Using IMS VDEX in Agrega

Sarasa. A¹, Canabal. M², Sacristán. J.C² and D.E. Jimenez. R²
¹ Universidad Complutense de Madrid, Madrid, Spain
² Red.es, Madrid, Spain

doi:10.3991/ijet.v4s1.806

Abstract—An essential element in a learning object repository is the meta-information associated with the resources housed by the repository, as this is what enables the objects to be retrieved. In many cases, the meta-information which is described consists of the classification of a resource in relation to a taxonomy or thesaurus. Agrega is a network of learning object repositories in which an LOM profile, called LOM-ES, has been used to label the objects. Part of the meta-information described corresponds to the classification of the object with respect to a group of taxonomies and a thesaurus, which have had to be described using the ims vDEX standard to enable them to be managed from the federation nodes. This article describes the way in which ims vDEX has been used in order to create this description and how these metadata instances are managed.

Index Terms—IMS VDEX, taxonomies, learning objects, digital repositories.

I. INTRODUCTION

The Public Corporate Entity Red.es [4], attached to the Ministry of Industry, Tourism and Trade, through the State Secretariat for Telecommunications and the Information Society, has general legal responsibility for carrying out certain functions that are designed to contribute to the promotion and development of the Information Society in Spain. A series of measures are being developed in the area of Education whose purpose is: a) To define a cataloguing reference standard, digital educational content package and publication, b) To generate a core group of digital learning objects, in some cases based on existing content, which serve to promote the technological environment created and c) To create a technological environment that will house content which adheres to the aforementioned standard, and which is accessible to the educational community under different modes of use. To meet this last objective a federation of digital repositories, called Agrega [1.10], has been implemented with nodes installed in each one of Spain’s Autonomous Regions. Each node permits the storage of SCORM 2004[11] digital objects labelled with LOM-ES metadata (application profile for metadata specifications LOM [8]). These nodes enable users to perform different operations in accordance with the profile they have defined, such as search, display or creation of new material. The purpose of the metadata described in LOM-ES [1] is to enable repository objects to be retrieved. Because it is a LOM application profile, it inherits the different categories defined in this standard to describe digital educational material. Each of these categories encapsulates information through which an object can be found. In particular, there is a category called Classification, the purpose of which is to describe how the labelled material can be classified with regard to numerous educational purposes. In order to carry out the classification, a group of taxonomies and a thesaurus, defined ad hoc by this profile, are taken as a reference. There are two aspects to its real utility in Agrega. On the one hand, each node enables the user to create and label their own objects, using these classifications as part of the labelling process and, on the other hand, the system, both in the validation and retrieval of objects, requires the grammatical management of these classification systems and related metadata instances. This article explains how the LOM-ES taxonomies and thesaurus have been implemented in ims vDEX, and also the Agrega functions which require the use and management of ims vDEX instances. The first section describes the structure of LOM-ES, classification category 9 and the taxonomies and thesaurus used in this category. Following is a synthetic description of the ims vDEX specification. One section is dedicated to explaining how the taxonomies and the thesaurus have been implemented using ims vex. This is followed by a description of its implementation and use in Agrega. Finally, proposals are given for future developments to be carried out.

II. THE LOM-ES APPLICATION PROFILE

The Learning Object Metadata (LOM) [8] is a specification which defines a group of labels for describing learning object metadata. However, the metadata described in LOM are general in nature and do not meet the needs of the Spanish education system. For this type of situation LOM provides an adaptation and extension mechanism called an application profile. Thus, LOM-ES is an application profile adapted to the Spanish educational sphere, in which extensions have been added to the labels and to the vocabularies used to fill out the labels. Because it is an application profile, LOM-ES has inherited the LOM category structure:

- Category 1. General: this groups together the general information described by a learning object as a whole.
- Category 2. Life cycle: this describes the history and current status of a learning object and also the entities involved in its creation and evaluation.
- Category 3. Meta-metadata: this describes the actual logging of metadata. It describes how a metadata instance can be identified, who created it, how, when and with what references.
- Category 4. Technical: this describes the requirements and technical characteristics of the learning object.
- Category 5. Educational Use: this describes the basic educational and pedagogical characteristics of the
learning object. Specifically, this is the teaching information which is essential for those agents involved in a quality educational experience. These agents include students, teachers, tutors and administrators.

