MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-610

An open repositories network development for medical teaching resources

Gérard Soula[a], Stefan Darmoni[b], Pierre Le Beux[c], Jean Marie Renard[d], Badisse Dahamna[b], Marius Fieschi[a]

[a]*LERTIM, Faculté Médecine, Université de la Méditerranée, Marseille, France*
[b]*GCSIS, Faculté Médecine, Université Rouen, Rouen, France*
[c]*LIM, Faculté Médecine, Université Rennes 1, Rennes, France*
[d]*CERIM, Faculté Médecine, Université Lille 2, Lille, France*

Abstract

The lack of interoperability between repositories of heterogeneous and geographically widespread data is an obstacle to the diffusion, sharing and reutilization of those data. We present the development of an open repositories network taking into account both the syntactic and semantic interoperability of the different repositories and based on international standards in this field. The network is used by the medical community in France for the diffusion and sharing of digital teaching resources. The syntactic interoperability of the repositories is managed using the OAI-PMH protocol for the exchange of metadata describing the resources. Semantic interoperability is based, on one hand, on the LOM standard for the description of resources and on MESH for the indexing of the latter and, on the other hand, on semantic interoperability management designed to optimize compliance with standards and the quality of the metadata.

Keywords:

Open Access, Open Archives Initiative, Open repositories, Interoperability, Metadata, Teaching resource.

Introduction

In view of the quality and cost of production of digital resources, reuse of digital resources is a major issue for producers wishing to enhance their digital heritage and for the bodies of players wishing to share them. Reuse of digital resources runs up against two major difficulties: the geographical dispersion of the data and their heterogeneity.

The open repositories approach offers some elements of response. The approach is based on two complementary initiatives:

- The Open Access (OA) initiative,
- The Open Archives Initiative (OAI).

The Open Access initiative [1] is at once a commitment to a principle, to a strategy and to a financial involvement which has laid down the rules of free access to the resources.

The Open Archive Initiative [2] has specified an interoperability technique for heterogeneous and widespread repositories. This technique defines the modes of exchange of the metadata files describing the resources. It thus deals exclusively with aspects of syntactic interoperability. On the other hand, it does not manage the semantic interoperability of the metadata and the added value services related to these metadata.

The techniques formulated by the OAI have been widely used in numerous projects [3]. However, the setting up of open repositories involves more than technology alone [4] and the research performed into the role, design and management of digital libraries falls within in the scope of this problem.

Research has evolved from a narrower view emphasizing enabling technology to one that encompasses the social, behavioral and economic context in which digital libraries are used. As suggested by Borgman [5], digital libraries should be much more than search engine portals. Search and access over a set of resources, while important to any digital library, are not sufficient. Digital libraries should provide advanced services that facilitate use of resources by their target community.

Lagoze [6] proposed that this added value consists of establishing context around those resources, enriching them with new information and relationships that express the usage patterns and knowledge of the library community. The digital library becomes a context for information collaboration and accumulation and much more than just a place to find information and access it.

Digital libraries are not static and require management and long-term evaluations to determine their quality and to identify new directions for growth. Zuccala's approach [7] is to carry out evaluation using a set of qualitative and quantitative research techniques, including webometric analyses and an online survey of repository users.

This paper describes the process we used to develop an open repositories network for the diffusion and sharing of digital teaching resources created by teachers at French-language Schools of Medicine.

The architecture and management of the open repositories network are described with emphasis on syntactic and semantic interoperability. The main results are presented and the benefits and limits of the system are discussed.

Objectives

The widespread adoption of digital technology in universities and, in France, an incentives policy adopted by the Ministry of Higher Education and Research has triggered a considerable increase in the number of teaching resources. These resources are easy to consult via digital workspaces located in universities. However, availability at national and international level raises frequent problems in the absence of interoperability between the different systems

In the medical field, the CISMeF team [8] set up a centralized system for the inventory and indexing of, and access to teaching resources in French.

This efficacious system is nevertheless faced with two constraints:

- It is based upon centralized indexing which takes no account of the ever greater autonomy of the different institutions regarding indexing,
- It uses proprietary standards and technologies which greatly hinder interoperability.

With the support of the UMVF, the French Virtual Medical University [9], we have been led to rethink the overall referencing, indexing and accessing model.

Our project is focussed on the creation of an open repositories network designed to enhance the visibility, accessibility and sharing of the teaching resources produced by medical faculty.

This new model takes account of the pooling of documents produced by different institutions while avoiding centralization of the indexing system.

The orientation of the project towards an Open Access approach has led us to define a number of priority objectives:

- Establishment of a network of interoperable repositories,
- Compliance with international standards in order to ensure syntactic and semantic interoperability between the different repositories,
- Strong management, at national level, in order to optimize compliance with semantic interoperability.

