# Learning Object Repositories with Federated Searcher over the Cloud

Fernando De la Prieta, Ana B. Gil, Antonio Juan Sanchez Martín, and Carolina Zato

Department of Computer Science and Automation Control, University of Salamanca,
Plaza de la Merced s/n, 37007, Salamanca, Spain
{fer,abg,anto,carol_zato}@usal.es

**Abstract.** The education sector is a significant generator, consumer and depository for educational content. Educators and Learners have access to technologies that allow them to obtain information ubiquitously on demand. The problems arising from the integration of educational content are usually caused by the vast amount of educational content distributed among several repositories. This work presents a proposal for an architecture based on a cloud computing paradigm that will permit the evolution of current learning resource repositories by means of cloud computing paradigm and the integration of federated search system.

**Keywords:** Learning Objects, digital repositories, information retrieval, cloud computing.

## 1 Introduction

The learning object paradigm is one of the main advances within the field of reutilization of educational resources. Formally, a Learning Object (LO) is defined [12] as *any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning*. In short, practically any educational resource (lesson, task, graph, subject, etc.), but there is a clear consensus that an LO must be the minimal reusable self-contained unit of learning content with a specific objective [2][14]. The paradigm is based in the fact of any education resource can be described by means for metadata, independently of its topic, format or size. The encapsulation of education resources in the form of metadata makes their digital distribution possible and therefore their reutilization, because this metadata allows making a first approach to the educational resource. The metadata schema is standardized. In fact, there are currently many standards such as DublinCore [5], IEEE LOM [12], etc. The existence of standards facilitates the management of the resources, enabling the interoperability among systems that use compatible standards.

LOs are stored in specific digital libraries, called Learning Object Repositories (LOR). Currently there is a significant growth of LOR as part of the hidden web in large databases. These systems typically provide a web interface to allow the

searching of education resources through the metadata. On of the main characteristic of these LORs is their heterogeneity [8] and therefore the interoperability among LORs is limited. However, to deal with this issue, they typically have a layer (interface) to makes possible the external access and hence, the interoperability. External search agent (a client or another LOR) can access. There are different standards or specifications that focuses on this interoperability layer, mainly OAI-MPH (*Open Archives Initiative for Metadata Harvesting*)[15] and SQI (*Simple Query Interface*) [6].

However, despite the theoretical advances done within this paradigm, the reality shows that its implantation in the real live is still limited [11]. From our point of view, there are two main problems. Firstly, from the usability viewpoint, the data that the authors assign to each descriptor of the metadata (independently of the specific standard) is very important, because this data that is used for searchers and if it is not correct, the results of the searches will be incoherent. To this end, it is necessary to follow a traceable process from the creation of an educational resource to the creation of its metadata in order to establish a metadata structure that is consistent, relevant and interpretable. However, the existence of many standards, the interoperability among them, the difficult to use the authoring tools and style of explanations by the authors, exacerbate the problem. And secondly, from a technological point of view, the heterogeneity of the repositories and their malfunction (as it is shown in the following section) constitute one of the main weakness of the paradigm.

This work presents the evolution of platform CLOR [3], by means of the integration of the federated searcher AIREH [9] in order to produce a clear advantage in the context of LO paradigm. On the one hand, CLOR (*Cloud-based Learning Object Repository*) is the present of a new generation of LOR because it is deployed into a cloud platform and makes use of the advantages of this computational paradigm (non-SQL databases, unlimited storage, etc.). On the other hand, AIREH (*Architecture for Intelligent Retrieval of Educational content in Heterogeneous Environments*) is a platform that makes possible the federated searches among many LORs and it integrates a recommender system [13].

This paper is structure as follow next section includes an study of the state of the art of current LOR and a real evaluation. Section 4 presents the proposed system and, finally, last section includes the results and conclusions.

## 2      Open Issues – Learning Object Repositories

LOs are commonly stored in repositories, which are characterized by their heterogeneity. The deployment infrastructure can basically be either distributed or centralized. Taking into account that an LO is formed by a digital resource and its metadata, there are four kinds of possible infrastructures [8]: (i) centralized resources and centralized metadata, (ii) centralized resources and distributed metadata, (iii) distributed resources and centralized metadata and (iv) distributed resources and distributed metadata. Furthermore, three kinds of storage strategies can be distinguished [8]: (i) *File-based,* which uses files with predefined formats and an

index-based management; (ii) *Database-based,* which uses any kind of database, and is the most extended method; and (iii) *Persistent objects-based,* where the LO are stored as serialized objects.

