Think ahead: evaluation and standardisation issues for e-learning applications

Vladan Devedžić

Department of Information Systems and Technologies,
FON - School of Business Administration, University of Belgrade,
POB 52, Jove Ilica 154, 11000 Belgrade, Serbia and Montenegro
Fax: +381-11-461221 E-mail: devedzic@galeb.etf.bg.ac.yu

Abstract: Developing an e-learning application requires much effort, even when sophisticated authoring tools are used. In this process, knowledgeable authors of learning material and knowledge engineers skilled in using e-learning technology are not the only important players; software engineers and web application developers often play major roles as well. The requirement of producing high-quality, effective learning systems is deeply related to their evaluation and to designing them so as to conform to the emerging standards for learning technology systems. This paper focuses on these two issues and surveys current major trends in the evaluation of e-learning systems and in bringing them as close as possible to standards. Both issues have recently attracted much more attention than before, especially in the context of web-based education.

Keywords: e-learning systems; evaluation; standardisation; metadata; web-based education.

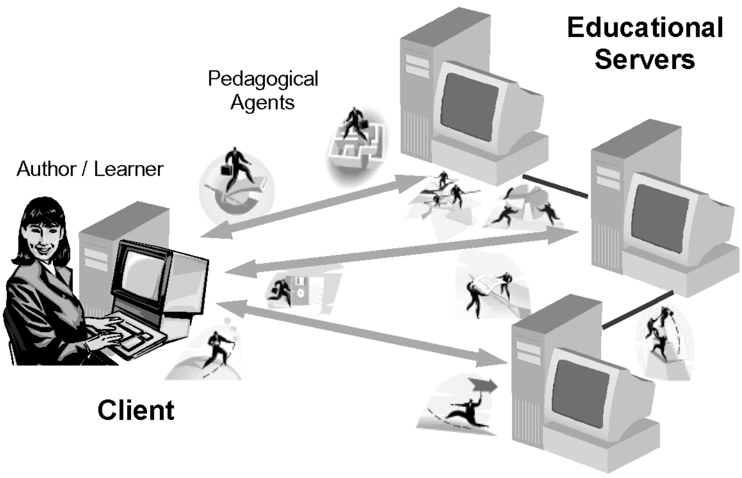
Reference to this paper should be made as follows: Devedžić, V. (2003) 'Think ahead: evaluation and standardisation issues for e-learning applications', *Int. J. Continuing Engineering Education and Lifelong Learning*, Vol. 13, No. 5/6, pp.556–566.

Biographical notes: Vladan Devedžić is an Associate Professor of Computer Science at the University of Belgrade, Serbia and Montenegro. His current research interests and efforts are oriented towards practical engineering aspects of developing intelligent educational systems on the web. He has written more than 190 papers (26 of them published in internationally recognised journals), three books on intelligent systems, and several chapters in books on intelligent systems and software engineering, edited by distinguished scientists. He is participating in several ongoing national and international R&D projects.

1 Introduction

In web-based education, educational material is generally distributed over a number of educational servers (see Figure 1) [1]. The authors (teachers) create, store, modify, and update the material working with an authoring tool on the client side. Likewise, learners use different learning tools to access, browse, read, and consult the material in completing their learning tasks. Pedagogical agents provide the necessary infrastructure for knowledge and information flow between the clients and the servers by interacting with students/learners and authors/teachers and by collaborating with other similar agents, in the context of interactive learning environments [2]. They help in locating, browsing, selecting, arranging, integrating, and otherwise using educational material from different educational servers. Pedagogical agents access educational content (both domain and pedagogical) on a server by using educational web services such as learning, assessment, reference, and collaboration services. Personalisation of the process is achieved by using various techniques of learner modelling, adaptive navigation support, and adaptive presentation planning (curriculum sequencing).

Figure 1 The context of web-based education



Authors develop educational content on the server in accordance with important pedagogical issues such as instructional design and human learning theories, to ensure educational justification of learning, assessment, and possible collaboration among the students. The way to make the content machine-understandable, machine-processable, and hence agent-ready, is to mark it up with pointers to a number of shareable educational ontologies. Every such ontology should provide a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for some particular topic or service [3]. Ontologies make up the knowledge infrastructure of the emerging Semantic Web [4].

