Design Criteria for the Development of an Institutional Learning Object Repository

Roberto J. Godoy*
Departamento de Informática, Universidad Nacional del Litoral
Santa Fe, 3000, Argentina
rjgodoy@fich.unl.edu.ar

Hugo Minni
Departamento de Informática, Universidad Nacional del Litoral
Santa Fe, 3000, Argentina
hminni@fich.unl.edu.ar

Gonzalo Zarza
Departamento de Informática, Universidad Nacional del Litoral
Santa Fe, 3000, Argentina
zagon@fich.unl.edu.ar

Horacio Loyarte
Departamento de Informática, Universidad Nacional del Litoral
Santa Fe, 3000, Argentina
hloyarte@fich.unl.edu.ar

Abstract

The production of educational contents in academic environments has been traditionally developed without an integral plan which allows their adequate distribution and access. In addition, it produces effort duplication and misuse of resources. As a possible solution, this work proposes Learning Objects and Knowledge Repositories to be used.

The authors state some design criteria for developing a web-enabled Learning Object Repository, according to the IEEE Learning Object Metadata standard. This repository would be used in a university educational context, replacing the disjointed procedures that are currently employed. Not only knowledge, but also human resources are the main entities in this proposal.

Interoperability, portability and efficiency are the non-functional requirements considered at this design stage. In order to satisfy all of requirements, the combined usage of standard-compliant file-repositories and a relational database is proposed. The data consistence between these storage methods is assured by a specific mechanism delineated with this intention; in addition, some intuitive operators are defined for the specification of search criteria in arbitrary metadata queries.

Keywords: Education Informatics, Learning Objects, Databases, Web Systems, Software Engineering.

* Group supported by the Facultad de Ingeniería y Ciencias Hídricas

1 INTRODUCTION

Many universities and academic units in their educational proposal produce much material by using Information and Communications Technologies (ICT). The production of these contents has been developed without an integral plan which allows their adequate distribution and access. In addition, it produces effort duplications and misuse of resources.

The vertiginous change and improvement of ICT have contributed to this disorder. It generated difficulties about revision, compatibility and distribution of this educational material. Infrequent using of standards for making contents and the currently poor web-semantic contribute negatively to this problematic situation. As a possible solution, this work proposes design criteria based in Learning Objects (LO) and Knowledge Repositories concepts.

A Knowledge Repository is an integral system for coordinating related educational material or learning objects (LO). This kind of system is essential for strategies that propose the elaboration of educational contents standardized for using in the different ways from learning: traditional, e-learning and blended learning. A Knowledge Repository has great importance in complex educative systems like our universities. This material, widely distributed and standardized, admits customizations of the learning process through student-centered learning models.

A learning object is defined as a "digital resource that can be reused to support learning" [11]. Because the poor semantic of de world wide web, metadata is required in order to describe learning objects, enabling learners and instructors to search, evaluate and utilize them. Standards compliance leads to a uniform style, enhancing the sharing, reuse, and exchange of metadata-described LO.

The authors belong to the Universidad Nacional del Litoral (UNL) at Santa Fe-Argentina. This Institution develops almost 60 undergraduate and postgraduate careers, and more than 40000 students course at its classrooms (and virtual classrooms). This Academic House presents the problematic situation exposed above) and the authors are defining the bases for implementing there a Learning Object Repository. The repository is intended to replace the currently usual procedures for generating and indexing educational contents.

Designing a Knowledge Repository is a complex multidisciplinary task. In this work, the ideas mentioned in the previous paragraphs are explicitly taken to a computational plane. IEEE Learning Object Metadata (LOM) Standard was chosen among several others because it emphasizes on the "minimal set of attributes needed to allow these Learning Objects to be managed, located, and evaluated" [9] and it is included by the Content Aggregation Model of the Sharable Content Object Reference Model (SCORM) [1] [2]. This work focuses into those concerns related to LOM Base Schema, which describes the conceptual structure of metadata instances for learning objects. In this development stage the non-functional requirements of interoperability (the system shall implement a standard semantic) and portability (data representations shall not be dependant of the system) are specifically addressed.

