Abstract
This paper presents the development of a controlled vocabulary for functional description in an educational repository project which has adopted a DC application profile. The vocabulary, organized according to the identified functions of educational documents and learning objects’ components, permits their retrieval and reuse to be improved.

Keywords:
Dublin Core; Application profiles; Educational vocabularies; Learning objects; Reusability; Educational digital libraries.

1. Introduction
Nowadays, the global educational community has a strategic need: to share and reuse educational contents. The main goal of the DOTEINE Project (“LIS and ICT for Education: Instruments for Information Literacy and the Organization of Educational Resources”) and its annex project, the IACORIE Project (“Content Analysis Instruments for the Organization of Educational Resources: LRC Knowledge for Society”), is to develop documentary tools for storing, representing, organizing and retrieving educational contents in educational settings. To accomplish this goal, we have developed a repository to gather educational digital documents that can be used at three educational levels: Secondary
Education, Undergraduate and Postgraduate. To describe these documents, we have adopted a Dublin Core application profile called MIMETA (IACORIE Metadata Model).

The most important information requirements of the educational community (1) correspond, on the one hand, to the need to retrieve educational resources and/or their components by their content matter as well as by the function they have in their original context, and on the other hand, to retrieve them at a high granularity level for reusability purposes.

To respond to these demands, several tasks must be undertaken. First, digital educational documents must be provided with analytical descriptions of a suitable depth and due preservation of context by means of proper identification of whole-part and part-whole relationships. Second, we should have at our disposal two specific controlled vocabularies: one to describe content matter and the other one to describe the different functions that document components can fulfill. However, vocabularies currently available do not take into account functional description, preventing us from carrying out functional retrieval.

This paper presents the application profile in which the new vocabulary has been integrated, previous analysis of components’ functions carried out, and the vocabulary structure developed. Finally, a summary of the conclusions reached and future works is given.

2. The MIMETA Application Profile

Among its 22 elements, MIMETA gathers most of DC basic elements as well as some other elements from educational schemes (mainly the IEEE LOM, AICC, EdNA, GEM and MERLOT). The application profile was developed after a detailed analysis of the main standards and ongoing educational digital library projects (2). The elements considered for general description of resources are:

<table>
<thead>
<tr>
<th>Element</th>
<th>Description and use</th>
</tr>
</thead>
<tbody>
<tr>
<td>identifier</td>
<td>Resource identification code</td>
</tr>
<tr>
<td>author</td>
<td>Person or entity associated with the resource</td>
</tr>
<tr>
<td>author type</td>
<td>Type of authorship</td>
</tr>
<tr>
<td>institution</td>
<td>Institutional authorship</td>
</tr>
<tr>
<td>title</td>
<td>Title or name assigned to the resource</td>
</tr>
<tr>
<td>date</td>
<td>Date associated with the resource</td>
</tr>
<tr>
<td>date type</td>
<td>Type of date</td>
</tr>
<tr>
<td>language</td>
<td>Language of the document content</td>
</tr>
<tr>
<td>description</td>
<td>Free-text description of the content</td>
</tr>
<tr>
<td>format</td>
<td>Type of data</td>
</tr>
<tr>
<td>type</td>
<td>Resource type description</td>
</tr>
<tr>
<td>location</td>
<td>URI and/or navigation path if applicable</td>
</tr>
<tr>
<td>rights</td>
<td>Property rights involved</td>
</tr>
<tr>
<td>subject</td>
<td>Subject matter, controlled vocabulary</td>
</tr>
<tr>
<td>keywords</td>
<td>Keywords, free text</td>
</tr>
<tr>
<td>relation</td>
<td>Relationships established between resources. Relationship considered: “Is part of”</td>
</tr>
</tbody>
</table>
The elements considered to describe the educational characteristics are:

<table>
<thead>
<tr>
<th>Element</th>
<th>Description and use</th>
</tr>
</thead>
<tbody>
<tr>
<td>user type</td>
<td>Kind of potential user</td>
</tr>
<tr>
<td>context</td>
<td>Educational context/level</td>
</tr>
<tr>
<td>difficulty</td>
<td>Degree of difficulty associated with the resource</td>
</tr>
<tr>
<td>prerequisites</td>
<td>Previous knowledge required</td>
</tr>
<tr>
<td>interactivity type</td>
<td>Kind of interaction supported by the resource (active, expositive or combined).</td>
</tr>
<tr>
<td>pedagogical</td>
<td>Pedagogical methods and goals</td>
</tr>
</tbody>
</table>

These educational elements are considered mandatory for learning objects and optional for informative documents until they acquire educational value in a context of use, following the document differentiation explained below.

