110 *Int. J. Cont. Engineering Education and Lifelong Learning, Vol. 16, Nos. 1/2, 2006*

The Knowledge Life Cycle for e-learning

David E. Millard,* Feng Tao, Karl Doody,
Arouna Woukeu and Hugh C. Davis

School of Electronics and Computer Science,
University of Southampton,
Southampton, SO17 1BJ, UK
E-mail: dem@ecs.soton.ac.uk E-mail: ft@ecs.soton.ac.uk
E-mail: kad202@ecs.soton.ac.uk E-mail: aw1@ecs.soton.ac.uk
E-mail: hcd@ecs.soton.ac.uk
*Corresponding author

Abstract: In this paper, we examine the semantic aspects of e-learning from both pedagogical and technological points of view. We suggest that if semantics are to fulfil their potential in the learning domain then a paradigm shift in perspective is necessary, from information-based content delivery to knowledge-based collaborative learning services. We propose a semantics driven Knowledge Life Cycle that characterises the key phases in managing semantics and knowledge, show how this can be applied to the learning domain and demonstrate the value of semantics via an example of knowledge reuse in learning assessment management.

Keywords: Semantic Web; Knowledge Life Cycle; knowledge reuse; e-learning.

Reference to this paper should be made as follows: Millard, D.E., Tao, F., Doody, K., Woukeu, A. and Davis, H.C. (2006) 'The Knowledge Life Cycle for e-learning', *Int. J. Continuing Engineering Education and Lifelong Learning*, Vol. 16, Nos. 1/2, pp.110–121.

Biographical notes: David Millard received his BSc (Hons) in Computer Science and his PhD in Contextual Information Systems from the School of Electronics and Computer Science at the University of Southampton, UK, where he is now a Lecturer. Prior to his lectureship he spent several years as a research fellow on the Equator IRC working on ubiquitous knowledge systems. His research interests now include web and grid services, narrative systems, physical hypermedia and mobile learning.

Feng (Barry) Tao is currently a research fellow in the School of Electronics and Computer Science, University of Southampton. He received his BSc in Computer Science and PhD in Data Mining and Knowledge Discovery. His research interests include the Semantic Web, Knowledge Management, web/grid services and their applications in e-science (GEODISE) and e-learning (ELEGI).

Karl Doody is studying his MEng in Computer Science in the School of Electronics and Computer Science at the University of Southampton, UK. His dissertation was on the application of knowledge technologies to collaborative web systems.

Arouna Woukeu is a Senior Research Assistant at the University of Southampton where he has also completed his PhD in Hypermedia Engineering and the Semantic Web. He holds a BSc and MSc in Computer Science and his current research interests include hypermedia and Semantic Web engineering, web/grid services, mobility grid computing and mobile learning.

Hugh C. Davis is Head of Learning Technologies group in the School of Electronics and Computer Science, University of Southampton. He has been involved with and published technology for learning since the mid-1980s. Currently his main research area is the architecture of systems for e-learning. He is a member of the X4L TOIA team and has grants such as the European Learning Grid Infrastructure (ELeGI) project, EU IST Framework VI, Digital Libraries support of Innovation Approaches to Teaching and Learning in Geography (Dialog Plus) JISC/NSF.

1 Introduction

As e-learning applications become more integrated and e-learning systems more distributed, there is an increased need to manage their software and data components (Anido et al., 2002). There is a trend in the distributed systems and middleware areas of computing towards Service-Oriented Architectures (SOA), these emphasise loosely coupled components that interoperate by providing distinct services through standardised interfaces. In particular the grid is evolving as an SOA for securely orchestrating and sharing stateful services and resources across distributed or virtual organisations (Foster, 2001).

Both web and grid service architectures have been applied to the e-learning domain (Okamoto and Kayama, 2004; Rodriguez et al., 2003), the argument is that they are advantageous as they are modular and extensible and offer increased interoperability to software producers. While grid services were originally conceived as a method of distributing high performance computation, they also offer benefits in distributed knowledge and information management, offering a guaranteed level of security that is essential for serious e-learning applications.