- Category 6. Rights: this describes the intellectual property rights and conditions of use which apply to the learning object.
- Category 7. Relationship: this describes the relationships which exist between one learning object and others. To define multiple relationships various instances from this category must be used. If it has relationships with a number of learning objects, each one of these will have its own instance in this category.
- Category 8. Annotation: this provides comments on the pedagogical use of the learning object and information on who created the comment and when it was created. This category enables educators to share their assessments of the learning object, recommendations for its use, etc.
- Category 9. Classification: this describes where the learning object is located within a specific classification system. To define multiple classifications multiple instances from this category must be used.

It is in category 9, therefore, where it is necessary to use classification systems.

![Figure 1. Category 9.](image)

The way to use category 9 is to set a cataloguing purpose, purpose label, (in LOM-ES it is recommended that you include the educational level, discipline, accessibility and competency purposes). For each purpose you must use an instance of the label classification. Within a classification label in which a purpose has been set, you use the label taxonPath to indicate the classification system source which covers the purpose set and describes the classification path in relation to the system defined in the source. This classification path is shown by listing the nodes, taxon labels, of the path. One object can have different classification paths in relation to a system and the way to reflect this is to use a taxonPath label for each of these paths. You can also add an auxiliary description and a group of keywords to a classification label. For each of the purposes recommended by LOM-ES a group of taxonomies has been defined and an existing thesaurus has been adapted:

- Purpose: Educational Level. This describes how an object can be classified in relation to the educational level for which it can be used. The taxonomy “Educational Level LOM-ESv1.0” has been defined.
- Purpose: Accessibility. This describes the characteristics relating to the accessibility of the object. The taxonomy “Accessibility LOM-ESv1.0” has been defined.
- Purpose: Discipline. This describes the way in which the object relates to curriculum content or the area of knowledge it covers. A taxonomy has been defined based on the Spanish educational curriculum described in the current LOE education act called “2006 LOE Curriculum Tree”, and the ETB thesaurus [5] has been adapted, extending its coverage to the Spanish education system.

The precise way in which to use these taxonomies can be demonstrated using the purpose 'discipline' and the taxonomy 'curriculum tree'. When an ODE is catalogued using this taxonomy, it needs to be associated with a branch of the curriculum tree. To do this you must select a curricular objective, which will be transferred to the corresponding LOM-ES fields as follows:

1) First of all a new classification label will be created and, associated with this, a pair of source-value sublabels will be created where the classification purpose "discipline" will be indicated.
2) A new TaxonPath label will be created for the chosen curricular objective.
3) The TaxonPath sublabel source is given the value “2006 LOE Curriculum Tree”
4) For each level/curricular objective which makes up the complete route of the tree, a taxon will be created within the TaxonPath generated.
5) The numerical identifier is stored in the Id field of the taxon created.
6) The value of the entry is stored in the Entry field of the taxon created.
7) After the TaxonPaths, the rest of the classification sublabels will be labelled, that is, description and keyword.

The order in which the taxons of the taxonomic route chosen in the curriculum tree are stored is from the root to the branch selected in the association. For example, suppose you have an object that you wish to associate with the following objective:

- 6 Artistic education: music and dance (MEC)
- 6.1 Music (MEC)
- 6.1.2 Professional (MEC)
- 6.1.2.48 Accompanied Repertoire (CM, MA)

When you associate the curricular objective it will be stored in the LOM-ES fields as follows (the classification label and the rest of the sublabels are omitted):
Lastly, the definition of the taxonomies and the adaptation of the thesaurus have been carried out in Work Group 9 belonging to Subcommittee 36 of ICT for learning (SC36) of the Spanish Association for Standardisation and Certification (AENOR). The taxonomies will be published next to the profile in AENOR. However, the ETB-MEC-CTA [5] will be published in the context of the European Committee for Standardisation (CEN) [2], as it is an extension and adaptation of the ETB thesaurus published in said institution.

