Computers in Human Behavior 51 (2015) 1134–1141

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Adopting the metadata approach to improve the search and analysis of educational resources for online learning

Maria-Cruz Valiente, Miguel-Angel Sicilia *, Elena Garcia-Barriocanal, Enayat Rajabi

Computer Science Department, University of Alcala, Polytechnic Building, Ctra. Barcelona km. 33.6, 28871 Alcala de Henares, Madrid, Spain

ARTICLE INFO

Article history:
Available online 8 April 2015

Keywords:
Collaborative learning
e-learning innovations
Knowledge Society
Metadata-supported learning
Teaching/learning strategies

ABSTRACT

The emergence of large collections of learning resources created through the harvesting and aggregation of metadata raises important concerns on the suitability of educational resource descriptions as provided in metadata schemas. For learning purposes, both teachers and students usually seek information on their own, and the vast majority of the search that they do in search engines like Google is driven by multiple keywords or classifications. Therefore this type of metadata-based learning resources could help them obtain better results related to the educational resources they are looking for and provide the basis for collaborative learning environments which enable knowledge sharing and reuse in terms of web-based search systems. This paper reports an exploratory study based on the availability and suitability of keywords and classifications in metadata-based educational resources to improve collaborative learning between teachers and students through the search and analysis of learning resources from a large sample obtained from the Global Learning Objects Brokered Exchange (GLOBE).

1. Introduction

The adoption and implementation of e-learning innovations in the context of the Knowledge Society have become more demanding in recent years (García, Colomo, & Lytras, 2012a; Lytras, 2010; Lytras & Ordóñez de Pablos, 2011) as well as the empowerment of teachers to work with their students and other teachers in order to share and reuse educational resources (Damiani, Lytras, & Cudré-Mauroux, 2010; García, Colomo, & Lytras, 2012b). In this context, the proliferation of educational resource repositories of different kinds has raised the need to aggregate the descriptions of resources into larger collections, thereby providing a critical mass for users, especially for learners, with educational needs that may not be confined to a single thematic repository. Since metadata defines the set of properties that educational resources should include for their retrieval, the use, for example, of the *Learning Object Metadata* (LOM) standard (IEEE 1484.12.1, 2002), combined with harvesting protocols (i.e. protocols for the collection of metadata from repositories) such as the *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) (Open Archives, 2008), has facilitated the deployment of such collections for collaborative learning purposes. LOM defines the structure of a metadata instance that a learning object should conform (IEEE 1484.12.1, 2002). In this context, a learning object can be defined as any digital resource that can be reused to support the learning process (Wiley, 2000). From here on, we use the term ‘learning object’ or ‘learning resource’ to refer to the LOM-based educational resource and the term ‘educational resource’ to refer to the resource itself. Fig. 1 shows a general overview of the learning object model defined by the LOM standard. OAI-PMH is a low-barrier mechanism for repository interoperability, that is, the standard used to harvest the metadata from other repositories where the data providers are the repositories that expose that structured metadata via OAI-PMH. Then, service providers will make OAI-PMH service requests to harvest the metadata involved in the process (Open Archives, 2008). Repositories in this situation represent network accessible servers that can process the different OAI-PMH requests, enabling the sharing and reuse of Web educational resources targeted toward the professional growth of teachers and improved learning for students.

However, the aggregation of heterogeneous collections provided by different learning communities for its application in a common collaborative learning environment (in our particular case, teachers and students) is a complex task to tackle. The variety of learning resources in granularity (i.e. how big a learning object is) and the different technical formats such as video, application and text, makes also difficult to use full-text indexing as commonly used in search systems on scholarly literature collections. In consequence, search engines tend to rely on indexing metadata instead

* Corresponding author.
E-mail addresses: maricruz.valiente@uah.es (M.-C. Valiente), msicilia@uah.es (M.-A. Sicilia), elena.garciab@uah.es (E. Garcia-Barriocanal), enayat.rajabi@uah.es (E. Rajabi).