We believe that the semantic aspects of learning content are the key to facilitating large-scale collaboration of e-learning activities over service-oriented infrastructures. To use explicit and accurate semantics, a consensus in the domain at the conceptual level is necessary, so that computer and human participants can understand and communicate. An ontology is the best vehicle in this context to formally hold a specification (of the conceptualisation) that can be shared within the community to describe semantics accurately and consistently. An ontology explicitly defines the domain concepts and their relationships and is similar to a dictionary or glossary, but with richer structure, relationship and axioms that describe a domain of interest more precisely.

These rich semantics offer both teachers and learners new opportunities for locating and reusing resources (Demetrios et al., 2004; Ronchetti and Saini, 2004; Sridharan et al., 2004). But defining the correct semantics for a learning application is difficult and maintaining ontologies can be problematic (akin to managing the evolution of a complex graph).

We propose a Knowledge Life Cycle for learning, to help define and maintain evolving semantics (Schreiber et al., 1999). Our intention is not to develop a definitive ontology or to promote a particular architecture, but to demonstrate how a semantic-driven Knowledge Life Cycle model can be applied to the learning domain.

In this paper, we present an overview of the semantics involved in learning, present the Knowledge Life Cycle and show the advantages of rich semantics via a demonstration of knowledge reuse.

2 A pedagogical view of semantics

In this section, we examine the affordances of semantics from a pedagogical point of view in an effort to answer the question: what can semantics do for the domain of learning?

2.1 How semantic enrichment can improve learning

Increased semantics offer students a more effective view of their learning and enables new learning opportunities (Demetrios et al., 2004; Ronchetti and Saini, 2004). There are a number of the ways in which reasoning about semantics could make this happen:

- *Connecting communities*: services can put people in contact with other people who are experts or learners with similar interests
- *Personalised content*: intelligent tutoring systems have for some time being delivering content that was personalised for the user, based on an understanding of their goals and previous knowledge
- *Personalised sequencing*: Adaptive Hypertext Systems attempt to provide pathways through materials by matching domain ontologies with dynamically evolving user models
- *Adaptive assessment*: systems may choose questions for the learner at the boundary of their understanding; thus, improving the efficiency of assessment and providing feedback that provides detail in critical areas
- *Feedback agents*: intelligent agents that observe student behaviour (e.g. assessment results, interactions with a virtual experiment, etc.) can attempt to provide feedback and links to suitable material to assist the learner
- *Recommender agents*: the system could recommend alternative resources based on user searching and studying patterns. In a formal setting, it could query the syllabus and timetable to recommend a plan of study
- *Annotation tools*: users could annotate information themselves, providing useful information for others and allowing both readers and other services the opportunity to process the information in alternative ways
- *Search engines*: when resources have been semantically enriched, then search engines can be much more powerful. Where services are semantically enriched, search engines can choose suitable services to manage the query
- *Analytic tools*: the e-science community is leading the way in the production of tools that harvest, store and analyse data from a range of sources.

2.2 How semantic enrichment can improve the management of learning

E-learning practitioners often comment that they believe they spend as much time organising materials as they spend on teaching and the production of materials. We believe that semantics may ease this problem in a number of ways:

- *Production of materials*: production of teaching materials is a notoriously time-consuming task and the ability to locate and to reuse existing materials is a primary motivation for providing metadata for learning resources. The next stage is to provide services to assist in the location of suitable materials from heterogeneous sources.
- *Student management*: an understanding of the roles of the actors (teachers, students, experts, assessors, etc.) makes the production of services for assigning students to the correct classes, discussion groups, experimental teams, etc., possible.
- *Timetable management*: an important task for teachers of online tasks is the timing of events, such as the release of some new materials, the closing date of some assessment, the exact time of a synchronous group chat session, etc. These events can be made to happen automatically when a course is described in some language such as IMS Learning Design.
- *Record keeping*: record keeping and quality assurance can be the bane of a teacher's life, requiring them to spend much time ensuring that all the results are kept in the correct places such as institutional enterprise systems, student portfolios as well as made available for QA purposes by whatever external authorities might be involved. All of this work is an obvious target for automation by services that understand the goals.
- *Quality assurance*: quality assurance often involves the maintenance of sample work and feedback/reflections, as well as ensuring that new programmes, courses and assessments have been through appropriate validation. Again, this is a task that could be assisted by intelligent services, which could guide such tasks through the set of other services involved.