III. THE IMS-VDEX SPECIFICATION

The specification IMS Vocabulary Definition Exchange (VDEX) [9] defines an XML-based format for the exchange of different types of value lists, which are used as a source for the vocabularies used for labelling metadata. Two different vocabulary categories are considered, differentiated by the key used to identify a concept:

- Vocabularies for which the key is some kind of token, which effectively references a human language term ("tokenized terms")
- Vocabularies for which the key is a human language term ("human language terms")

The type of data used in LOM (and therefore in its application profiles as is the case for LOM-ES), and in the majority of the IMS specifications are tokenised terms. However, human language terms are also used in the case of classifications. Thus, the specification supports the description of the longest metadata labelling value defining forms:

- The description of controlled vocabularies/terms which are expressed as pairs of source-value.
- The description of hierarchical vocabularies or taxonomies. These are characterised by using structured identifiers which express the position of a term in an inheritance (this is a path with a tree-shaped structure). The identifiers are tokens which are independent of the language and which can be associated with any number of human language references. For this purpose the ims vDex model has a hierarchical structure which enables a taxonomy to be represented.

- The cataloguing of resources into collections and the indexing of these uses groups of preferred terms which are normally human language terms. These have some standardised relationships which make up the thesauruses. Ims vDex supports this type of vocabulary description although the structure of the information model is not prepared explicitly for representing relationships. These relationships generate a graph structure. Thus, a classification in relation to a thesaurus is a path within a graph. To express these relationships, ims vDex provides terms which represent the possible relationships defined between pairs of terms.

For each one of these forms of describing a vocabulary there is a specific ims-vdex profile.

<table>
<thead>
<tr>
<th>Figure 2. Example of labelling category 9.</th>
</tr>
</thead>
</table>

The specification also offers mechanisms for two needs which arise in the use of the metadata specification vocabulary values:

- Application profiles [7]. A normal phenomenon in specifications is the defining of application profiles which adapt and extend the original versions for use by a particular user community. As a part of these extensions, it may be the case that one vocabulary is replaced by another or is extended or reduced. Ims vDex offers a standardised mechanism for carrying out this adaptation in a way which is consistent with the original specification.

- Specific application of a vocabulary. One difficulty which often arises in the use of a metadata specification is knowing the right way to use its labels' vocabulary values. This is often associated with the community in which it has been defined. For this reason, the specification enables additional information to be included with the definition of the values domain with notes that indicate, and help to ensure, the correct use and interpretation of the terms. For the same reason it is also possible to define vocabularies separately from the source of the term value vocabulary.

IV. MANAGEMENT IN AGREGA OF VOCABULARIES, TAXONOMIES AND THESAURUSES WITH IMS VDEX

In the context of Agrega, the ims-vDex specification has been used to describe the controlled vocabularies, taxonomies and thesaurus which are recommended in the LOM-ES application profile, with which the repository objects are labelled:
• **Controlled Vocabularies.** The values of the labels defined in the profile can correspond to controlled vocabularies (this is a closed list of values) or to a free text vocabulary. To specify the former, IMS VDEX offers a specification mechanism which has been adapted in Agrega as follows:
  - All the controlled vocabularies are held in one single .xml file.
  - Given that Agrega is multilanguage, VDEX descriptions have been generated for all the languages supported, with a "master" file specified in English (in accordance with the vocabularies considered in the LOM-ES v1.0 schemas), with all the other files being translations of the master file. Included in the name of each file is the corresponding suffix for the language in which the vocabularies are expressed ("vocab_lomes_en.xml" for English, "vocab_lomes_es.xml" for Spanish, etc.).
  - Although the order in which the terms are presented is not significant, since it is not necessary for some vocabularies to appear before others, it has been decided to mark the attribute orderSignificant="true", to ensure that the terms in the corresponding files in other languages are found in the same position and are thus symmetrical. It is also recommended that the vocabularies appear in the same order in which they appear in the LOM-ES specification categories.
  - The profile of the VDEX vocabulary type for the controlled vocabularies file is hierarchical and therefore the attribute profileType takes the value "hierarchicalTokenTerms".
  - The "language" attribute indicates the language in which the controlled vocabularies of the .xml file are expressed.
  - The label "vocabName" with its corresponding label <langstring> takes the value "LOM-ES v1.0 Vocabularios" translated to the language indicated in the attribute "language".
  - Both the vocabularies and the terms which form these are given the label "term", with an identifier and a name specified through the labels "termIdentifier" and "caption"."langstring" respectively.
  - As with the vocabularies, the order of the words which make up each vocabulary is not important according to the LOM-ES standard. However, the attribute orderSignificant="true" has been associated with each vocabulary to ensure that the 'offspring terms', that is the words of the actual vocabulary, appear in the same position in the corresponding files in other languages.
  - The vocabulary identifier ("termIdentifier") is the same as the one which corresponds to the associated field in the LOM-ES v1.0 specification, which means it depends on the category and the position of the element with which the vocabulary is associated. So, for example, the identifier which corresponds to the vocabulary "status" in LOM-ES, is “2.2” (category 2, element 2).
  - The identifier for each term belonging to a vocabulary ("termIdentifier") is made up of the vocabulary identifier followed by another figure, making it possible to identify a word within a vocabulary. For example, if the identifier for the term "draft" of the vocabulary "status" of LOM-ES is “2.2.1". The order assigned to the terms within the vocabulary is not important, but the order does need to be the same in all the different language files in order to create symmetry between the numerous .xml files.