The open repositories network

The metadata providers and service providers-based architecture

The architecture of the environment (Figure 1) is based on a set of shared protocols and interoperability standards aimed, firstly, at using shared metadata to describe the resources and, secondly, to allow exchange of these metadata.

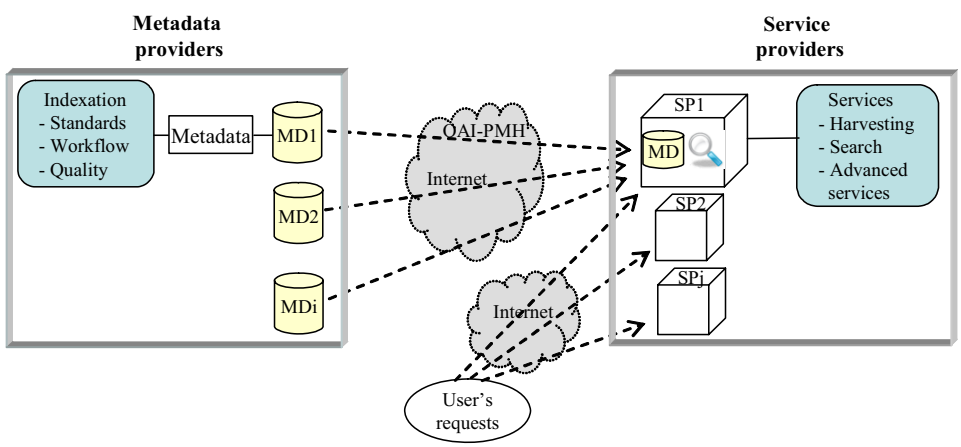

Figure 1 - Open repositories network architecture

The urbanization of this system calls on two types of players:

- The metadata providers who expose metadata passively in standardized repositories and which can be queried by service providers,
- The service providers who first harvest metadata actively by querying the metadata providers before providing services using these metadata.

Any structure can serve as a metadata provider or a service provider, or both simultaneously.

Successful implementation of this environment centered on metadata exchanges is dependent on the syntactic and semantic interoperability of the metadata.

Syntactic metadata interoperability

Syntactic interoperability is provided by the OAI-PMH protocol (Protocol for Metadata Harvesting). OAI-PMH [10] is a simple and powerful framework for metadata harvesting. With OAI-PMH, metadata providers are never directly accessed by end-users. Service-providers harvest metadata-providers and the result is used to create services covering the content of several metadata repositories. OAI-PMH specifies the syntax of 6 query verbs (Identify, ListSets, ListMetadataFormats, ListIdentifiers, ListRecords, GetRecord) used for metadata harvesting. Queries via HTTP allow recovery of metadata XML files.

Semantic metadata interoperability

The quality of the environment is conditioned upstream by the semantic interoperability of the metadata. We chose to describe the teaching resources using the LOM (Learning Object Metadata) standard proposed by the IEEE Learning Technology Standards Committee [11]. The purpose of the LOM standard is to facilitate search, evaluation, acquisition and use of learning objects, for instance, by learners or instructors or automated software processes. This multi-part standard also facilitates the sharing and exchange of learning objects, by enabling the development of catalogs and inventories while taking into account the diversity of cultural and lingual contexts in which the learning objects and their metadata are reused.

LOM defines a hierarchy of data elements for learning object metadata. At the top level of the hierarchy are nine categories: General, Lifecycle, Meta-metadata, Technical, Educational, Rights, Relation, Annotation and Classification. Each category contains sub-elements which, in turn, contain elements or sub-elements. The LOM standard includes more than 80 data elements and it was established as an extension of the Dublin Core metadata element set (15 elements) [12].

Implementations of the LOM are normally based on application profiles to meet community context-specific needs, without losing interoperability. Despite its many elements, the LOM conceptual data schema may be extended by adding new vocabularies and by adding new elements. We used this possibility and, to meet the specific needs of medical teaching resources, we defined an LOM application profile integrating the MeSH thesaurus. Each medical teaching resource is indexed in MeSH using the Describer/Qualifier/Major-Minor triplet. We used the classification category for the MeSH extension. Figure 2 is an example of an LOM XML MeSH integration.

```
<lom:taxonPath>
  <lom:taxon>
  <lom:id>D008581/Q000150/N</lom:id>
  <lom:entry>
    <lom:string language="fre"> méningite/complications/N
    </lom:string>
    <lom:string language="eng"> meningitis/complications/N
    </lom:string>
   </lom:entry>
  </lom:taxon>
</lom:taxonPath>
```

Figure 2 – Example of LOM MeSH integration

Adoption of a standard such as LOM is an essential, but insufficient, precondition. To achieve optimum metadata quality, it is vital that all players share the same understanding of the semantics of the items in the LOM and use the same vocabularies. In France, the SupLOMFR group, set up by the Ministry, is the reference providing guidance in the use of the LOM. The number of items defined in the LOM standard is very large and SupLOMFR specifies whether LOM data elements are optional, recommended or obligatory. SupLOMFR also:

- Provides information on how elements are interpreted in the French teaching context,
- Specifies the use of vocabularies and of taxonomies,
- Specifies the classification schemes.

