Open Research Online

The Open University's repository of research publications and other research outputs

A model for the creation of human-generated metadata within communities

Journal Item

Version: [not recorded]

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.1080/17439880500093869

oro.open.ac.uk

A Model For The Creation Of Human-Generated Metadata Within Communities

Andrew Brasher & Patrick McAndrew
The Open University, UK

This paper considers situations for which detailed metadata descriptions of learning resources are necessary, and focuses on human generation of such metadata. It describes a model which facilitates human production of good quality metadata by the development and use of structured vocabularies. Using examples, this model is applied to single and multiple communities of metadata creators. The approach for transferring vocabularies across communities is related to similar work on the use of ontologies to support the development of the semantic web. Notable conclusions from this work are the need to encourage collaboration between the metadata specialists, content authors, and system designers, and the scope for using accurate and consistent metadata created for one context in another context by producing descriptions of the relationships between those contexts.

Introduction

E-learning systems offer opportunities to match appropriate e-learning resources to an individual's learning needs. However, suggested ways to achieve such matching often require detailed metadata descriptions of both e-learning resources and individuals. As an example of a situation in which these detailed descriptions are needed, consider these extracts from a description of students using mobile devices within an art gallery (adapted from Vainio, 2003). Two phrases have been rendered in bold text; it is these phrases which will be discussed below.

> Two people, Anu and Tiina, visit an art museum. Anu is a student on a university course on European art history of 1400-1500. Tiina has a general interest in art from the same period. They each wish to learn more about the works of Sandro Botticelli. They sit in the art museum's study room for a while and examine the exhibition with their mobile devices. They request a map from the gallery's network. **From their learner profiles the devices recognise that they are interested in 15th century art** and indicate areas of the gallery with works from that era. They select Botticelli's works, and the devices propose a route through the gallery. …….
> **Whenever they stop in front of a painting, the device offers a brief audio description of it and displays its image on their displays**.

In order that a computer system can deliver the most appropriate resources to these individuals the metadata describing the available audio descriptions must be sufficiently accurate and detailed so that it can be used to decide which audio description to deliver from a database of available audio descriptions of the paintings. The same comment also applies to the accuracy and consistency of learner profiles for the individuals. Many factors can affect what is an appropriate resource such as location, language, and prior learning. For this paper we focus on the information that can be stored with the object, with reference to the IEEE Learning Object Metadata (LOM) (IEEE, 2002), rather than with the person, e.g. using the IMS Learner Information Profile (LIP) (IMS, 2001), and look at the concept of difficulty in particular. This concept presents problems in interpretation and scope for subjective judgement and so is an example for a range of factors for which similar implementation challenges will occur.

An example of functionality that meeting these challenges could support is to enable an actor (either a person or computer algorithm) to navigate through a database of learning objects via an indication of how difficult each learning object is, e.g. to answer queries such as 'I need a LO easier than this one'. In this paper, we use the term 'learning object' to mean "any digital entity which can be used, re-used or referenced during technology supported learning".

The notion of educational difficulty is an indication for how hard it is to work with or through a learning object for the typical intended target audience. The "typical target audience" can be characterized by data elements in the LOM 5.6:Educational.Context and 5.7:Educational.TypicalAgeRange (IEEE, 2002).

Thus, for the relevant audio description to be delivered to the students, the metadata for each audio description should include accurate metadata describing its

- educational difficulty
- educational context
- typical age range.

Furthermore, to support the functionality indicated above (i.e. the delivery of a relevant audio resource), the descriptions of each of these characteristics should not only be accurate, but also either consistent for all the resources to be compared or the system must be able to cope with inconsistencies in the descriptions of characteristics such as difficulty.

To illustrate what we mean by consistency in this case, consider the vocabulary recommended for educational difficulty (IEEE, 2002):

very easy;

easy;

medium;

difficult;

very difficult.

If someone now assigns metadata from a different vocabulary which included the term 'intermediate' (instead of 'medium') i.e.:

very easy;

easy;

intermediate;

difficult;

very difficult;

and if the term 'intermediate' (instead of 'medium') is applied to several learning objects which are of 'medium' difficulty for the 'typical intended target audience', then this is an inconsistency. However, if the term 'very difficult' was assigned to these learning objects, then this would be inaccurate. In general it is possible to cope with inconsistencies in metadata, however, this will require increased sophistication in any system that exploits the metadata. The approach we outline in this paper seeks to avoid such problems by integrating the design and evaluation of vocabularies into the overall user-centred design process.

It is possible to automate the generation of some metadata. For instance there are some metadata descriptors which can be generated from an analysis of the learning resource itself, i.e. they can be generated from *intrinsic sources* (Brasher & McAndrew, 2004). However, in this paper we will focus on human generation of metadata. In the following sections we present a model to facilitate the production of accurate and consistent human created metadata descriptors by individual communities. The model focuses on the development and use of structured vocabularies (see e.g. CEN, 2003; ISO, 1986). Furthermore, we explore methods

which enable metadata created by more than one community to be interpreted consistently.

The model

Introduction

As discussed in our paper (Brasher & McAndrew, 2004), the factors affecting the quality of human produced metadata include motivation, as well as accuracy and consistency. For the design of a metadata creation system it is recommended that a user-centred design process is used because of the critical role of the user in the performance of these systems. The method described here should thus be used within a user-centred design framework such as the socio-cognitive engineering method (Sharples et al., 2002); the working practices we describe fit into the cycle of iterative design described in that paper.

One challenge in the design process for a human metadata creation system described above is to be able to identify or establish a community that will generate the metadata. For reasons identified in our paper (Brasher & McAndrew, 2004), a community which is involved in the design and creation of the educational resources (e.g. authors, teachers, tutors) is a community which designers of a metadata creation system will look to in order to fill the role of creators.

We assume that a metadata schema has been selected prior to initiating our model, e.g. through comparing publicly available schemas to the overall system requirements. In summary the initial phases of our model are

1. iterative development of a vocabulary and tools to enable prototypical use of the vocabulary to create metadata by a community: the metadata sets thus created are then used to test the utility of the vocabulary and tools, until a final version of the vocabulary is agreed upon;

2. analysis of the tasks carried out by a community of potential metadata generators to establish a practical way to generate metadata using the vocabulary.

Furthermore, we suggest that the interpretation of vocabulary data can be considered for two cases – one where a single community can be expected to work on describing resources, and the other where multiple communities need to collaborate to build up and share vocabularies. For the first case (i.e. single community) phases 1 and 2 can satisfy the requirements and we suggest an implementation of these phases using a semi-formal approach based on exchange of vocabularies and shared access. However, in the second (multiple communities) we look to the work on the semantic web to offer a formal method based on encoding individual communities' understanding in an ontology shared between multiple communities. Thus the last phase of our model is

3. linking communities via ontology development and instantiation.

Development of Structured vocabularies

Introduction.

