Exploring the Relevance of Europeana Digital Resources: Preliminary Ideas on Europeana Metadata Quality

Paulo Alonso Gaona García

Doctor en Ingeniería de la Información y del Conocimiento. Universidad de Alcalá, España. Magíster en Ciencias de la Información y de las Comunicaciones. Ingeniero de sistemas. Universidad Distrital Francisco José de Caldas. Profesor titular de la Facultad de Ingeniería. Universidad Distrital Francisco José de Caldas, Bogotá-Colombia.
pagaonag@udistrital.edu.co
orcid.org/0000-0002-8758-1412

Ana Fermoso García

Doctora en Informática y licenciada en informática. Universidad de Deusto, España. Profesora titular en la Facultad de Informática. Universidad Pontificia de Salamanca, España.
afermosoga@upsa.es
orcid.org/0000-0001-7204-4414

Salvador Sánchez Alonso

Doctor en Informática. Universidad Politécnica de Madrid, España. Ingeniero en informática. Universidad Pontificia de Salamanca. Profesor titular del Departamento de Ciencias de la Computación, Universidad de Alcalá, España.
salvador.sanchez@uah.es
orcid.org/0000-0002-9949-4797

Abstract

Europeana is a European project aimed to become the modern "Alexandria Digital Library", as it targets providing access to thousands of resources of European cultural heritage, contributed by more than fifteen hundred institutions such as museums, libraries, archives and cultural centers. This article aims to explore Europeana digital resources as open learning repositories in order to re-use digital resources to improve learning process in the domain of arts and cultural heritage. To carry out this purpose, we present results of metadata quality based on a study case associated to recommendations and suggestions that provide this type of initiatives in our educational context in order to improve the access of digital resources according to a specific knowledge areas.

Exploración de la relevancia de los recursos digitales de Europeana: ideas preliminares sobre la calidad de los metadatos Europeana

Resumen

Europeana es un proyecto europeo destinado a convertirse en la moderno Biblioteca Digital de Alejandría, dado que se orienta a facilitar el acceso a miles de recursos del patrimonio cultural europeo, soportados por más de mil quinientas instituciones como museos, bibliotecas, archivos y centros culturales. Este artículo tiene como objetivo explorar los recursos digitales de Europeana como repositorios abiertos de aprendizaje con el fin de reutilizar recursos digitales para mejorar procesos de aprendizaje en el dominio de arte y patrimonio cultural. Para llevar a cabo este propósito, se presentan los resultados

Rev. Interam. Bibliot. Medellín (Colombia) Vol. 40, número 1/enero-abril 2017 pp. 59-69   ISSN 0120-0976

59

de la calidad de metadatos a partir de la definición de un caso de estudio junto con recomendaciones y sugerencias que proporcionan este tipo de iniciativas en nuestro contexto educativo, con el fin de mejorar el acceso a los recursos digitales de acuerdo con un área de conocimiento específico.

Palabras clave: Europeana, Tesauro AAT, cobertura, metadatos, análisis de datos, calidad metadatos.

1. Introduction

Europeana is a project oriented towards centralizing the largest amount of digital resources possible. These resources are stored in external repositories so as to be catalogued, centralized and easily accessed by means of metadata indexation. However, by having a centralized catalogue built through collecting metadata from specialized repositories, the results obtained when looking up digital resources vary considerably and also lack accuracy; this leads to a waste of time in terms or search and selection of a particular digital resource.

This study allowed us identify firstly, to what extent Europeana covers certain topics or areas of knowledge, permitting to reuse and develop learning objects of knowledge areas related to the European cultural heritage. In addition, we examined the relationships of part of the metadata elements of these digital resources, to verify the integrity of the search results of the users, and the availability of accessible digital resources. As this process revealed, the metadata of the majority of the Europeana digital resources we browsed, were insufficient. This enabled us to identify a number of deficiencies in search processes of digital resources of a specific field of knowledge. This study was carry out between periods 2012 and 2014.

The purpose of this study is to analyze how well or how poorly Europeana covers certain topics or knowledge areas in the AAT, mainly those topics most used by teachers and student in high schools. In order to achieve this, the coverage study aims to analyze if Europeana, is a digital library that teachers could use for the development of learning objects in a specific knowledge areas, through the reuse of free/open access digital resources. In conclusion, the aim of this research is to analyze the metadata quality of Europeana's digital resources related to knowledge areas as arts and cultural heritage in order to re-use in learning process.

The motivation behind the present article is to explore if Europeana digital library is a well-structured library in order to provide resources for learning process in different knowledge areas such as heritage culture, architecture and arts. As well lies in the analysis of metadata quality that has been found so far from a set of digital resources that were extracted over Europeana by using data visualization techniques (Gaona-García, Sánchez, & Fermoso, 2012). One of the sections of this article focuses on analyzing the data exchange models that are used by Europeana. Subsequently, the article discusses the process associated to digital resource exploration and coverage analysis through a set of terms used in the Art & Architecture Thesaurus (AAT, 2015). The following sections present the results obtained from metadata quality analysis in terms of completion; these metadata were identified through digital resource extraction. The final section presents the result of this analysis and provides some recommendations in order to use this type of projects in learning process and so on in our educational context.