Unfortunately, this widely present context of e-learning systems development and deployment requires much effort. The chances of a system being successful increase proportionally with its quality. Although the quality of an e-learning system is a vague concept, there are some approaches, methods, and techniques for measuring different system aspects, which enable the evaluation of such systems. The development efforts can be notably reduced with careful planning for system evaluation from the very beginning of the project, as well as by designing the system to support some of the emerging e-learning technology standards. This paper surveys these two important aspects of e-learning.

2 Quantitative aspects of e-learning systems evaluation

Evaluation strategies and (numeric) performance measures help to quantify and rank a given e-learning environment and assess how different environments compare to each other. Note that evaluation is concerned with a system's performance and its decision-making capabilities, as opposed to assessment, which is about determining the effectiveness of the system by measuring learner outcomes [5]. This section briefly overviews different aspects of e-learning systems evaluation, such as common problems, goals, methodology, strategies, approaches, and techniques, and lists some specific metrics typically used in evaluation, both web-based and traditional. It also covers specific issues related only to evaluation of intelligent e-learning systems, showing a neutral view of problems and failures of this technology.

2.1 Common problems of e-learning systems evaluation

Evaluating e-learning systems is difficult because the underlying theories are either new or still under development, and there is no widespread agreement as to how the fundamental tasks (student modelling, adaptive pedagogical decision making, generation of instructional dialogues etc.) should be performed [6]. Although the e-learning community puts much effort into research related to evaluation, systematic e-learning-specific evaluation frameworks are still lacking.

Moreover, it is extremely difficult to compare two or more e-learning systems because realistic evaluation would require comparing systems in the same domain and aimed at the same target group of learners; however, there are not many truly operational e-learning systems in the same domain. Hence evaluation is typically done for a single system and is combined with the assessment of learning effectiveness and quality of the learning material.

2.2 Evaluation goals

The goal of an evaluation can be to measure the performance of an e-learning system as a whole, or to evaluate a certain learning strategy/theory or model underlying the system. For example, Mitrović et al. have analysed the performance of a constraint-based student modelling approach used in their SQL-Tutor [6], as did Weber and Brusilovsky for their ELM-ART tutor in the domain of LISP programming [7]. The goal can also be to evaluate individual components of the system (e.g., the domain knowledge base and the user interface), or the level of its adaptivity with respect to a selected learning goal(s) and/or parameter(s), as well as to measure the system's navigation support for the learner or some of its built-in features and options. Other possible goals of e-learning system evaluation include analysis of the system's feedback to the student (such as hints, error messages, and problem-solving help), obtaining an analytical model of the quality of online courses, indicators of how well the system matches the difficulty level of problems and exercises to the learner's knowledge level, statistics for using different parts of the domain knowledge from the expert module in the learning process, the time learners take in order to complete solutions to some problems, and so on.

2.3 Evaluation methodology, strategies, approaches and techniques

Oliver et al. suggested a 6-step methodology for e-learning system evaluation [8]:

- identification of the audience for the evaluation
- selection of an evaluation question
- choice of an evaluation strategy
- choice of data collection methods
- choice of data analysis methods
- selection of the most appropriate format(s) for reporting the findings.

Generally, the quality of an intelligent e-learning system can be evaluated against conventional teaching, as well as against a 'dumbed-down' version of the system (obtained by removing its intelligent components), using pre- and post-tests with real or simulated students. Simulated students are computer models (simulations) of human learners. The advantage of using simulated students is a higher level of control of the experimental setting, but it should be just a first step towards a full evaluation, since simulated students are too simplistic in comparison to real students in realistic environments [6]. Using real students generates more realistic behaviour and data and lowers the evaluation costs, but is by far less controllable.

A frequently used approach to evaluation of e-learning systems is conducting some form of sensitivity analysis: values of some features of the system and/or system parameters are varied in order to determine which ones make the most difference in the system performance and learning effectiveness [9]. For example, the accuracy of mastered knowledge in the learner model can be varied, or the amount of feedback the system offers to the learner can be different in each run of the experiment. Since sensitivity analysis requires repeating the experiments with a high degree of control, it is best conducted with simulated learners.