Greenberg et al. (Greenberg, Pattuelli, Parsia, & Robertson, 2002) have described how a simple Web form can assist authors in generating 'good quality' Dublin Core metadata. We describe below an approach to improving accuracy and consistency that has been developed drawing from the experiences at the Open University, Greenberg's work (ibid), developments such as the IMS Vocabulary Definition

Exchange (VDEX) specification (IMS, 2004a) and our own research within the GUARDIANS and MOBilearn projects (GUARDIANS, 2002; MOBIlearn, 2004). Our work considered how to assist those completing metadata with the aim of allowing a larger group of people, including authors and editors, to supply metadata of sufficient quality. The approach uses the current range of XML editors in conjunction with vocabulary schema to address the issues of accuracy and consistency. Three aspects are tackled:

1. vocabularies are identified and/or developed to control the terminology that can be used in completing a metadata instance;
2. the vocabulary information is augmented with descriptions to help people to understand the metadata requirements;
3. the vocabularies are stored in way that facilitates their prototyping and iterative refinement during the design and development of systems that will create and exploit the metadata.

The use of vocabularies was formalised in two ways: (i) by utilisation of a schema for vocabulary structure and (ii) to explicitly refer to the location of vocabularies used within a metadata instance via a URL.

The structure schema (see e.g. (IMS, 2004a), http://guardians.open.ac.uk/schemas/thesaurus/) offers the facility to define a vocabulary with the properties of a thesaurus (ISO, 1986), which provides features such as the explicit definition of relationships between terms in the vocabulary. Making such relationships explicit aids users of the vocabulary (in this case metadata authors), because it helps to reduce uncertainty about which is the most appropriate term.

The *location* means that vocabularies are stored and referenced via a URL. This offers access to the **complete** vocabulary through the interpretation of the metadata, and also provides a record for vocabulary decisions as they are made. The IEEE standard LOM does recommend the use of URIs to reference vocabularies as good practice (IEEE, 2002) but does not explain how to code the vocabularies in a way that facilitates their exploitation. Information about how to code vocabularies is provided by IMS in the aforementioned VDEX specification (IMS, 2004a) which also gives some use cases indicated how vocabularies encoded in the VDEX format could be exploited. In the next section we propose that such structured vocabularies can facilitate the design of systems for creating and exploiting metadata.

Tools and working practices.

A diagram showing the working relationship of the tools for rapid prototyping and evaluation of specific vocabulary and metadata schema combinations is shown in Figure 1. In this diagram the 'Standard Metadata schema' will be a metadata schema that has been selected to meet the overall system requirements.
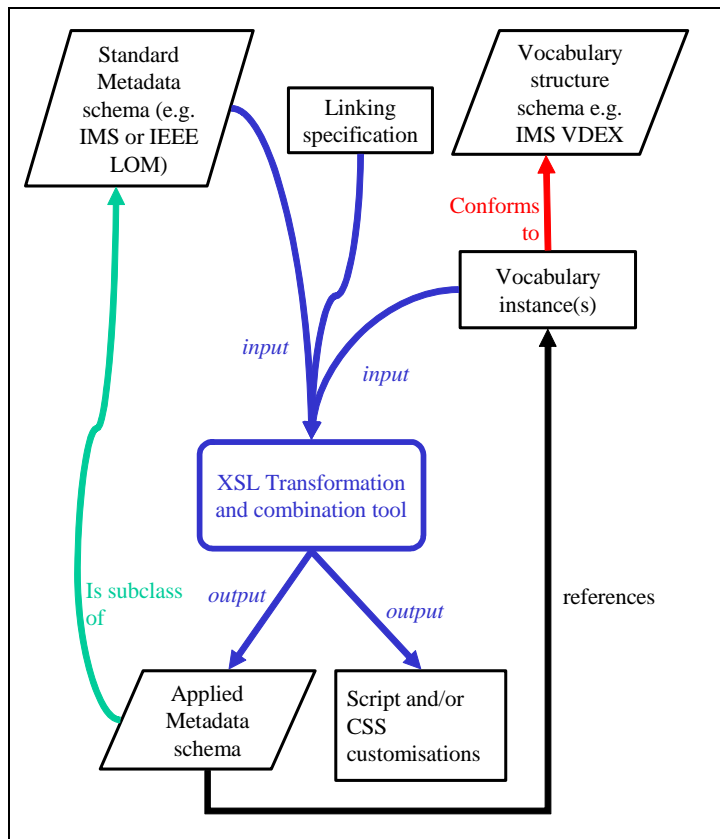
Figure 1: Relationship of tools, schema and data for rapid prototyping

This example assumes that the initial requirements for the functionality of the creation tool have been gathered, and that an initial specification of user interface objects, tools and additional resources has been made. The tools, schemas and data shown in Figure 1 may then be used to develop demonstration prototypes in the following way

1. Use the *Vocabulary structure schema* (e.g. (IMS, 2004a), http://guardians.open.ac.uk/schemas/thesaurus/) to implement vocabulary instances.

2. Run the *XSL Transformation* (Clark, 1999) *and combination tool.* This takes as inputs the vocabulary instances created in step 1, the standard metadata schema and a *linking specification* (an XML document that specifies which vocabulary instances should be used with which elements in the Standard Metadata schema). Its outputs are:

   i. an Applied Metadata schema which limits the values permitted in the relevant elements to values in the relevant vocabularies (as specified by the linking specification);

   ii. a set of script and/or CSS and/or user interface object customisations for the chosen XML instance editor or browser-based prototype;

3. Use outputs (i) and (ii) to customise the XML instance editor or browser-based prototype so as to produce the demonstration prototype.

4. Evaluate demonstration prototype by producing metadata instances; i.e. evaluate both the vocabulary and the tool itself with the community of users.

The Standard Metadata schema and Vocabulary structure schema (see Figure 1) will usually be invariant during the process (although it is conceivable that recommendations for changes to one or both could be an outcome of the evaluation in step 4). It is the contents of the vocabularies (i.e. the vocabulary instances) and the

nature of the user interface of the XML instance or browser based editor which will usually be required to change during iterative development. A practical approach can be to produce only an Applied Metadata schema for the first iteration of the demonstration prototype. This, used in conjunction with an off-the-shelf XML instance editor can be used to evaluate the contents of the vocabulary resources specified. A screen shot illustrating how the use of such an Applied Metadata schema within an off-the-shelf XML instance editor produces drop-down menus for choosing vocabulary values is presented in Figure 2. Indeed, a variant of this approach has been described in the IMS Vocabulary Definition Exchange Best Practice and Implementation Guide (IMS, 2004b, section 6.1).