Semantics can also help orchestrate services to achieve more complex goals. For example, an assessment system might call a service to handle some marks. This service might then ask an enterprise system service to store the marks in a database; it might call a service to annotate the student records with the new information and then might call an e-mail service to inform the students of the need to update their personal development plans accordingly. We believe that appropriate semantic enrichment of the elements in the learning domain should make possible the automatic creation of workflows by the composition of appropriate services.

3 Paradigm shift

To fully realise the potential of semantics in the pedagogical domain (as described above), it is necessary to make a paradigm shift in the way we deal with semantics (Sridharan et al., 2004), this shift happens in two ways: from metadata to ontologies and from information to knowledge.

Ontologies are a more sophisticated way of modelling *metadata* and *knowledge* is relevant *information* delivered at the right time and context.

3.1 From metadata to ontology

Metadata has been widely used to structurally describe learning resources so that they can be better reused. Example standardisations are Dublin Core,[1] which is a general purpose metadata standard and the IMS Metadata and IEEE LOM[2] (Learning Object Metadata) standards.

While metadata is a starting point to describe content, recent developments in the Semantic Web area have focused on the use of ontologies for richer semantics. An ontology is 'a specification of a conceptualisation' (Gruber, 1993). Ontologies can be seen as an improvement over metadata as they formally define not only keywords (as concepts) but also relationships among them. A simple example of teacher/student modelling shows how an ontology is constructed. The relationships might be:

Student → *assignedWith* → Course
Teacher → *delivers* → Course

As well as these main relationships, each concept would also have its own properties like 'name', 'course ID', etc. Ontologies enable us to make the second shift, from information to knowledge.

3.2 From information to knowledge

Using ontologies enables machines to move from dealing with information to dealing with knowledge (well structured, relevant resources, both content and services, available at the right time and context). The core principle is that knowledge is sharable, reusable and part of a larger context.

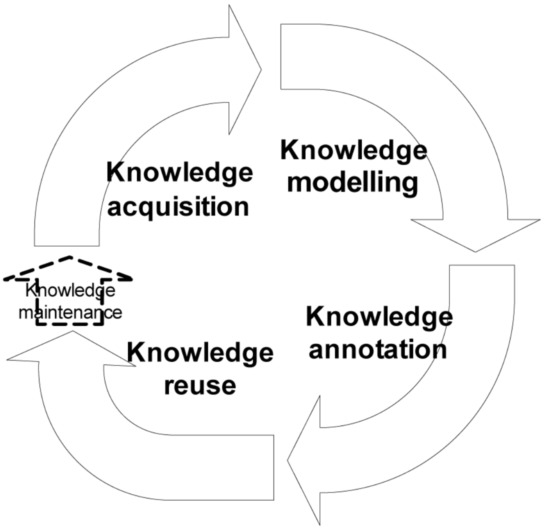
When a system has a shared ontology it knows how to handle the semantically enriched resources consistently. For example, when a student wants to search for a course, the course query service knows from the shared ontology what the search criteria are and these will match with course delivery services even if the two services were developed separately and are deployed independently (e.g. they could be grid services developed by different software developers and running on different operation systems). The services can understand each other by following the shared ontology.

4 The Knowledge Life Cycle

We have claimed that knowledge is information about structured, relevant resources that is sharable and reusable. To ensure this, resources must be annotated with rich semantics that are agreed by the members of the domain community. The development and maintenance of ontologies that capture this rich meaning is the subject of knowledge engineering. In this section, we present the different stages of the *Knowledge Life Cycle*, a model that describes how knowledge is captured, applied and reused (shown in Figure 1):

- *Knowledge acquisition (KA)*: the first stage is to acquire the knowledge from the domain experts. The objective is to develop a domain vocabulary and a sense of the most important concepts. This can be done in a variety or ways including scenario construction and interviews
- *Knowledge modelling (KM)*: the next stage is for this description to be formalised as an ontology. Classes are defined based on the concepts identified in the KA stage and the relationships between these classes are specified
- *Knowledge annotation*: once an ontology has been defined, it is tested through an application. To do this, example resources from the domain are annotated with the ontological metadata. This enables the KM stage to be evaluated and revised
- *Knowledge reuse*: reuse is achieved when new applications reuse the resources (made possible by the shared ontology), for example, by incorporating existing learning objects into a new course design.