The main problem is that LORs still do not implement any abstraction layer that can encapsulate the internal logic of the repository. Consequently, the search process and LO harvesting is a slow process, which require the manual intervention of users who must reuse the learning resources.  In this sense, to isolate this internal heterogeneity (storage techniques and models), there are interoperability layers which serve as a middleware layer between the repository and the clients, are (i) OAI-MPH (*Open Archives Initiative Protocol for Metadata harvesting*) [15] which is a protocol that provides a technology-independent framework for retrieving documents or resources, thus enabling interoperability among systems; and (ii) SQI (*Simple Query Interface*) [6] that is formed by a set of abstract methods based on web services. SQI is also is neutral in terms of the format of results as well as query language. This interfaces supports synchronous/asynchronous and stateful/stateless queries.

The state of the art shows a high heterogeneity in existing standards. Therefore, a study of LOR has been performed in order to analyze the real situation. In general, the systems in which this layer is included, suffer from various problems such as:

- The problems associated with the monolithic structure of LOR, which does not allow external management with the flexibility and power necessary to ensure easy interoperability, and dispersed and heterogeneous sources.
- the absence of automatic mechanisms that control the technical quality, semantics and syntax of LO, ensuring the correct specification of such LOs in any of the metadata schemas that describe them.

The study includes the analysis of the following LORs: *Acknowledge*, *Agrega*, *Ariadne*, *AriadneNext*, *CGIAR*, *EducaNext*, *LACLO-FLOR*, *LORNET*, *MACE*, *Merlot*, *Nime*, *OER Commons* and *Edna Online*. It consists of performing 60 queries to each LOR through an SQI layer that the repositories provide.  All of them use IEEE LOM [12] as metadata schema and VSQL [1717] as query language. Additionally, the majority of them are stateless (65%), and all of them have synchronous interfaces, but only 4 have the asynchronous interface.

The test shows that 6 of the 14 repositories do not work or are unavailable and they have to be removed from the scope of this study (Ariadne, AriadneNext, EducaNext, Nime and EdNa Online). MACE and LOCLO-FLOR produce an error in the authentication. After this step, this test is reduced to only four repositories Acknowledge, Agrega, LORNET and Merlot. The latter three are perfectly valid and all SQI methods work perfectly; however the repository Acknowledge only implements the essential methods to perform queries (Note that SQI specification does not force the implementation of all methods of the specification).