• **Taxonomies** [3]. In category 9, as was mentioned previously, the metadata description with respect to different purposes is created using taxonomies which are defined ad hoc in the profile. IMS VEX also offers a standard form of describing them, which has been adapted as follows:
  - The multilanguage aspect is managed in different files, generating as many different VDEX descriptions for the taxonomy as there are languages.
  - The files are named in the same way, including in each case the name of the taxonomy in question followed by the suffix corresponding to the language in which it is expressed. For example, in the case of the curriculum tree it is "curriculum_tree_es.xml", "curriculum_tree_en.xml", etc.
  - The order of the terms is marked with the value "true" if it is significant and "false" if it is not. For example, in the specific case of the curriculum tree, the terms are ordered since the meaning of these terms allows for a temporal sequence. Therefore, we mark the vocabulary attribute orderSignificant="true".
  - The profile of the VDEX vocabulary type for the controlled vocabularies file is hierarchical and therefore the attribute profileType takes the value "hierarchicalTokenTerms".
  - The attribute "language" indicates the language in which the taxonomy of the corresponding .xml file is expressed.
  - The label "vocabName" with its corresponding label <langstring> contains the name of the taxonomy represented.
  - All the taxons are indicated with the label "term", specifying an identifier and a name through the labels "termIdentifier" and "caption"."langstring" respectively.
  - All the taxons have the same number of nested "term" terms as there are offspring in the specified taxonomy.
  - The identifier of each taxon ("termIdentifier") is based on that of its predecessors. So, for example, for the curriculum tree the identifier for the different stages is a single figure, the identifier for the cycles/courses is made up of two numbers, the
identifiers for the areas/subjects is at a third level and lastly the blocks are at nesting level 4, for which four figures are used. A specific example is: “Nursery School” for which the identifier must be “1”, the identifier for “Stage Two” will have to be “1.2.” after which would be the area “Self-knowledge and independence” with the identifier “1.2.1” and so on.

- Thesauruses [6]. In LOM-ES it is recommended that the thesaurus be used to cover one of the classification purposes, namely discipline. Ims vex also offers a standard form of describing them, which has been adapted as follows:
  o The multilanguage aspect is managed in different files, generating as many different VDEX descriptions for the thesaurus as there are languages.
  o The files are named in the same way, including in each case the name of the thesaurus in question followed by the suffix corresponding to the language in which it is expressed. “etb_es.xml”, “etb_en.xml”, “etb Ga.xml”, etc.
  o The order of the terms is marked with the value “true” if it is significant and “false” if it is not.
  o The profile of the type of VDEX vocabulary for the thesaurus files is profileType=“thesaurus”.
  o The attribute “language” indicates the language in which the thesaurus of the corresponding .xml file is expressed.
  o The thesaurus file contains the label “vocabName” with its corresponding label <langstring> and the value is the name of the thesaurus represented in the corresponding language. For example, “ETB - Teaching methods”.
  o If you wish, you can include the label “vocabIdentifier” to include an identifier for the thesaurus.
  o All the taxons are indicated with the label “term”, specifying an identifier and a name through the labels “termIdentifier” and “caption”-“langstring” respectively.
  o Nesting is not permitted between terms.
  o The relationships between terms are formed through the label “relationship”, with one label for each relationship within the thesaurus. Each “relationship” label is composed of two terms, “sourceTerm” and “targetTerm”, which contain the identifiers (specified previously in the labels “termIdentifier”) of the related terms. The label “relationshipType” is optional and does not need to be used.