Another popular approach to e-learning systems evaluation is to conduct a specifically designed learning experiment with the system and human learners, such as having the learners solve a specific problem put by the system, and asking a human expert to observe the experiment and evaluate the results and the interaction between the learners and the system [10]. A suitable technique here is to track and post-evaluate what the students have done in solving the problem by analysing their log files and the transcript of their chat sessions in collaborative work. Then the human expert's solution/advice is compared to that provided by the system.

Quantification of parameter values, system features, the quality of the advice generated by the system and other measurements are typically done using predefined discrete values on a certain numeric scale, or a set of descriptive values (e.g., the quality of the advice can be ranked as 'worth saying', 'so-so', 'not worth saying'). Descriptive values are counted and a tabular, spreadsheet, or graphical representation such as a pie chart is used to show the percentage of each.

Other techniques used for evaluation of e-learning systems include online questionnaires, postings, assignments, e-mail discussions, tracking (in the log files) the paths through the nodes in the knowledge base that the learners use to solve problems, recording and categorising the errors they make, and varying in a controlled way the level of difficulty of the exercises the system generates for the learners.

2.4 Specific metrics used in evaluation of e-learning systems

It is important to stress that no single metric is ideal; many metrics are subjective, many are best suited to specific environments, and many are technology-dependent. Very often, it is actually a metrics suite that is applied to measure a certain e-learning system in a more comprehensive way. Some typical metrics include [6,7,9]:

- total interaction time the learner spent with the system [in minutes]
- the number of attempted problems
- the number of solved problems
- total number of attempts to solve the problems
- the number of problems solved at the first attempt
- the number of attempts to solve problems that could not be solved at the first attempt [attempts/problem]
- problem solving speed [problem/min]
- posterior probability of mastering a certain part of the domain knowledge encoded in the expert module (useful for updating the learner model)
- accuracies of mastered and unmastered parts of domain knowledge encoded in the expert module
- simulated student competence (represented as a number on a certain scale)
- simulated student guessing factor (represented as a number on a certain scale)
- assessor guessing (represented as a number on a certain scale)
- observability of problem solving, teaching strategies, and student models
- negative evidence
- the number of learner's preferences that can be configured
- the number of rules tapped by the assessment divided by the estimated total number of rules in the task domain (a quantitative measure of content validity).

2.5 Evaluation of adaptive e-learning systems

In addition to the general e-learning systems evaluation approaches and metrics listed above, there are some specific means of measuring effectiveness in adaptive e-learning systems. These may be specifically related to the adaptive decisions that the system makes, the efficiency and performance of the system and/or the algorithms that the system employs [11]. However, there is no standard or agreed evaluation framework for measuring the value and the effectiveness of adaptation yielded by adaptive e-learning systems [12].

Weber and Brusilovsky evaluated the ELM-ART system using feature-by-feature comparison of ELM-ART with other adaptive e-learning systems for teaching programming-related subjects [7]. The features used for comparison were:

- adaptivity in organising the course material, recommending examples, auto-evaluated questions, and knowledge-based navigation support for examples and solutions analysis (the systems compared were categorised as either adaptive or non-adaptive with respect to this metric)
- automated intelligent diagnosis of submitted and evaluated programs
- personalised curriculum sequencing
- the possibility of adaptive testing
- adaptive link annotation (such as colour fonts and special symbols) and link removal mechanisms (e.g. removing links to the pages note ready to be learned)
- communication and collaboration support (such as e-mail, chat, forums).

Sensitivity analysis is used for evaluating adaptive e-learning systems in terms of keeping or removing a certain adaptive feature (such as adaptive navigation support or student modelling) and comparing the two versions of the system with respect to learning effectiveness [13]. On the other hand, Smith and Blandford note that adaptivity should preferably be an inherent part of a system, and so if it is removed from the system, the system may not be fully functional [12]. Moreover, in their study of adaptivity of their MLC tutor they found that using only superficial measures of performance may be misleading: the enhanced results alone may simply be the consequence of an alternative interface and not a direct consequence of any adaptation. The alternative way of adaptation they recommend is to use machine-learning techniques to dynamically generate rules, which hold generalised information about the learner's current area of interest and are used to construct a suggestion list of links to follow at any point in time during the learning session. This suggestion list is adaptive, reflecting the on-going browsing activity.