2. Background

The success of resource location, depends largely on the quality with which the metadata has been designed, making it an essential condition for the results being produced by search engines in the repositories (De la Prieta & Gil, 2010; Muñoz-Arteaga, Calvillo-Moreno, Ochoa-Zezzatti, Santaolaya-Salgado, & Álvarez-Rodríguez, 2010) and therefore, it is important to improve indexing strategies for learning objects stored in them (Ochoa, Cardinaels, Meire, & Duval, 2005; Stuckenschmidt, Vdovjak, Houben, & Broekstra, 2004; Wiley, 2002). Moreover, the lack of metadata constitutes a poor classification of the resources. This condition is a key factor that affects search results in a specific knowledge area. As the quality of contents is an indicator that permits us to evaluate digital resources, there are several studies that refer to methods on how to evaluate quality in digital resources (Chuanjun, 2004; Downes, 2007; Gonçalves, Moreira, Fox, & Watson, 2007), as well as the quality of the contents existing in collections of digital resources (Chuanjun, 2004; Downes, 2007).

[Paulo Alonso Gaona García - Ana Fermoso García - Salvador Sánchez Alonso]

2.1. Data-exchange Models Used by Europeana

Europeana is a project aimed to collect and make available the largest possible amount of cultural resources in digital form; these resources, stored in external repositories, are compiled, catalogued and organized through a central access portal. The purpose is to offer access to millions of digital resources registered by external providers. To date, Europeana has allowed registering a series of digital resources for linking and cataloging more than 50 millions of resources (Europeana.pro, 2016). These registries through using data models, permit deal with the definition of various guidelines and policies intended to exchange, normalize, store, manage and deploy the registered metadata.

2.2. Europeana Semantic Element (ESE)

The ESE Model (Clyphan, 2013) represents the current Europeana's production model, which maintains a set of elements identified through 15 metadata. These metadata are defined by standard Dublin Core (DC, 2008) together with a set of 13 metadata that have been created by Europeana so as to fulfill some of the project's own needs.

During model implementation, a series of difficulties had to be addressed at both expression and extension levels towards other models. However, one of the most representative problems lay in original metadata losses for the metadata defined through their content providers (Doerr et al., 2010). This encouraged the development of different proposals written by some of the content aggregators so as to improve communication processes (Houssos et al., 2011), and also to improve metadata exchange (Koulouris, Banos, & Garoufallou, 2011). At the same time, these proposals provided a proper atmosphere for the creation of a new data exchange model called EDM.

2.3. Europeana Data Model (EDM)

EDM is a data exchange pilot model whose main purpose is to preserve the original metadata as well as serving as a more flexible and expressive model (when compared to ESE). This pilot model allows multiple-register access on a single digital resource. EDM is based on Semantic-web good practices, supporting standards

such as Open Archives Initiative Object Reuse and Exchange (OAI-ORE) (Lagoze et al., 2007) for Internet resource exchange and interoperability. The model also uses semantic description languages through Resource Description Framework (RDF) and Simple Knowledge Organization System (SKOS). Additionally, it links resources by means of the Linked Open Data (LOD) project (Berners-Lee, 2006); which is an initiative that has involved 8 providers from 15 different countries and adds up to 2.4 million released metadata using standard Linked Data recipes (Europeana.lab, 2016). Table 1 presents the EDM elements.

Table 1. Europeana Semantic Element present in EDMmodel.

Source: The ESE Model (Clyphan, 2013).

According to EDM model (Doerr, et al., 2010) actually the present version of EDM integrates the ESE elements. The integration of ESE into EDM is expressed in RDF, this offers the additional advantage of exploiting the Web architecture for linking resources (Haslhofer & Isaac, 2011). For that reason, and taking into consideration the Europeana data model we have just described, in the next sections we will focus our coverage study in the five already mentioned elements of Dublin Core, that have been integrated into EDM: *language*, *type*, *country*, *content provider* and *rights*. First we will describe the methodology for coverage analysis and then, the results analysis.

However, as described in the transformation process (from ESE into EDM [Haslhofer & Isaac, 2011]), when taking the same set of metadata described by ESE, only access to a minimum digital resource metadata set is allowed; thus exhibiting poor performance in terms of digital-resource expression, which limits the scope of the model to serve various purposes such as indexation, search and access.

Rev. Interam. Bibliot. Medellín (Colombia) Vol. 40, número 1/enero-abril 2017 pp. 59-69   ISSN 0120-0976

61

3. Methodology

To carry out this research, we used a collection of digital resources of Europeana digital library as a case study. Europeana was choosen because of the following characteristics: i) it uses a semantic model for data exchange, the Europeana Data Model/EDM (Doerr et al., 2010), ii) it includes a large number of digital resources related to the European cultural heritage (over 53 million digital resources to date), iii) it uses the largest group of content providers on a European level, and finally iv) it allows reusability of open digital resources. Thus, according to the purpose of this study, Europeana is an initiative that offers the opportunity to reuse digital resources, allowing teachers or professionals of the sector of cultural heritage to use them as learning objects for educational purposes. On the basis of these characteristics, we browsed through the Europeana digital resources to determine its level of coverage concerning a branch of knowledge of the Art and Architecture Thesaurus (AAT). This coverage allows us to determine if Europeana is a good option in order to re-use digital resources in learning process.

