An interoperability model based on ontologies for Learning Object Repositories

Sandobal Verón, Valeria Celeste
Grupo de Investigación en Educación Sobre Ingeniería (GIESIN)
UTN- Facultad Regional Resistencia
Chaco, Argentina
vsandobal@frre.utn.edu.ar

Ale, Mariel Alejandra
Gutiérrez, María de los Milagros
Centro de Investigación y Desarrollo en Sistemas de Información (CIDISI)
UTN – Facultad Regional Santa Fe
Santa Fe, Argentina
{male; mmgutierr}@frsf.utn.edu.ar

***Abstract*—Currently, most of the repositories implement Dublin Core (DC) as metadata standard, allowing the application of Protocol for Metadata Harvesting (OAI-PMH: Open Archives Initiative - Protocol for Metadata Harvesting). However, DC is not the most appropriate standard for the description of learning objects, which makes necessary resort to other standards. The Learning Object Metadata (LOM) standard emerges as the most suitable for the description of learning objects. Also, other standards such as Common European Research Information Format (CERIF) Metadata Object Description Schema (MODS) arise among others. This variety of standards makes the interoperability between repositories will become increasingly complex. Most solutions, until now, propose to adopt a metadata standard and include the necessary metadata to be harvested. This paper presents a solution based on ontologies for interoperability between repositories that use different metadata in the description of its objects**

Keywords—Interoperability model; metadata standard; repositories; learning object

I. INTRODUCTION

Currently, institutional repositories (IR) have experienced tremendous growth, both at national and international level. Learning Objects Repositories (LOR) arise, as a specialization of the RI, which are characterized because they contain learning objects (LO). The LOR is a "collection of LO that have information (metadata) detailed which is accessible via Internet. In addition to housing the LO, the LO can store the locations of those objects stored on other sites, both online and in local locations "[1]. The Committee of Standardization of Educational Technology (IEEE 2002) [2] provides that "learning objects (LO) is a digital or non-digital entity that can be used, reused and referenced during the learning supported by technology. " On the other hand, Wiley [3] defines the LO as elements of a new type of instruction based on the object-oriented paradigm, which are available for the LOR fulfill its goal, LO should be described by metadata.

The vast majority of institutional repositories use metadata standard Dublin Core (DC), while the LOR continued with this standard through the Internet and can be reused in multiple educational settings in order to describe learning objects. Over time, DC metadata standard were insufficient considering that an LO should be defined from the pedagogical point of view and not just as a resource. For example, it is necessary to describe the type of interaction, type of educational resource, the educational level to which it is directed, the degree of difficulty, among others. The report of the Confederation of Open Access Repositories (COAR) 2015 [4] mentions that it is increasingly important to the community adopting common metadata identifiers (both authors, institutions, organizations that fund research and publications) vocabularies and taxonomies. In this report, interoperability problems have been identified, that must be solved by classifying them according to their relevance and in high, moderate and low complexity. In this regard, it has been identified as highly relevant and moderate complexity using additional metadata formats; and relevant and high complexity using metadata quality. Given the above-mentioned report, and in relation to metadata, it was suggested adding new metadata standards that are more suitable for the description of learning objects. While DC is the most widely used standard and the mandatory for implementation of the protocol Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH), it presents certain vagueness in the interpretation of some labels. Some standards that are mentioned as possible to correct these shortcomings are: Metadata Object Description Schema (MODS), Common European Research Information Format (CERIF), Machine –Readable Cataloging (MARC), among others. The diversity of metadata standards and adoption in different LOR hinders interoperability between repositories. Although, there is an interoperability standard that allows metadata harvesting learning objects, this happens at the syntactic level. Currently, semantic interoperability is one of the challenges facing the community which intends to implement repositories as a form of publishing production.

In this paper, a model based on ontologies that supports semantic interoperability between LOR is introduced, independently of the metadata standard to adopt.

