Developing controlled vocabularies for educational resources sharing: a case study

Daniela Giordano[1], Isaak Kavasidis[1], Concetto Spampinato[1], Panagiotis Bamidis[2]

[1] Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, Università di Catania, Viale A. Doria 6, 95125 Catania, Italy
[2] Lab of Medical Informatics, Medical School, Aristotle University of Thessaloniki, PO Box 323, 54124 Thessaloniki, Greece
{dgiordan, ikavasidis, cspampin}@dieei.ing.unict.it
{bamidis}@med.auth.gr

Abstract. We present a novel controlled vocabulary for the classification of Educational Resource Type and Media type that has been developed within the mEducator Best Practice Network, and discuss the rationale behind its design and its SKOS implementation. Several available controlled vocabularies were analyzed, but none was found entirely satisfactory for the project's purposes. Although the driving motivation of this effort was to account for and deal with multi-type content variety that is especially characteristic of the medical field, the result of this process and the proposed organization generalise fairly well to broader educational/learning contexts. We discuss the proposed vocabularies pointing out the multifaceted role they may fulfil in the linked learning era.

Keywords: Resource Type, Media Type, controlled vocabularies, SKOS, Linked Data, metadata reuse

1 Introduction

Thesauri and taxonomies (which extend simple lists of terms by expressing also the relationship between terms) are currently acknowledged as an essential tool in any successful knowledge management effort [1]. In a taxonomy terms are arranged or linked in a tree, so that narrower, more specific “children” terms fall under broader, more generic “parent” terms; in thesaurus relationships of affinity, synonyms and relatedness among terms can be expressed. Controlled vocabularies, more simply, consist of a flat list of terms used for indexing or categorizing; their goal is to ensure consistency in indexing, tagging, or categorizing and to guide the user to the desired information. In this respect, controlled vocabularies can be used to assist the users either in filling metadata for resource description, and conversely, can be proposed in the user search interfaces to assist during the search process. This latter practice somehow addresses a recent criticism of taxonomies/controlled vocabularies, i.e. that often they are too far from the user language, and therefore are unlikely to be used spontaneously for searching. Hence the support for effective resources retrieval is questioned. On the other hand, “folksonomies” emerging from social tagging of

resources do provide a collection of terms closer to the end-user language. Taxonomies and folksonomies are not mutually exclusive or incompatible, and there is a fast growing body of literature pointing out the benefit of their integration [2], [3].

In systems that resort to user generated metadata, as content sharing solutions often do, there is a tension between the design choice of enforcing controlled vocabularies versus allowing free tagging and filling of the metadata fields. This tension originates from considering that in the indexing phase, when a resource is published, and should be described as effectively as possible, the user, who is not a professional indexer, should be aware of the vocabularies, and know how to use them. Often this is not the case, and even when the user is aware of the existence of a classification system, its actual usage, especially in the case of vocabularies with a very large number of terms (e.g., MeSH) will depend on how easily it can be browsed and understood. On the other hand, filling with unconstrained vocabularies metadata fields, often results in poor quality because of semantic ambiguity in the name of the fields [4].

From a Linked Data (LD) perspective [5] some of these issues can be solved or ameliorated at a fundamental level. In fact, the LD paradigm places emphasis on the reuse of available metadata fields even from different schemas, which inherently favours the reuse of the most useful and meaningful fields; whereas concerning the actual choice of controlled terms, there is the freedom to choose any system that suits the need of the community, as long as it is treated consistently with the DL principles and links are created with other data sets dealing with similar terminologies. An even better practice is to create through RDF links explicit mappings, by resorting to RDF or OWL properties between similar terms in different classification systems (e.g., through the use of OWL's "same as", or "relates to" or SKOS's "closeMatch", "exactMatch", etc.). One clear advantage of this approach is the enabling of automatic metadata enrichment by resorting to the LD Cloud (e.g., [6], [7]): this practice may result into more robust search and retrieval systems also by assisting the users with dynamical suggestion of terms during metadata publishing or during the searches.

The goal of this paper is twofold. First, we present a controlled vocabulary concerning Educational Resource type and Media type that has been developed in the context of the mEducator project, funded within the eContentPlus Programme and concerned with best practices in innovative technical solutions for educational content sharing and repurposing. This vocabulary has a specific raison d'être in the context of a very wide and open network of stakeholders broadly concerned with medical education and also with the practical issue of repurposing educational resources that are often overspecialised and very expensive to develop. Then we discuss generalisation issues and, based on the experience accrued within the project we derive some implications in the context of the Linked learning movement.