http://dx.doi.org/10.1016/j.chb.2014.12.059
0747-5632/

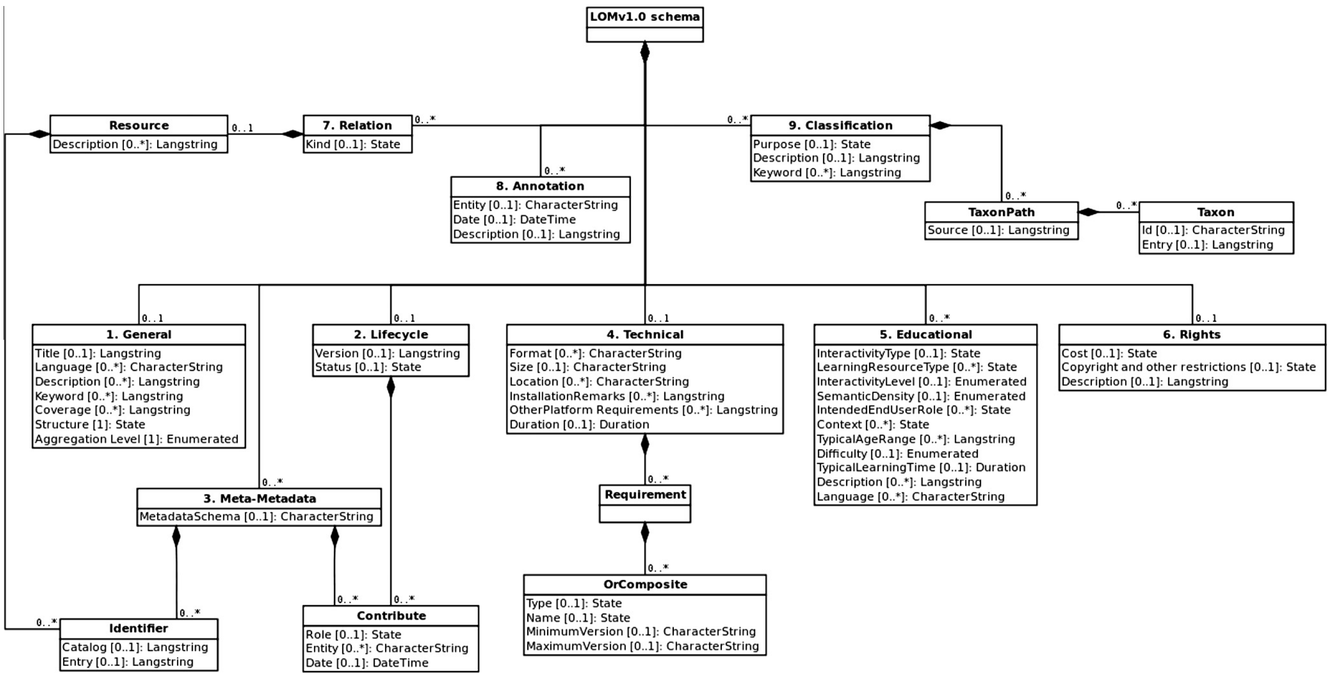

Fig. 1. UML class diagram representing an overview of the learning object model defined by the LOM standard. *Source:* http://en.wikipedia.org/wiki/Learning_object_metadata.

of the contents themselves. This raises the question on how effectively different metadata elements properly describe and categorize the educational resource space. Therefore, we focus on gathering empirical evidence on the keyword and classification spaces of large aggregated collections, how that spaces compare to other description mechanisms and to which extent they could be effectively used to interlink with other Web resources in the context of collaborative learning environments.

In doing so, a large portion of the Global Learning Objects Brokered Exchange (GLOBE[1]) collection was subject to analysis using different methods. GLOBE enables the sharing and reuse of learning objects between different learning resource repositories worldwide, and it is nowadays the most diverse and large collection available openly (GLOBE, 2011). LOM has the problem that it does not establish the elements that should be present in each learning object (i.e. fields are not marked as required, recommended or optional). Therefore, an empty LOM record is a valid LOM record. This characteristic creates several problems when the metadata is shared among different repositories. For this reason, GLOBE defines an application profile that enhances the LOM standard by defining a set of required fields (GLOBE, 2011). An application profile specifies a set of metadata elements selected from one or several metadata schemas which are combined for the definition of a new domain-specific schema (GLOBE, 2011). In this scenario, the LOM-based general keyword field that describes the learning content is established as a recommended field, and the LOM-based classification space is basically optional apart from the textual label of the classification taxon which is defined as a mandatory field. Both, general keywords and the classification space are the two key items that we want to analyze in order to determine the availability and suitability of keywords and classifications as a searching tool for collaborative learning domains where teachers and students are involved.

There have been some empirical studies on the actual use of LOM metadata. For example, Friesen (Friesen, 2004) collected samples from several international repositories to study the real use of LOM remarking the fact that the potential value was not being realized. After that one, Ochoa, Klerkx, Vandeputte, and Duval (2011) became the study with a largest empirical base, up to 630,317 metadata instances from GLOBE. In their study, the most comprehensive study to date on the use of LOM for heterogeneous resource collections, the *Keyword* element was found to be used in more than 55% of the metadata instances, and with the *Taxon* element around 60%.

However, none of the existing studies on the use of LOM metadata analyzes the availability and suitability of metadata for educational purposes. This paper represents an extension of the work done by Sicilia, Sanchez-Alonso, Garcia-Barriocanal, Minguillón, and Rajabi (2013). Our intention is to carry out an enhanced and deeper study aimed at the improvement of collaborative learning in terms of metadata-based educational resources.

1.1. Metadata

Metadata is generally defined as "data about data" although is better understood as "any statement about an information resource", regardless of any specific domain (Garshol, 2004). In computer science, the concept of metadata is usually understood as the description of information regarding objects on the network. In our context, these objects will be learning objects that will support collaborative learning through web-based search systems.

One of the most well-known vocabularies for metadata is Dublin Core (Weibel, Kunze, Lagoze, & Wolf, 1998), which is composed by 13 properties that can be used to describe information resources. Examples of these properties are "title", "creator", "subject", "description", "publisher", "date" and "language" as shown in Table 1.

Besides, metadata could be categorized into five types depending on their functionality (Baca, 2008): *administrative*, *descriptive*, *preservation*, *technical* and *use*. Metadata categories with the corresponding descriptions and some examples are given in Table 2.