978-1-5090-4596-9/16/$31.00 ©2016 IEEE

The paper is organized as follows. Section II presents the preliminary concepts used for the development of this proposal. Then Section III presents the proposed interoperability model. Finally, conclusions and future work are described.

II. INTRODUCTORY CONCEPTS

A. Interoperability

Open Access (AA) born as an initiative to solve the so-called crisis of the traditional model of scientific communication. This crisis was characterized by high costs for publication in prestigious scientific journals, as well as for access to published articles. The initiative of AA from Budapest, seeks that the results of scientific research articles have free internet availability. This should allow researchers and general public to read, download, distribute, and print documents published in AA.

Interoperability is one of the main features that make possible the implementation of the AA. According to the COAR report, the real value of the repositories lies with the potential interconnect to create a network of repositories, which can provide a unified access to research results and which may be (re) used both by machines and by researchers, interoperability being a key factor [5]. In this work, it is used the concept of interoperability given in [6]: Interoperability is the ability of two or more systems or components to exchange information and use the information exchanged. Rodrigues [7] applies this concept to the IR defining interoperability as the ability of systems to communicate with others by exchanging information, metadata and digital objects including a round trip in a usable format. Levels of interoperability for IR can be classified according Garrido Arenas [8] in:

• *Infrastructure*: by the use of protocols such as ISO-OSI and TCP / IP to perform data exchange.

• *Syntax*: to provide information systems so they can read data from other similar systems, allowing to obtain a representation that can be supported. An approach towards interoperability from syntactically is the OAI-PMH [9] protocol, which provides the functions necessary for collecting metadata, not full text of the documents that are referenced. This protocol requires that the repository adheres to the Dublin Core metadata for its use.

• *Structure*: existence of common logic models that allows information systems to communicate with each other through protocols.

• *Semantics*: the ability of information systems have a common understanding of the terms that will be exchanged. The level of semantic interoperability ontologies emerge as a solution to mediate the problems of semantic heterogeneity. An ontology is an explicit specification of a conceptualization [10]. In this sense, an ontology is developed to give meaning to terms of a given context, bridging the semantic gap between heterogeneous systems. Taking the case of LOR, each defines learning objects with different metadata standards. Then, you need to identify what metadata such as LOM standard is equivalent in meaning to the DC tag: creator. If the standard is analyzed it is possible to realize that this equivalence is established with LOM metadata: LifeCycle \ Contribute \ Role \ Author LOM category: LifeCycle-Contribute-Entity. Similarly, in the MARC standard can be found this equivalence with the metadata Personal Name within the category Fields main entrance, which has indicators such as: Own Name and Last Name; so the name and surname of the author is separated unlike DC and LOM. Thus it is noted that the same meaning can be represented in different ways depending on the standard metadata using the LOR. The use of ontologies can establish relationships between concepts, beyond providing a common meaning between these heterogeneous systems. At present there are ontologies to most of the standards mentioned here such as DC, LOM, CERIF, MOD; as well as definitions, the specifications related to these standards, as in the case of ISO / IEC MLR (Learning Resource Metadata) that in Part 2 related metadata MLR DC; or IMS Learning Resource Metadata that makes minimal changes from LOM, by mapping metadata for both.

B. Related Work

Currently, there are several attempts in the search for semantic interoperability between repositories. Among which we can mention the made by [11]. It is consisting of a mechanism designed to recover LO from heterogeneous LOR using a framework of semantic interoperability through metadata elements. LOR includes two main operations: get metadata from different LOR and bring them to a central one resulting centralization in the search from one place. Among the LOR with which it is connected extract the most widely used metadata as title, keywords, description, location which are used as input for the LORIuMET (Learning Object Repositories interoperability using metadata). Another proposal is defined by [12] where a model of interoperability between systems of information from Colombia is proposed, which includes Institutional Repositories and Digital Libraries. The proposed model includes a description of standards and norms in which BDCOL (Biblioteca Digital Colombiana) will be based for the exchange and collection of metadata and digital objects. For achieving syntactic interoperability it uses a character encoding scheme UTF-8 and metadata standards according to collections. DC standards, qualified DC, ETD-MS among others were analyzed. As regards semantic interoperability, document type was normalized and it is suggested the use of controlled vocabularies. For structural interoperability, the use of OAI-PMH is suggested. A proposal between analyzed is the one of [13] that poses a model in which the main components are: the repositories, the multi-agent system that is composed of an Indexer Agent and Search Agent, mapping metadata (that uses of test two repositories, one using DC and the other LOM) which is a database designed to perform the correlation of the different patterns of metadata that were processed by the system; the domain ontology which is a set of ontologies from the types of learning object repositories (such as an ontology of information security) and a search service that provides an interface.