3.1. Coverage Analysis

In this study, we examined the coverage of digital resource in Europeana according to a subset of terms related to topic "styles and periods" of AAT thesaurus. The terms were selected according to aspects like their taxonomic reference, classification, depth level, and the theme-relation level of each term regarding to topics of heritage cultural digital resources to be explored within Europeana.

Firstly, we identified the area of knowledge that would be chosen in order to perform a coverage analysis of the Europeana digital resources. We adapted the term coverage (or thematic coverage), as it is defined by Codina (2000) and Whitehall (1992, 1995) to refer to the number of available digital resources related to a topic or knowledge area. Thus, coverage is analyzed to explore the completeness of the Europeana digital library in terms of topics or knowledge areas that include a larger number of digital resources. Therefore, to study the coverage of the library, we take as a basis a set of terms related to the topics of "*styles and periods*", specifically the "*styles and periods by general era*"

and "*styles and periods by region*", i.e. knowledge branches of the AAT thesaurus. Next, we present the criteria according to which we selected the AAT thesaurus: i) the comprehensive and detailed representation of its hierarchical structure (Tudhope, Binding, Blocks, & Cunliffe, 2006), ii) the domain of knowledge (art and architecture) of the AAT thesaurus is one of the most complete and widely renowned (Aitchison, Gilchrist, & Bawden, 2000), and finally, iii) the variety of conceptual descriptors it offers, which enables indexing and very large and precise queries (Soergel, 1995).

The coverage study was raised as proof of concept, therefore it does not only aims to show results of empirical analysis of the coverage of Europeana, but it is intended to provide a method for conducting future studies on Europeana coverage in other knowledge areas.

In order to examine coverage in the Europeana project, a set of 118 terms were analyzed; these terms were defined through the "*Styles and Periods Facet*" taken from the AAT Thesaurus. The terms were selected according to aspects like their taxonomic reference, classification, depth level, and the theme-relation level of each term regarding the digital resources to be explored within Europeana.

Once we identified the domain of knowledge to be explored, it was crucial to know the Europeana data representation model and to identify the metadata elements used in each digital resource to proceed with the extraction process. Therefore, we used as reference the data exchange model defined by Europeana, known as the Europeana Data Model/EDM (Doerr, et al., 2010). This data exchange model has as its main objective to maintain original metadata as provided by the content providers, and to be a model with a wide semantic expression capability. It is based on the best practices of the Semantic Web, and is compatible with standards such as the Open Archives Initiative - Object Reuse and Exchange protocol (OAI-ORE) (Lagoze et al., 2007) for Internet resource sharing and interoperability.

3.2 Recognition of Digital Resources Through Data Extraction Process

In this second stage, we developed a strategy to link the Europeana digital resources to a specific knowledge

[Paulo Alonso Gaona García - Ana Fermoso García - Salvador Sánchez Alonso]

area using selected terms of the AAT thesaurus. For this reason, we developed a Web Crawler to browse digital resources discovered in Europeana for each AAT thesaurus term. A Web Crawler analyzes the syntax of a Web page and extracts information according to its structure (metadata) (Tadapak, Suebchua & Rungsawang, 2010). This extraction strategy was selected according to a series of features defined in a preliminary study we carried on relevant tools. These criteria were related to whether the tool: provides a clear definition of its architecture (Baeza-Yates & Castillo, 2004; Tripathy & Patra, 2008), is highly scalable (Boldi, Codenotti, Santini & Vigna, 2004), favours optimizing processes (Edwards, McCurley & Tomlin, 2001) and is highly efficient for content extraction processes (Castillo, 2005). The tool itself supports its operation by using computational algorithms that allow covering every page by means of search methods, and also by using specialized libraries to define extraction elements.

On the other hand, metadata extracted from each digital resource is selected according to the elements of the Europeana Data Model (EDM). The metadata extracted from digital resources are based on the EDM model —Table 1, section 2.3. In the Figure 1, is shown an example of one the resources obtained as result. This resource presents 11 representative metadata elements describing some characteristics of this digital resource:

- *Title*: A name given to the resource.
- *Creator*: An entity primarily responsible for making the resource.
- *Contributor*: An entity responsible for making contributions to the resource.
- *Date*: Date of creation of the resource.
- *Type*: The type of the original analog or born digital object as recorded by the content holder, this element typically includes values such as photograph, painting, sculpture, etc.
- *Description*: A description of the original analog or born digital object.
- *Data provider*: The name or identifier of the organization that contributes data to Europeana.