A similar approach can be used to generate prototype user interface components to customise an off-the-shelf XML instance editor or for inclusion in browser based interfaces. There is a finite number of user interface objects that may be chosen for use within such interfaces (objects such as text, text input boxes, list boxes etc.). In this case the *linking specification* will also specify the output object required, and the *XSL Transformation and combination tool* would generate e.g. prototype Javascript user interface components containing the data from the relevant vocabulary resources. Figure 3(a) shows an example of a structured vocabulary implemented in the IMS VDEX format. Figure 3(b) shows a HTML form which has been generated from this example.
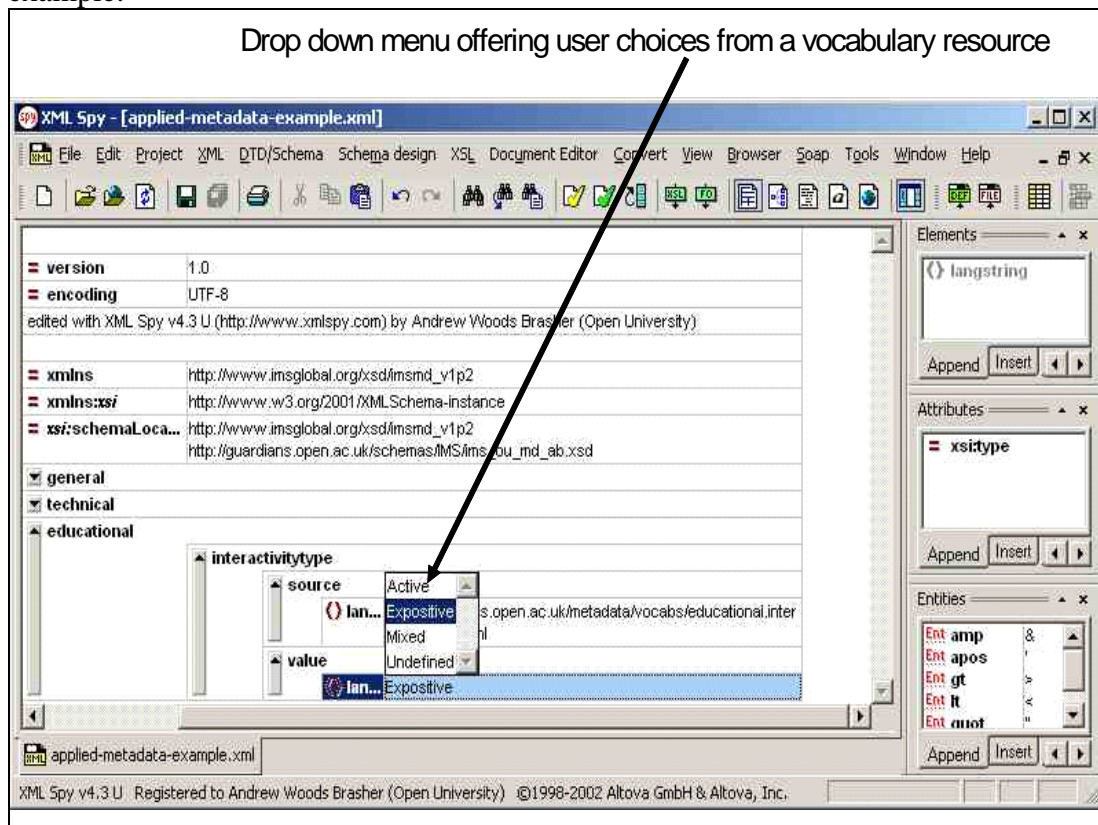

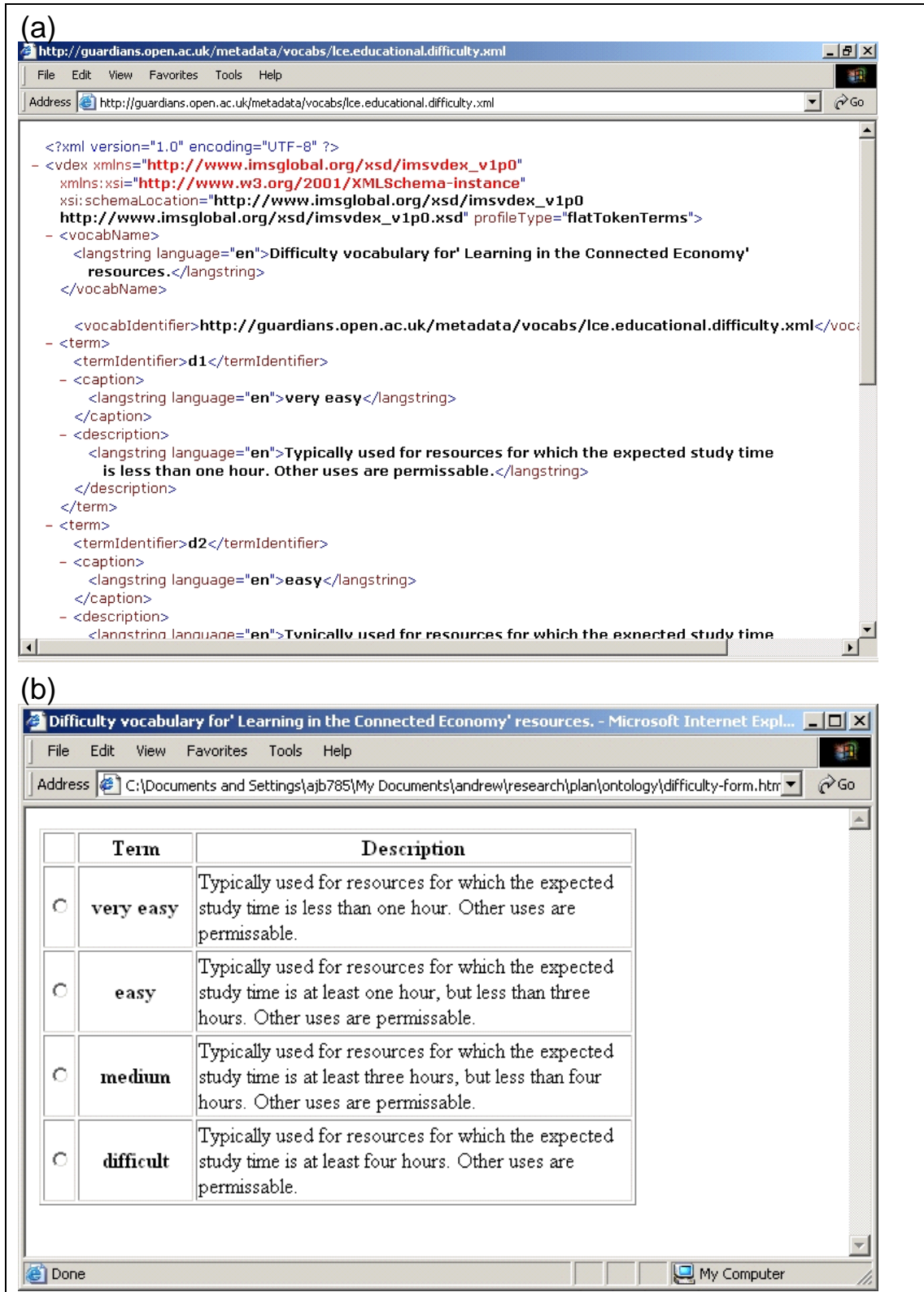**Figure 2: Screen shot of XML Editor**

Figure 3: **(a) VDEX source for difficulty vocabulary (b) HTML form generated from this vocabulary**

To summarise, the approach to rapid prototyping of metadata creation systems that we have proposed is based on the creation and iterative development of XML encoded structured vocabularies. These vocabularies are exploited in the prototyping process via transformations which facilitate evaluation and iterative development of

1. vocabulary terms and vocabulary structure and
2. 'Help' information for metadata creators

Both these features (1 and 2 above) add semantic information relevant to the specific user community targeted by the prototyping process to the syntactic information provided by the metadata schema. The prototyping process should then enable the design and implementation of metadata creation systems that use this semantic information to promote the generation of accurate and consistent metadata by the user community in question. This type of information has been shown to improve consistency amongst metadata creators, in particular Kabel et al. report the positive effect of vocabulary structure on consistency (Kabel, Hoog, & Wielinga, in press).