SupLOMFR is making a major contribution to the management of semantic interoperability.

Services

The function of the service providers is to trigger relevant harvesting of metadata, i.e. to target metadata providers according to the profile of potential users. Each service provider offers users a search engine to exploit the harvested metadata.

Results

Technical implementation

Under the aegis of the Ministry, the French ORI-OAI [13] consortium is assisting in the technical deployment of the open repositories network by integrating open source tools (Orbeon form builder, OSWorkflow, LIUS, Nuxeo). Seven packaged modules are available: document storage, OAI repository, OAI harvester, workflow manager, vocabulary manager, indexing engine and search engine. These modules dialogue with one another using web services. Each module is highly configurable to meet specific contexts.

Metadata production workflow

Good quality metadata are a key component in the successful implementation of an open repositories network. The need for good quality metadata must be fully understood by the community.

A solely teacher-generated metadata model has many limitations. Teachers may lack knowledge of indexing principles and are more likely to generate insufficient and poor quality metadata. Both teachers and metadata specialists play important roles in the metadata production process. This collaborative approach needs to specify the metadata production workflow.

Metadata creation should be identified as an incremental process and as a shared responsibility. This responsibility should be distributed in a practical and reasonable way within the institution. The creation and refinement of a workflow is dependent on operational and strategic factors. Several more or less complex workflows can be established. The main difference between workflow models is in the number and the roles of the players involved in the process.

A four-level workflow model including the teacher, information and communication technology specialist, librarian and legal department seems to fit the French university context. Each player produces metadata in his/her area of competence. This workflow can be represented using a state transition diagram as in Figure 3.

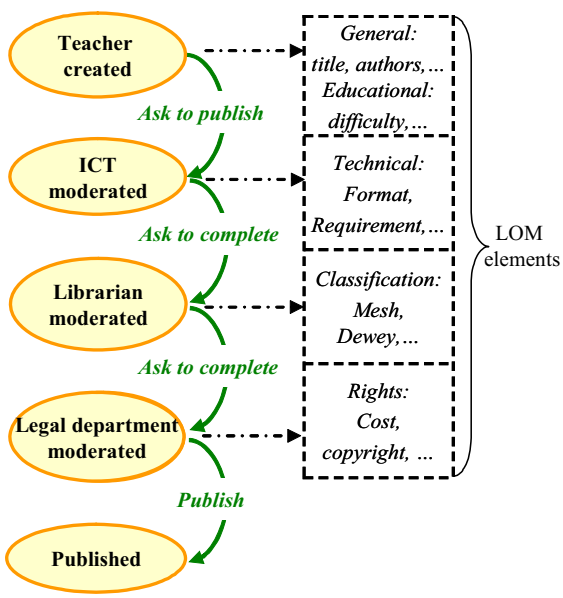

Figure 3 - A four level workflow model

Metadata creation comprises five successive states from the teacher-created state to the published state. The Metadata creation workflow starts with a teacher initiative in which general elements are produced. ICT and librarian complete the metadata. The Metadata creation workflow ends with a legal department decision after validation of the rights aspects.

Operational providers

Today, the network is based on several operational metadata providers and service providers.

During the last quarter of 2008, 4200 resources indexed by CISMef and 2000 resources indexed by Canal U, the French Higher Education and Research Web TV channel, were integrated into the first repositories.

To allow integration of the CISMeF 4200 teaching resources in the ORI-OAI repository, the CISMeF team developed a mapping function to restructure the original XML files according to the LOM metadata schema. CISMeF information about indexations (MeSH keywords, MeSH qualifiers and CISMeF resources types) and CISMeF metaterms for categorization are mapped to LOM metadata elements. The transformation from one metadata schema to another is ensured by XSLT processing. Figure 4 shows an example of mapping.

In 2009, other universities (Aix-Marseille, Paris, Rennes,…) joined the project. The CISMeF ORI-OAI repository will now receive an additional 4000 teaching resources.