- *Provider*: Name of the organization that delivers data to Europeana.
- *Identifier*: An unambiguous reference to the resource within a given context.
- *Format*: The file format, physical medium or dimensions of the resource.
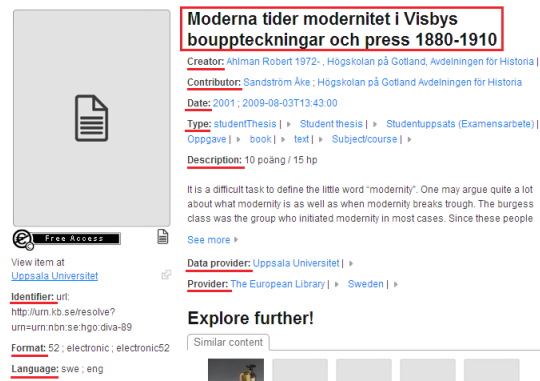- *Language*: A language of the resource.

Figure 1. Metadata properties extracted from a digital resources.

4. Result of Digital-Resource Exploration

In the third stage, we analyzed the relationship between the extracted digital resources and the thematic areas selected for browsing. For this, with the support of a group of experts in cultural heritage, we reviewed the extracted results and found that most of the extracted resources did not correspond to the knowledge domain defined by the terms of the AAT thesaurus. For this reason, we designed a browsing strategy that used refined searches and keywords related to the domain of knowledge of each term of the AAT thesaurus. This strategy produced more relevant results as far as the knowledge area of each term of the thesaurus is concerned.

From the exploration stage, we found 44,280 digital resources identified in Figure 2(a). In order to conduct a coverage analysis, such digital resources were classified according to format type, country of origin, language, content provider and copyright. The results of this analysis are shown in Figure 2(b), where 23,431 (53%) digital resources were found to lack enough attributes to describe the type of copyright of the digital

resource itself; the second biggest category was that of free-access digital resources, which accounted for 40% (17,802) of the resources analyzed; finally, 2,838 (6.5% on average) digital resources required payment for open search.
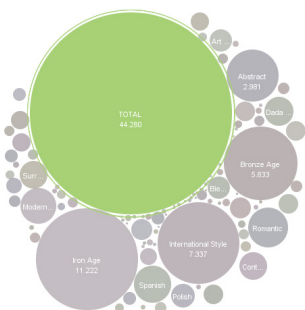

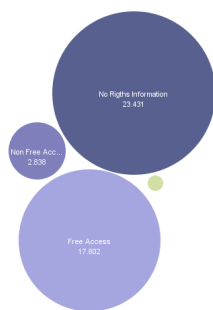**Figure 2(a).** Total number of digital resources. (Own elaboration)

Figure 2(b). Number of digital resources by Rights. (Own elaboration)

As a result, we found that a large set of topics associated to the AAT thesaurus knowledge area "Styles and Periods" was not fully covered by Europeana. On the other hand, the variety of results displayed while browsing suggested that Europeana did not support search methods to relate resources to a specific knowledge area. In Figure 3 we can see this distribution.
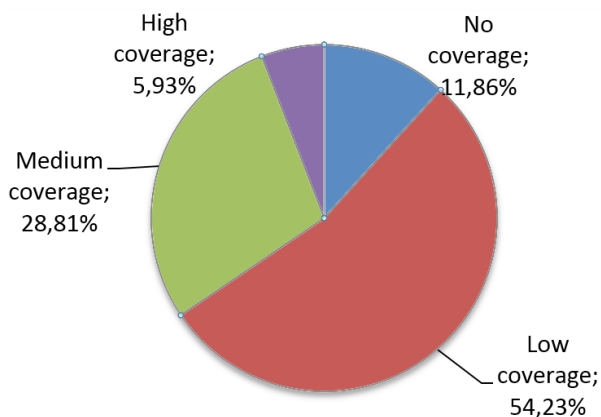

Figure 3. Percentage of coverage of AAT terms according to digital resources explored in Europeana.

According to the classification of Figure 5, can be identified a high set of AAT terms that have a low coverage (54.23%) and only a small group of AAT terms with high-level coverage of digital resources (5.93%). Below, we describe these results in more detailed according to *language*, *type* of format and *copyrights*.

4.1. Coverage by Language

With the purpose of conduct a coverage analysis, this digital resource was classified according to format type, country, language, content provider and copyright. The coverage according to language on the digital resource, predominantly English with (36%), "*Polish*" with (23%) and "*French*" with (21%). However, (7%) of these classification, describes a language called "*mult*", this means that this percentage of digital resources has support for multiples languages.

4.2. Coverage by Format

In the results of coverage according to type of format, the predominant format in digital resources found is "*Image*", with a percentage of 86%, followed by "Text" with 7%, and "Sound" with 4%. In other side, the countries that supply a large number of digital resources, are: "*United kingdom*" (30%), "*France*" (27%) and "*Poland*" (20%).