The proposal of [14] propose the handling of semantic interoperability between heterogeneous repositories using

Semantic Web-based approaches, where each Digital Library preserves the specific and requires no modifications are made to perform exchange services, reuse and harvesting of digital resources. A technological framework and method for the publication and link digital bibliographic data are presented and include the following activities: selection of data sources, harvesting metadata from repositories, modeling vocabulary or ontologies, data conversion to RDF format, data binding through their semantic relationships and publication and use of data. This framework was implemented in a set of Ecuadorian digital repositories and it was checked that "the developed activities ensure the reproducibility of the release cycle of linked data on any other OAI repository ".

Furthermore, Agosti [15] takes two models as a basis that provide a framework for digital libraries, "DELOS References Model" and "Streams, Structures, Spaces, Scenarios, Societies (5S)". Considering that, in the opinion of the authors, these models do not sufficiently improve the interoperability of systems, model through ontologies, Semantic Web technologies, and Linked Data to the reference models. Ontology designed allows modeling and map high-level concepts of the 5S DELOS model. Thus, a model for the user domain, content, functionality, quality, policy and architecture is presented. Authors make special reference to the user domains, functionality and content that enable them to obtain a high level of interoperability between the actors and the digital information/objects in the digital library, as well as the functions and services.

III. PROPOSED MODEL

The above proposals made a very good approximation to the search for semantic interoperability, but have a number of limitations. In some cases, only two of the most widely used standards are taken such as DC and LOM; in others, the metadata standards used in libraries of a particular country are elected; or extracting heterogeneous repositories metadata; regardless of metadata standard but only taking the most used tags such as title, keyword, description and location.

The model that is proposed in this article want to find a general solution, including any standard metadata using interoperability that can be obtained with the use of ontologies. To do this, the implementation of a hybrid approach [16] is proposed. This approach combines the advantages of simple ontologies methodologies and multiple ontologies. Where the original data heterogeneous sources have their own ontology (local ontology) that are represented independently; in turn these local ontologies are related based on the development of a global ontology through a shared vocabulary. The shared vocabulary contains basic or primitive terms of the domain. For the latter, current proposals which make contributions that intend to add meaning to educational metadata standards as DC, were analyzed. Among them is the proposed [17], where it is raised a profile for educational metadata (EMP: Educational Metadata Profile) to be used in the description of digital educational resources, and particularly those used for the distance mode. The profile is provided based on the metadata used by LOM with the aim of highlighting the educational metadata; completing the description with metadata as expected learning outcomes and instructional context (with values such as distance learning, blended learning, classroom education); as well as the added value for the label Type of Educational Resource. The authors also raise an ontology for the EMP in order to capture and process the semantic relationships between educational resources. In this regard, other metadata adaptation experience for its correct use for the description of learning objects it is proposed by [18], where a correlation between metadata used in qualified DC and LOM is performed. Seeking to include, in this way, educational information, digital skills and learning style. The test is performed on GREDOS institutional repository of the University of Salamanca. Although these proposals would include what is intended to represent the shared vocabulary, in this article it is proposed to build on the work done by [19], where an ontology based on standard DC and the guidelines of the National Digital Repositories System (SNRD) is modeled, Figure 1.