Due to interoperability concerns, all meta-information has to be stored in an application-independent format. On the other hand, allowing the final user to efficiently perform arbitrary metadata queries is one of the most important goals of the repository. For meeting both requirements the following decision is taken: a subset of the standard metadata is going to be stored in a relational database. The proposed mechanism is in no way restrictive, because complete LOM instances are also going to be stored as specified by the chosen standard binding.

From final users' perspective, the usability of the repository is improved if information exchange is done in terms of LOM base schema metadata and not in terms of the relational model. In addition, search conditions are specified by using appropriately defined operators for each LOM datatype.

Finally, the authors believe that the goals of reusing, which motivate LO methodology, should not be based only on the content itself, but also in the human resources supporting it. Thus, a noteworthy effort is dedicated to storage and management of personal information, considering that interrelations among contents are a direct consequence of the interrelation among their authors.

2 OVERVIEW

The repository previously introduced, will be implemented according to the LOM specification [7]. The XML Schema Definition Language Binding for LOM [8] is chosen for a canonical storage of this metadata.

LOMv1.0 base schema specifies a hierarchy of metadata categories. All elements in this hierarchy are optional, and some of them repeatable. In turn, the repeatable ones may be ordered or unordered. Each bottom-level element has a metadata type: duration (an interval in time), datetime (a point in time), character string (a language-independent string), "langstring" (a set of strings expressed in natural language) or vocabulary (whose values are taken from a predefined set, with an associated semantic). Vocabulary terms may be state or enumerations. The values of the latter have total order relations defined among them.

Some of the most remarkable metadata according LOM v1.0 are: life-cycle description, technical requirements, educational characteristics (context, interactivity type and difficulty level), copyright information and relation with others objects (e.g.: version, prerequisites, inclusion). In particular, LOM specification states that all the information about entities related to the life cycle of a LO must be stored in vCard 3.0 [4] format (a text-based representation of personal information as property name-value pairs, which is explained in Section 4).

3 RELATIONAL STORAGE

Simultaneous storage (i.e. storage in relational normal form as well as original format) implies an additional effort to be taken for assuring data consistence. This drawback is fully compensated not only by a natural compliance to standards, but also by the fact that it alleviates the necessity of parsing either LOM instances or vCard files when a search involving metadata is performed.

3.1 Relational Database Structure

Each learning object will be referenced through a tuple in `LObject` table, identified by an integer which will be used as LOM identifier entry for local catalog purposes. LO's attributes with multiplicity *0..1* will be stored as nullable attributes in the table representing the immediate parent category (in LOM schema) with *0..n* multiplicity (e.g: `Educational.Difficulty`) or, if such category does not exists, as a nullable attribute in `LObject` itself (e.g.: `General.Structure`). On the other hand, LOM attributes with *0..n* multiplicity will be stored as tuples in a dedicated table, with a foreign key referring the container category. Special cases of *0..n* ordered LOM attributes will be handled by augmenting the corresponding tuple with an "order" attribute, which will designate the precedence relation among siblings. Part of the described model is shown in Fig. 1.