4.3. Coverage by Right Information

Figure 4 shows the coverage of digital resources found, according to copyrights. It identifies that a large majority of them (53%), had no copyright description in digital resources, and only (40%) of the total of digital resources found, are resources of free access.
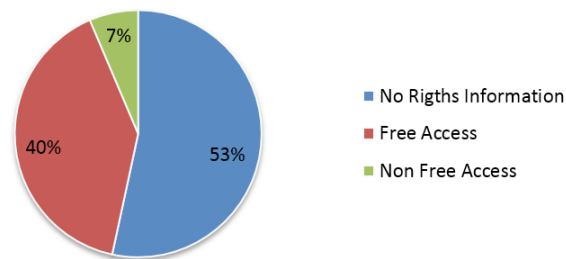

Figure 4. Coverage of digital resources by right information.

According to the results of coverage by copyright, below, in Figure 5, is presented a detailed analysis of digital resources found, that are free access, non-free access, and do not have a copyright information, classified by type of format.
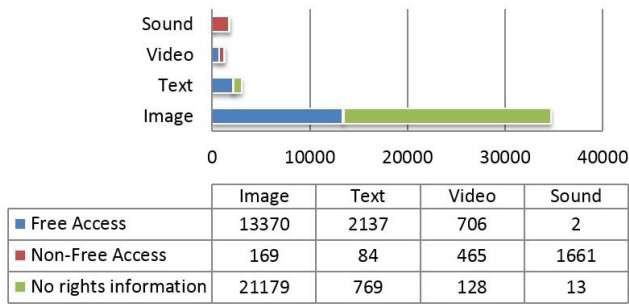
[Paulo Alonso Gaona García - Ana Fermoso García - Salvador Sánchez Alonso]

	Image	Text	Video	Sound
■ Free Access	13370	2137	706	2
■ Non-Free Access	169	84	465	1661
■ No rights information	21179	769	128	13

Figure 5. Description of coverage by right information vs type of format.

The absence of description for copyright in metadata elements of digital resources, and the variety of results, led to a second study related to metadata quality of the digital resources explored. With this aim, we present a complementary study in the following section, where a detailed analysis to examine metadata quality according to completeness of digital resources explored was performed. Through this analysis we will study if the level of quality of metadata can influence the study of coverage.

According to these results of coverage, is necessary to perform a detailed analysis to examine metadata quality of digital resources explored. To carry out more evidence, the study of metadata quality was made to 555.0000 Europeana digital resources extracted. Next section we analyses this results.

5. Results of Quality and Metadata Completion Analysis

The metadata quality are essential to search digital resources. Without them, the results of searches of digital resources are poor and inefficient (Cechinel, Sánchez-Alonso & Sicilia, 2009). The problem of low quality in repositories has been mentioned by other researches. To determine the quality of the metadata, our study will be based on the following areas: completeness and accuracy (Bui & Park, 2013; Manouselis, Vuorikari, & Van Assche, 2010). We use some analysis criteria strategies of quality assessment dealt with in (Bui & Park, 2013; Manouselis, et al., 2010; Palavitsinis, Manouselis & Sánchez-Alonso, 2014). In

the following sections, we will present the results of this study.

As a result, we found that a large set of topics associated to the AAT thesaurus knowledge area "*Styles and Periods*" was not fully covered by Europeana. On the other hand, the variety of results displayed while browsing suggested that Europeana did not support search methods to relate resources to a specific knowledge area.

According to the categories defined by the specification ESE: *Mandatory*, *Recommended and Optional* elements, has a varied definition of completeness in the metadata. Table 2 describes the case for analyzing completeness of Mandatory elements.

Table 2. Completeness of mandatory elements.

MANDATORY ELEMENTS	DIGITAL RESOURCES	
	Records filled	%
dc:title	547980	98,47
dc:description	280467	50,40
dc:language	57707	10,37
europeana:dataProvider	0	0,00
europeana:isShownAt	523060	93,99
europeana:isShownBy	263924	47,42
europeana:provider	556514	100,00
dc:subject	446337	80,20
dc:type	527807	94,84
dc:coverage	36253	6,51
dcterms:spatial	152677	27,43
europeana:rights	401373	72,12

In the case of mandatory elements can be evidenced a low level of 62,71% completeness for metadata elements defined by Europeana, but the metadata elements that have been defined by external providers, have in total a low level of completeness for an average of 56,81%. Table 3 presents results of recommended elements.

In the case of recommended elements above, identifies a level of completeness with an average half 38.40%. Finally, in Table 4 presents completeness of optional elements.

Rev. Interam. Bibliot. Medellín (Colombia) Vol. 40, número 1/enero-abril 2017 pp. 59-69 ISSN 0120-0976

65

Table 3. Completeness of recommended elements.

RECOMMENDED ELEMENTS	DIGITAL RESOURCES	
	Records filled	%
dcterms:alternative	71293	12,81
dc:creator	348142	62,56
dc:contributor	114476	20,57
dc:date	276118	49,62
dcterms:created	77638	13,95
dcterms:issued	97690	17,55
dcterms:temporal	24345	4,37
dc:publisher	512922	92,17
dc:source	358713	64,46
dcterms:isPartOf	255670	45,94

Table 4. Completness of optional elements.

