Bridging the semantic gap in standards-based learning object repositories

Elisa Rodríguez, Miguel-Angel Sicilia, Sinuhe Arroyo

Information Engineering Research Unit
Computer Science Dept., University of Alcalá
Ctra. Barcelona km. 33.6 – 28871 Alcalá de Henares (Madrid), Spain
elisa.rodriguez@alu.uah.es, msicilia@uah.es,
sinuhe.arroyo@alu.uah.es
http://www.cc.uah.es/ie

Abstract. Semantic learning object repositories are essentially characterized for the storage of object metadata in formal languages and linked to ontologies, which becomes the key asset for advanced retrieval and composition functionality. The mismatch between current practice based on XML records and the richer semantics required for semantic systems raises the need for bridging both kind of components. This practical matter requires an examination of two important technical aspects. On the one hand, a re-examination of current interoperability specifications is required for compatibility with the semantic approach. And on the other hand, there is a need to develop effective mapping techniques from metadata elements in the form mandated by established standards and the ontological representations required by semantic repositories. This paper provides an initial exploration of these two aspects. Concretely, overall architectural issues are examined taking as a point of departure the IMS DRI specification and the design of a generic and flexible mapping mechanism from IEEE LOM XML records to a WSML ontological representation of learning objects is provided. These two elements are being used as the foundation for an approach to harvesting metadata from existing LOM repositories to an *ad hoc* ontology of learning objects based on LOM.

Keywords. Learning objects, learning object repositories, ontologies, metadata, IMS DRI.

1 Introduction

The application of the Semantic Web vision (Berners-Lee et al., 2001) to the concept of learning objects repositories results in the requirement for learning object metadata with formal semantics and expressed in terms of one or several ontologies. Several accounts for such core learning object ontology can be approached (Sicilia, Sánchez-Alonso and Soto, 2005), but in any case, a problem of mapping existing "plain", non-semantic metadata records to formal ontology representations exists. This problem is relevant both for harvesting and search, and the fact that a standard ontology of learning object is not available requires a generic solution capable of accommodating different ontology languages and present and future technologies.

The problem of integrating semantic repositories with standard ones can be approached from different perspectives. Here we focus on interoperability in two inter-related aspects that address two important levels for the problem:

- Architectural issues, from a functional perspective.
- Concrete language-to-language issues, where the input language is typically an XML-based representation, and the target language is one of the available ontology languages as OWL.

Architectural issues must be approached from existing standards. Since there are a variety of specifications for data interchange between digital repositories, the more generic option has been selected as a first target. Concretely, the IMS DRI specification is examined in this paper from the viewpoint of interoperability with current practice.

Language translation issues are specific to the input and target language, but a large commonality between them exists in practice, so that ideas and techniques used for a given pair can be useful to others. The concrete design described here comes from the pragmatic requirements of a project, but the main ideas can be used for other representations.

The rest of this paper is structured as follows. Section 2 describes how the IMS DRI specification can be extended to deal with semantic representations, and which new components are required for it. Then, a concrete example of bridging from IEEE LOM metadata records to an ontology representation in WSML is described. Finally, conclusions and future research directions are sketched in Section 4.

2 Architectural issues in the integration of semantics in digital repositories

The architecture of the repository is based in the IMS DRI ***Phase 1*** specification version 1.0 (IMS 2003) which aims to "*provide recommendations for the interoperation of the most common repository functions*". IMS DRI 1.0 allows for the definition of metadata-only repositories: "Repositories may hold actual assets **or the meta-data that describe assets**". This is the typical arrangement of many current LO repositories, and it is considered as the option in the rest of this paper.

DRI defines the interactions between core functional components (*resource utilizers* and *repositories*) that support interoperability, including:

- SEARCH, GATHER, (ALERT)/EXPOSE
- REQUEST/DELIVER
- SUBMIT/STORE

Note: ALERT is a core function, but is not addressed within this version of the DRI specification. The DRI Project Group is focusing on these core interoperability functions within the functional architecture. The following functional diagram of the IMS DRI specification depict the core interaction addressed (the rest of the elements are blurred since they are not covered by Phase 1).