2. Materials and methods

As mentioned earlier, Dublin Core (Weibel et al., 1998) is one of the most well-known ways to describe information resources. The

[1] http://www.globe-info.org.

Table 1
Metadata example (adapted from Garshol (2004)).

Property	Value
Title	Curing the Web's Identity Crisis
Creator	Steve Pepper & Sylvia Schwab
Subject	RDF, topic maps, subject indicator
Description	This paper describes the crisis of identity facing the World Wide Web and, in particular, the RDF community. It shows how that crisis is rooted in a lack of clarity about the nature of "resources" and how concepts developed during the XML Topic Maps effort can provide a solution that works not only for Topic Maps, but also for RDF and semantic web technologies in general
Publisher	IDEAlliance
Date	May 2003
Language	English

eXtensible Markup Language (XML) Schema (Peterson, Shudi, Malhotra, Sperberg-McQueen, & Thomson, 2012; Thomson, Mendelsohn, Beech, & Maloney, 2012) and Extensible Stylesheet Language (XSL) (Adler et al., 2001) are two examples of well-known general-purpose metadata tools. For the purposes of this paper, the Learning Object Metadata (LOM) standard (IEEE 1484.12.1, 2002) specifies several metadata elements for providing descriptions of learning objects. LOM is an international open standard that proposes different metadata elements grouped into nine categories: *General*, *Lifecycle*, *Meta-Metadata*, *Technical*, *Educational*, *Rights*, *Relation*, *Annotation* and *Classification*.

The *General* category groups the general information that describes a learning object as a whole. This category is composed of the next elements: *Identifier*, *Title*, *Language*, *Description*, *Keyword*, *Coverage*, *Structure* and *Aggregation level*. Since the *Keyword* element is intended for the description of topics in a learning object using keywords or phrases in any existing language, this is one of the fields that we need to use in our study. Other interesting field for our purposes is *Coverage*, which specifies the time, culture, geography or region to which a specific learning object applies.

The *Lifecycle* category describes the history and current state of a learning object and those entities that have affected the learning object during its evolution. This category is composed of the next elements: *Version*, *Status* and *Contribute*. The *Contribute* element defines the entities (i.e. people or organizations) that have contributed to the current state of the particular learning object. A *Contribute* element is composed of the *Role*, *Entity* and *Date* fields that provide information regarding the contribution. For our purposes, the useful field is the *Entity* element. An *Entity* identifies the person or organization contributing to the learning object. It is worth noting at this point that authors and contributors related to educational resources are not the same. The author is the original creator of the content of a specific educational resource itself. However, it is very usual that the contributor is not the original author but the entity that is involved in the generation of the LOM-based metadata associated with the learning object. Therefore, in our context, teachers will be usually those of the contributors who provide the learning objects for students or other teachers with an interest in the subject under study.

The *Meta-Metadata* category groups information about the metadata instance itself, not regarding the learning object that the metadata instance describes. This category is composed of the next elements: *Identifier*, *Contribute*, *Metadata schema* and *Language*. For the purpose of our study, this category is not subject of interest.

The *Technical* category describes the technical requirements and characteristics of a learning object. This category is composed of the next elements: *Format*, *Size*, *Location*, *Requirement*, *Installation remarks*, *Other platform requirements* and *Duration*. The *Format* element that classifies the learning objects into technical categories (video, text, application, ...) is a value of interest for collaborative learning environments. The *Location* element is the character string assigned to represent the access of a learning object such as the *Universal Resource Locator* (URL). Since we want students and also teachers to learn together through LOM-based educational resources, it is important that the access to resource contents be present in the learning object.

The *Educational* category describes the key educational or pedagogic characteristics of a learning object. This category is composed of the next elements: *Interactivity type*, *Learning resource type*, *Interactivity level*, *Semantic density*, *Intended end user role*, *Context*, *Typical age range*, *Difficulty*, *Typical learning time*, *Description* and *Language*. In this case, the *Learning resource type* element is the field of interest since it specifies the kind of learning object we are retrieving (lecture, exercise, lesson plan, presentation, blog, textbook, ...).

The *Rights* category describes the intellectual property rights and conditions of use for a learning object. This category is composed of the next elements: *Cost*, *Copyright and other restrictions*, and *Description*. For the purpose of our study, this category is not subject of interest.

The *Relation* category defines the relationship between learning objects. This category is composed of the next elements: *Kind*, *Resource* and *Identifier*. For the purpose of our study, this category is not subject of interest.

Table 2
Metadata categories (Baca, 2008).