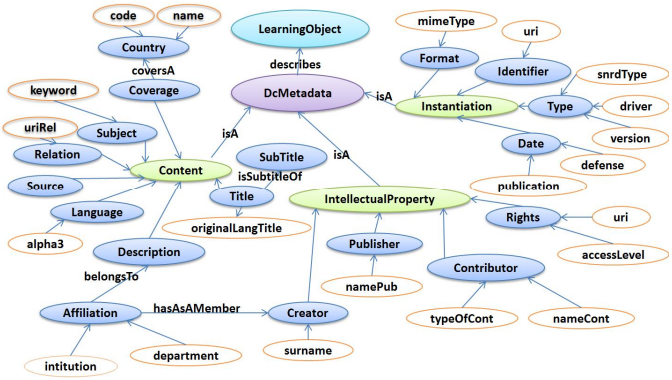

Fig. 1. DCOntoRep Ontology

The DCOntoRep ontology describes the metadata involved in the DC standard, classified into three categories: content, instantiation and intellectual property. On this basis, SNRD recommendations have been added as part of the enrichment process of ontology, such as:

- The use of subtitle, where the concept of subtitle related with title is added, through the *iSubTitleOF* relationship, where repetition and mandatory restrictions are implemented through cardinality constraints.

- The use of standards such as ISO 639 and ISO 3166 for labels and coverage language respectively. In this case, ontologies conceptualizing such rules were imported.

- In the case of "type" concept, attributes were added, to respond to the DRIVER controlled vocabulary, and a subtype agreed by the SNRD for scientific results; as well as a digital version object. The binding of these attributes is implemented through cardinality constraints.

- The "description" concept should be extended indicating the affiliation of the authors involved in the digital object. To meet these recommendations are included: a new label for membership and the necessary relationships so that they remain linked with description and author tags. In addition to axioms of integrity.

• Incorporation of Semantic Web Rule Language rules (SWRL) which clarifies certain business rules that could not be expressed through classes, attributes and relationships.

Taking into account these considerations, in this paper it is proposed the conceptualization of metadata using ontologies, which not only represent a common vocabulary but also defining constraints, axioms, inference and matching concepts, providing a solution to the syntactic and semantic interoperability information.

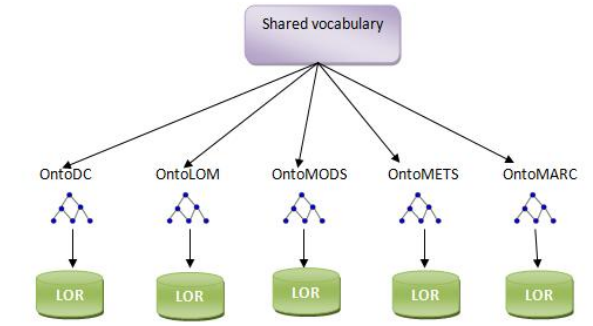

Fig. 2. Hybrid Approach proposed

A. *Local or multiple ontologies*

Metadata standards used in each of the repositories will be represented by local ontologies. In this case, existing and available ontologies on the network as MODS[1], LOM[2], DC[3], CERIF[4] y MARC[5](Figure 2). Proposing to repositories using these ontologies if the selected metadata standard is among those proposed. In the case of requiring the implementation of another standard, it will be important to maintain interoperability of repositories development of local ontology.

B. *Shared Vocabulary*

For the implementation of shared vocabulary, as it is mentioned in the previous section, the work [19] is taken, where the DC standard is taken as basis adding metadata needed to comply with the guidelines SNRD.

Likewise, the incorporation of metadata directly related to the educational/pedagogical level is considered appropriate, since our implementation will be done in the LOR.

Neither the standard DC nor guidelines SNRD take into account these particular characteristics for the description of LO. In order to achieve this shared labels vocabulary related to the educational environment, above mentioned standards were compared, and it concludes on adding metadata that are part of the LOM standard, since it is the most appropriate considering that is developed specifically for learning objects.