Generating metadata using vocabularies

The following simple example shows how useful metadata can be generated from extrinsic sources (Brasher & McAndrew, 2004) with negligible additional effort by content authors. This is achieved by consideration of the tasks normally carried out by authors and the example demonstrates the role that a consideration of motivational and task-related factors can play within the design process for a metadata creation system.

The example extends the one shown in Figure 3, and is based on a collection of learning objects drawn from a masters-level course 'Learning in the connected economy'. This course was part of a Postgraduate Certificate in Online and Distance Education offered by the UKeU pilot (see Weller, Pegler, & Mason, 2003 for a description of the learning object structure of the course).

In the case of the learning objects created for the course 'Learning in the connected economy', the course team that created the material had a primary 'typical target audience' in mind (note: we use the term 'course team' to refer to the team of authors and editors responsible for creating material for a particular course). This target audience is described on the Postgraduate Certificate in Online and Distance Education website as shown in Figure 4, and for the purposes of our argument, this also represents the educational context of the learning objects.

Figure 4: Entry requirements for PGC course in 'Online and Distance Education'

In this course, one of the pedagogic requirements that the course team for this course have to comply with is that students are given an indication of the 'estimated study time' that each learning object should take to complete, as shown in Figure 5(a).

It is apparent then that this 'estimated study time' figure was generated by the content authors in accordance with the guidelines suggested in the IEEE standard for the 'Typical Learning Time' metadata element:

'Approximate or typical time it takes to work with or through this learning object
for the typical intended target audience.'(IEEE, 2002).

Thus the 'estimated study time' information shown in Figure 5 (a) could be automatically generated from learning object metadata, if this metadata included the correct figure within the 'Typical Learning Time' element. This then provides the motivation for authors to enter a reasonable figure for 'Estimated study time'. For information a view of the actual metadata that is available to authors is shown in Figure 5 (b).

(a)
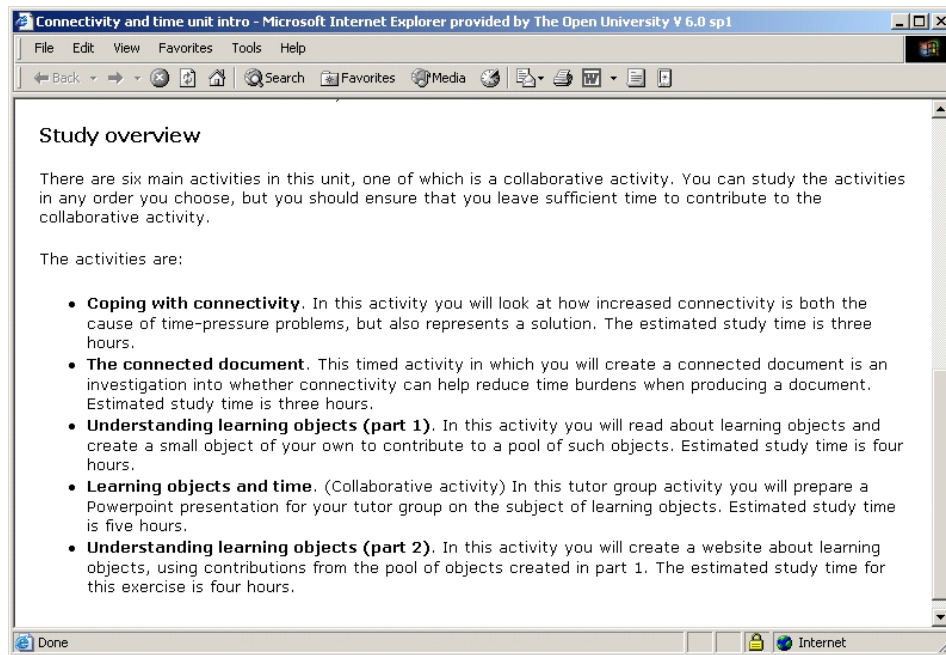
(b)
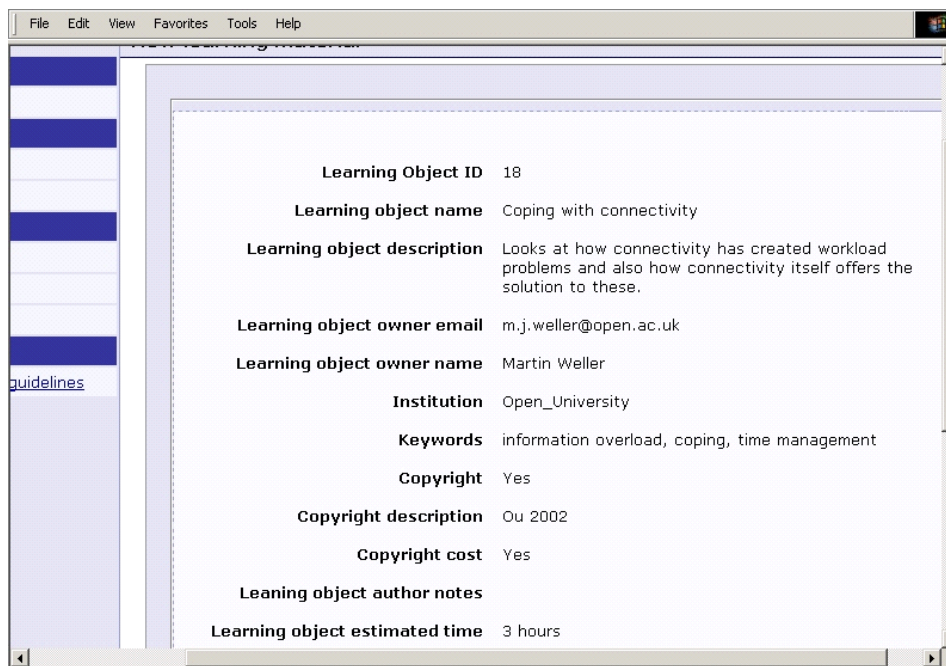
Figure 5: Learning time shown in (a) Information presented to students (b) view of metadata accessible to authors

With respect to the 'Difficulty' element, the explanatory note in the IEEE standard states: 'How hard it is to work with or through this learning object for the typical intended target audience.' (IEEE, 2002), and recall that with respect to both the 'Typical Learning Time' and 'Difficulty' elements the standard states

'NOTE—The 'typical target audience' can be characterized by data elements

5.6:Educational.Context and

5.7:Educational.TypicalAgeRange.' (IEEE, 2002).

Next, we assume that within this context (i.e. the educational context of the Postgraduate Certificate in Online and Distance Education, the context described by the website shown in Figure 4) that 'Typical learning time' will be related to 'Difficulty' for the typical target audience. This is a realistic simplification in this case as task-based learning objects form the major part of the content of the course. For our purposes the exact nature of the relationship is not important – we merely assume that as 'Typical learning time' increases, so does 'Difficulty'. Now, assume that the authors of the course unit for which the overview is shown in Figure 5 (a) think that the range of levels of 'Difficulty' for this course is 'very easy' to 'difficult', i.e. they agree, as a community, that this range of descriptors of 'Difficulty' are apt. Then by analysing the 'Typical learning time' information given to students (and entered into the metadata) for the complete course, this community of authors proposes a categorisation based on this 'Typical learning time' as shown in Figure 6.