OPTIONAL ELEMENTS	DIGITAL RESOURCES	
	Records filled	%
dc:format	197246	35,44
dc:identifier	543221	97,61
dcterms:extent	168786	30,33
dcterms:medium	74211	13,33
dc:rights	338706	60,86
dcterm:provenance	9267	1,67
dc:relation	244220	43,88
dcterms:conformsTo	0	0,00
dcterms:hasFormat	3120	0,56
dcterms:isFormatOf	4406	0,79
dcterms:isReferencedBy	0	0,00
dcterms:references	0	0,00
dcterms:isReplacedBy	1767	0,32
dcterms:replaces	1	0,00
dcterms:requieres	15	0,00
dcterms:tableOfContents	0	0,00
europeana:unstored	0	0,00
dcterms:hasVersion	5	0,00
dcterms:isVersionOf	12973	2,33
dcterms:isrequiredBy	0	0,00
dcterms:editor	0	0,00

In the case of optional elements, the level of completness is rather low with an average 14,66%.

Figure 6 shows the detailed analysis results obtained from the explored metadata, based on the mapping and normalization guide defined by ESE model through a set of mandatory elements, recommended elements, and Europeana's own elements.
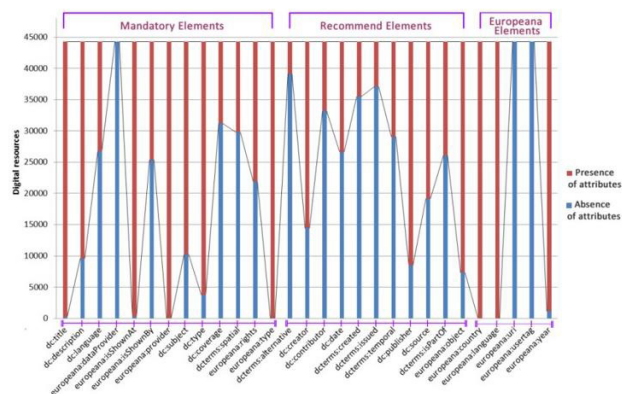

Figure 6. Analysis of metadata used in Europeana digital resources explored.

In the first set, namely "Mandatory elements", there is a (38.13%) attribute absence from the total amount of explored resources. For the second set, namely "recommended elements", attribute absence reaches (61.60%). Finally, the set of elements belonging to Europeana exhibits attribute absence of (50.52 %) for the metadata. This means that the majority number of metadata elements classified in "Mandatory elements" has a high quality of completeness. Therefore, we can conclude that we have found some important deficiencies in Europeana metadata on the basis of our accuracy analysis: redundancy, absence, ambiguity and inconsistency of its metadata, among other.

6. Conclusions

Europeana's data exchange model defines a digital-resource search within the same set of general metadata that describe the resource. This yields considerably varied search results (thus not very accurate). Therefore, the model provides a limited set of metadata that do not permit classifying digital resources coming from content providers and aggregators according to a specific knowledge domain.

As a conclusion to this preliminary study of metadata quality associated with the branch of knowledge "styles

[Paulo Alonso Gaona García - Ana Fermoso García - Salvador Sánchez Alonso]

and periods" of AAT, and in general on any knowledge area, it is important to define metadata to classify the digital resource from a topic or subject area. Offering a limited set of metadata descriptions of digital resources within a digital repository, generate a variety of irrelevant results in search processes. The definition of these metadata and quality themselves, offer a variety of search criteria (thematic area, language, resource type, copyrights, etc.) can be integrated into the development of visual interfaces trough visualization techniques (Gaona-Garcia, Martín-Moncunill, Sánchez-Alonso, & Fermoso, 2014; Gaona-Garcia, Sánchez-Alonso, & Montenegro, 2014).

The results of this analysis showed a lack of completeness of the metadata defined according to the Europeana Data Model/EDM. However, such results are not new, since other studies concerning digital repositories such as ARIADNE (Ternier et al., 2009), the National Science Digital Library (Fox, Gonçalves, & Kipp, 2002) and other collections of digital resources (Bui & Park, 2013), also found deficiencies concerning the definition of metadata. In fact, this deficiency, the lack of precision in the definition of metadata elements, is one of the main factors that directly influence digital resources search. Similarly, the absence of metadata elements for the classification of a digital resource according to topics or knowledge areas was reflected in the poor search results that we obtained through the study of thematic coverage. Those are factors that influence search results, and have also been mentioned by Cechinel et al. (2009).

Despite the deficiencies in the metadata quality associated to completeness, Europeana presents great opportunities to be a digital library with reusability of digital resources for the development of learning objects. Not only by the results of coverage and the large volume of digital resources available to this library, but by the EDM exchange data model. Strategy, which is emerging as a model that fully cover future issues related to semantic search, and thus facilitate better quality of search process by linking them through Linked Data (Bizer, Heath, & Berners-Lee, 2009; Dietze et al., 2012; Haslhofer & Isaac, 2011). However, in order to improve the quality of them, Europeana should make great efforts not only to improve the quality of the metadata of digital resources, but also define

strategies by implementing alternative metadata to facilitate the search for digital resources from a subject or area of knowledge. The key issue is the integration of knowledge representation schemes such as the use of ontologies or thesauri, to classify digital resources associated with a branch of knowledge.