Figure 1 The Knowledge Life Cycle



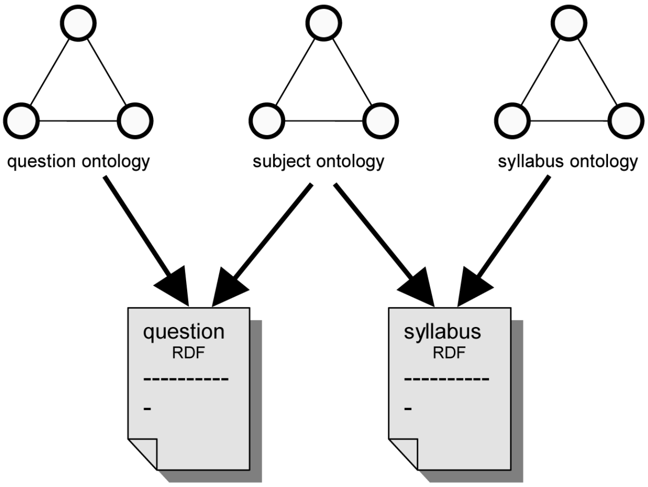
The Knowledge Life Cycle is intended to iterate over several generations. This means that the ontologies are expected to evolve and maintenance is necessary. Annotations from earlier generations will need to be updated for their reuse to continue. Doing this automatically is still the subject of much research (Maedche et al., 2003) but the formal nature of the life cycle (and its audit trail) ensures that this is at least manually possible.

5 Technical view of learning semantics

In this section, we will describe how we have used the Knowledge Life Cycle to develop reusable semantics for learning. It is worth repeating that this is not an attempt to create a definitive ontology, but a demonstration of the life cycle within the domain of learning. Throughout this work we have used key mark-up languages, such as XML, RDF and OWL,[3] which are often chosen to represent semantics via ontologies. This formalised expression makes the ontologies machine accessible and interpretable.

Figure 2 shows the e-learning information we chose to model and the three ontologies that we developed to do this.

Figure 2 Example ontologies and RDF documents



The information is contained in two RDF documents. The first contains descriptions of questions, similar to those that would normally be found in a Question Bank. The second contains descriptions of syllabi, as would normally be found in a course prospectus.

We modelled these two sets of information using three ontologies. The first is a question ontology that describes all the properties of a question's structure and purpose. The second is a subject ontology that describes the key areas within a particular topic that is to be taught and how they are related. The third describes the structure and properties of a syllabus.

To create these ontologies, we needed to go through the KA process and build up concepts and relationships around the concepts of questions and syllabi.

5.1 Building a learning vocabulary

The first part of the Knowledge Life Cycle is the process of KA, which has the objective of forming a formal, explicit and shared consensus in the domain. In the learning domain this translates to a process of interviewing learning domain experts and examining teaching and learning materials to create a common vocabulary and identify key concepts. For our own KA the authors examined various sources to create the three different ontologies.

The question ontology was based on the metadata standards employed by existing question bank systems such as E3AN (Davis et al., 2002), and standards such as IMS QTI, 2.0.[4] The domain ontology was based on existing work on Topic Maps (Dicheva and Dichev, 2005), the Simple Knowledge Organisation System (SKOS)[5] ontology developed by the W3C and the IEE taxonomy for digital electronics.[6] The syllabus ontology was based on the syllabus requirements and records of the University of Southampton. Our ontology work has also grown out of a series of interviews with domain experts, which resulted in a list of key learning concepts, attributes and relationships and helped us choose the types of resources we wanted to model in the first place.

5.2 Building learning ontologies

The next part of the Knowledge Life Cycle is modelling, where key concepts and terms identified at Acquisition are formalised into an ontology.

Protégé 2000[7] is an ontology building and KA tool that has been frequently used for KM purposes. It allows knowledge engineers to focus on modelling without worrying about the underlying language and syntax. The modelling work can be saved in various formats including RDF and OWL.

We built our initial ontology in Protégé with an OWL plug-in. ‘Question’, ‘Topic’, ‘Course’, etc. are key concepts under which the taxonomy is further expanded to express hierarchical relationships (parents/children) among concepts. Figure 3 shows a screenshot of Protégé being used to examine the Question class. The ontological information is then saved in OWL format so that at the next stage it is possible to validate the annotations that are made. OWL models the permitted concepts and relationships, but also defines extra semantics concerning constraints, for example, by limiting the cardinality of relationships.