The labels were incorporated are: type of interactivity: ones defined in LOM are considered possible types: expository, active, mixed, not defined; educational resource type, where values can be: exercises, simulation, questionnaire, diagram, figure, graph, index, slide, table, descriptive text, test, experiment, problem presentation, self-assessment; end Users could be selected from: teachers, authors, students, administrators; context where the possible values are: cycle primary education, secondary education, higher education first, second cycle higher education, vocational training, continuing education, adult education; age range to which is addressed; difficulty: the defined degrees of difficulty are: no difficulty, easy, medium difficulty, difficult, very difficult.

With these additions the shared vocabulary will appear as shown in Figure 3.

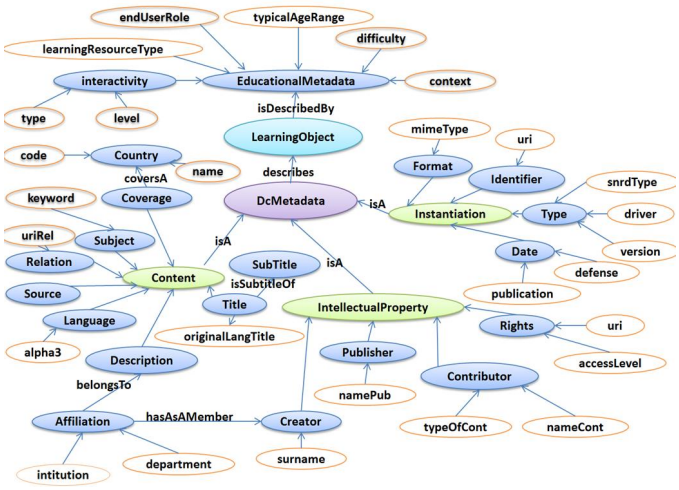

Fig. 3. DCOntoRep enriched with educational metadata LOM

To the ontology, the Educational Metadata class is added to describe the metadata specific to the educational / teaching area. This class is related to the class *learningObject* through *isDescribedBy* relationship. The built class has as attributes: *learningResourceType*, *endUserRole*, *typicalAgeRange*, context and difficulty. The values of each of these attributes are incorporated as defaults. It was also added one class interactivity with type and level attributes that describe both the type of interactivity and the level of interactivity of the learning object. The interactivity class is related to the Educational Metadata class through the *hasInteractivity* relationship. By performing the above inclusions, it is necessary to add SWRL rules which allows clarify some rules within the types that cannot be modeled through classes, attributes and relationships. Examples are shown below. A case to consider is that if an LO is the *driver type*: article, book, bookPart, conferenceObject, doctoralThesis, masterThesis, bachelorThesis, patent, review, workingPaper, report, other (dataset) and other (research project) corresponds to the type of article, book, bookPart, conferenceObject, doctoralThesis, masterThesis, bachelorThesis, patent, review,

[1] Metadata Object Description Schema. MODS RDF Ontology. https://www.loc.gov/standards/mods/modsrdf/

[2] Ontología LOM. http://slor.sourceforge.net/ontology/lom.owl

[3] Dublin Core in OWL 2- http://bloody-byte.net/rdf/dc_owl2dl/dcterms

[4] The Common European Research Information Format Ontology. CERIF Ontology 0.2. http://eurocris.org/ontologies/cerif/1.3/index.html

[5] MarcOnto – Integration Ontology for Bibliographic Description Formats - http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.155.201&rep=rep1&type=pdf

workingPaper, report, other (dataset) and other (research project) corresponds to the *type* of "*expositive*" *interactivity* and a "*very low*" level, which is specified in the following rule (1):

Type (?t) ^ (driver (?t, "article") ∨ driver (?t, "book") ∨ driver (?t, "bookPart") ∨ driver (?t, "conferenceObject") ∨ driver (?t, "doctoralThesis") ∨ driver (?t, "masterThesis") ∨ driver (?t, "bachelorThesis") ∨ driver (?t, "patent") ∨ driver (?t, "review") ∨ driver (?t, "workingpaper") ∨ driver (?t, "report") ∨ driver (?t, "other")) -> type(?i, "expositive") ^ level(?i, "very low") (1)