7. References

1. AAT. (2015). *Art & Architecture Thesaurus (AAT)*. Retrieved from http://www.getty.edu/research/tools/vocabularies/aat/, (last access: 24 february 2015).

2. Aitchison, J., Gilchrist, A., & Bawden, D. (2000). *Thesaurus construction and use: a practical manual*. London: Psychology Press.

3. Baeza-Yates, R., & Castillo, C. (2004). Crawling the infinite Web: five levels are enough. In S. Leonardi, (Ed.). *Lecture Notes in Computer Science, 3243: Algorithms and Models for the Web-Graph* (pp. 156-167). New York: Springer.

4. Berners-Lee, T. (2006). *Linked Data - Design Issues*. Retrieved from http://www.w3.org/DesignIssues/LinkedData.html.

5. Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. In T. Heath, M. Hepp, & C. Bizer (Eds.). *Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS)*. Retrieved from *http://linkeddata.org/docs/ijswis-special-issue*

6. Boldi, P., Codenotti, B., Santini, M., & Vigna, S. (2004). Ubicrawler: A scalable fully distributed web crawler. *Software: Practice and Experience, 34*(8), 711-726.

7. Bui, Y., & Park, J.-R. (2013). *An assessment of metadata quality: A case study of the national science digital library metadata repository*. Paper presented at the Proceedings of the Annual Conference of CAIS/Actes du congrès annuel de l'ACSI.

8. Castillo. (2005). *Effective web crawling*. Paper presented at the ACM SIGIR Forum.

9. Cechinel, C., Sánchez-Alonso, S., & Sicilia, M. Á. (2009). Empirical analysis of errors on human-generated learning objects metadata. In F., Sartori & M. Á., Sicilia (Eds.) *Metadata and semantic research* (pp. 60-70). New York: Springer.

10. Clyphan. (2013). *Europeana Semantic Element ESE v3.4.1*. Retrieved from http://pro.europeana.eu/share-your-data/data-guidelines/ese-documentation, (accessed: 25 feruary 2015).

Rev. Interam. Bibliot. Medellín (Colombia) Vol. 40, número 1/enero-abril 2017 pp. 59-69 ISSN 0120-0976

67

11. Codina, L. (2000). Evaluación de recursos digitales en línea: conceptos, indicadores y métodos. *Revista española de documentación científica, 23*(1), 9-44.

12. Chuanjun, S. (2004). On the Evaluation of the Quality of Digital Collections. *The Journal of The Library Science In China, 4*, 45-48.

13. DC. (2008). *Dublin core metadata element set, version 1.1: Dublin Core Metadata Initiative (DMCI)*. Retrieved from http://dublincore.org/documents/dces/ (last access: 24 february 2015).

14. De la Prieta, F., & Gil, A. (2010). A multi-agent system that searches for learning objects in heterogeneous repositories. In P. Pawlewski, V. J. Julián, F. Fdez Riverola, E. Corchado, R. Corchuelo, J. Bajo, J. M. Corchado Rodríguez, F. Dignum, Y. Demazeau, & A. Campbell (Eds.). *Trends in Practical Applications of Agents and Multiagent Systems* (pp. 355-362). New York: Springer.

15. Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N., & Taibi, D. (2012). *Linked Education: interlinking educational Resources and the Web of Data*. Paper presented at the Proceedings of the 27th annual ACM symposium on applied computing.

16. Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., & van de Sompel, H. (2010). The Europeana Data Model (EDM). *IFLA 2011: World library and information congress: 76th IFLA general conference and assembly. Gothenburg, Suecia: IFLA.*

17. Downes, S. (2007). Models for sustainable open educational resources. *Interdisciplinary Journal of Knowledge and Learning Objects, 3*, 29-44.

18. Edwards, J., McCurley, K., & Tomlin, J. (2001). *An adaptive model for optimizing performance of an incremental web crawler*. Paper presented at the Proceedings of the 10th international conference on World Wide Web.

19. Europeana.lab. (2015). *Experimental datasets of Europeana*. Retrieved from http://labs.europeana.eu/api/linked-open-data/data-downloads/, (last access: 24 February 2015).

20. Europeana.pro. (2016). *Europeana pro*. Retrieved from http://pro.europeana.eu/share-your-data/how-to-contribute-data (last access: 24 february 2015).

21. Fox, E. A., Gonçalves, M. A., & Kipp, N. A. (2002). Digital libraries. In H. Adelsberg, B. Collins, & J. Pawlowski (Eds.). *Handbook on Information Technologies for Education and Training* (pp. 623-641). New York: Springer.

22. Gaona-Garcia, P., Martín-Moncunill, D., Sánchez-Alonso, S., & Fermoso, A. (2014). A usability study of taxonomy visualisation user interfaces in digital repositories. *Online Information Review, 38*(2), 284-304.