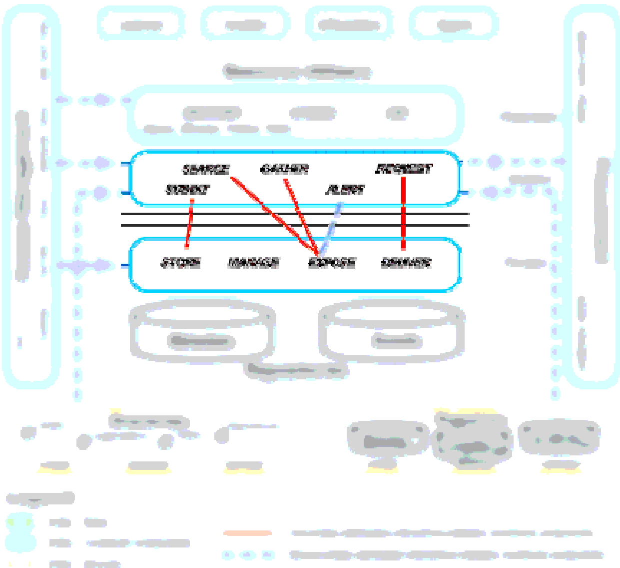

Fig 1. IMS DRI first phase functional model

In what follows, the results of analyzing the implications of semantics in the overall architecture are described.

The *Search* reference model defines the searching of the meta-data associated with content exposed by repositories. Compatibility of SEARCH/EXPOSE in semantic repositories must be provided by some kind of mediation layer. This raises the need for additional elements:

— A **query mediator**, which takes as input either Z39.50 or XQuery queries and transform them to a search in the internal format of the semantic repository.
— A **SEMANTIC-SEARCH** function to directly search in semantic terms.

Since our model for semantic learning object repositories focuses on the storage of (semantic) metadata, REQUEST/DELIVER and SUBMIT/STORE will be a normal IMS DRI function. However, the SUBMIT function is actually a metadata-transfer function, since IMS packages usually contain metadata descriptions for learning objects. This raises the need for two elements in the architecture:

- An import/export facility from/to IMS metadata to the semantic representation of metadata.
- An extended function specific to the storage of semantic metadata. A basic function ASSERT is defined for that. This affects requisites (i), (ii) and (iii) for this prototype.

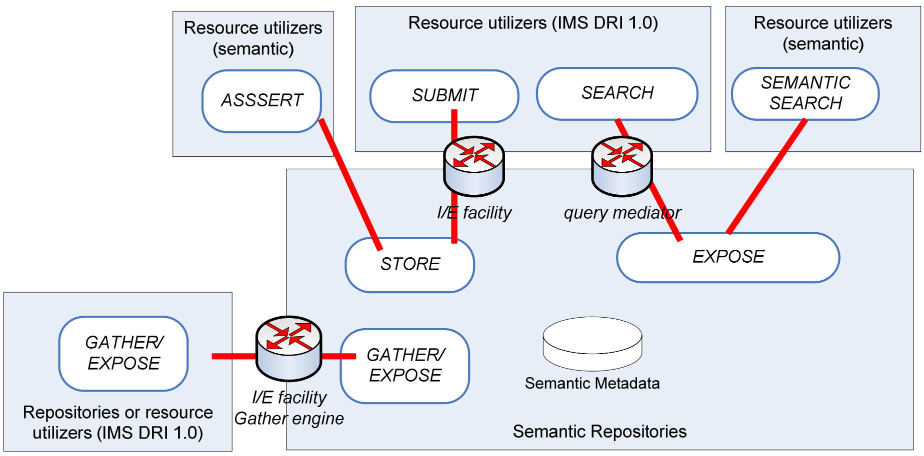

Fig 2. IMS DRI combined with a semantic repository

The IMS DRI states: "The Gather reference model defines the soliciting of metadata exposed by repositories and the aggregation of the meta-data for use in subsequent searches, and the aggregations of the meta-data to create a new meta-data repository". The aggregations of metadata are to be handled by the same import/export facility described above. However, the soliciting in "Pull" mode requires an additional consideration. Concretely:

- A gather engine with the added functionality of invoking the corresponding import/export facility.

Semantic gather engines with functionality extended from the recommended OAI functions could be devised, but they are left for a future investigation. The "Push" gather does not require any further functional element in the architecture.

Fig. 2 illustrates the concrete points in which conventional search, gather and submit functions require specific components to bridge from the non-semantic to the semantic representation.

3 Mapping metadata to ontology instances

Current metadata schemas as IEEE LOM or DCMI provide XML mappings in terms of a number of metadata elements or "fields". In IEEE LOM, these are organized in categories, and a number of data types as *LangStrings* and *Durations* are defined in the specification. However, these elements in the ontology not always have a direct mapping to a type in the language. The possible mappings define the types of translations that must be provided. The examples in Table 1 illustrate some common mapping types. The types in the ontology are concepts that surround the central `LearningObject` concept, which is the domain of the WSML attributes defined. WSML[1] is an ontology language aimed at the description of Semantic Web Services with capabilities for defining ontologies with different combinations of description logics and logic programming, so it is a more general case than other ontology languages as OWL.

LOM element	**LOM type**	**WSML attribute**	**Type in the ontology**	**Mapping type**
1.2. title	LangString	`LOMtitle`	`concept LangString`	LS2LS
1.3. language	Character-String	`LOMlanguage`	`concept Language`	V2I
2.2. status	Vocabulary	`LOMstatus`	`concept LOMstatus`	V2I
5.7.Typical age range	LangString	`LOMtypicalAgeRange`	`concept LangString` **or** `concept AgeRange`	LS2LS For numerical ranges: LS2AR
4.7. Duration	Duration	`LOMduration`	`Duration`	D2D
5.9. Typical Learning Time	Duration	`LOMtypical-LearningTime`	`Duration`	D2D

Table 1. Example IEEE LOM to WSML ontology mapping types

Table 1 shows how some elements as 4.7 and 5.9 have a direct mapping – in this case "duration to duration", since both LOM and WSML have similar types. However, mapping vocabulary types to instances (V2I) may in some cases come from character types – provided that a specific collection of possible values is provided as in 1.3. Language strings are mapped to a purposefully created `LangString` class in the ontology for 1.2, to account for multiple languages. The case of typical age ranges is special since it allows for mapping as strings in the general case, but in the cases that the strings provided are strict numerical ranges, a translation to `AgeRange` is used. This allows for using these ranges in complex queries involving numerical comparisons.

Other mappings are even less straightforward. The following are examples:

— Relations in LOM require mappings to attributes between `LearningObjects` yet available as instances, and these could entail the checking of some basic requisites of relations – see (Sánchez-Alonso and Sicilia, 2004). In addition, the mapping of LOM aggregation levels require the checking through axioms of basic requirements, e.g. that an aggregation level 1 resource has no "hasPart" relationships with others.

---

[1] WSML Final Draft 5 October 2005 D16.1v0.21 The Web Service Modeling Language WSML, available at http://www.wsmo.org/TR/d16/d16.1/v0.21/

- Some elements as 4.4. Requirement in LOM include logical clauses, in that case, “OrComposite” that require a translation to axioms if the semantics of the mapping are to provide the better formal representations.

A design that provides a degree of flexibility in the translation requires the specification of the mappings through configuration files, so that changes or additions that accommodate to existing mapping types can be done without additional development. The following is an example fragment of such schema.

```
<ontFile>Basic_LOM.wsml</ontFile>
<mappings>
<mapping>
       <type>D2D</type>
              <category>General</category>
              <attributeName>LOMDuration<attributeName>
              <element>duration</element>
       </mapping>
<mapping>
       <type>D2D</type>
       <category>Educational</category>
       <element>TypicalLearningType</element>
      <attributeName> LOMTypicalLT<attributeName>
</mapping>
</mappings>
```

The schema above indicates first the base ontology with the overall WSML definitions, as those described in Table 1. The D2D mapping requires only the specification of the elements in both sides of the mapping; more complex mapping would require more complex schemas.

Fig. 2 provides an overview of the main interfaces involved in the mapping, using Java as programming language. The main ideas behind that design are the following:

- A generic, ontology-language neutral interface is provided in the first layer, as a hook for abstractions and possible refactoring when developing different implementations.
- In a second layer, the specifics of the LOM to WSML approach are modeled in a generic way.
- The third layer is specific to existing `lom-j` and `org.wsmo` libraries used in the implementation. A *command* design pattern is used for the purpose of separating the concerns of each type of mapping as those exemplified in Table 1. This allows also for the mapping process to be structured in two phases. In a first phase, the translator instances are created from the instructions in the configuration input stream. Once the structure holding the translator instances is prepared, they are fired as commands, allowing for filtering in the case it would be required.