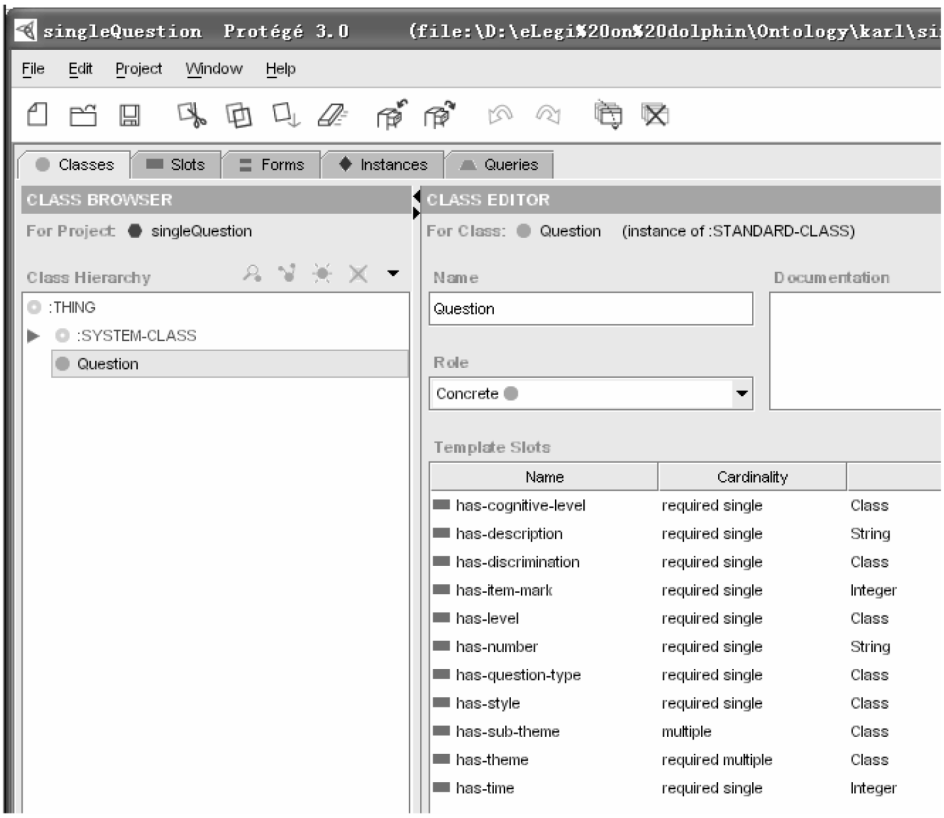
5.3 Annotating resources

The next step is knowledge annotation, the process of binding together relevant learning resources with instances from the ontology so that raw content is enriched with more formal meanings predefined in the shared ontology. This is also termed *knowledge binding* (Tao et al., 2003) and often depends upon human effort to tag the resources. For this work we used two methods to generate instances called ‘resource annotation’ and ‘ontology instantiation’.

Resource annotation describes the process where resources are directly annotated by domain experts. In e-learning this would mainly be done by teachers and learners. The task of ontology instantiation is a specialised skill that requires knowledge engineers to translate resource information into the ontology. This is often too complicated or time consuming for resource providers, so an annotation tool would be preferable to make the process easier.

Ontology instantiation describes the use of such a tool to *directly* build instances out of an ontology. Protégé can also be used in this way to create and store ontological instances. It may then be treated as a knowledge base or the instances can be saved as independent files or exported from Protégé as RDF.

Figure 3 Constructing the question ontology in protégé



For Project: singleQuestion

For Class: Question (instance of :STANDARD-CLASS)

Name: Question

Role: Concrete

Name	Cardinality	
has-cognitive-level	required single	Class
has-description	required single	String
has-discrimination	required single	Class
has-item-mark	required single	Integer
has-level	required single	Class
has-number	required single	String
has-question-type	required single	Class
has-style	required single	Class
has-sub-theme	multiple	Class
has-theme	required multiple	Class
has-time	required single	Integer

5.4 *Reusing resources*

Once the resources are enriched with semantics, we enter the knowledge reuse stage. There are several common ways in which the semantics can be reused:

1. *Resource discovery*: recommend learners learning materials according to the semantic matching between semantics of the resources
2. *Process automation*: as demonstrated by web services and the grid, as services have their interface, parameters and effects semantically described, automation becomes possible (an example might be an assessment service that automatically grades sets of questions)
3. *Service integration*: this is about exploiting semantics to assist the service-oriented architecture where simpler services can be combined together to realise more complex customised functionalities (an example might be a live course system automatically assembled and run based on a learning design).