Category	Description	Examples
Administrative	Metadata used in managing and administering collections and information resources	– Acquisition information – Location information – Selection criteria for digitization
Descriptive	Metadata used to identify and describe collections and related information resources	– Cataloging records – Specialized indexes – Annotations by creators and users
Preservation	Metadata related to the preservation management of collections and information resources	– Documentation of physical condition of resources – Documentation of actions taken to preserve physical and digital versions of resources
Technical	Metadata related to how a system functions of metadata behaves	– Hardware and software documentation – Technical digitization information (e.g. formats) – Authentication and security data (e.g. passwords)
Use	Metadata related to the level and type of use of collections and information resources	– Use and user tracking – Search logs – Rights metadata

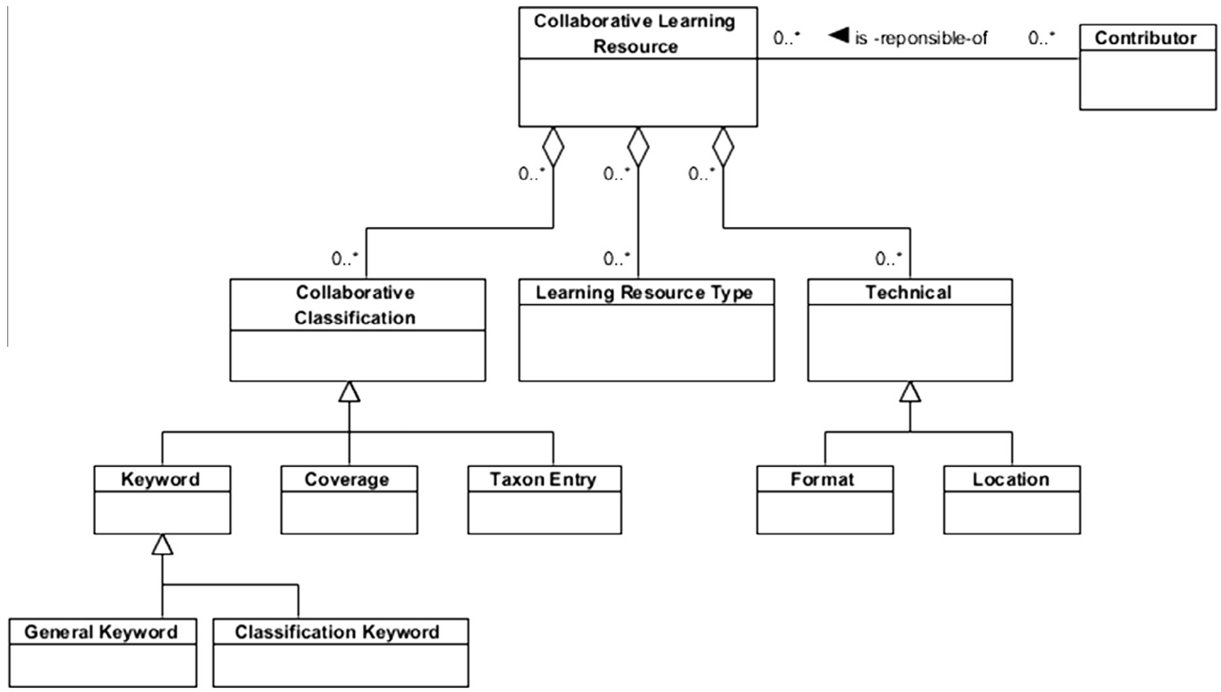

Fig. 2. UML class diagram representing the metadata of interest in our collaborative learning environment.

The *Annotation* category provides comments on the educational use of a learning object, and information on when and by whom the comments were created. This category is composed of the next elements: *Entity*, *Date* and *Description*. For the purpose of our study, this category is not subject of interest.

Finally, the *Classification* category describes the specific classification for a learning object. This category is composed of the next elements: *Purpose*, *Taxon path*, *Description* and *Keyword*. In this case, useful fields for the purpose of our study are *Taxon path* and *Keyword*. If some form of controlled vocabulary or knowledge organization system (KOS) is applicable, *Taxon paths* provide room for referencing their elements, also introducing typing by means of classification purposes. A *Taxon path* is composed of an ordered list of *Taxon* elements which describe terms within a taxonomy. Each *Taxon* includes the next elements: *Id* and *Entry*. The *Entry* element defines the textual label of a *Taxon* and this is the value that we need. Fig. 2 summarizes the LOM fields selected for our study.

3. Results

The sample of LOM metadata instances used in our study was obtained by using an OAI-PMH harvester tool. The total number of LOM instances that were harvested from GLOBE repositories that provide a metadata harvesting service based on the OAI-PMH protocol was 815,223.

Table 3
Language code groups occurring most frequently in the sample.

Language code	Total
en, en-gb, en-US, eng, english	392673.00
nl	97976.00
x-none, none, blank	77555.00
de, de-AT, de-DE	49807.00
es, es-EC, es-es	47828.00
it, ITA	23102.00
hu, hu-HU	20316.00
is	8804.00
ca	8066.00
fr	6414.00

As keywords are lexical elements, one first important analysis is that of the language of the educational resources. Ochoa et al. (2011) pointed out of an evolution in time, being in any case English the most frequent language. Table 3 summarizes the use of the languages declared in 502,062 learning resources in the sample (61.59% of the total resources available), grouping language variants. However, the results are not as relevant as expected. There are resources in the sample with language codes such as "ITA-ENG-DEU-FRE", which appears to mean a combination of different languages, and cannot be therefore, processed directly by a system in their current state.

As shown in Table 3, the English language is clearly dominant, whereas the Dutch language is the second one, maybe due to some particular reasons already pointed out by Ochoa et al. (2011). Besides, although the resources with language are enough significant four our purposes, it is worth noting that a large amount of resources has no language declared.