The graph on Fig. 1 shows the average of the time response a federated search is performed to these four alive LOR, the total average in the 60 queries is 3,718 seconds. The following graph in Fig. 2 shows the number of results retrieved per federated search during the test. The average of results is 24,24 result/querie.
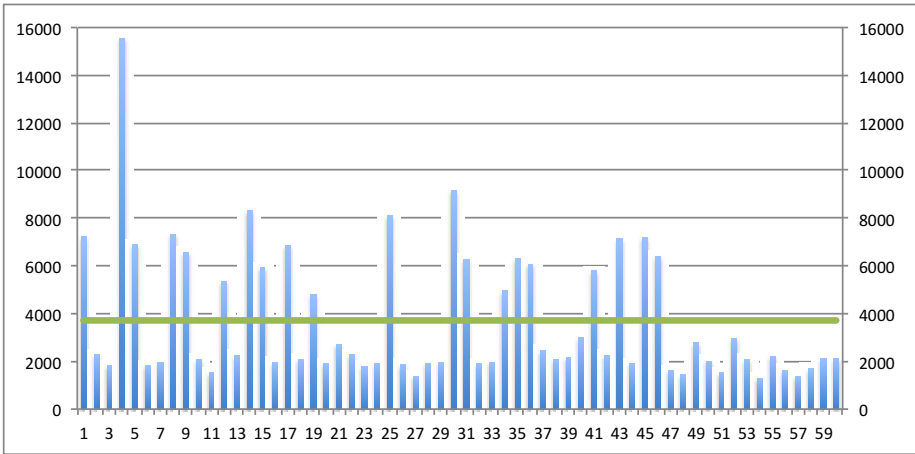
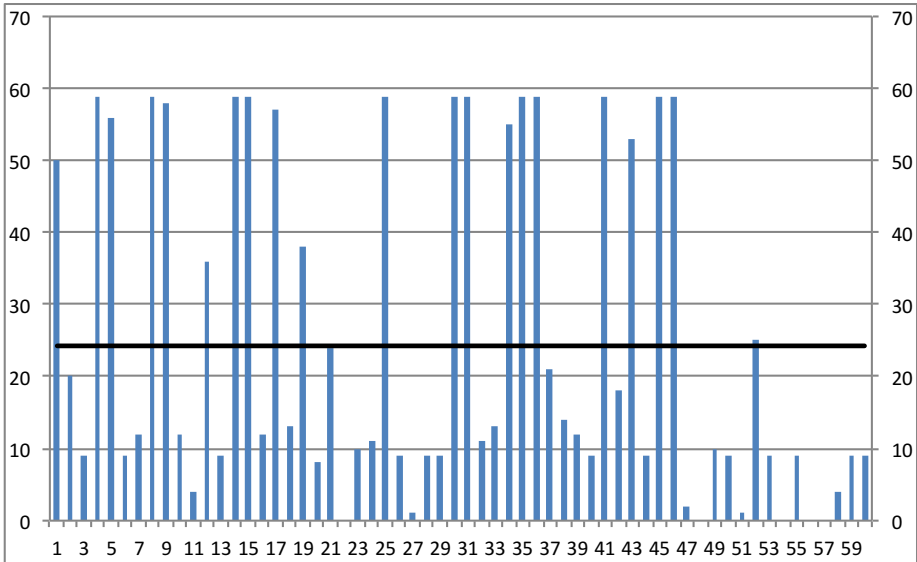**Fig. 1.** Time average of a federated search in 4 repositories



**Fig. 2.** Number of results of a federated search in 4 repositories

As it is possible to observe, the performance of the LOR is not appropriate. In order to deal with this problem, new LOR architectures have to be proposed and developed. This new generation of LOR must ensure the availability of resources and interoperability, permitting federated searches from external clients.

# 3 Proposal Architecture

These problems require solutions that are adapted to the heterogeneity. The solution should enable a centralized global search and the effective reuse of resources by the end user. As it is advanced on the introduction of this work is the integration of AIREH and CLOR in order to establish a system that not only allow the federated search among several LOR, but also the retrieve and safe storage of the retrieved LO.
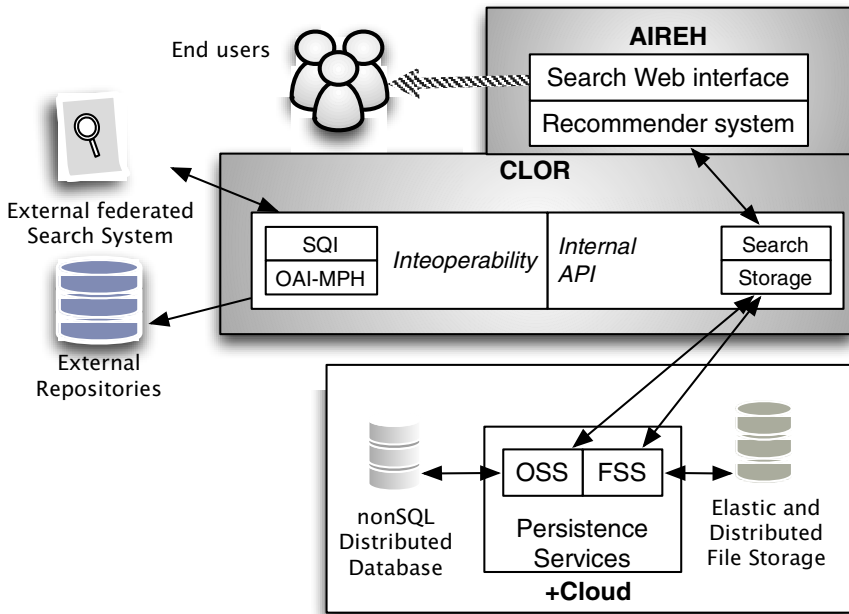