'Typical learning time' / hours	'Difficulty'
< 1	Very easy
1 <= 'Typical learning time' < 3	Easy
3 <= 'Typical learning time' < 4	Medium
'Typical learning time' > 4	Difficult

Figure 6: Categorisation of 'Typical learning time'

This categorisation enables 'Difficulty' metadata to be automatically generated for **this particular educational context**, and inserted into the relevant metadata record as in Figure 7. Such a metadata record provides the system with a reference to the source vocabulary itself, hence enabling algorithms to utilise the vocabulary. For example, the caption data shown in Figure 3(a) can be retrieved via the source URL in Figure 7.

```
- <difficulty>
  - <source>
      <langstring xml:lang="x-
        none">http://guardians.open.ac.uk/metadata/vocabs/lce.educational.difficulty.xml</langstring>
    </source>
  - <value>
      <langstring xml:lang="x-none">d2</langstring>
    </value>
  </difficulty>
```

Figure 7:   Fragment from example metadata record

This metadata is sufficient to be useful within this particular educational context. It can enable learning objects to be sorted and compared in terms of 'Difficulty' for the context which is described in Figure 4; knowledge of this context is implicit in the community of authors decision about the range of applicable descriptors of 'Difficulty', and their description of the 'Difficulty' of individual learning objects.

However, difficulty metadata that is generated by the mechanism described hereto is only valid for one educational context (i.e. that described in Figure 4), and can only be correctly interpreted by people and systems that are 'aware' of this fact. With the resources described so far (i.e. the vocabulary and metadata resources) it is not possible for an algorithm to make useful comparisons between the difficulty of a learning object on 'Connectivity' from this post graduate course with a learning object on the same topic from an undergraduate course in another subject area. To make such comparisons possible, extensions to the metadata and the creation and

exploitation systems are necessary. For example, the IEEE LOM (IEEE, 2002) permits any number of Educational elements, hence a system could implement a 'Difficulty' element for every educational context perceived to be of interest, and a system to exploit this could include an algorithm to perform the necessary comparisons. However, there are problems with this approach, not least the burden of creating this additional metadata (note that although the difficulty metadata can be automatically generated, the 'Typical learning time' metadata is still a prerequisite, i.e. for every context someone will have to ascribe a value for this, hence allowing the 'Difficulty' value to be generated).

In the next section we propose a method to reduce the burden placed upon communities of authors, yet still enable the difficulty of learning objects to be compared across contexts.

Linking multiple communities

Introduction.

An ontology is a formal explicit specification of a shared conceptualisation (Gruber, 1993). In this definition the term 'conceptualisation' refers to an abstract model of how people think about things in the world; 'shared conceptualisation' means that this conceptualisation is shared by a community, i.e. it reflects a common understanding of the domain in question; 'explicit specification' means that the concepts and relationships within the abstract model are given explicit terms and definitions, and 'formal' means that these concepts, relationships and definitions are expressed in a formal language (e.g. as in a syntactically correct program in a computer programming language). Given this definition, available implementations of the metadata schemas referred to earlier (IEEE, 2002; IMS, 2001, 2004a) can all be considered as ontologies, e.g. the XML schema implementation of VDEX (IMS, 2004c) .

However, these implementations do not provide a mechanism by which the difficulty of learning objects can be compared across educational contexts, without incurring the cost of creating additional metadata specifically for those contexts.

This section shows how 'Difficulty' metadata such as that generated by the mechanisms described in our previous examples can be applied and exploited in other contexts via the application of languages specifically designed to encode ontologies (e.g. (McGuinness & Harmelen, 2003)). For example, we will show how to enable such metadata to be used to allow an actor (person or computer program) to navigate through learning objects via an indication of how difficult each learning object is, e.g. to answer queries such as 'I need a LO easier than this one' **and** to answer queries such as 'I need a LO like this one, I should perceive it as being as difficult as this one, but it should be from a Mathematical view point;'. As before the emphasis is on identifying endemic motivation and enabling individuals to create accurate and consistent metadata with a minimum of effort.

Tools and working practices.

We first describe an example working practice, then describe tools that, by supporting this working practice, can enable the functionality described above.

Firstly, consider two educational contexts: context1 and context2. For example, context1 could be 'PGC in Open and Distance Education' and context2 could be a level 1 course 'Understanding E-Learning: A Guide for Teachers and Learners' (OU,

2004). Please refer to Figure 4 for a description of context1 (PGC) and Figure 8(a) for a description of context2 (level 1). To explain further the differences between these two contexts Figure 8(b) shows advice provided to (prospective) students on The Open University's 'Help with Registration' page (OU, Not known) which explains the meaning of 'level 1 course' i.e. 'Level 1 courses do not usually assume that you have an academic background in the subject area, but most other undergraduate courses and professional and postgraduate courses expect you to have some knowledge (and perhaps some experience), whether or not there are formal entry requirements.'.

(a)

Entry ⓘ   back to top

The course will appeal to a wide variety of people, for example:

- Those who have already taken the plunge into computing and have some experience of the Internet and e-mail.

- Those who would like to learn more about how the Internet affects learning.

This is emphatically *not* a course for 'techies'. Its tone is friendly and accessible as is that of the book. You will encounter a certain amount of educational terminology but this is explained for the lay reader. You need to be a fairly confident computer user, able to install software, use 'Office' software, and browse the Internet. If you don't feel comfortable with this, you might want to take an introductory computing course first.

Level 1 courses provide core subject knowledge and study skills needed for both higher education and distance learning, to help you progress to courses at Level 2. If you have any doubt about the level of study, please seek advice from your Regional Centre.

(b)

Academic requirements

It is important that you are adequately prepared for any course that you decide to take. There are no formal entry requirements for most undergraduate courses, but you must have the basic skills needed for study at university level.

Level 1 courses do not usually assume that you have an academic background in the subject area, but most other undergraduate courses and professional and postgraduate courses expect you to have some knowledge (and perhaps some experience), whether or not there are formal entry requirements. You will find details in the 'Entry' section of each course description.

Most of our master's degrees have formal entry requirements, which you will find in the qualification descriptions.

If you have any doubts about whether you will be adequately prepared, please look at the Learner's Guide to Course Choice. However tempting it may be, please do not try to study at a higher level than you are prepared for. You will find the going hard, and you will risk failure.

Figure 8: (a) Entry requirements for level 1 course in 'Understanding E-Learning: A Guide for Teachers and Learners' (b) Advice provided on The Open University's 'Help with Registration' page

Typically the course team responsible for creating material for context1 is team1, and that responsible for creating material for context2 is team2. These course teams can meet, discuss and agree how the difficulty of learning objects they have created for their own context will be perceived by students in the other context. For example, the teams may agree that learning objects created for context1 (PGC level) will be perceived as 'more difficult' by students in context2 (level 1) than the difficulty descriptors that have been applied by team1 for context1 appear to indicate. This agreement establishes a relationship between descriptions of difficulty in one context and perceptions of difficulty in the other. We propose that in general (i.e. for any pair of contexts and no matter what the exact nature of the contexts in question is), that the gamut of useful descriptors of this contextual relationship is:

- *Assignments of difficulty made in this context **context1** are perceived as 'more difficult' than assignments of difficulty made in this context **context2;***
- *Assignments of difficulty made in this context **context1** are perceived as' less difficult' than assignments of difficulty made in this context **context2;***
- *Assignments of difficulty made in this context **context1** are perceived as 'as difficult' as assignments of difficulty made in this context **context2.***

(Note that in this case 'assignment' should be interpreted as 'the action of assigning', not as ' a task allocated to somebody as part of a course of study'.)