2 Background: multi-type content sharing and repurposing

The mEducator project ("Multi-type Content Sharing and Repurposing in Medical Education, www.meducator.net) is a Best Practice Network funded by the European Programme eContentPlus and is concerned with the evaluation of two innovative ways to share medical educational contents across European institutions and across

the community of educators (i.e., one solution utilizing semantic web services technology and one solution based on mashup technology). Among the project's goals is the provision of recommendations to relevant standardization bodies (e.g. to Medbiquitous, that currently employs the HLOM, the Healthcare extension of the IEEE LOM, [8]) regarding suitable, additional extensions to the metadata schema to effectively support both the sharing and the repurposing process. In this regard, an action of the project has been concerned with the actual design and development of these extensions. After several rounds of polling the partners about inclusion of the available fields from the HLOM, and critical re-examination of the HLOM from the Technical Reference Group of the mEducator Consortium, an RDF-based meducator schema was developed [9], reusing existing metadata fields and including new meducator fields dealing specifically with the repurposing history of a resource and other pedagogical aspects of the resource, e.g. medical educational outcomes.

One line of investigation of mEducator that is at the basis of the work presented in this paper concerns the identification of educational content types that are specific to medical education and to test the descriptive adequacy of the currently available e-learning metadata standards against these content types. For example, a Virtual Patient (VP) is a specific medical content type that employs a problem-based learning approach and allows the learners to follow either linear or branched paths to diagnose the patient and suggest treatments [10]. Similarly, interactive medical images (e.g., X-rays) where the learners can draw tracings to identify anatomical structures or pathologies/abnormalities and obtain visual feedback are a novel resource type especially suitable for the development and refinement of visual perception skills [11]. Initially, the distinction between content type and resource type was blurred, and the only agreement within the Consortium on the semantics of the field was that it should try to capture the nature of the educational resource without any reference to topics or disciplines. Thus, during the first iteration of testing and validation of the proposed meducator metadata schema, the field "content type" was included, and it was left as free text to fill (i.e., the initial decision was not to use any controlled vocabulary). The rationale of this decision was to gather data about the variability of the type of educational items that would be contributed by the content providers of the Consortium, and use this information to inform subsequent decisions about the meducator schema and the potential need to resort to controlled vocabularies.

3 Methodology

3.1 Deriving the requirements

The first version of the schema, where content type was a free text field, was used by the Consortium Partners to describe the educational resources that each one had agreed to contribute to the project. This was done by resorting to MetaMorphosis, a social network based on ELGG platform [12], where the users from the participating institutions could register and contribute their resources by filling the metadata forms. The data from 350 forms contributed by about 100 users form the various institutions were analysed by means of content analysis.

From the analysis of how the users filled the field it was apparent a conflation between format (e.g. PPT presentation), educational content type (e.g., a tutorial, or a case study) and tool/interaction type. For example, we had users who referred to "webTraces" to indicate that their contents were images annotated with graphical feedback, since such content had been developed within the LCMS "WebTrace"[11]. This conflation of several aspects into the same field, pointed to the need to clearly differentiating "media/format" aspects from "content type", and to attempt a differentiation between the pedagogical role of the resource and the rhetoric of presentation/interaction.

Some semantic overlap with notion of “subject/topic” was noted in the way users filled the form, although this was not entirely unexpected. In addition, the analysis of the content to be shared revealed that, consistently with the aim of the mEducator project, the type of materials to be shared amongst were broader in scope, and would include design documents such as curricula, or elaborations over official documents (typically, clinical guidelines) to be considered jointly (i.e. as companion resources); other cases would include referral to general resources available on the web, where up to date materials would be issued regularly on a given topic or discipline.

The broad typology of sharing needs from the mEducator community target users was an additional reason to support the choice of resorting to a controlled vocabulary.

The general requirements for the controlled vocabulary were:

- Maximise reuse from existing vocabularies, also in the light to ensure machine processability and compatibility with existing standards.
- Cater for the need of sharing materials that might not be "strictly" educational. This implies complementing a focus on the instructional/pedagogical aspect of the resource (e.g. a tutorial) with a focus on the informational model (e.g. a "news", regardless of whether it is being delivered through TV, Internet, or Newspaper). A clear separation of the informational model from the media of delivery would facilitate repurposing across different media.
- In the characterization of the media, novel types emerging from the Web 2.0 paradigm should be considered (e.g., Wikis, blogs, and the like).
- Take into account the variety of pedagogical approaches that are in place in the community and are reflected in community specific "content types". This requirement might be referred to as pedagogical flexibility [13].
- Completeness, with respect to the specific "content types" in use in the community of medical educators.
- Support efficiency in the retrieval, but also in the annotation process, i.e., the vocabulary should be as lean as possible and understandable at a glance.