Similarly, when we have *type driver* other with its corresponding instance SNRD: photograph, plan, map, slide, poster, satellite imaging, x ray, transparency, microscope slide, film, documentary and video recording also corresponds to the *type* of *expositive* interactivity, but as to the *level*, it would be *low* because there is little user intervention, the rule would be as follows (2):

Type (?t) ^ driver (?t, "other") ^ (snrd (?t, "fotografía") ∨ snrd (?t, "plano") ∨ srnd (?t, "mapa") ∨ snrd (?t, "diapositiva") ∨ snrd (?t, "póster") ∨ snrd (?t, "imagensatelital") ∨ snrd (?t, "poster") ∨ snrd (?t, "radiografía") ∨ snrd (?t, "transparencia") ∨ snrd (?t, "diapositiva de microscopio") ∨ snrd (?t, "pelicula") ∨ snrd (?t, "documental") ∨ snrd (?t, "videograbación")) -> type(?i, "expositive") ^ level(?i, " low") (2)

So, when we have a document of *driver type* as conference object and its corresponding snrd conference document, in this case will need to add the rule (3) indicating that this *type* of object is the *learningResourceType* as *lecture*.

Type(?t) ^ driver(?t, "conference object") ^ snrd("?t, "document de conferencia") -> learningResourceType(?t, "lecture") (3)

C. *Structure and services proposed*

In order that the shared vocabulary can be used, relate it to local ontologies and also provide search services and warehousing, considering they are two of the main functions of the LOR, the structure of Figure 4 is proposed.

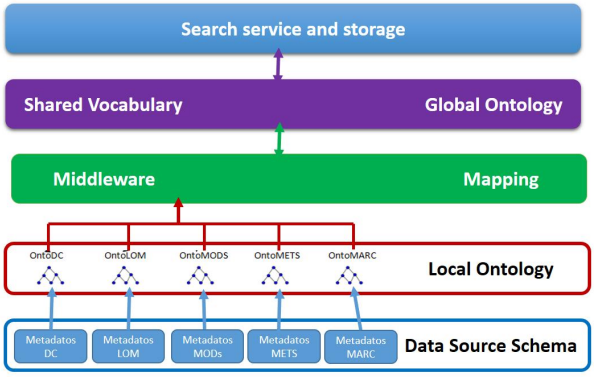

Fig. 4. Structured and services proposed

In the above structure, we can distinguish different levels that make up the model proposed on the basis of a hybrid approach. On the first level is the data source schema, represented by the metadata of different metadata standards implemented by LOR. Each of these schemes are represented locally through local ontologies. Among the latter and the global ontology, which represents the shared vocabulary, it is an intermediate layer called "middleware" that aims to make the mapping between the concepts of local and global ontologies, in order to respond to the proposed search services and deposit located services in the Upper level.

This mapping intends to handle the semantic heterogeneity, i.e. a correct interpretation of the criteria, for example, regardless of metadata standard that uses a LOR. Also, completing the description of LO stored in repositories, such as the deposit of an LO repository takes the standard DC and it requires more detail including among its metadata the SNRD recommendations and educational metadata, suggested by LOM

IV. Conclusions and Future Work

In this paper it was presented a model to achieve interoperability between institutional repositories using different metadata standards. For this, it is presented an approach based on ontologies where local ontologies are combined with an ontology that represents a common vocabulary.

Such vocabulary is based on metadata DC considering the recommendations of SNRD including them through new concepts, rules and relationships; as well as metadata suggested by LOM for the specific description of the educational part. The choice of DC as a basis for building the common vocabulary is based on the need to comply with the OAI-PMH protocol so that the repository can be harvested.

While the shared vocabulary is implemented through the corresponding ontology, it remains to be done the mapping between local ontologies and global ontology considering that today specific ontologies DC standards, LOM, MODS, MARC and CERIF are available.

References

[1] JORUM+ Project. (2004). The JISC Online Repository for [learning and teaching] Materials.