23. Gaona-Garcia, P., Sánchez-Alonso, S., & Montenegro, M. (2014). Visualization of information: a proposal to improve the search and access to digital resources in repositories. *Ingeniería e Investigación, 34*(1), 83-89.

24. Gaona-García, P., Sánchez, S., & Fermoso, A. (2012). Análisis de cobertura del tesauro AAT en la biblioteca digital Europeana: ideas preliminares para su empleo en la educación. *SPDECE 2012.*

25. Gonçalves, M. A., Moreira, B. L., Fox, E. A., & Watson, L. T. (2007). What is a good digital library? – A quality model for digital libraries. *Information Processing & Management, 43*(5), 1416-1437. doi: 10.1016/j.ipm.2006.11.010

26. Haslhofer, B., & Isaac, A. (2011). *data. europeana. eu: The europeana linked open data pilot*. Paper presented at the International Conference on Dublin Core and Metadata Applications.

27. Houssos, N., Stamatis, K., Banos, V., Kapidakis, S., Garoufallou, E., & Koulouris, A. (2011). Implementing enhanced OAI-PMH requirements for Europeana. *Research and Advanced Technology for Digital Libraries, 396-407.*

28. Koulouris, A., Banos, V., & Garoufallou, E. (2011). *Aggregating metadata for Europeana: the Greek paradigm*. Retrieved from http://vbanos.gr/wp-content/uploads/2011/11/icininfo2011_koulouris_banos_garoufallou_preprint.pdf

29. Lagoze, C., Van de Sompel, H., Johnston, P., Nelson, M. L., Sanderson, R., & Warner, S. (2007). *Open Archives Initiative Object Reuse and Exchange (OAI-ORE): Technical report, Open Archives Initiative.*

30. Manouselis, N., Vuorikari, R., & Van Assche, F. (2010). Collaborative recommendation of e-learning resources: an experimental investigation. *Journal of Computer Assisted Learning, 26*(4), 227-242.

31. Muñoz-Arteaga, J., Calvillo-Moreno, E., Ochoa-Zezzatti, C., Santaolaya-Salgado, R., & Álvarez-Rodríguez, F. (2010). Use of Agents to Realize a Federated Searching of Learning Objects. In Y. Demazeau, F. Dignum, J. M. Corchado, J. Bajo, R. Corchuelo, E. Corchado, F. Fernández-Riverola, V. J. Julián, P. Pawlewski, A. Campbell (Eds.). *Trends in Practical Applications of Agents and Multiagent Systems* (pp. 1-8). New York: Springer.

32. Ochoa, X., Cardinaels, K., Meire, M., & Duval, E. (2005). *Frameworks for the automatic indexation of learning management systems content into learning object repositories*. Paper presented at the Proceedings of World

68

Rev. Interam. Bibliot. Medellín (Colombia) Vol. 40, número 1/enero-abril 2017 pp. 59-69 ISSN 0120-0976

Conference on Educational Multimedia, Hypermedia and Telecommunications 2005.

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

34. Soergel, D. (1995). The art and architecture thesaurus (AAT): A critical appraisal. *Visual Resources, 10*(4), 369-400.

35. Stuckenschmidt, H., Vdovjak, R., Houben, G. J., & Broekstra, J. (2004). I*ndex structures and algorithms for querying distributed RDF repositories.* Paper presented at the Proceedings of the 13th international conference on World Wide Web.

36. Tadapak, P., Suebchua, T., & Rungsawang, A. (2010). *A machine learning based language specific web site crawler.* Paper presented at the Network-Based Information Systems (NBiS), 2010 13th International Conference on.

37. Ternier, S., Verbert, K., Parra, G., Vandeputte, B., Klerkx, J., Duval, E.,... Ochoa, X. (2009). The ariadne infrastructure for managing and storing metadata. *Internet Computing, IEEE, 13*(4), 18-25.

38. Tripathy, A., & Patra, P. K. (2008). *A Web mining architectural model of distributed crawler for Internet searches using PageRank algorithm.* Paper presented at the Asia-Pacific Services Computing Conference, 2008. APSCC'08. IEEE.

39. Tudhope, D., Binding, C., Blocks, D., & Cunliffe, D. (2006). Query expansion via conceptual distance in thesaurus indexed collections. *Journal of Documentation, 62*(4), 509-533.

40. Whitehall, T. (1992). Quality in library and information service: a review. *Library management, 13*(5), 23-35.

41. Whitehall, T. (1995). Value in library and information management: a review. *Library management, 16*(4), 3-11.

42. Wiley, D. A. (2002). Connecting learning objects to instructional design theory: A definition, a metaphor and a taxonomy. In The instructional use of learning objects. *Bloomington, Indiana, 2830*(435), 3-24.

[Exploring the Relevance of Europeana Digital Resources: Preliminary Ideas on Europeana Metadata Quality]

Rev. Interam. Bibliot. Medellín (Colombia) Vol. 40, número 1/enero-abril 2017 pp. 59-69 ISSN 0120-0976

69