**Fig. 3.** CLOR and AIREH integration

The components of system, which are showed in Fig. 3, are described as follow:

— AIREH unifies the search and retrieval of objects, thus facilitating the learning search process by filtering and properly classifying learning objects retrieved according to some rules. The framework is based of a virtual organization [10] of intelligent agents that allow to deal with the heterogeneity of the environment. This new architecture will solve the problems of distribution, the integration of different repositories, the abstraction of the internal logic of each repository, and the classification, storage and retrieval of LOs, in a completely transparent way. In addition to adding capacities, such as simple scalability, to possible situations involving the use of new protocols, the architecture also adds internal logical repositories, cataloging or heterogeneous applications designed to cover services related features. AIREH implements Case Based Reasoning (CBR)[7] [1] that uses previous search information to rank the items that best suit the needs of the

application user based on previously obtained information. It uses the profile information of each user as well as their educational information (content-based filtering).

— CLOR provides to AIREH the capacity to store the profile of each user as well as the persistence of retrieved LO (not only the metadata, but also the education resource). It is complemented with different interoperability layers, such as SQI or OAI-MPH, which will ensure the communication with other LORs and federated searches from external clients. It is framed at the platform level within Cloud services. Its main task is to encapsulate the communication with the lower layers of the Cloud platform that provides the need computational resources for the storage the educational resources in the web service file storage system and the metadata (in JSON format) associated with each resource into a non-SQL database. The main advantage is that it permits storing any kind of metadata independent of its structure or schema, that is, its standard. Furthermore, queries about the LO will be performed very quickly thanks to the use of a document-oriented database [16].

— +Cloud platform [4] provides, such as storage and databases. This platform is based on the Cloud Computing paradigm. This platform allows offering services at the PaaS and SaaS levels. The SaaS layer is composed of the management applications for the environment (virtual desktop, control of users, installed applications, etc.), and other more general third party applications that use the services from the PaaS layer. The components of this layer are the identity Manager, a File Storage System base on Web services, and an Object Storage Service ch provides a simple and flexible schemaless data base service oriented towards documents.

## 4      Results and Conclusions

The proposed system has been evaluated by performing a battery of tests to validate their efficiency in real environments. Evaluation metrics from information retrieval fields were adopted. The two most commonly used evaluation measures are precision (the fraction of documents retrieved by the system that are also relevant to the query) and recall (the fraction of the relevant documents present in the database that are retrieved by the system). Fig. 4 shows the number of relevant LO retrieves per query in comparison with other repositories, and Fig. 5 shows an average of time expended in retrieve LO per query.
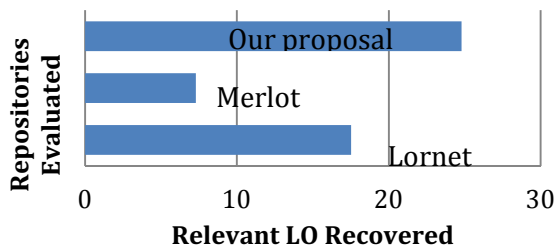


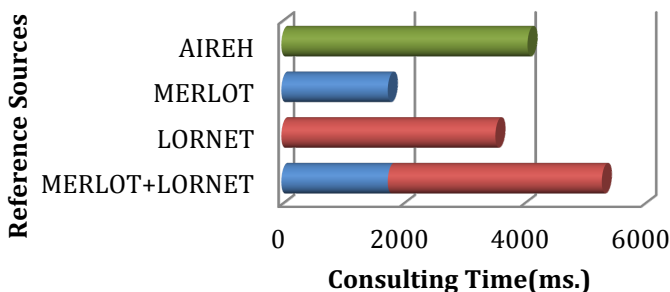**Fig. 4.** Relevant LO recovered per query

**Fig. 5.** Average time of queries

The first problem of LO paradigm are the incoherence in the medatada because the labeling process, which is basically done by hand, generates documents with serious shortcomings as there are many deficiencies related to the lack key attributes in the description. This makes it difficult, or impossible in some cases to study this aspect.

The second problem is the heterogeneity of the repositories and their malfunction. The proposed system tries to minimize this second problem, because it deals with the existing open issues:

1. The proposed model allows dealing with the heterogeneity of current and future standards since it is based on a non-relational database.
2. Cloud computing paradigm allows offering services with the same level of quality independently of its demand.
3. The low linkage among components permits implementing many interoperability layers without needing to upgrade to other modules.
4. The integration of AIREH allows filtering the results according to certain parameters related to the quality of the retrieved metadata documents; the user context information is then integrated into the use of the LOs.

# References

1. Bajo, J., Corchado, J.M.: Evaluation and monitoring of the air-sea interaction using a CBR-Agents approach. In: Muñoz-Ávila, H., Ricci, F. (eds.) ICCBR 2005. LNCS (LNAI), vol. 3620, pp. 50–62. Springer, Heidelberg (2005)
2. Chiappe, A., Segovia, Y., Rincon, H.Y.: Toward an instructional design model based on learning objects. Educational Technology Research and Development 55, 671–681 (2007)

3. de la Prieta, F., Bajo, J., Marín, P.A.R., Méndez, N.D.D.: A New Generation of Learning Object Repositories Based on Cloud Computing. In: Casillas, J., Martínez-López, F.J., Vicari, R., De la Prieta, F. (eds.) Management Intelligent Systems. AISC, vol. 220, pp. 99–106. Springer, Heidelberg (2013)
4. De la Prieta, F., Rodríguez, S., Bajo, J., Corchado, J.M.: A Multiagent System for Resource Distribution into a Cloud Computing Environment. In: Demazeau, Y., Ishida, T., Corchado, J.M., Bajo, J. (eds.) PAAMS 2013. LNCS, vol. 7879, pp. 37–48. Springer, Heidelberg (2013)
5. Dublin Core Metadata Initiative. DCMI Metadata Terms
6. European Committee for standardization – Cen Workshop Agregament. A simple Query Interface Specification for Learning Repositories. Ref. No.: CWA 15454:2005 E (2005)
7. Fdez-Riverola, F., Corchado, J.M.: CBR based system for forecasting red tides. Knowledge-Based Systems 16(5), 321–328 (2003)
8. Frango, I., Omar, N., Notargiacomo, P.: Architecture of Learning Objects Repositories. Learning Objects: standards, metadata, repositories & LMS, pp. 131-155 (2007)
9. Gil, A.B., De la Prieta, F., Rodríguez, S.: Automatic Learning Object Extraction and Classification in Heterogeneous Environments. In: Pérez, J.B., Corchado, J.M., Moreno, M.N., Julián, V., Mathieu, P., Canada-Bago, J., Ortega, A., Caballero, A.F. (eds.) Highlights in PAAMS. AISC, vol. 89, pp. 109–116. Springer, Heidelberg (2011)
10. Gómez-Sanz, J.J., Pavón, J., Garijo, F.: Meta-models for building multi-agent systems. In: Proceedings of the 2002 ACM Symposium on Applied Computing, pp. 37–41 (2002)
11. Hodgins, H.W.: The future of learning objects. Educational Technology-Saddle brook then Englewood cliffs NJ 46(1), 49 (2006)
12. IEEE Learning Objet Metadata (LOM). Institute of Electrical and Electronics Engineers (2002), http://ltsc.ieee.org
13. López, V.F., de la Prieta, F., Ogihara, M., Wong, D.D.: A model for multi-label classification and ranking of learning objects. Expert Systems with Applications 39(10), 8878–8884 (2012)
14. Lujara, S.K., Kissaka, M.M., Bhalaluseca, E.P., Trojer, L.: Learning Objects: A new paradigm for e-learning resource development for secondary schools in Tanzania. World Academy or Science, Engineering and Technology, 102–106 (2007)
15. Lagoze, C., Van de Sompel, H., Nelson, M., Warner, S.: The Open Archives Initiative Protocol for Metadata Harvesting. Open Archives Initiative (2002)
16. Oren, E., Delbru, R., Catasta, M., Cyganiak, R., Stenzhorn, H., Tummarello, G.: Sindice. com: a document-oriented lookup index for open linked data. International Journal of Metadata, Semantics and Ontologies 3(1), 37–52 (2008)
17. Simon, B., Massart, D., Van Assche, F., Ternier, S., Duval, E.: Authentication and Session Management. Version 1.0. (2005)