3.2 Related work

A widely shared criticism of current e-learning standards is that they fail to address adequately the "learning" [14], and, in general, miss crucial information required for efficient searching and automated processing [13]. From the perspective of the e-

learning community, two works are especially relevant. In [13] the LOM "learning-resource type" field is criticised on the ground that the controlled values mix instructional (e.g., *exercise, simulation, experiment*) and format information (e.g., *diagram, figure, slide, table*) that need to be separated. This is exactly the situation in which we incurred by leaving the field free to be filled by the users. It should be noted that the same problem occurs in the HLOM extension, where both instructional (e.g., *tutorial, Virtual Patient, simulation*) vocabulary values and format values (e.g., *narrative text*) have been added. The solution proposed in [12] is an ontology, with root class Instructional Object. The philosophy of this work is to subsume under the subclass "Concept" the specific domain of interest (definition, facts, laws and processes) and to complement "Concept" with "Satellite" elements, a subclass that captures the instructional information associated to any concept of the domain (e.g., *real world problem, exercise, exploration, example non-example, explanation, remark, evidence, etc*). Format and media aspects are not addressed in this ontology.

On a similar note, but with a different goal, i.e., supporting automated annotation, [15] proposes the ALOCOM ontology to capture both the content structure and the content type of a Learning Object (LO). The need to express the Structure of the LO, stems from having to deal effectively with the different granularity levels of a LO, which is seen as consisting of various "content fragment" (non further decomposable media); "content object" (an aggregation of fragments with added navigational features), and "learning object" (an aggregation of content object on a learning objective) The proposed classes for learning object are: *lesson, course, chapter, test and tutorial*. From the content type perspective, the pedagogical roles are captured in the "supporting" subclass of content object, and at the moment include: *example, exercise, reference, description, illustration, question, and answer*. This is somehow similar to the "satellite" subclass in [13].

There exist several other Resource Types and Media Type vocabularies. Some have been developed within broader metadata standard specifications, e.g., Dublin Core [16], some have been specifically developed by educational communities, such as the RDN/LTSN (Resource Discovery Network/Learning and Teaching Support Network) [17]; a comprehensive listing of the variety of vocabularies developed in the attempt to capture pedagogical aspects is provided in [18], where some general requirements that should inform their development are also provided.

Concerning the media, in addition to the MPEG-7 ontology [19] there are specialisations in narrower domains, e.g., the RDA/ONIX framework for resource categorization in the media publishing domain [20]; whereas another set of relevant terminology is available through the categories "publication format" and "publication component" widely accepted medical thesauri MeSH [21]. Another relevant work is the SIOC ontology [22], providing terms especially relevant to characterize emergent social media.

3. 3 Design considerations

The methodology used to craft the taxonomies or the controlled vocabularies for mEducator was to analyze first existing vocabularies, taxonomies and ontologies, to evaluate their fit for reuse. Selection of the most appropriate terms was then

performed based on coverage of the meducator requirements, and then adding the missing concepts emerged from the first filed testing of the mEducator schema.

Critical analysis of the available resources, however, pointed out some shortcomings with respect to mEducator needs. A common problem was lacking expressivity in the description of learning/teaching activities implied or supported by the resource, other ones were too biased towards lecturing-based pedagogy or tutorial expository modes; others were too much oriented toward media fragments (e.g., [19]). Others would be too focused on various granularity levels in content organization (e.g. [15]) whereas other used the more practical approach of condensing in one term more than one aggregation level (e.g.,, "course/module/unit" in the RDN/LTSN vocabulary [17]). This latter approach was adopted as a design criteria, since it was deemed a good practice, alternative to pursuing generalisation into more abstract, catch all terms, that often tend to be semantically ambiguous. The expected benefit is lessening the cognitive load on the user without sacrificing precision of retrieval, at least as far as concerning identification of the nature of the resource.

Analysis of the identified sources mentioned in section 3.2 resulted in the selection of 45 concepts obtained by picking up from the sources, merging with mEducator community specific terms, and in some cases rewording some concepts. Nearly none of them were mutually exclusive; thus, from the user annotation perspective, the classification grid must enable multiple selection to characterise reasonably the resource to be shared.

In general, the term selection and refinement process was carried out taking into consideration that in the end, the resources will be indexed by the end-users and not by professional cataloguers, therefore attempting to obtain manageable lists (e.g., a too long list should be categorised in a way meaningful to the end-user), with terms readily visualised and familiar (ideally, self-explaining).