3. **Learning objects’ functional structure**

We consider there to be two types of documents which are of use in education: on the one hand, informative documents and, on the other, the strictly speaking educational documents (3). Among the latter –educational documents oriented towards the acquisition of knowledge, skills and competencies- we find learning objects (LO). According to the principles of instructional design and the typical usage in educational communication, LO should be provided with the following generic structure (typical semantic superstructure):

- **Presentation**: Components that set the content in context for the learner and present the educational goals aimed for.
- **Exposition**: Components that expound the educational contents that have motivated the creation of the document.
- **Illustration**: Components that illustrate the theoretical contents, supporting them and complementing them.
- **Evaluation**: Components used to assess the fulfillment of the learning objectives sought.

4. **The type and function controlled vocabulary**

DC and IEEE LOM propose two different solutions for populating their “Type” and “Educational resource type” elements. DC recommends using its own vocabulary, the DCMI Type Vocabulary, comprising 12 terms: Collection, Dataset, Event, Image, Interactive Resource, Moving Image, Physical Object, Service, Software, Sound, Still Image and Text. However, the DCMI has also encouraged the adoption of specialized vocabularies developed within communities of practice (4). For its part, LOM proposes using the 1989 Oxford English dictionary and any other terms used in the different communities of use.

The need for an available specific, organised and limited value space for the defined retrieval purposes has led us to analyse the vocabularies available, for which the exhaustive JISC vocabulary report for CETIS was very useful (5). Having observed the lack of a convenient vocabulary for our goals, we decided to use them as a guide and terminological basis for a vocabulary constructed *ex novo*. To do this, the structural entity of documents
(units or components) and the functions identified for their components were used as a basis. It was, then, decided to organise terms in a faceted structure (6) in an attempt to reflect the multiple dimensions and diversity of educational documents, thus creating a flexible and semantically rich descriptive language.

![Fig. 1. Document type and function vocabulary. Hierarchical display (Extract)](image)

The thesauri management system used by our project, Tematres (7), enables our vocabulary to be exported in Zthes and SKOS-Core format. Facets have to be solved by means of ordered collections, as shown below:

```xml
<?xml version="1.0" encoding="ISO-8859-1" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <skos:ConceptScheme rdf:about="http://doteine.uc3m.es/tesauros/tipoindex.php">
    <dc:title>Tipos y funciones de documentos educativos</dc:title>
    <dc:creator>Grupo DOTEINE</dc:creator>
    <dc:subject>TESAUROS DE EDUCACIÓN</dc:subject>
    <dc:description>Tesauro de tipos y funciones de documentos</dc:description>
  </skos:ConceptScheme>
</rdf:RDF>
```
The vocabulary developed has been used to populate the `<type>` element, which is by definition, a repeatable element. A sample XML record for a demonstrational video (LO component with an illustrational function) gathered in the repository is shown in the following figure:

```xml
<dc:format xsi:type="dcterms:IMT">Video</dc:format>
<dc:type xsi:type="dcterms:DCMIType">Moving image</dc:type>
<dc:type xml:lang="es" xsi:type="mimeta:TFT">digitales</dc:type>
<dc:type xml:lang="es" xsi:type="mimeta:TFT">componentes</dc:type>
<dc:type xml:lang="es" xsi:type="mimeta:TFT">ilustraciones</dc:type>
<dc:type xml:lang="es" xsi:type="mimeta:TFT">demostraciones</dc:type>
<dc:subject xml:lang="es" xsi:type="mimeta:TD">Buscadores</dc:subject>
<dc:subject xml:lang="es" xsi:type="mimeta:TD">Operadores booleanos</dc:subject>
```

This description enables the corresponding document to which the record refers to be retrieved. This can be useful, for instance, for teachers in need of digital videos (<format>) that illustrate (<type>) how boolean operators are used in search engines (<subject>) to integrate them into their own LOs.

5. Conclusions and future work

After analysing type, structure and function of educational documents we have constructed a controlled vocabulary that enables teachers to retrieve LO and their components with the specificity and granularity required in their tasks of content reuse for the creation of new educational content.

From this point on, several research lines must be explored. On the one hand, research is focusing on implementation of the thesauri’s graphic display in a concept map manner – with richer semantic relationships linked to the corresponding resources- (8). On the other hand, query log analysis (9) will move us closer to the vocabulary actually used by users, which will surely enrich our thesauri, improving their retrieval power. Finally, we are planning to develop specific methods to evaluate the impact of the vocabulary on content reuse and, therefore, on teaching work.

Acknowledgements
The DOTEINE Project has been financed by the Spanish Interministerial Commission of Science & Technology (CICYT, ref. BSO2003-04895). The IACORIE Project has been financed by the Regional Government of Madrid (Comunidad de Madrid, ref. 06/HSE/0165/2004). We would like to thank Dr. Jian Qin, from the School of Information Studies at Syracuse University, for the review of an early version of this paper.

6 References: