12), 1720-1733.

Shreeves, S. L., Knutson, E. M., Stvilia, B., Palmer, C. L., Twidale, M. B., & Cole, T. W. (2005). Is' quality'metadata'shareable'metadata? The implications of local metadata practices for federated collections.

Vinagre, M. H., Pinto, L. G., & Ochôa, P. (2011). Revisiting digital libraries quality: a multiple-item scale approach. Performance Measurement and Metrics, 12(3), 214-236.

Waaijers, L., & van der Graaf, M. (2011). Quality of Research Data, an Operational Approach. D-Lib Magazine, 17(1/2).

Yen, B., & Park, J. (2006). An assessment of metadata quality: A case study of the national science digital library metadata repository. In Haidar Moukdad (ed.), CAIS/ACSI 2006 Information Science Revisited: Approaches to Innovation from http://www.cais-acsi.ca/proceedings/2006/bui_2006.pdf.

Zhang, J., & Dimitroff, A. (2005). The impact of metadata implementation on webpage visibility in search engine results (Part II). Information processing & management, 41(3), 691-715.