4 The mEducator Resource Type Taxonomy

Resource Type in the context of the mEducator project is defined as:

> *A classification of the Resource based on its informational qualities, taking into account the nature of the information provided, how it is organized, presented or collected (e.g., a tutorial, a textbook, a simulation) and the nature of the interactions that are expected to take place between the users and the content (e.g., reading, practicing, experimenting, assessing, etc.). Resource type includes all the artifacts that traditionally have a role in the context of educational practice, and any document, product or tool that has a role during the practice of healthcare professions and that can be either the focus or a supporting resource for the design of an educational experience (e.g., a laboratory test, or a clinical practice guideline).*

On purpose, aspects strictly related to media, format and implementations were not considered, to give primacy to the nature of the information provided by the resource. The 45 concepts were classified according to the following three broader categories:

1) Educational practice artefact

Definition: any document, product or tool designed and developed *specifically* for use in any phase of the teaching/learning process, including the supporting institutional and administrative activities, or generated as a result of the educational activity.

Comment: this category includes all the artefacts that have a role strictly within the context of educational practice, and as such, would be of no particular value in the context of practicing a healthcare profession (e.g., a Game, a Study Guide, a reading list)

2) Professional practice artefact

Definition: any document, product or tool that is used during the practice of healthcare professions.

Comment: The emphasis is on the artefacts that have their "raison d'être" in the professional context (e.g. a clinical record or diagnostic test) but that can be effectively introduced in the educational activities to design realistic, authentic learning experiences. Resources of this type might not be readily available in digital format; therefore the capability to identify and retrieve them would facilitate their repurposing to customize learning/teaching activities.

3) Reference Material

Definition: any collection of information that is not routinely used in professional/or educational practice, but it is used only for reference purpose.

Comment: for this category a complete mapping to a selection of MeSH terms was possible, and the Mesh definitions were reused.

To facilitate reuse, the vocabulary has been further organised into "Core Resource Type " and "Medical Resource Type" vocabulary, this latter being an extension of the Core Resource Type. Table 1 lists the terms under each category, and definitions for each term are provided in Appendix. It must be noted that a Resource can be indexed according to multiple terms from the same or distinct category; for example, a "course/module/unit" might contain a "simulation", "problem/exercise with feedback" and "reading list", and therefore should be indexed with these four terms.

4.1 Comments on the vocabulary values: towards generalisation

It is interesting to note that a good proportion of the items that have been identified are quite general and would apply to any teaching/learning setting, regardless of the specific domain, although this analysis has net been done in the light of completeness with respect to other domains. It is easy to see that for many items collected under the Medical resource type, equivalent ones can be found in other domains.

In the educational practice artefacts, a Virtual Patient could be seen as a type of Simulation, and Clinical Case Study as a domain-based specialisation of Case Study; making these specialisations explit as medical type extensions increases precision.

"Teaching file" was an interesting case of semantic differences across communities, since medical groups such as radiologists refer to a well established practice of creating files with Images to be discussed/diagnosed by the students, whereas other communities (paediatrics, general educators) used the term to refer to a set of organized resources meant for the teacher. Thus the approach was to keep the term as used by radiologists and include it in Medical Resource Type, and introduce the notion of Resource Pack that generalizes the second sense of "Teaching file" (see definition in Appendix).

Table 1. The mEducator Resource Type taxonomy. In appendix 1 definitions are listed.

mEducator Resource Type Taxonomy			
	Educational Practice Artefact	**Professional Practice artefact**	**Reference material**
Core Resource Type	Course/Module/Unit Lecture (recording) Lecture Slide/Presentation Lecture Notes/Handout Textbook/Chapter Tutorial Reading List Worked example Demonstration Simulation Case Study Practical Problem/Exercise Problem/exercise (with Feedback) Game/Serious Game Assessment Item/Instrument Student generated content Resource Pack Study Guide Syllabus/Programme/Curriculum Educational Policy	Professional Practice Guideline Scientific Journal Article Diagnostic algorithm Table Dataset Database Software	Atlas Bibliography Dictionary Encyclopaedia Handbook Index Legislation Laboratory manuals Terminology
Medical Resource Type	**Virtual Patient** **Teaching File** **ClinicalCaseStudy**	**Clinical Practice Guideline** **Diagnostic/Laboratory test** **Diagnostic/Laboratory test** (annotated) **Medical/Diagnostic algorithm** **Medical Form** **Clinical record** **Patient Education Handout**	**Anatomical Atlas** **Pharmacopoeias**