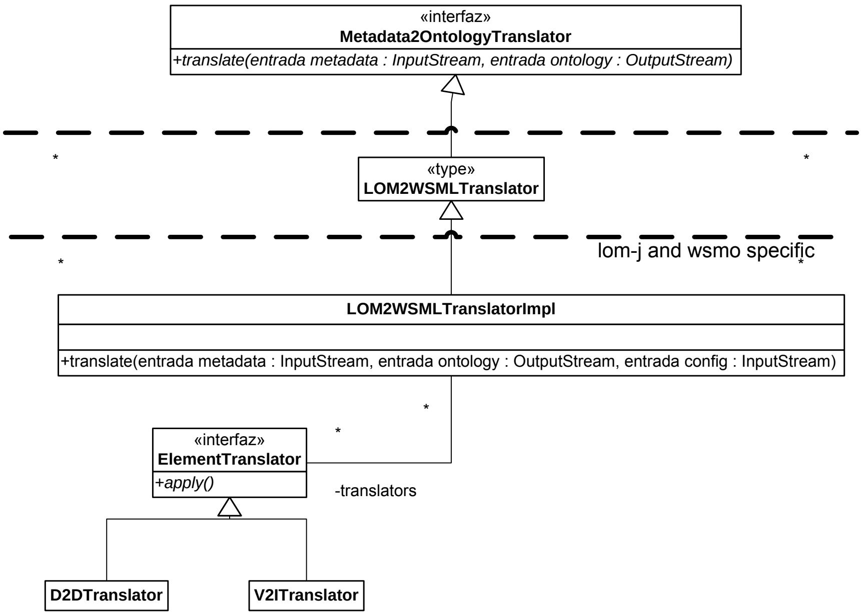

Fig 3. Main elements of the mapping design

In cases as the one of typical age range in Table 3, a class with the implementation on how to decide among alternative mappings is provided and referenced in the configuration file.

The library sketched in Figure 3 can be used standalone for mappings but it can be also used as part of a GATHER functionality as described in the previous section. The facility for gathering metadata would use the interfaces provided as filters for the input-output streams, allowing for different configurations. Since `lom-j` libraries provide support for it, the same design can be used for DCMI metadata if desired.

4 Conclusions and Outlook

Repositories of learning object metadata require extensions to current digital repository standards to deal with the connection of metadata to ontologies and the use of semantic query languages. Further, compatibility with existing XML-based representations requires generic and flexible mechanisms that are able to translate plain metadata elements to instances and predicate values inside learning object ontologies. In that direction, this paper has described an extension of the overall IMS DRI architecture for semantic repositories, and a generic mapping mechanism from IEEE LOM metadata to WSML instances. The main elements of that paradigm have been recently described by López-Cobo, Sicilia and Arroyo (2006).

Further work should deal with the integration of a Semantic Service approach to semantic repositories that enable the discovery of repositories that provide a logics-based description of the kind of resources they provide. Further, the concepts of choreography and orchestration should be included in these architectures to deal with scenarios that are typical of the e-learning domain (Sicilia and Lytras, 2005). In other direction, there is still a need for research in ontologies for learning objects that provide representations for different accounts of learning (Sicilia and Lytras, 2005b), and that include a recording of the pedagogical rationale of their design (Sicilia, 2006).

Acknowledgements

This work has been supported by project LUISA (*Learning Content Management System Using Innovative Semantic Web Services Architecture*), code FP6−2004−IST−4 027149 and by project ELSEM, UAH-PI2005-070 – University of Alcalá.

References

Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), pp. 34-43.

IMS DRI (2003) IMS Digital Repositories Interoperability Core Functions Information Model, Version 1, Available: http://www.imsglobal.org/digitalrepositories/

López Cobo, J.M., Sicilia, M.A., Arroyo, S. (2006). Making Learning Objects useful with Semantic Web Services. In *Proceedings of the e-Challenges Conference* (to appear).

Sánchez-Alonso, S. and Sicilia, M. A. 2004. How learning object relationships affect learning object contracts: commitments and implications of aggregation. In *Proceedings of ED-MEDIA 2004 - World conference on educational multimedia, hypermedia and telecommunications*, pp. 331-336.

Sicilia, M. A. and Lytras, M. (2005). Scenario-oriented reusable learning object characterizations.. *International Journal of Knowledge and Learning*, 1(4), pp. 332-341.

Sicilia, M. A. and Lytras, M. (2005). On the representation of change according to different ontologies of learning. *International Journal of Learning and Change*, 1(1), pp. 66-79.

Sicilia, M. A., Sánchez-Alonso, S. and Soto, J. (2005). Resource descriptions for semantic element repositories: addressing flexibility. In *Proceedings of m-ICTE 2005 - Third International Conference on Multimedia and Information & Communication Technologies in Education*, pp. 1045-1049.

Sicilia, M.A. (2006). Semantic learning designs: recording assumptions and guidelines. *British Journal of Educational Technology* 37(3), pp. 331-350.