Among the taxonomies which have been described in IMS-VDEX, is the 2006 LOE curriculum tree, which describes a classification system based on subjects and disciplines and which reflects the Spanish non-university curriculum, based on current legislation (LOE 2006). This taxonomy has been specifically created for the Agrega project and covers both the common curriculum and the specific curriculum of each autonomous region.

Figure 4. Ims-vdex instance for the thesaurus.

V. MANAGEMENT OF IMS-VDEX IN AGREGA

The learning objects which are managed in Agrega, are SCORM 2004 objects with metadata labelling based on LOM-ES. It is for this reason that, in order to make its management and use easier, the IMS-VDEX formalisation of the controlled vocabularies and classification systems recommended by LOM-ES has been used. Although both are in the IMS-VDEX format, only the classification systems can be configured in Agrega. To do this an administrator profile is required, from which you have access to an area called Taxonomies and Thesauruses, which permits access to a group of forms in which you can add or remove IMS-VDEX instances from the curriculum trees, the taxonomies and the ETB thesaurus. These instances are used by Agrega as a source for:

- Metadata editor. The labels which contain the values corresponding to the controlled vocabularies or those where the values of a taxonomy, thesaurus or curriculum tree need to be entered, have either dropdown elements associated, in the case of controlled vocabularies, or a graph navigation system, in the case of taxonomies, thesauruses and curriculum trees, for selecting the appropriate value. The content of these dropdown elements or navigation systems is extracted from the content of the ims-vdex instances managed by Agrega. This way, the changes in the classification system only affect the corresponding ims-vdex instance and not the code.

- Instance Validation System. Agrega allows externally-produced learning objects to be imported. In order to be imported, these objects must be checked against the LOM-ES schemas. In particular, checks are made to ensure that the values of the
labels corresponding to the controlled vocabulary or classification systems are correct. The base schema used for parsing in this process are the LOM-ES xsd schemas and the IMS-VDEX instances of controlled vocabularies and classification systems.

- **Browser-based search.** One function provided by Agrega is a repository object search, performed by browsing the thesaurus structure or the curriculum tree. For this purpose the user navigates through the structure of a tree or a graph and, in each node of this structure, they will be shown the repository objects available for the associated values and the value represented by that node within the taxonomy or curriculum tree. To implement this search-navigation system, you need, on the one hand, the instances of the taxonomy and the curriculum tree in IMS-VDEX format so that the navigation system can get the content from these and, on the other hand, all the objects stored within the repositories need to be labelled in category 9 of LOM-ES in accordance with these classification systems. Internally, the objects are retrieved in accordance with the value represented by the node of the structure the user is in.

- **Interoperability** [12]. Another use of having the controlled vocabulary and classification system instances in IMS-VDEX is that it enables interoperability with other learning object repositories. Given that, beside the objects which can be exchanged between heterogeneous repositories, it is possible to encapsulate both the structure of the metadata, through the xsd, and the content through the instances in IMS-VDEX, and it is also possible to create tables for mapping the values of vocabularies (particularly between objects which use application profiles of the same origin, for example LOM-ES metadata objects with LOM-FR metadata objects).

- It facilitates management and processing as it is an XML-based format.
- It facilitates interoperability with other learning object repositories.
- It facilitates reuse.

However, there is also a disadvantage with regard to the current extension of the specification. This is due to the fact that it is a recent specification in the process of being adopted by other initiatives similar to Agrega.

**REFERENCES**


**AUTHORS**

Sarasa, A., is full time professor at the Facultad de Informática at Universidad Complutense de Madrid and IT consultant at red.es (e-mail: asarasa@sip.ucm.es).

Canabal, M., is Area Responsible - Application, Content, Training and Digital Context Projects at red.es (e-mail:JoseManuel.Canabal@red.es).

Sacristán, J.C., is Projects Coordinator at Red.es (e-mail: JuanCarlos.Sacristan@red.es).

Jimenez, R., is Projects Coordinator at Red.es (e-mail: Raquel.Jimenez@red.es).

Agrega is a project which is being developed and financed within the framework of the Avanza plan through the Internet in the Classroom Programme, by means of a collaborative agreement between the Public Corporate Entity Red.es, the Ministry of Industry, Tourism and Commerce, the Ministry of Education and the Autonomous Regions. The article was modified from a presentation at X International Symposium on Computers in Education (SIIE2008) 1st-3rd October 2008, Salamanca, Spain. Manuscript received 20 November 2008. Published as submitted by the author(s).