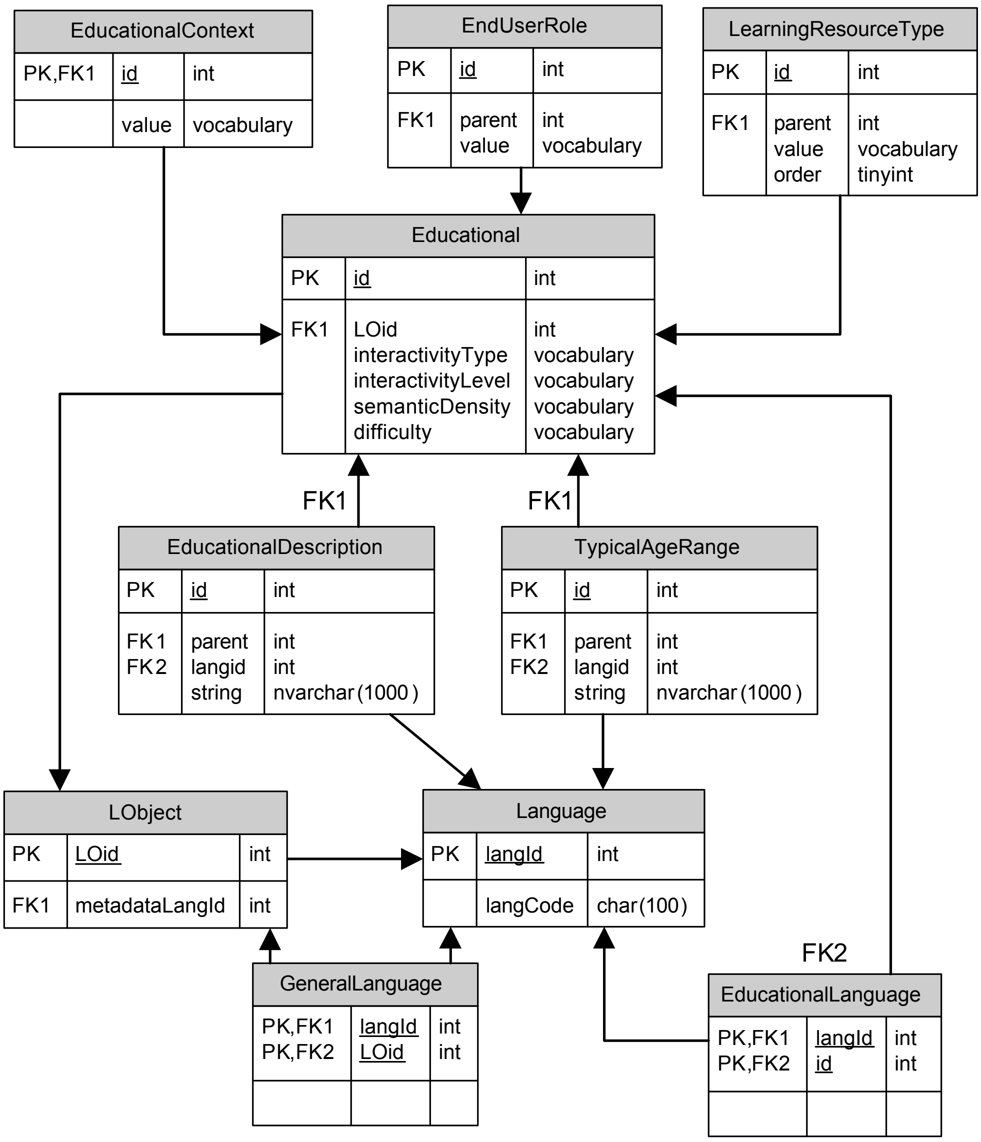

Figure 1: `LObject` table and main related entities.

LOM standard specifies that strings are comprised of 32-bit characters, but such encoding is unsupported by most database engines, so Unicode (16-bit character set) will be used as a general rule (SQL *national character* or *national character varying* type**)**. For some attributes the standard specifies a restricted character set, hence 7-bit US-ASCII encoding (i.e.: SQL *character* type) may be safely applied. A string regarded as part of a langstring will be stored as a tuple in some specific table. The components of such tuple will be the Unicode string itself, plus a reference to the specified language and a reference to the container where the langstring occurs.

There is a problem concerning the datetime type, because it can not represent dates with reduced precision, which are allowed by ISO 8601 [10]. Since precision is significant in queries (as detailed in Section 3.3) another storage type must be used. Two possibilities arise: employing separated fields for each date component (year, month, day, etc) or storing dates as alphanumeric strings in ISO 8601 format. The latter yields simpler query expressions and it is going to be used in the repository.

This also solves the problem that would have happened if datetime had been used: the DB engine provides no encoding for dates older than 1753. It is reasonable to expect date values before this year (e.g.: dating the "author" contribution for an old book).

LOM Duration values must be specially considered because the lack of a standard datatype for time intervals. These values will be stored as integer attributes, expressing the equivalent amount of seconds that would correspond considering 365 days per year and 30 days per month. By this approach it is possible to represent as much as 68 years, which clearly is long enough for the uses mandated by the standard.

Vocabulary terms will be stored as references to a tuple of either `Vocabulary` or `Enumeration` tables (according to whether the value space is a state or enumerated vocabulary, respectively). Both tuples-set include a reference to the source (separately stored) and `Enumeration` adds an integer component for designating the total order relation.

A `domain` attribute is included in both tuples, in order to maintaining a single tuple for each vocabulary term, allowing a term to be applied to many attributes and reducing the amount of auxiliary tables. This attribute is a binary mask signalling the columns where the term is permitted. Insert and update operations affecting a vocabulary term will be required to validate the actual `domain` value against the attribute to be set. This approach was suggested taking into account that queries and not insertion will be the most frequent operation over the database.