In the general keyword space, we have found the total of 5,454,068 keywords in the sample. However, we have to discard all the general keywords starting with "x-mt-" due to they are generated through machine translation. Therefore, the total of valid general keywords is reduced to 3,194,140 giving us an average rate of 6 keywords per resource (502,062 learning objects which include general keywords). Fig. 3 illustrates a summary with the number of resources that share the same number of keywords.

Each resource in the sample has at least one keyword that will classify the learning resource into a specific educational category available for search engines aimed at collaborative learning. An excerpt of the English keywords with more occurrences in the sample is shown in Table 4.

If we look for these keywords in the classification section of the resources, we find that the number of resources that define classifications is 482,184 (59.15% of the total number of resources). However, only 16,162 resources (1.99% of the total number of resources) include classification keywords in their definition that amounts to 17,495 keywords, where the 99% of these keywords are in the English language (see Table 5) and the rest has no language value. Besides, as illustrated in Fig. 4, the vast majority

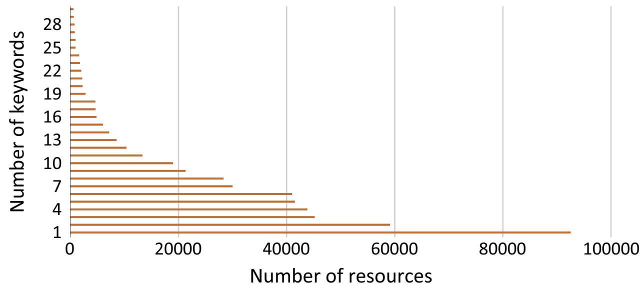

Fig. 3. Analysis of general keywords by resources in the sample.

Table 4
General keywords occurring most frequently in the sample.

Keyword	Total
Science	29664.00
Astronomy	17724.00
Earth Science	16367.00
Physical Sciences	15965.00
Space sciences	13627.00
Physics	9769.00
Ukoer	7554.00
Geography	6835.00
Biology	6765.00
People	5756.00
Science as Inquiry	5579.00
Icelandic language	5317.00
Mathematics	5314.00
Image	5296.00
Geology	5243.00
Literature	5229.00
History	5180.00
Space Science	4823.00
Teaching materials	4758.00
Environmental science	4611.00
Vocabulary	4569.00
Meteorology	4420.00

of these resources only include one keyword in the *Classification* category.

On the other hand, the total of unduplicated keywords in the *Classification* section amounts to 63, and 47 of them are used as a general keyword, too (from a total of 343,351 different general keywords). However, it is worth highlighting at this point that classification keywords are never redundant with respect to the general keywords of a specific resource. This category defines a more general classification than general keywords, as can be seen in Table 6.

Another type of classification is the coverage of the learning resource. According to LOM (IEEE 1484.12.1, 2002), the coverage field will typically include spatial location, temporal period of jurisdiction. In the sample, we find 88,818 values distributed across 58,633 resources (7.2% of the total resources available). Table 7 illustrates the most frequent values of coverage in the sample.

Since this field defines the time, culture or region to which a learning resource applies, coverage values never are redundant with respect to the general keywords or classification keywords of a specific resource (see example in Table 8).

As mentioned earlier, taxons amount to particular terms within a taxonomy, where entries represent the textual labels of these taxons. If we analyze the classification taxon entries, we find that the sample contains 1,072,149 distributed across 256,898 resources (31.51% of the total number of resources). The number of entries that can be found in the different languages is shown in Table 9. In this case, the English language is once again the most popular language, but surprisingly the unspecified and Catalan languages are second and third in the list over the Spanish and Dutch languages.

Table 5
English classification keywords occurring in the sample.

Keyword	Total
Science and Technology	6900.00
Mathematics and Statistics	3958.00
Social Sciences	2348.00
Humanities	2149.00
Arts	1291.00
Business	625.00
Test/Draft	136.00
Art	1.00

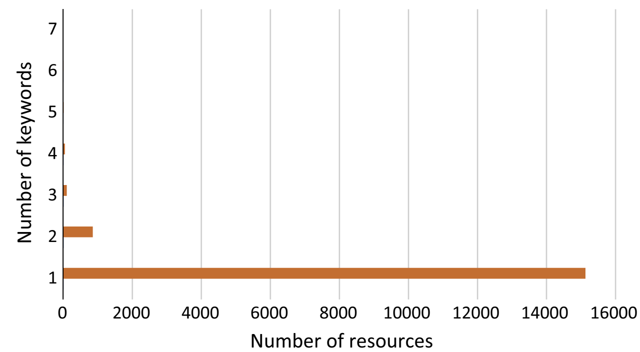

Fig. 4. Analysis of classification keywords by resources in the sample.

A sample of English entries with more than 3000 occurrences in the sample is shown in Table 10. Taxon entries categorize learning objects in a different way as keywords and coverage do. Therefore, there is no redundancy among them but a complement. Furthermore, as the LOM standard suggests, since the identifier element is also present in the sample, both the taxon identifier and the taxon entry could be used to designate a particular taxon in a learning object.