[2] IEEE. Learning Object Metadata Working Group . Online Version 2002. Available on Web site: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1032843

[3] D. A. Wiley, "Connecting learning objects to instructional designa theory: A definition, a metaphor, and a taxonomy" Utah State University. Digital Leaning Enviroment research Group, The Instructional Use of Learning Objects: Online Version. 2000. Disponible en: http://reusability. org/read/chapters/wiley.doc.

[4] COAR Roadmap. Future Direction for Repository Interoperability. Online Version 2015. Disponible en: https://www.coar-repositories.org/activities/repository-interoperability/coar-interoperability-project/coar-interoperability-roadmap/

[5] COAR The Current State of Open Access Repository Interoperability. Online Version October, 2012. V2. Disponible en: https://www.coar-repositories.org/es/activities/repository-interoperability/coar-interoperability-project/the-current-state-of-open-access-repository-interoperability-2012/.

[6] IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. Online Version 1995. Disponible en: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=467542

[7] E. Rodrigues, “Challenges and Opportunities of Interoperability for Open Access Repositories”, China Open Access Week. Beijing, October 2012.

[8] H. Garrido Arenas, M. Lisowska Navarro, “¿Estamos preparados para trabajar en red? Interoperabilidad: desafíos para la región latinoamericana”. XX ISTEC (Ibero-American Science & Techonology Education Consortium) General Assembly. Puebla, México. Marzo 2014.

[9] T. R. Gruber, “A translation approach to portable ontology specification. Knowledge acquisition”, 5(2), 199-200. 1993

[10] [The Open Archive Initiative Protocol for Metadata Harvesting. Protocol Version 2.0. Disponible en: http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm#Introduction

[11] M. H. Sarip Maarof and Y. Yahya, "LORIuMET: Learning Object Repositories interoperability using metadata," Information Technology, 2008. ITSim 2008. International Symposium on, Kuala Lumpur, 2008, pp. 1-5.

[12] Gómez-Dueñas, L. F., “Modelos de interoperabilidad en bibliotecas digitales y repositorios documentales: Caso Biblioteca Digital Colombiana-BDCOL”, 2009.

[13] J. Vian, R. L. R. Campos, C. E. G. Palomino and R. A. Silveira, "A Multiagent Model for Searching Learning Objects in Heterogeneous Set of Repositories," Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on, Athens, GA, 2011, pp. 48-52.

[14] Piedra, N., Chicaiza, J., Quichimbo, P., Saquicela, V., Cadme, E., López, J., ... & Tovar, E. (2015). Marco de trabajo para la integración de recursos digitales basado en un enfoque de web semántica. RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação, (SPE3), 55-70.

[15] Agosti, M., Ferro, N., & Silvello, G. (2016). Digital library interoperability at high level of abstraction. Future Generation Computer Systems, 55, 129-146

[16] H. Wang and Z. Ye, "Building Multi-Level Data Warehouse Based on Hybrid-Ontology," *Computer Network and Multimedia Technology, 2009. CNMT 2009. International Symposium on*, Wuhan, 2009, pp. 1-4.

[17] Solomou, G., Pierrakeas, C., & Kameas, A. (2015). Characterization of educational resources in e-learning systems using an educational metadata profile. Journal of Educational Technology & Society, 18(4), 246-260

[18] Morales Morgado, E. M. M., Campos Ortuño, R. A. C., Yang, L. L., & Ferreras-Fernández, T. (2014). Adaptation of Descriptive Metadata for Managing Educational Resources in the GREDOS Repository. International Journal of Knowledge Management, 10(4), 50–72

[19] Sandobal Verón, Valeria C.; Ale, Mariel; Gutiérrez, María de los Milagros.; “DCOntoRep: hacia la interoperabiliad semántica de Repositorios Institucionales de Acceso Abierto”, Proceedings of the 1st Argentine Symposium on Ontologies and their Applications (SAOA 2015) Rosario, Argentina, September 2-3, 2015