3.2 Natural Language Strings

A major simplification in the database comes from weakening the information for elements of type langstring with multiplicity higher than one. It was found that one-to-many relations between a container and langstring and between langstring and natural language strings have no practical difference. Then, it is proposed to obliterate the association in the relational model. E.g.: if the "keyword" element is considered, for the purpose of querying, the difference between examples 1 and 2 is very subtle. Indeed, it would be unnoticed, except for a query criterion joining diverse strings included in the same keyword. This kind of query seems to be improbable and useless.

```
<KEYWORD>
      <LANGSTRING>
            (“en”, “Object oriented programming”)
            (“es”, “Programación orientada a objetos”)
            (“en”, “OOP”)
            (“es”, “POO”)
      </LANGSTRING>
</KEYWORD>
<KEYWORD>
      <LANGSTRING>
            (“en”, “Learning object”)
            (“es”, “Objeto de aprendizaje”)
      </LANGSTRING>
</KEYWORD>
```

Example 1: Two keywords, with four and two strings in natural language, respectively.

```
<KEYWORD>
      <LANGSTRING>
            (“en”, “Object oriented programming”)
            (“es”, “Programación orientada a objetos”)
            (“en”, “OOP”)
            (“es”, “POO”)
            (“en”, “Learning object”)
            (“es”, “Objeto de aprendizaje”)
      </LANGSTRING>
</KEYWORD>
```

Example 2: A single keyword with six strings in natural language.

3.3 Metadata Query

Metadata queries, as considered here, are targeted to final users who are data producer or data consumer. As stated before, query criteria will be specified in terms of LOM categories. This allows taking full advantage of the Standard-defined information, although it was necessary to perform a careful analysis about admissible queries, and the datatypes and operators to be used.

On the subject of character strings, they are suitable for querying about partial or complete matches. In some metadata categories the condition of matching against a regular expression may be useful. Since regular expressions are more complex than SQL LIKE clauses, a two-phase strategy may be conceived: a preliminary filtering stage implemented over the database itself; and a refinement stage at application level. With reference to langstring, these possibilities will be enhanced by optional restrictions according to the natural language in which the string is expressed.

Language values themselves are represented via character strings composed of a primary tag from ISO 639 codes for representation of languages, followed by optional tags identifying country, dialect or script variations. As stated on [3] “languages whose tags start out with the same series of subtags (…) are NOT guaranteed to be mutually comprehensible”. Nevertheless, in English and romance languages (e.g: Spanish, French, Portuguese) people is used to assume such a relationship; for this reason, the application should provide a method for recognition of prefixes, allowing the identification of sets of mutually comprehensible languages.

An additional search criterion is determined by some elements when allowing ordered lists of values. The alternatives considered are: to search the first element of a list (it is the most significant, according to the standard) and, in a more general way, to search among the *n* first (*n* is either a maximum number or a percentage of the existing values in each instance).

According to ISO 8601, the LOM standard allows dates with arbitrary precision to be specified. Because a total order relation does not exist over these representations, the conventional operators are not applicable (e.g.: the precedence between 2006 and 2006-08-20 is not defined). This complication is solved considering two relations over the representations of dates with arbitrary precision: an inclusion relation (considering a date with greater precision included into its representations with lower precision) and a partial order relation (in the traditional sense of temporal precedence).

4 PERSONAL METADATA

The information about entities (i.e., people, organizations) contributing to a learning object are expressed in vCard 3.0 format. A vCard is a set of property name-value pairs, mainly concerning with identification, addressing, geographical location, and organization membership (e.g.: first name, last name, address, telephone number, email, title, etc.).

Only a subset of the vCard data is going to be stored on the relational model, although this subset is enough for satisfying final user requirements of locating entities. Fig. 2 shows the `vCard` table and some related tables, note that `Language` table refers to the one introduced in Figure 1. For standard conformance, the complete entity information will be available by the means of retrieving and parsing the corresponding vCard file.

The only properties relationally persisted should be those concerning with people identification or location. Exceptions are made when it is desirable to hold a normalized representation of some property, or when the disk-space overhead is too small in comparison with the time-overhead of parsing the raw vCard for getting the value.