Then, for classification purposes, we have found in the sample that none of the total learning objects have the four involved fields filled: general keywords, classification keywords, entries and coverage. Only a total of 10,721 learning objects, less than the 2% of the total resources, include general keywords, entries and coverage; and only 9576 resources have both fields, general keywords and classification keywords. However, we can find in the sample that the number of learning resources including general keywords and entries are 288,113. Table 11 summarizes the number of resources that have the fields filled for classification purposes.

As regards contributors of learning objects, they are another point of reference in a collaborative learning domain since they select useful educational resources from the Web and generate the corresponding educational objects to be available for search engines in a specific repository. In our environment where teachers and students collaborate with each other to learn efficiently, in many cases, educational resources will be selected depending on the contributor that created or published them (all of us have reference people/organizations when we seek information in the web). In this case, about the 70% of the total resources in the sample have registered the contributor of the learning object.

Regarding the categorization of educational resources, the LOM fields of *technical format* and *learning resource type* may help learners get an idea at first glance of the type of the educational resource they are looking into. The number of resources with the *technical format* and *learning resource type* fields filled amounts to 362,765

Table 6
Excerpt from the sample regarding general keywords and classification keywords.

Resource id	General keyword	Classification keyword
1325	continuous time	Mathematics and Statistics
1325	laplace transform	
1361	algebra	Mathematics and Statistics
1361	chart	
1361	elementary	
1361	real numbers	
2569	actionscript	Science and Technology
2569	attribute	
2569	class	
2569	color	
2569	element	
2569	extensible markup language	
2569	flex	
2569	flex mxml	
2569	gradient	
2569	MXML	
2569	property	
2569	style	
2569	transparency	
2569	xml	

Table 7
Coverage values occurring most frequently in the sample.

Coverage	Total
All	12465.00
Nederland	5741.00
North America	3382.00
World	3319.00
Switzerland	1375.00
Null	1282.00
Austria	1244.00
United Kingdom	1117.00
Amsterdam	895.00
Europa	860.00
Germany	825.00
Contemporary architecture	732.00
Central and Eastern Europe	684.00
1930d	637.00
Universal	576.00
VS	503.00
Europe	454.00
1949	420.00
Rotterdam	369.00
Milano	357.00

(44.5% of the total number of resources) and 466,564 (57.23% of the total number of resources), respectively. An excerpt of the *technical format* terms and *learning resource type* terms with more occurrences are shown in Tables 12 and 13.

On the other hand, as mentioned earlier, if we want to obtain useful web-based search systems for collaborative learning, the availability to access the resource contents is of major importance. Oddly enough, the number of learning resources is about the 95% (777,759 resources) but not the 100%. However, this field should be present in any resource in order to be able to know the original source of the learning object and determine whether it is actually a reliable educational resource.

Finally, apart from the fields of interest for collaboration learning environments, another interesting point of analysis concerns the conformance of the metadata with the LOM standard. For example, the *Structure* element of the *General* category defines the underlying organizational structure of a learning object. LOM limits the values of this field to the next: *atomic*, *collection*, *networked*, *hierarchical* and *linear*. However, we can find in the sample that, from a total of 155,369 resources (19.1% of the total number of resources) with the general structure filled, 18,525 have other values than the values defined by the standard. These values are *mixed* and *branched*. Furthermore, the *Aggregation level* element of the *General* category assigns the functional granularity of a specific learning object. The LOM standard remarks that a learning object with the general structure field equals to 'atomic' should have the general aggregation level field equals to 1 (smallest level of aggregation, e.g. raw media data of fragments), and equals to 2 (e.g. a lesson), 3 (e.g. a course) or 4 (e.g. a set of courses that lead to a certificate) for the rest of the structure values. In this case, we find in the sample that, from a total of 50,066 resources with these two fields filled, only about the 6% of the total number of resources, 12,972 have a value different from 1 when the structure value is 'atomic' (from a total of 28,500 resources), and 1666 resources have a value different from 2, 3 or 4 when the structure field is different from 'atomic' (from a total of 21,566 resources). Furthermore, we find 105,303 resources with the aggregation level field blank but the structure field has a value, and we find 79,955 resources with the structure field blank but the aggregation level field has a value.

Table 8
Excerpt from the sample regarding general keywords and coverage.

Resource id	General keyword	Coverage
1238	association building	Zurich Switzerland
1238	building data	
1238	facade construction	
1238	building shape	
1238	prestressing	
1238	insulation value	
1238	U-value	
1238	steel mesh	
1238	glare eliminator	
1238	cloth sheet	
1238	element facade	
1238	profiling	
1238	ventilation flap	
1238	emission reduction	
1377	Ancient Civilizations	All
1377	Ancient Near East	
1377	Assyria	
1377	Babylonia	
1377	Classics	
1377	Egypt	
1377	Mediterranean	
1377	Nubia	
1377	Persia	
1377	Religion	
1377	Sumer	
1377	Western Civilization	

Besides, in order to be conformant to the LOM standard, the *Source* element of the *Classification* category should specify the name of an existing classification system (''official'' or any user-defined taxonomy). In the sample, we find 85 different classification sources from a total of 791,890 sources distributed across 420,320 learning objects (51.56% of the total resources available). Among these sources, the number of official classification systems amounts to 45, more than half of the total number.