Providing there is metadata available which describes the difficulty of learning objects for these two educational contexts context1 and context2, then what is required is a mechanism for encoding the statements describing the relationships presented above so that they can be interpreted by machines and people. A way of doing this is described below.

An ontology for describing contextual relationships.

The table presented in Figure 9 describes the classes of concepts needed to realise the example of use described above. This simple ontology has been realised in the Web Ontology Language OWL (McGuinness & Harmelen, 2003), and a UML class diagram illustrating the relationships between the classes is shown in Figure 10. Note that for clarity in this diagram (Figure 10) some slot names for the ContextRelationship class have been shortened:

'assignmentsMadeInThisContextAreAsDifficult' shortened to 'asDifficult',
'assignmentsMadeInThisContextAreLessDifficult' shortened to 'lessDifficult',
'assignmentsMadeInThisContextAreMoreDifficult' shortened to 'moreDifficult'.

Difficulty class
Individuals of class Difficulty represent a level of difficulty that can be assigned to individuals of LO. A particular individual (i.e. level of difficulty) can be related to other individuals via the slots 'less_difficult_than' and 'more_difficult_than'.

Assignment class
The Assignment class describes assignments of individuals of Difficulty to individuals of LO by a particular actor in a particular Context. For example, this individual of Assignment identified as **Henry Hall***: 'Henry Hall says that learning object individual lo2 is 'difficult' for students in the 'Physics' context.'.*

Context class
Individuals of the class Context represent the context (e.g. educational level and study skills), that individuals of class Assignment maybe assigned to.

ContextRelationship class
The ContextRelationship class describes relationships between assignments of difficulty in different contexts. An example of how data contained within should be read is given by this individual **R1***: "For students in the Maths context, the assignments of difficulty made for students in the Physics context are less difficult."*

LO class
Individuals of the class LO represent learning objects.

Figure 9: Description of classes in the ContextRelationship ontology
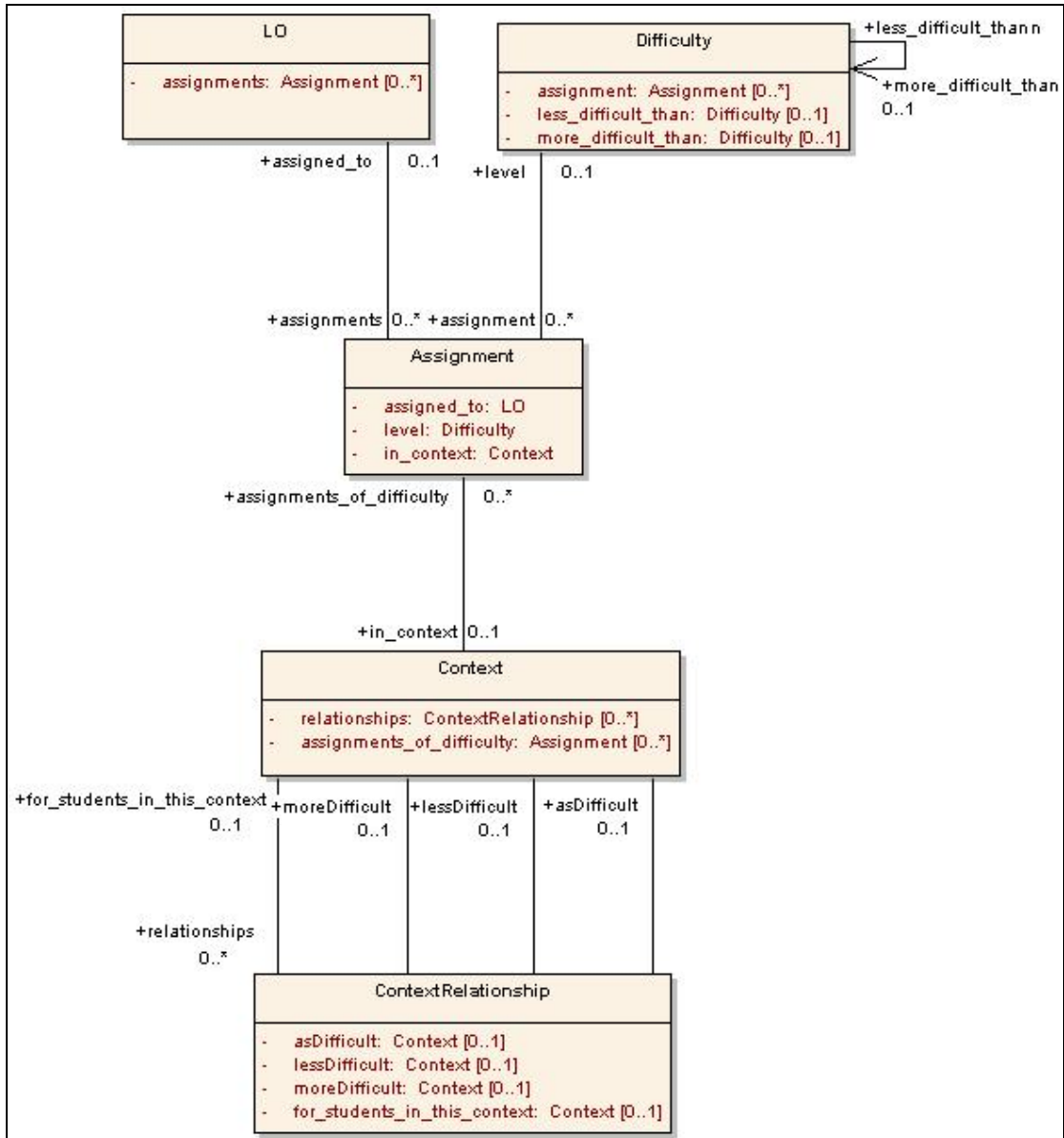
Figure 10: UML class diagram showing relationships between the classes and slots in the ontology

Figure 11 is a schematic which shows the relationship between individuals in this ontology, and entries in the metadata records for two learning objects. One learning object (loLevel1) belongs to context1 (Level 1) whilst the other (loPGC) belongs to context2 (PGC). Again, for clarity in this diagram some slot names for the ContextRelationship class have been shortened as for Figure 10.