Concerning the "Professional Practice" collection of terms, it can be noted that, many specific terms can have a counterpart in other, highly specialized professional domains (e.g., law, economics, psychology, etc.) that have professional codes, ethics, practice guidelines, data collection tools, software to support the job activity. For example, a "Diagnostic algorithm" or a "Diagnostic/laboratory test" makes sense also in most engineering/technical disciplines, and, as such, could be part of other domain-specific extensions. Similar considerations apply for the Reference materials. As mentioned before, the reason why one would index and share these types of resources, that are somehow "raw materials" (not re-elaborated for educational purposes) is that they can be well suited to be the building blocks of constructivist and "authentic" learning experiences. Also, it might be argued, they can indirectly foster the process of community building (that is most often mediated by sharing tools and tricks of the trade) that can make a system for sharing resources successful or not.

5 The mEducator Media Type Taxonomy

The property “Media type” of the mEducator learning resource refers to the media type of the resource, or the media types embedded in it, if this is a complex one. It is possible to have multiple values for this field, e.g., an image sequence accompanied by an audio clip, or a blog post embedding a video clip. Media type may refer also to the packaging standards that are used to interoperate the resource across different learning platforms (e.g., SCORM package, MVP-Medbiquitous Virtual Patient) but leaves out details such as file digital format and size.

In order to facilitate identification of “fragments” for repurposing, and of classic e-learning “packages” for interoperability, Media Type has been organized in four categories: Primary Media, MultiMedia, Web/Social Media and Media Package. The Web/Social Media category reuses terms from the SIOC ontology [22].

Table 2 lists the terms selected for each category, whereas Table 3 provides the definitions for Primary Media and Multi Media

Table 2. The mEducator Media Type taxonomy.

mEducator Media Type Taxonomy			
Primary Media	**MultiMedia**	**Web/Social Media**	**Media Package**
Text	Interactive Multimedia	Site	SCORM
Image	Immersive Environment	Forum	IMS
Sketch/Graphical Annotation		Blog	DICOM
Animation		Post	MVP (VirtualPatient)
Audio		Podcast	
Video		Webinar	
3D Model		Wiki	

Table 3. mEducator Resource Type Taxonomy: definitions for Primary Media and MultiMedia

Definitions for mEducator Media Type Taxonomy	
Primary Media Sub-headings	
Text	A resource consisting primarily of words for reading. Example: an e-book without interactive features or multimedia elements. Comment: same as http://purl.org/dc/dcmitype/Text
Image	A static visual representation other than text Same as: http://purl.org/dc/dcmitype/StillImage
Sketch/Graphical Annotation	An image or text augmented by layer information. May combine attaching text labels to graphical elements is a natural visual notation that appears in many kinds of hand drawn diagrams, such as those appearing in user manuals, to indicate part-whole relationships. Narrower than Image.
Animation	A series of visual representations imparting an impression of motion when shown in succession. Include visual output from a simulation. Note: specializes http://purl.org/dc/dcmitype/MovingImage
Audio/Sound	A resource primarily intended to be heard. Same as http://purl.org/dc/dcmitype/Sound
Video	A camera-based recording of visual and audible components. Audio might not be included. Note: specialises http://purl.org/dc/dcmitype/MovingImage
3D Model	A three-dimensional digital representation of an object.
MultiMedia - Subheadings	
Interactive Multimedia Resource	A resource that aggregates any combination of text, audio, still images, animation, video, and requires interaction from the user to be understood, executed, or experienced. Examples include AJAX Web pages, Applets, e-books with multimedia elements.
Immersive Environment/Virtual Reality	An artificial, interactive, computer-created scene or "world" within which a user can immerse themselves. May resort to gestural controls, motion tracking, and computer vision respond to the user's actions and movements. Examples include Second Life, or a virtual reality simulation of surgical procedures. Narrower than Interactive Resource

6 Implementation

IMS Vocabulary Definition Exchange (VDEX) [23] is a standard currently being used for the interoperability of eLearning vocabularies; however, since mEducator places emphasis on compatibility with Semantic Web application, and one of the solutions that is going to be tested is based on Semantic Web Services and Linked Data, SKOS [24] was chosen to represent mEducator vocabularies/taxonomies, also in accordance with the JISC recommendations on terminology services and technologies [25]. SKOS (Simple Knowledge Organization System) is the W3C specification for representing thesauri, classifications, subject headings, taxonomies, and folksonomies, and fully supports multilingualism.

The vocabularies namespaces are:

- *http://purl.org/meducator/resourceType/*
- *http://purl.org/meducator/mediaType/*

whereas the RDF files are available at:

- *http://www.meducator.net/mdc/resourceType.rdf*
- *http://www.meducator.net/mdc/mediaType.rdf*

respectively.