4. Discussion

The study was focused on reporting the availability and suitability of keywords and classifications in metadata-based learning objects to use in collaborative learning environments. GLOBE provided us with a sample big enough to do this and the study reveals that LOM-based search offers plenty of opportunities and a real advantage over search based on the resource's contents.

This study described above reveals that the English language maintains its dominance among the distribution of languages in GLOBE, supported by the fact that international sectors use

Table 9
Classification taxon entry language code groups found in the sample.

Language code	Number of entries
en, en-US	539568.00
unspecified	180643.00
ca	158340.00
es	88261.00
fr	19734.00
sv	17488.00
de	13687.00
nl	12954.00
ro	12038.00
it	10352.00
ko	7391.00
sl	7102.00
lt	2786.00
ja	1665.00
hu	76.00
gr	33.00
el	11.00
da	10.00
lv	6.00
fi	4.00

Table 10
Classification taxon entries occurring most frequently in the sample.

Entry	Total
Brick	35734.00
Housing	21480.00
History	19082.00
Cross-curricular education	15602.00
Wood	15068.00
Concrete	14473.00
Asphalt	13225.00
Engineering Competency Classification	11897.00
Limestone	6882.00
Metal	6794.00
Architecture Competency Classification	6345.00
Geography	5839.00
Stucco	5738.00
Design	4865.00
Graphic arts	4738.00
Granite	4359.00
Slate	4268.00
Chemistry	4046.00
Primary education	4010.00
Book	3970.00
Primary school	3970.00
Transportation	3861.00
Government	3837.00
Sandstone	3828.00
Biological sciences	3772.00
Illustration	3586.00
Mathematics	3434.00
Iron	3342.00

Table 12
Technical formats occurring most frequently in the sample.

Technical format	Total
Text/html	119089.00
Application/pdf	52493.00
Video/x-msvideo	50915.00
Text-html	24715.00
Video/x-ms-asf	12298.00
Application/msword	11688.00
Html	10562.00
Graphics-photos	10173.00
Application/vnd.ms-powerpoint	9418.00
Downloadable-docs	8158.00
Image/pjpeg	7639.00
Audio/Visual	7171.00
Video	6690.00
Image/jpeg	4615.00
Application/http	4409.00
Video/swf	4227.00
Video/mp4	3275.00
Lesson/Lesson Plan	3002.00
Audio	2881.00
Application/vnd.ms-excel	2394.00

English as their official language. Besides, all the resources in the sample tend to be conformant to the LOM standard guidelines. The average of general keywords describing the topic of each learning object is fair, but classification keywords usually are not considered in the sample. In this case, although we can find a same term in the general and classification keyword fields, for a specific resource these fields always have different values.

Classification taxon entries are the right complement to the keywords that enable us to label learning objects into a specific taxonomy with no redundancy among them, even when the number of entries is much lower than general keywords. Regarding other aspects in the classification of the resources, the general coverage field is also present in the sample and can also be used for classification purposes. Coverage allows us to describe the region, culture or period of time to which a learning object applies. Therefore, as a whole, using all these four metadata fields together will enable the categorization of educational resources in terms of keywords and classification terms providing a basis for the improvement in their analysis and search applied to collaborative learning domains.

In this vein, we can complement this collaborative classification using the location, format and learning resource type of a learning object as additional and useful information that help users decide whether the educational resource is relevant for them. Besides, the contributor of the learning object can be considered as the reference expert for specific educational resources (e.g. learning resources provided by teachers for their students).

Therefore, although our empirical study was not as successful as we expected due to the usual low percentage of fields filled in the GLOBE sample, and the use of keywords for searching in not anything new, the LOM-based educational resources does not have

Table 11
Number of resources per combination of classification fields.

	Classification entries	Classification keywords	General coverage	General keywords	Classification entries + General keywords	Classification keywords + General keywords	General coverage + General keywords
Classification entries	256,898	0	25,226	121,161	121,161	0	10,721
Classification keywords	0	16,162	0	9576	0	0	0
General coverage	25,226	0	58,633	43,654	10,721	0	43,654
General keywords	121,161	9576	43,654	502,059	121,161	9576	43,654

Table 13
Learning resource types occurring most frequently in the sample.

Learning resource types	Total
Other	95464.00
Image	90180.00
Narrative text	53203.00
Presentation	38261.00
Data	33113.00
Web page	20309.00
Figure	17237.00
None	12938.00
Project	10918.00
Readings	8904.00
Drill and practice	8792.00
Reference material	8669.00
Activities and Labs	7932.00
Open activity	7826.00
Syllabi	5864.00
Reference	5763.00
Lesson plan	5662.00

the problem of being poorly structured and it is an innovative approach to collaborative learning environments aimed to share and reuse knowledge in a structured way. We have proven the validity and usefulness of the content of selected fields from the LOM standard targeted to effective retrieval of educational resources.