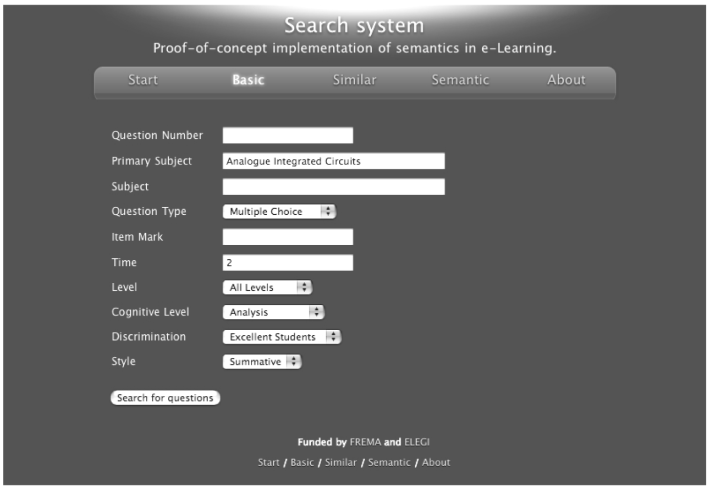
For our demonstrator we have concentrated on the first of these and designed a set of web services concerned with resource discovery (in this case questions for learning assessment) that leverage semantics in an increasingly complex way.

Figure 4 shows the search interface. There are three search services, a basic search (shown), a similarity search and a semantic search.

5.4.1 A basic search service

One of the advantages of using the knowledge lifecycle is that the resulting ontological metadata has a better chance of being well designed. The basic search exposes this high-quality metadata directly; it allows a searching application to look for questions based on a simple pattern match based on the question attributes in the question ontology.

Figure 4 The search interface (basic search selected)



Question Number	
Primary Subject	Analogue Integrated Circuits
Subject	
Question Type	Multiple Choice
Item Mark	
Time	2
Level	All Levels
Cognitive Level	Analysis
Discrimination	Excellent Students
Style	Summative

5.4.2 A question similarity service

The second service increases the complexity of the search interface as it takes a question as the search parameter. The service then returns other questions that are similar to the one submitted. Although the service is more complex, it does not reuse any semantics. Instead it uses the ontological metadata in a straightforward way and produces a heuristic based on weighted attributes.

5.4.3 A semantic search service

The final service demonstrates knowledge reuse. This service relies on the annotated syllabus data, which shares some of its ontological markup with the questions (both questions and syllabi are annotated by the subject ontology). Because of this it can reuse the question data in a search based on a syllabus. The user specifies the syllabus topics that they are interested in and the system returns a set of questions that address that topic. Reasoning rules encoded in Jena[8] ensure that topics that are *narrower* than the target topic still match (i.e. if we are seeking a question on 'mobile communication' then those described as 'cellular communication' still match).

This not only demonstrates the reuse of the original knowledge, but also the added value that semantics can bring to e-learning applications. More complex versions of this same service could add extra relationships between topics and additional reasoning rules to ensure that the search service understands the relationships between topics and returns questions that are fit for a target topic, even if this is not immediately obvious from the original annotations.

6 Conclusions

In this paper, we have looked at the semantic aspects of learning from the pedagogical and technological views. Sophisticated semantics can enrich resources and enable a paradigm shift from information-based content delivery to knowledge-based, context-aware collaborative learning services. Ontologies can be used as an improvement over existing metadata to add the semantics needed for enriched services and resources.

We have also proposed the use of the Knowledge Life Cycle to manage the key phases in modelling learning semantics. We have described our efforts to follow a life cycle model within the learning domain – namely by performing an acquisition exercise, building a learning ontology and creating semantic instances in Protégé and RDF – and have constructed three search services that show how semantics can be reused and information reasoned about to provide complex functionality.