The schema used for attaching Resource and Media types to the educational resources, is the mEducator schema that can be found at *www.purl.org/meducator/ns*. In Figure 1 an excerpt of an instance of a resource that uses the implemented vocabularies is shown.

Resource "Biomolecular Structure Function analysis practicals" type is "Practical", "Simulation", "Software", "Resource Pack"

```
<mdc:resourceType rdf:resource="http://purl.org/meducator/resourceType#practical"/>
<mdc:resourceType rdf:resource="http://purl.org/meducator/resourceType#simulation"/>
<mdc:resourceType rdf:resource="http://purl.org/meducator/resourceType#software"/>
<mdc:resourceType
rdf:resource="http://purl.org/meducator/resourceType#resourcePack"/>
```

Resource "Biomolecular Structure Function analysis practicals" media is "animation","3D model":

```
<mdc: mediaType rdf:resource="http://purl.org/meducator/mediaType#3DModel"/>
<mdc: mediaType rdf:resource="http://purl.org/meducator/mediaType#animation"/>
```

Fig. 1. An example of the instance section characterizing the learning resource "Biomolecular Structure Function analysis practicals". **mdc** is the alias of the mEducator schema.

7 Discussion

Resource Type and Media Type have been designed to work jointly to profile the resources. Another option could have been to include in the medical resource type extensions some other specific cases where the Resource Type implies the Media, as is the case for VPs (to a certain extent), or for traceable medical images (that imply both images and graphical annotation). This might lead to some redundancy, but the trade-off could be achieving more immediate recognition of a concept from the users. This is possibility will be re-evaluated after completing the testing of the vocabulary in its current form with the end-users.

Resource Type and Media Type are not the only (mandatory) controlled vocabularies that were adopted in the mEducator schema. A richer profiling of the learning resource is achieved, with respect to controlled vocabularies by properties regarding the IPR licensing schema (property reuse and values from Dublin Core), educational level of the intended audience and learning outcomes taxonomy. This latter one was developed anew and given its specificity for the medical domain its description is outside the scope of the paper. However, it should be noted that this learning outcomes taxonomy plays an important role, similar to the characterization

of the content with respect to the domain (e.g., fact, procedure, law, etc.) included in the ontologies in [13] and [15].

The development of the proposed vocabularies can be considered from two angles: 1) an exploration of what constitutes a content type from an educational perspective, and 2) an attempt to a characterization that is meaningful for the community of medical educators and that is useful in terms of assisting the retrieving of content with some repurposing goal in mind. To this aim it should be remarked that the level of granularity that has been adopted has been very much affected by the notion of *repurposing* that is key to mEducator. In fact, repurposing is a middle ground between the notion of reusing a learning object in its entirety, and reusing some of its media fragments or content fragments [15], this latter requiring a much more fine-grained approach to annotation. On the other hand, *repurposing* presupposes that a fairly complex and elaborated resource keeps its general connotation (e.g., a virtual patient), but is repurposed to take into account different contexts, e.g., different cultures, different languages, different measurement systems, or to cover yet an additional learning outcome. In this sense it seems reasonable to annotate at a level that does not explicitly take into account content structure, and yet may allow specifying the nature of the components included in the resource.

Concerning the practical ways to support the adoption and the correct use of controlled vocabularies (to reap their benefits), issues more directly related to the users must to be taken into account. One lessons from the development of the vocabulary activity was that any classification system, no matter how much "self-explaining" its categories and terms are, becomes increasingly meaningful to the user engaged in the indexing process as long as he or she gains a full view of the facets used in the classification. Interestingly, this leads us into the realm of user-interface design, where the recommendation is to show contextually all the top and first level categories from which the user can select their indexes. This requirement is even more important when the coexistence of the folksonomic approach is allowed (i.e., a user who does not recognize a suitable controlled term to describe the resource can enter a new term), as is the case in our project. For example, if the user thinks that resource content type is a WIKI s/he should be immediately aware that that option has been contemplated in the Media Type section.

8 Concluding remarks

We are currently involved in two lines of activity. The first one is concerned with evaluating and validating the proposed vocabularies, as implemented in the two mEducator solutions to share medical content. The proposed vocabulary will be then presented to Medbiquitous (www.medbiq.org), the international consortium dedicated to develop information technology standards for healthcare education and competence assessment, as part of the goals of mEducator.