Query criteria about most vCard values may be expressed as if they were either LOM natural language strings or character strings.

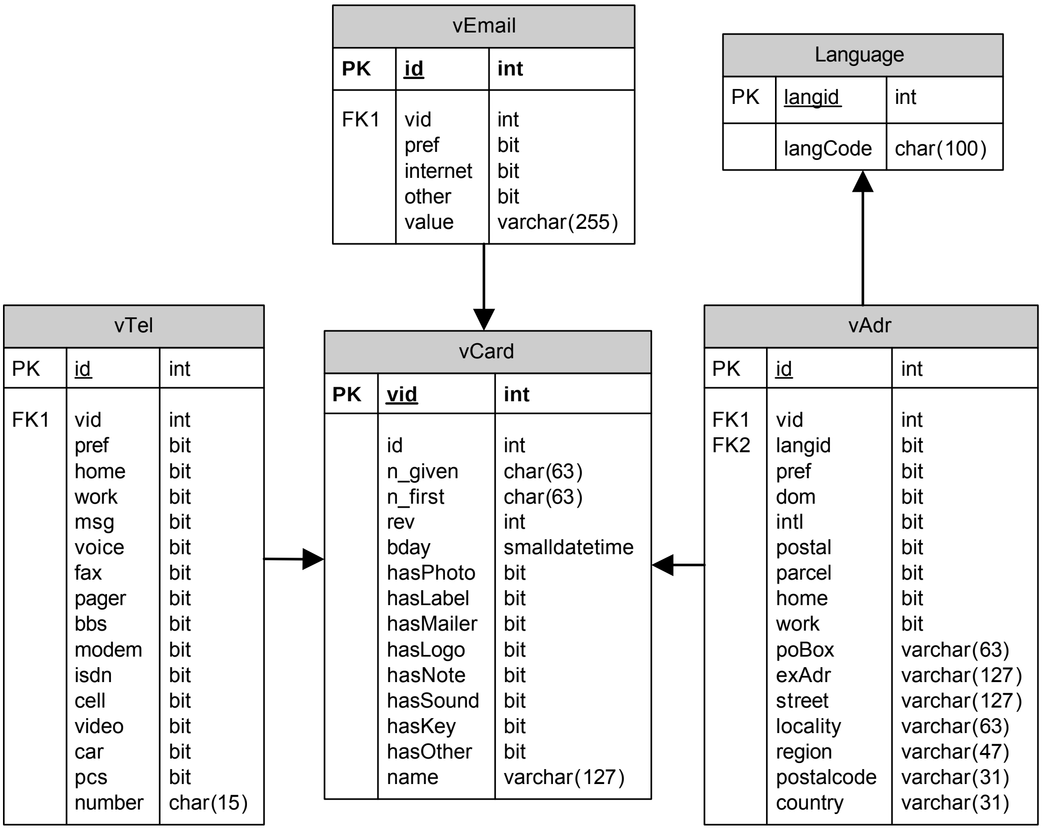

Figure 2: `vCard` table and main related entities.

4.1 vCard Externalization

Considering any change in the vCard as an action that has to be globally reflected, personal information from LOM will be externalized via system-specific sources (see Example 3). This minimal representation is a valid vCard in terms of RFC 2426; hence, it is permissible in the so-defined value space of LOM. Note that *Formatted name* (`FN`), *name* (`N`) and *version* types are required by the vCard 3.0 specification, while the *source* property references the URI of a resource that provides "additional or more up-to-date information" [*RFC 2426*].

```
BEGIN: VCARD
VERSION:3.0
FN:Mr. John Doe Jr.
N:Doe;John;;Mr.;Jr.
SOURCE;CONTEXT=LDAP:ldap://ldap.host/cn=John%20Doe,
 %20o=UNL,%20c=AR
END: VCARD
```

Example 3: Usage of source type.