Figure 11: **UML object diagram showing the relationship between metadata and instances of concepts in the ontology**
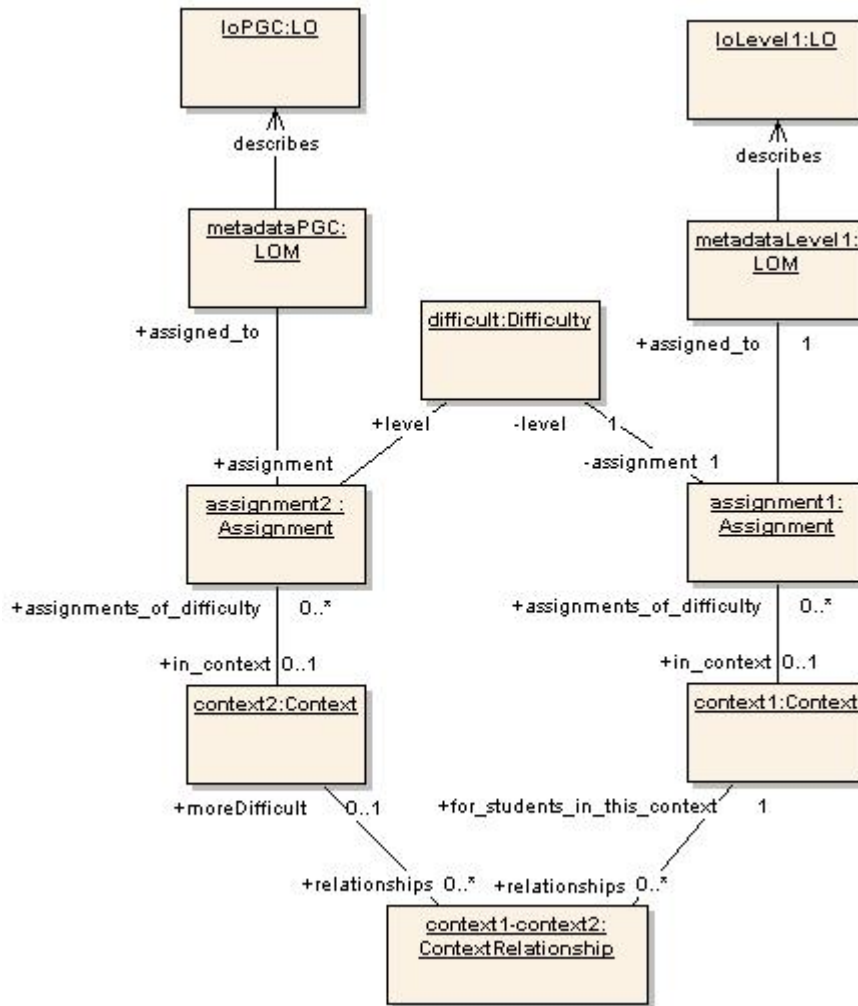
This ontology permits queries such as 'Find LOs of difficulty 'difficult' or greater (for this context) or of equivalent difficulty but from other contexts', and it allows many pairs of contexts to be related to each other in a chain (we note possible issues with tractability and uniqueness of solutions but these are beyond the scope of this paper).

Generating relationships between contexts.

We have shown how to enable metadata to be used which allows an actor (person or computer program) to navigate through learning objects via an indication of how difficult each learning object is, for both single educational contexts and pairs of educational contexts. However, there remains a question of how the data describing such contextual relationships will be generated. Earlier we described how an agreement on the nature of the relationship could be reached via discussions between course teams. We now discuss factors which would motivate such discussions and agreements, and the creation of the necessary individuals in an ontology such as that described in Figure 10. Note that here we use the term 'individual' to mean 'an instance of a class in the ontology' (as in McGuinness & Harmelen, 2003).

Firstly, we focus on the 'Assignment' and 'Context' individuals shown in Figure 11. It is feasible that if the course teams creating the learning objects in question use a content management system (Browning & Lownden, 2001), which they sign in to, and then choose or create course modules, that the 'Assignment' and 'Context' individuals can be automatically generated from session information generated by the content management system. Secondly, we have already shown how 'Difficulty' metadata may be automatically generated from learning objects, thus provided a content management system is used there are means by which the 'Assignment', 'Context' and 'Difficulty' individuals may be automatically generated.

Returning to motivational factors related to the contextual relationship, we think that the discussions and agreements required need only involve an individual manager from the course team for each context. There are two reasons for this

1. the nature of the relationships are simple, and the assignment of them is unlikely to be controversial, hence agreements reached between two individual managers are likely to be supported by their respective teams

2. the choice of which other contexts to establish relationships with is likely to be a management decision, involving consideration of issues such as departmental, organisational and/or government policy.

To illustrate point 2 we return to the example described in the introduction. In this example, two visitors to an art gallery want to receive relevant audio descriptions of paintings that they stop to view. Consider two potentially relevant contexts, i.e. context3 being a postgraduate course in art history, and context4 being the context in the art gallery in which information is provided by the gallery targeted at the general public. It can be imagined that the establishment of a relationship between these contexts could fit with the goals of both organisations involved, and more specifically with goals of managers within relevant departments in (or related to) the specific contexts in question. For example, enabling museum visitors to retrieve relevant and suitable resources from the post graduate course in art history could be of interest to an institution pursuing a policy similar to the 'OpenCourseWare' initiative of the Massachusetts Institute of Technology (MIT, 2003), and thus a motivating factor for a manager within the team responsible for a relevant postgraduate course at such an institution to participate in discussions. Secondly, enabling museum visitors to retrieve relevant and suitable resources from other providers would improve e.g. the visitors' impression of the museum, hence being a motivating factor for a manager within the museum to participate in such discussions.

Finally, we realise that organisational motivation is needed to create and/or utilise an ontology such as the one described. One factor in the choice of the OWL language to implement the relatively simple (compared with what is possible with OWL) constraints and relationships necessary for this ontology is the expectation of the availability of tools that can reason about them (Abecker & Tellmann, 2003), which should enable widespread exploitation of context relationship instances implemented in OWL.

Conclusions

This paper proposes a methodology for the design and implementation of metadata systems which aims to exploit a human community's endemic motivations and shared understanding to ensure the quality of metadata produced by that community. By including task and motivational aspects in the system design process the methodology seeks to improve the accuracy and consistency of metadata descriptors which are generated by people, and hence improve the performance of systems which exploit

this metadata. It includes processes for the creation and iterative testing of structured vocabularies by the communities that will utilise them, including the evaluation of publicly available vocabularies.

Currier et al. (Currier, Barton, O'Beirne, & Ryan, 2004) have stated that collaborative creation of metadata by resource authors and metadata specialists, and design of tools and processes are key issues for further research to ensure metadata of sufficient quality. The methods in this paper support the idea of a collaborative approach and we observe that to ensure the success of the methods requires collaboration between resource authors, metadata specialists, software engineers, and knowledge engineers. For example, the knowledge of the metadata specialists is codified by the knowledge and software engineers so as to support content creators in their metadata creation tasks.

We have shown how accurate and consistent metadata created for one context can be exploited further by the creation of machine interpretable descriptions of the relationships between contexts, i.e. by implementing ways to support communities shared understanding of relevant aspects using the Web Ontology Language, OWL. In this case, the existence of a relevant ontology has not been used to attempt to improve the quality (in terms of accuracy and consistency) of metadata created by people, but to enable further exploitation of human-created metadata, thus increasing its value. We recognise that the organisation as a whole must bear the cost of enabling this, in terms of ontology development and instantiation, but through our methods the impact on individual content authors should be minimised.