The second line of action regards the formalisation of the mappings of the chosen terms towards DC, HLOM, MeSH and other sources used in developing the vocabulary. This will be done taking into account also the output of a recent JISC project on the Vocabulary Mapping Framework [17], where a mapping of

vocabularies from source standards is provided, with the aim to support the building of transformations between any of them. It is expected is that through this activity and through the RDF linking of related vocabularies, a valuable resource to assist in the tailoring of vocabularies for specific communities will be made available.

From a research perspective, we plan to investigate how the network of mappings that becomes realised by connecting vocabularies can complement the information that can be extracted by mining the free text fields of the meducator schema, as proposed in [27], to further facilitate the resource retrieval process.

Acknowledgements

This work was supported by the project mEducator (Multi-type Content Sharing and Repurposing in Medical Education), funded by the eContentplus Programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable (Grant ECP 2008 EDU 418006)."

References

1. Lambe, P.: Organising knowledge: taxonomies, knowledge and organisational effectiveness. Chandos Publishing, Oxford (2007)
2. Kiu C., Tsui, E.: TaxoFolk: a hybrid taxonomy–folksonomy classification for enhanced knowledge navigation. Knowledge Management Research & Practice 8, 24--32 (2010)
3. Bamidis P.D., Kaldoudi E., Pattichis C.: From Taxonomies to Folksonomies: a roadmap from formal to informal modeling of medical concepts and objects". In: 9th IEEE International Conference on Information Technology and Applications in Biomedicine, Larnaca, Cyprus, 5-7 November (2009)
4. Heath B.P., McArthur, D.J., et al.: Metadata lessons from the iLumina digital library. *Commun. ACM* 48, 7, 68--74 (2005)
5. Bizer, C., Heath, T. et al.: Linked data - The Story So Far. Special Issue on Linked data, International Journal on Semantic Web and Information Systems (IJSWIS) (2009).
6. Mannens, E., Troncy, R. et al.: Automatic metadata enrichment in news production. In: 10th Workshop on Image Analysis for Multimedia Interactive Services, pp. 61--64, IEEE (2009)
7. Dietze, S., Benn, N. Yu, H. et al.: Comprehensive service semantics and light-weight Linked Services: towards an integrated approach". In: 8th International Semantic Web Conference (ISWC 2010), Shanghai, China, November (2010)
8. http://www.medbiq.org/std_specs/standards/index.html#HCLOM
9. Mitsopolou, E., Taibi, D., Giordano, D. Dietze, S., Yu, H.Q., Bamidis, P., Charalampos, B., Woodham, L: Connecting medical educational resources to the Linked Data cloud: the mEducator RDF Schema. In: 1st Linked Learning Workshop, ESWC, Heraklion, Crete, May 29, (2011)
10. Kaldoudi, E., Bamidis, P., Papaioakeim, M.,Vargemezis, V.: Problem-Based Learning via Web 2.0 Technologies. In: 21st IEEE International Symposium on Computer-Based Medical Systems, pp. 391--396, IEEE (2008)
11. Giordano, D., Leonardi R.: Web-trace and the learning of visual discrimination skills. In:1st International Workshop on Pen-based Learning Technologies, Catania, Italy. 24-25 May, pp. 80-85.. IEEE CPS (2007)

12. Kaldoudi, E., Dovrolis, N., Konstantinidis, S., Bamidis, P.D.: Social networking for learning object repurposing in medical education", J. Inform Techn Healthcare, 7, 4, 233--243 (2009)
13. Ullrich, C.: The learning-resource-type is dead, long live the learning- resource-type!," Learning Objects and Learning Designs, 1, 1, 7--15 (2005)
14. Jonassen, D., Churchill, D.: Is there a learning orientation in learning objects?" International Journal on E-learning 3(2): 32-41 (2004).
15. Jovanovic, J., Gasevic, D., Devedic, V.: Ontology-based Automatic Annotation of Learning Content. International Journal on Semantic Web and Information Systems, 2, 2, 91--119 (2006)
16. http://dublincore.org/documents/dcmi-type-vocabulary/
17. Barker P., Cross, P., Fernandez, A. et al. : RDN/LTSN resource type vocabulary- v.1.0. Available at: http://www.intute.ac.uk/publications/rdn-ltsn/types/
18. Currier S., MacNeill S., et al.: Vocabularies for describing pedagogical approach in e-learning: a scoping study. In: DC-2006 International Conference on Dublin Core and Metadata Applications: metadata for knowledge and learning.
19. MPEG-7 Ontology, http://www.w3.org/2005/Incubator/mmsem/XGR-mpeg7/
20. Dunsire, G.: Distinguishing Content from Carrier: The RDA/ONIX Framework for Resource Categorization. D-Lib Magazine 13, 1/2 (2007)
21. MeSH, http://www.ncbi.nlm.nih.gov/mesh
22. The SIOC ontology. http://sioc-project.org/
23. http://www.imsglobal.org/vdex/vdexv1p0/imsvdex_bestv1p0.html
24. SKOS, Simple Knowledge Organization System, www.w3.org/2004/02/skos
25. Tudhope D., Koch, T., Heery R.: Terminology Services and Technology - JISC state of the art review, http://www.ukoln.ac.uk/terminology/JISC-review2006.html (2006)
26. The vocabulary mapping project, http://cdlr.strath.ac.uk/VMF/index.htm
27. Giordano, D., Faro, A., et al.: Feeding back learning resources repurposing patterns into the "information loop": opportunities and challenges. In: 9th IEEE International Conference on Information Technology and Applications in Biomedicine, Larnaca, Cyprus, 5-7 November (2009)