In spite of *source* inclusion, entities' information may be replicated in each LO for integrity reasons (as shown in Example 4). Consequently, even if sources became unavailable, there would be always an inline vCard stating the last well-known version. A timestamp for the embed vCard might be given through last revision type (REV).

```
BEGIN: VCARD
VERSION:3.0
FN:Mr. John Doe Jr.
N:Doe;John;;Mr.;Jr.
SOURCE;CONTEXT=LDAP:ldap://ldap.host/cn=John%20Doe,
 %20o=UNL,%20c=AR
REV:2006-06-19T03:23:00
ADR;TYPE=work:;;Ruta Nac 168.;Santa Fe;Santa Fe;3000
EMAIL:JDoe@fich.unl.edu.ar
END: VCARD
```

Example 4: vCard enriched with the addition of inline data and timestamp.

5 CURRENT DEVELOPMENT

Two possibilities are considered for loading and retrieving data from the system. Web interfaces are going to be provided with the goal of facilitating these actions. At the same time, uploading and downloading LOM and vCard files will be allowed for interoperating with other applications.

Currently, the LOM information from standard-compliant files is processed via an XML parser, while personal information, coming from an uploaded vCard file, is parsed according to [6], and then mapped to an internal vCard object model.

This mapping includes validation according to RFC 2426. On the other hand, information gathered via user interfaces is validated by means of a direct conversion to the internal model. Then, a LOM or vCard representation is obtained from the internal model and stored in the file repository.

Figure 3 outlines the activity diagrams for different possible inputs. Note that web-form submitting (a) and vCard uploading (b) cases are straightforward. On the contrary, LOM file uploading requires not only parsing the XML file itself but also the vCards embed into its `Entity` elements, and storage is preceded by an externalization step, as described in 4.1. The termination of the process is similar in all cases: file representation is stored into a file repository, and the internal model is persisted through the object-relational mapping (ORM) layer. The process fails whenever a parsing error occurs.