As we stated in our recent paper (Brasher & McAndrew, 2004), most of the descriptors of the educational properties of learning objects specified in the IEEE LOM (IEEE, 2002) must be generated from extrinsic sources. We recognise that one possible extrinsic source is recordings of actual use of a resource. It is possible that such data can be analysed and hence enable the automatic generation of metadata descriptors e.g. the 'Typical Learning Time' of users of a learning object. However, even if it is possible to generate relevant descriptors from data about actual use of learning objects, it may be better to create such descriptors at the time the resource itself is created for two reasons:

    (i)   data about previous use will not exist when the resource is first made available, hence precluding its exploitation until enough usage data has been recorded;

    (ii)   users of the Metadata Exploitation System may want to retrieve objects according to the 'expected' (or desired or intended) use i.e. the use that the creators of the resource envisioned, which may be different from the uses that actually occur.

We now consider if the method described for linking multiple communities could be applied to other descriptors in the IEEE LOM education category. Recall that we have used the descriptor 'Difficulty' as an example. The descriptors 'Interactivity Level' and 'Semantic Density' have similar characteristics to 'Difficulty' in that their interpretation also depends on a shared understanding, e.g. the same comment is made with respect to the suggested value space for both 'Interactivity Level' and 'Semantic Density' i.e. 'NOTE—Inherently, this scale is meaningful within the context of a community of practice.' (IEEE, 2002). Thus whilst it would be possible to implement a process which would generate 'Interactivity Level' or 'Semantic Density' metadata from intrinsic sources *for some types of learning objects*, the use of the context relationship approach would be the only way to enable algorithms to exploit this

metadata in multiple contexts. The context relationship approach could also be used with other descriptors e.g. 'Typical Age Range'.

Finally, we remark that the situations which necessitate human-generated metadata for educational resources are similar for resources from all other domains. Thus any semantic web application which has a reliance on metadata created from these sources could make use of the methods described in this paper i.e. to consider and exploit the motivational and community aspects of human metadata generation.

Notes on contributors

Andrew Brasher is a Research Fellow in the Institute of Educational Technology at The Open University. Since joining the OU in 2001 his research work has focused on the development of novel tools and working practices for creation of educational materials. From 2002 he has been working on roadmap studies aimed at defining future research activities related to educational technology, and in particular to learning supported by mobile devices. Prior to joining the OU he was a researcher and consultant in the field of electronic publishing.

Patrick McAndrew is a senior lecturer in the Institute of Educational Technology at The Open University where he teaches and researches in the use of technology in support of learning. His work examines ways to design for active engagement by learners working together. This has involved studies in task based approaches to learning and their representation as learning designs within knowledge sharing environments. In 2001 he co-founded the UserLab research team to undertake projects in e-learning. He has a degree in Mathematics from the University of Oxford and a PhD in Computer Science from Heriot-Watt University, Edinburgh.

References

Abecker, A., & Tellmann, R., Eds. (2003), Analysis of Interaction between Semantic Web Languages, P2P architectures, and Agents, SWWS – Semantic Web Enabled Web Services project IST-2002-37134. Available online at http://swws.semanticweb.org/public_doc/D1.3.pdf (accessed 12/12/2003).

Brasher, A., & McAndrew, P. (2004). Human-Generated Learning Object Metadata. Paper presented at the *On the Move to Meaningful Internet Systems 2004: Workshop on Ontologies, Semantics, and E-learning (WOSE)*, Agia Napa, Cyprus.

Browning, P., & Lownden, M., Eds. (2001), JISC TechWatch Report: Content Management Systems, JISC. Available online at http://www.jisc.ac.uk/uploaded_documents/tsw_01-02.pdf (accessed 26/08/2004).

CEN, Ed. (2003), Controlled Vocabularies for Learning Object Metadata: Typology, impact analysis, guidelines and a web based Vocabularies Registry, CEN/ISSS Learning Technologies Workshop. Available online at ftp://cenftp1.cenorm.be/PUBLIC/CWAs/e-Europe/WS-LT/cwa14871-00-2003-Oct.pdf (accessed 19/05/2004).

Clark, J., Ed. (1999), XSL Transformations (XSLT) Version 1.0, W3C. Available online at http://www.w3.org/TR/xslt (accessed 21/05/2004).

Currier, S., Barton, J., O'Beirne, R., & Ryan, B. (2004) Quality Assurance for Digital Learning Object Repositories: Issues for the Metadata Creation Process, *ALT-J, research in Learning Technology,* 12, 6-20.

Greenberg, J., Pattuelli, M. C., Parsia, B., & Robertson, W. D. (2002) Author-generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization,

Journal of Digital Information (JoDI), 2. Available online at
http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Greenberg/ (accessed 12/05/2004).

Gruber, T. R. (1993) A translation approach to portable ontology specifications, *Knowledge Acquisition,* 5, 199-220.

GUARDIANS (2002) GUARDIANS project. Available online at
http://www.fdgroup.co.uk/guardians/home.html (accessed 12/05/2004).

IEEE (2002) *IEEE Standard for Learning Object Metadata*, IEEE Std 1484.12.1-2002 (pp. i-32), (The Institute of Electrical and Electronics Engineers, Inc.).

IMS (2001) IMS Learning Information Package Specification. Available online at
http://www.imsglobal.org/profiles/ (accessed 12/05/2004).

IMS (2004a) IMS Vocabulary Definition Exchange. Available online at
http://www.imsglobal.org/vdex/ (accessed 13/05/2004).

IMS (2004b) IMS Vocabulary Definition Exchange Best Practice and Implementation Guide. Available online at
http://www.imsglobal.org/vdex/vdexv1p0/imsvdex_bestv1p0.html#1527082
(accessed 14/05/2004).

IMS (2004c) Vocabulary Definition Exchange XML schema. Available online at
http://www.imsglobal.org/vdex/vdexv1p0/imsvdex_v1p0.xsd (accessed 19/05/2005).

ISO (1986) *Documentation - Guidelines for the establishment and development of monolingual thesauri* (2nd ed ed. Vol. ISO 2788). Geneva: International Organization for Standardization.

Kabel, S., Hoog, R. d., & Wielinga, B. J. (in press) Indexing learning objects: vocabularies and empirical investigation of consistency, *Journal of Educational Multimedia and Hypermedia*.

McGuinness, D. L., & Harmelen, F. v., Eds. (2003), OWL Web Ontology Language Overview, W3C. Available online at http://www.w3.org/TR/owl-features/ (accessed 21/05/2004).

MIT (2003) MIT OpenCourseWare | OCW Home. Available online at http://ocw.mit.edu/ (accessed 10/6/2004).

MOBIlearn (2004) MOBIlearn Project. Available online at http://www.mobilearn.org/ (accessed 12/05/2004).

OU. (2004) Understanding E-Learning: A Guide for Teachers and Learners. Available online at http://www3.open.ac.uk/courses/bin/p12.dll?C02T186 (accessed 19/05/2004).

OU. (Not known) C&Q - Help with registration. Available online at
http://www3.open.ac.uk/courses/essential/info.shtm (accessed 20/5/2004).

Sharples, M., Jeffery, N., du Boulay, J. B. H., Teather, D., Teather, B., & du Boulay, G. H. (2002) Socio-Cognitive Engineering: A Methodology for the Design of Human-Centred Technology, *European Journal of Operational Research,* 136, 310-323.

Vainio, T. (2003). *Visit to an art gallery*.Unpublished manuscript.

Weller, M., Pegler, C., & Mason, R. (2003) Working with learning objects - some pedagogical suggestions. Available online at
http://iet.open.ac.uk/pp/m.j.weller/pub/altc.doc (accessed 14/05/2004).