The paradigm shift from information to knowledge offers serious advantages to the next generation of distributed learning systems. This manifests itself not only in immediate advantages for well-formed metadata and interoperability, but also in the way in which semantically annotated resources can be reasoned about. We believe that a Knowledge Life Cycle model is critical to successfully managing learning and teaching semantics and achieving the goals of resource sharing, collaboration and automation.

Acknowledgements

This work was funded by The European Commission under the E-Learning Grid Infrastructure project (ELeGI), IST-0002205, Sixth Framework Programme and also the e-Framework Reference Model for Assessment (FREMA) project funded under the UK JISC e-Learning Programme.

References

Anido, L., Rodriguez, J., Caeiro, M. and Santos, J. (2002) 'Towards standards-driven educational content providers', *Second IEEE International Conference on Advanced Learning Technologies (ICALT'02)*, Kazan, Russia, pp.139–143.

Davis, H.C., White, S.A. and Dickens, K.P. (2002) 'Engineering a testbank of engineering questions', *International Conference on Engineering Education (ICEE 2002)*, Manchester, UK.

Demetrios, G.S., Lytras, M.D., Wagner, G. and Diaz, P. (Eds) (2004) 'Ontologies and the Semantic Web for e-learning', *Special Issue of the IEEE IFETS Journal of Educational Technology and Society*, Vol. 7, No. 4, Available at: http://www.ifets.info/others/issues.php?id=25.

Dicheva, D. and Dichev, C. (2005) 'Authoring educational topic maps: can we make it easier?', *Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*, Kaoshung, Taiwan, pp.216–218.

Foster, I.T. (2001) 'The anatomy of the grid: enabling scalable virtual organisations', *Seventh International Euro-Par Conference on Parallel Processing*, Manchester, 28–31 August, UK, pp.1–4.

Gruber, T.R. (1993) 'A translation approach to portable ontologies', *Knowledge Acquisition*, Vol. 5, No. 2, pp.199–220.

Maedche, A., Motik, B., Stojanovic, L., Studer, R. and Volz, R. (2003) 'An infrastructure for searching, reusing and evolving distributed ontologies', *12th International Conference on World Wide Web (WWW'03)*, Budapest, Hungary, pp.439–448.

Okamoto, T. and Kayama, M. (2004) 'The learning GRID for knowledge building', *Fourth IEEE International Conference on Advanced Learning Technologies (ICALT'04)*, Joensuu, Finland, 30 August–01 September, pp.149–152.

Rodriguez, J., Anido, L. and Fernandez, M.J. (2003) 'How can the web services paradigm improve e-learning?', *Third IEEE International Conference on Advanced Learning Technologies (ICALT'03)*, Athens, Greece, 9–11 July.

Ronchetti, M. and Saini, P. (2004) 'Knowledge management in an e-learning system', *Fourth IEEE International Conference on Advanced Learning Technologies (ICALT'04)*, Joensuu, Finland, 30 August–01 September, pp.365–369.

Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., Van de Velde, W. and Wielinga, B. (1999) *Knowledge Engineering and Management: The CommonKADS Methodology*, MIT Press.

Sridharan, B., Tretiakov, A. and Kinshuk, A. (2004) 'Application of ontology to Knowledge Management in web based learning', *Fourth IEEE International Conference on Advanced Learning Technologies (ICALT'04)*, Joensuu, Finland, 30 August–01 September, pp.663–665.

Tao, F., Cox, S., Chen, L., Shadbolt, N., Xu, F., Puleston, C., Goble, C. and Song, W. (2003) 'Towards the semantic grid: enriching content for management and reuse', *Delivering e-Science, UK e-Science All-hand Conference*, Nottingham, UK, pp.695–702.

Notes

[1] The Dublin Core Metadata Initiative, http://dublincore.org/.

[2] IEEE Learning Object Metadata, http://ltsc.ieee.org/wg12/.

[3] XML, RDF and OWL are W3C standards, which can be found in http://www.w3c.org.

[4] IMS QTI Specification available from: http://www.imsglobal.org/question/.

[5] SKOS Development Page: http://www.w3.org/2004/02/skos/.

[6] IEE Authors Guidelines: http://www.iee.org/Publish/Support/Auth/elhier.pdf.

[7] The Protégé homepage, http://protege.stanford.edu/index.html.

[8] Jena development site: http://jena.sourceforge.net/.