The LOM-based search approach is a great opportunity for teachers and students teaching/learning together in a collaborative learning environment. The selected fields from the LOM standard enable us to access to educational resources in different web-based search systems in a formal and effective manner. For example, there are several learning portals in the Web that could match perfectly with this approach such as Khan Academy, available in English and Spanish (Khan Academy, 2014a, 2014b), Open Discovery Space (ODS) (Open Discovery Space, 2014) and the teachers' corner of the European Space Agency (ESA) (European Space Agency, 2014). Indeed, ODS is using metadata on the search of educational resources. These different learning systems instead of working independently, they could be integrated into a common learning environment in terms of LOM-based educational resources which enables the sharing and reuse of the learning objects within different areas of interest fostering collaborative learning between teachers and students from over the world.

Acknowledgements

This research has been partially supported by ICT PSP Project Open Discovery Space and project MAVSEL funded by the Spanish Ministry of Economy and Innovation.

References

Adler, S., Berglund, A., Caruso, J., Deach, S., Graham, T., Gross, P., et al. (2001). *Extensible Stylesheet Language (XSL) version 1.0*. W3C Recommendation 15 October 2001. <http://www.w3.org/TR/2001/REC-xsl-20011015/> Retrieved November 2014.

Baca, M. (2008). *Introduction to metadata* (2nd ed.). Los Angeles: Getty Research Institute.

Damiani, E., Lytras, M. D., & Cudré-Mauroux, P. (2010). Guest editorial: Special issue on human-centered web science. *World Wide Web, 13*(1–2), 1–2.

European Space Agency (2014). *Teachers' corner education*. <http://www.esa.int/Education/Teachers_Corner> Retrieved November 2014.

Friesen, N. (2004). The international learning object metadata survey. *The International Review of Research in Open and Distance Learning, 5*(3).

García, F. J., Colomo, R., & Lytras, M. D. (2012a). Outcomes of international research projects on technology applied to education. *Journal of Universal Computer Science (UCS), 18*(1), 1–4.

García, F. J., Colomo, R., & Lytras, M. D. (2012b). Informal learning in work environments: training with the Social Web in the workplace. *Behaviour & IT, 31*(8), 753–755.

Garshol, L. (2004). Metadata? Thesauri? Taxonomies? Topic Maps! *Journal of Information Science, 30*(4), 378–391. <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html> Retrieved November 2014.

GLOBE (2011). *GLOBE metadata application profile specification document. Version 1*, February 2011. <http://www.globe-info.org/images/ap/globe_lom_ap_v1.0.pdf> Retrieved November 2014.

IEEE 1484.12.1 (2002). *Draft standard for learning object metadata*.

Khan Academy (2014a). *Khan Academy English version*. <https://www.khanacademy.org/> Retrieved November 2014.

Khan Academy (2014b). *Khan Academy Spanish version*. <https://es.khanacademy.org/> Retrieved November 2014.

Lytras, M. D. (2010). From the Special Issue Editor: Information systems research for a sustainable knowledge society. *IS Management, 27*(3), 196–197.

Lytras, M. D., & Ordóñez de Pablos, P. (2011). Software technologies in knowledge society. *Journal of Universal Computer Science (UCS), 17*(9), 1219–1221.

Ochoa, X., Klerkx, J., Vandeputte, B. & Duval, E. (2011). *On the use of Learning Object Metadata: the GLOBE experience*. In Proceedings of the 6th European Conference on Technology Enhanced Learning (EC–TEL 2011). Springer-Verlag, Berlin Heidelberg, LNCS 6964, (pp. 271–284). <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.392.3603&rep=rep1&type=pdf> Retrieved November 2014.

Open Archives (2008). *The open archives initiative protocol for metadata harvesting. Protocol Version 2.0. Document Version 2008-12-07T20:42:00Z*. <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm> Retrieved November 2014.

Open Discovery Space (2014). *The open discovery space portal*. <http://portal.opendiscoveryspace.eu/> Retrieved November 2011.

Peterson, D., Shudi, G., Malhotra, A., Sperberg-McQueen, C. M., & Thomson, H. S. (2012). *W3C XML schema definition language (XSD) 1.1 Part 2: Datatypes*. W3C Recommendation 5 April 2012. <http://www.w3.org/TR/xmlschema11-2/> Retrieved November 2014.

Sicilia, M.-A., Sanchez-Alonso, S., Garcia-Barriocanal, E., Minguillón, J., & Rajabi, E. (2013). Exploring the keyword space in large learning resource aggregations: the case of GLOBE. *Workshop on learning object analytics for collections, repositories & federations (LACRO'13)*.

Thomson, H. S., Mendelsohn, N., Beech, D., & Maloney, M. (2012). *W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures*. W3C Recommendation, 5 April 2012. <http://www.w3.org/TR/xmlschema11-1/> Retrieved November 2014.

Weibel, S., Kunze, J., Lagoze, C., & Wolf, M. (1998). *RFC 2413: Dublin core metadata for resource discovery*. The Internet Society. <http://www.ietf.org/rfc/rfc2413.txt> Retrieved November 2014.

Wiley, D. A. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor and a taxonomy. In: Wiley, D. A. (Ed.), *The instructional use of learning objects*. <http://wesrac.usc.edu/wired/bldg-7_file/wiley.pdf> Retrieved November 2014.