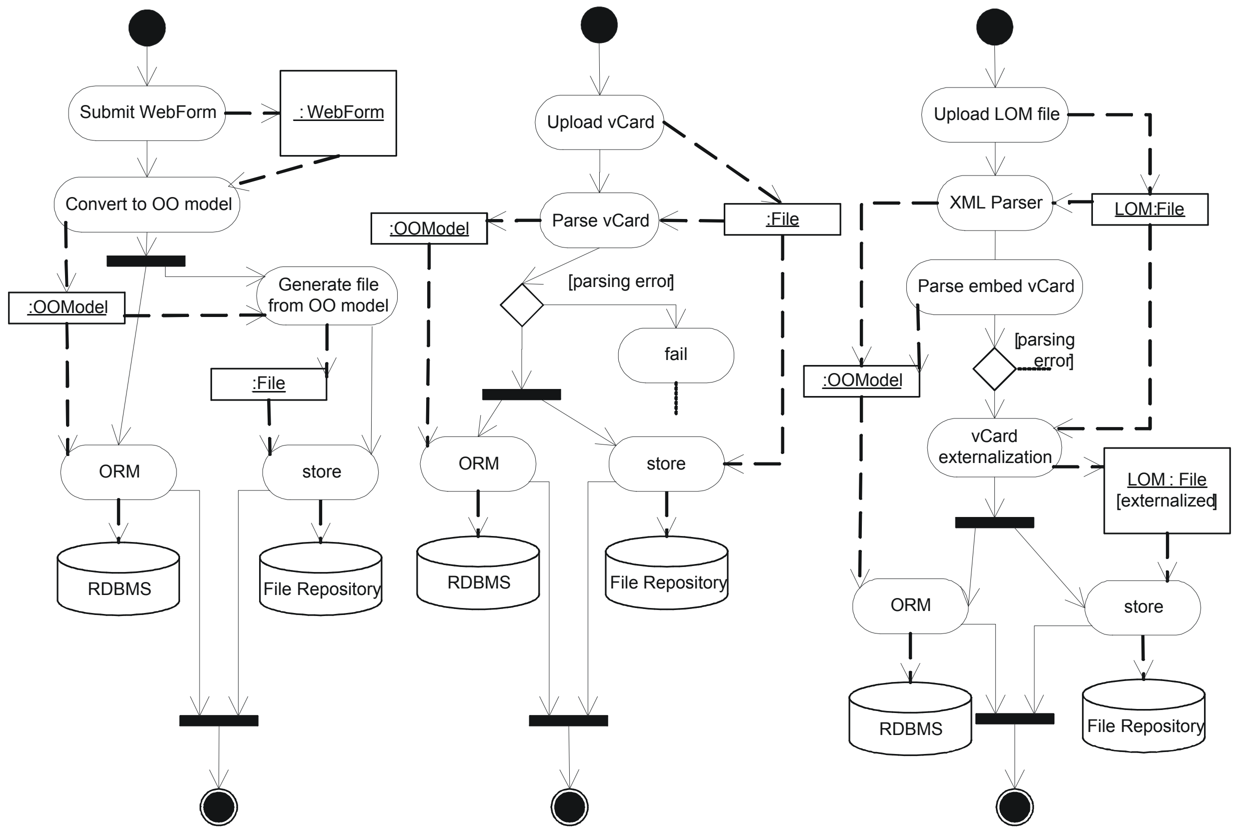

Figure 3: Activity diagrams for information input: (a) by submitting a web-form, (b) by uploading a vCard file, and (c) by uploading a LOM metadata file

6 CONCLUSIONS

Based on the IEEE LOM Standard, interoperability and portability concerns are main guidelines in this process. Interoperability is possible because the repository design is based on a widely accepted standard semantic. Portability is achieved by keeping the underlying Standard-based repositories fully independent of the application object model and its relational schema.

In this early stage, the system design is focused into the usage of the repository from the perspective of a final-user. Administrative aspects will be covered in further stages. Ideally, in relation to the mentioned issue, it should be possible to perform arbitrary queries involving LO metadata. Efficiency concerns led to store a metadata subset in a relational database.

Identical requirements made necessary to store vCard information in the relational database. This information is considered an essential repository component, because of its future institutional usage. Storing and management of these data are tightly integrated in the application design.

It was proposed that entities be located by the means of database searches, and the corresponding vCard file parsed for retrieving the complete information. An externalization mechanism was suggested for avoiding personal data duplication among several LOs.

This mixed approach (standard-compliant files repository / relational database) meets all the system requirements, even though an additional effort at implementation level must be spent for guaranteeing data consistence.

Since a relational database engine will be used for locating data, a careful analysis was done for outlining the admissible queries for each LOM datatype. Additionally, operators were defined with a user-friendly semantic. This analysis, as well as the modelling of interrelations among LOM categories and elicitation of stakeholders' requirements, is the basis of further development of a flexible intuitive user-interface.

The authors are due to discuss and adopt pedagogical methodologies and techniques for using the maximum of the shareable resources of the system. For the success of the project, the specific and permanent qualification of all the actors is essential.

The argentine universities, have always performed in low-resources environments; for them, the LO model and the LO Repository is a concrete method to improve the quality of educative systems and to optimize the available resources.

REFERENCES

[1] ADL. *Sharable Content Object Reference Model, Overview. Advanced Distributed Learning,* 2nd edition, 2004.

[2] ADL. *Sharable Content Object Reference Model, Content Aggregation Model.* Advanced Distributed Learning, version 1.3.1, 2004.

[3] Alvestrand H. "Tags for the Identification of Languages". *RFC 1766*. Internet Society. 1995.

[4] Dawson F., Howes T., "vCard MIME Directory Profile", *RFC 2426*. Internet Society. 1998.

[5] Freed N., Borenstein N. "Multipurpose Internet Mail Extensions". *RFC 2045-2046*. Internet Society. 1996.

[6] Howes T., Smith M., Dawson F. "Content-Type for Directory Information", *RFC 2425*. Internet Society .1998.

[7] IEEE. "IEEE Standard for Learning Object Metadata". *IEEE Standard 1484.12.1,* IEEE Computer Society, New York. 2002.

[8] IEEE. "IEEE Standard for Learning Technology - Extensible Markup Language (XML) Schema Definition Language Binding for Learning Object Metadata". *IEEE Standard 1484.12.3,* IEEE Computer Society, New York. 2005.

[9] IEEE Learning Technology Standards Committee. *Learning Object Metadata* WG12. http://ieeeltsc.org/wg12LOM. Last accessed in 2006-07.

[10] ISO.Data elements and interchange formats – Information interchange – Representation of dates and times. *International Standard 8601*. International Organization for Standarization, Switzerland, 1988.

[11] Wiley D, "Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy". *The Instructional Use of Learning Objects: Online Version*. 2000. Retrieved in 2006-07 from http://reusability.org/read/chapters/wiley.doc