Appendix: Definitions for Educational practice artifacts

Definitions for Educational Practice Artifacts	
Course/Module/Unit	A sequence of activities designed to advance student skills, knowledge, and attitudes in a particular discipline and to help students meet requirements as prescribed in a curriculum. Example: a course delivered through a LCMS (e.g., Moodle)
Lecture (Recording)	The video or audio recording, or the transcript of the exposition of a given subject delivered before an audience, for instructional purposes. Example: a speaker presentation at a conference
Lecture Slide/Presentation	Materials that are projected to support the delivery of lectures or presentations in a module or a course. Example: an annotated PPT presentation, an image, picture or diagram.
Lecture Notes/Handout	Notes containing topical information prepared as study/reference material for lecture(s) or a course. Example: a document distributed to the participants in a seminar
Textbook/Chapter	Book or chapter of a book explicitly meant for the study of a specific subject, characterised by the systematic presentation of knowledge about the subject.
Tutorial	A resource that provides guided, practical information about a subject

	Example: on-line primer to a software for statistical analysis
Reading List	A list of recommended or required sources which provide additional information on the subject being studied
Worked Example	A written problem or exercise designed to illustrate step-by-step how to perform a task or how to solve a problem.
Demonstration	A video or audio recording, or a transcript, demonstrating a skill or a procedure in practice.
Simulation	A representation of a process, activity or situation. Designed to support problem-based learning or exploratory learning, or the comprehension of the dynamics of a bio-physical or other complex process. Example: a simulated patient interview, a software simulating a spring oscillation at the varying of the spring parameters and the force applied to the spring.
Virtual Patient	An interactive computer simulation that allow the learner to take the role of a health care professional and develop clinical skills such as making diagnoses and therapeutic decisions.
Teaching file/Clinical Case Study	The presentation of a clinical case through diagnostic imaging modalities. The images contained in the file are typically supplemented with the following information; case title, history/presentation, findings, diagnosis, discussion.
Case Study (Non Clinical)	A detailed account or a process or activity, prepared to assist the student in determining what factors led to its success or failure. Example: a detailed analysis of an institutional change/innovation, a description of a critical incident.
Practical	An activity to develop the practical skills of a subject, or an examination of such practical skills. Example: an activity to be carried out in the laboratory.
Problem/exercise (no feedback)	A task, or other effort to be performed by the student to develop, maintain, or increase skill or cognitive abilities. May include Projects and Fieldwork
Problem/exercise (with feedback)	A task, or other effort to be performed by the student to develop, maintain, or increase skill, including problem-solving, and for which feedback is readily available to the student.
Game/Serious Game	A structured activity, usually undertaken for enjoyment, used to teach about a subject or to develop specific skills while playing.
Assessment Item/Instrument	An item, activity, system or instrument designed to measure student learning. May include question/answers pairs or practical skill demonstration. Example: Self-assessment questions, Multiple choices, Objective Structured Clinical Examination (OSCE)
Student generated content	Any content generated by a student of a group of students as a result of coursework. Can be used to exemplify good and bad practices, or shared for reference. Examples: entries in a Wiki, final coursework, a discussion thread in a Forum
Resource Pack	A complete package for a course, module or unit including learning resources of several types and supporting documentations (e.g., teacher/learning instructions). Sometimes referred to as “teaching file” in educational contexts other than medicine.
Study Guide	A written guide created to provide direction, and point out critical information to the students. It may include techniques for problem solving and hints to manage the study process (e.g. topic sequencing and timing, learning and testing strategies)
Syllabus/Programme/Curriculum	A document describing the contents of units, courses, and courses of study. May include information about access requirements, delivery, activities and assessment modalities.
Educational Policy	A document containing statements about how an educational system should operate.