An overall picture, at national level, of the network deployment covering all the themes, is provided by the French thematic digital universities portal [14]. The 12300 resources listed in July 2009 reflect the deployment status of the two themes involving medicine (4200) and science and techniques (8100). All the other themes (law, economy,…) will gradually be added to the network.

```
<CISMeF:learningresourcetype>
  <CISMeF:descriptionRessType>
    educationalcourses
  </CISMeF:descriptionRessType>
</CISMeF:learningresourcetype>
```

```
<lom:educational>
  <lom:learningResourceType>
    <lom:value>
      lecture
  </lom:value>
  </lom:learningResourceType>
</lom:educational>
```

Figure 4 – An example of mapping from CISMeF schema to LOM schema

Discussion

The choice of the open repositories approach for the diffusion and sharing of medical teaching resources has provided us with elements of response towards a solution of the heterogeneity issue.

The OAI-PMH protocol has enabled us to achieve complete control of syntactic interoperability. On the other hand, semantic interoperability is more difficult to ensure. Despite the adoption of a standard such as the LOM, the players in the community are not all equally demanding regarding observance of standards and control of the quality of the metadata. The increase in the number of players and indexing facilities raises the issue of indexing quality. Improvements in quality should be attainable by defining a supervised workflow designed to harmonize indexing practices. The management of semantic interoperability is a key factor in the creation of a high quality open repositories network.

For users, one of the added values is the search engines integrated in the service providers. These search engines are often general in scope and do not take sufficient account of user profiles. This partly explains the utilization rate which fails to match the efforts devoted to the system [15]. One avenue of research is to establish connexions between the resources in order to take account of the context [16]. The OAI-ORE (Object Reuse and Exchange) [17] aims to describe and exchange compound and semantically linked objects and is working in that direction.

Conclusion

This paper described an open repositories network development for the diffusion and sharing of medical teaching resources.

Driven by the Ministry and thanks to the efforts of the ORI-OAI consortium, the open resources approach is being rolled out in France, and in all sectors, and not only in medicine.

The next steps will seek to extend the network of players, develop services which take into account user profiles and evaluate usage.

The open repositories approach can be applied to documents other than teaching resources such as theses, training curriculum, or scientific publications. The aim is thus to enhance a whole digital heritage backed by high-quality semantic interoperability management.

Acknowledgments

This work was partially funded by the French Virtual Medical University.

References

[1] Budapest Open Access Initiative, http://www.soros.org/openaccess (accessed June 25, 2009)

[2] Open Archives Initiative, http://www.openarchives.org (accessed July 3, 2009)

[3] Heery R. and Anderson S., Digital repositories review, UKOLN and Arts and Humanities Data Service, University of Bath, 2007

[4] Jacobs N.(Ed.), Open Access: Key Strategic, Technical and Economic Aspects, Oxford : Chandos Press, 2006

[5] Borgman CL., What are digital libraries? Competing visions, Information Processing & Management, 35, (2009), 227-243

[6] Lagoze C., Krafft DB., Payette S. and Jesuroga S., What Is a Digital Library Anymore, Anyway? Beyond Search and Access in the NSDL, D-Lib Magazine, 11 (2005)

[7] Zuccala A., Oppenheim C. and Dhiensa R., Managing and evaluating digital repositories, Information Research, 13 (1), (2008)

[8] Darmoni SJ., Leroy JP., Baudic F., Douyère M., Piot J. and Thirion B., CISMeF: a structured health resource guide. Methods Inf Med. 39 (2000), 30-5

[9] Le Beux P., Le Duff F., Fresnel A., Berland Y., Beuscart R., Burgun A, Brunetaud JM., Chatellier G., Darmoni SJ., Duvauferrier R., Fieschi M., Gillois P., Guille F., Kohler F., Pagonis D., Pouliquen B., Soula G. and Weber J, The French Virtual Medical University, Stud Health Technol Inform 77 2000;554-62

[10] OAI-PMH protocol, http://www.openarchives.org/OAI/openarchivesprotocol.html (accessed July 3, 2009)

[11] LOM, http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html (accessed Sept 3, 2009)

[12] Dublin Core, http://dublincore.org (accessed, Sept 3, 2009)

[13] ORI-OAI, http://www.ori-oai.org (accessed Sept 3, 2009)

[14] French thematic digital universities portal, http://www.universites-numeriques.fr (accessed Sept 3, 2009)

[15] McMartin F., Iverson E., Wolf A., Morrill J., Morgan G. and Manduca C., The use of online digital resources and educational digital libraries in higher education, Int J Digit Libr 9 (2008), 65–79

[16] Aschenbrenner A., Blanke T., Flanders D., Hedges M. and O'Steen B., The future of repositories? patterns for (cross-)repository architectures, D-Lib Magazine, 14 (2008)

[17] OAI-ORE, http://www.openarchives.org/ore/ (accessed Sept 1, 2009)

Address for correspondence

Gérard Soula, LERTIM, Faculté de Médecine, Université de la Méditerranée, 27 boulevard Jean Moulin, 13385 Marseille Cedex 05, France
E-mail: gerard.soula@ap-hm.fr