A two-layer approach to evaluation of adaptive and personalised educational applications and services is proposed by Karagianidis et al. [14]. In the first layer, the goal is to assess if the conclusion drawn by the system concerning characteristics of user-computer interaction is valid and if the learner's characteristics (e.g., his/her current learning goals, current knowledge, preferences, and experience) are successfully detected by the system and stored in the learner model. The actual evaluation can be done by performing learning tests and comparing the system's conclusions to those of a human expert. In the second layer, the validity and meaningfulness of the system's adaptation decision are evaluated for selected system assumptions (e.g., evaluation of various adaptive features such as adaptive presentation and/or navigation technique for a given learning goal, or for specific learner's interests and knowledge). This evaluation can also be done by giving learners a particular goal and assessing how specific kinds of adaptivity help with that goal. The benefits of this layered evaluation approach are most obvious when adaptation is unsuccessful, since problem findings can be done in both layers and hence better localised.

2.6 Evaluation of collaborative e-learning systems

Issues specific to collaborative e-learning systems are also evaluated by having a human expert judge how 'reasonable' is the advice given by the system, in this case to the group of learners. This can be done by analysing events in which several candidate advice with different rankings exist, and comparing them to human expert rankings using a suitable Euclidean measure of distance between the system's and the expert's rankings [10].

Specific metrics in collaborative learning scenarios include voting tracking, waiting for feedback timeout, participation balance for the group members (e.g., the number of contributions to the problem solution made by each member and the quality of contributions themselves), chat tracking, the time the group took to develop the case study solution, the difficulties in the prerequisite knowledge the group members should have acquired beforehand, the number of misunderstandings concerning the case study, matching of the group members' learner models, and discussion encouragement intensity. Note that researchers in the e-learning community stress that the ability to use and match the models of multiple learners connected to a collaborative e-learning system is the key to adaptive collaboration support [7].

2.7 Towards ontology-enhanced evaluation of e-learning systems

Standard educational ontologies must cover a number of areas and aspects of teaching and learning, such as curriculum sequencing, student modelling, pedagogical issues, grading, and many more [11]. However, the work in that direction is still in the early stages and there are not many ontologies available. This is partially due to the lack of standard vocabulary in the domain of education and instructional design. There are several working groups and efforts towards such a standard vocabulary (IEEE Learning Technology Standards Committee (http://grouper.ieee.org/groups/ltsc/index.html), Technical Standards for Computer-Based Learning, IEEE Computer Society P1484 (http://www.manta.ieee.org/p1484/), IMS Global Learning Consortium, Inc. (http://www.imsproject.org/), and ISO/IEC JTC1/SC36 Standard (http://jtc1sc36.org/), but no official standard yet.

Since ontologies and the semantic web in general are a young field of research, and educational ontologies themselves are just starting to emerge, ontology-enhanced evaluation of e-learning systems is still to come. However, there are already clear indications of some directions in which development of such evaluation methodologies, approaches, and techniques can be expected:

- ontology-based evaluation of an e-learning system's knowledge validity (evaluating the contents of the system's knowledge as to how well it is supported by various domain and educational ontologies)
- evaluation of educational ontologies themselves (this will require sound methods and techniques for evaluating the quality of ontologies in general, specific learning- and teaching-oriented methods for evaluating educational ontologies, as well as a wider acceptance of standards in the area of education)
- evaluation of ontology-enhanced educational services
- evaluation of automated ontology learning for educational ontologies.

3 Standardisation issues

Although full-scale standardisation of e-learning technologies is still to come, several working groups and standardisation committees have already taken steps in this direction and created widely recognised standards documents. In spite of the fact that many of these documents are not raised to the level of official standards yet, awareness of standardisation efforts is an important parameter of e-learning systems design.

A part of the IEEE Learning Technology Standards Committee (LTSC) is the P1484.12 Learning Object Metadata (LOM) working group that has recently developed a Draft Standard for Learning Object Metadata [15]. The draft LOM standard addresses the problem of web-based educational metadata (information about educational objects on the web) by defining a structure for interoperable descriptions of learning objects. In LOM, a learning object is defined as any entity (digital or non-digital) that may be used for learning, education or training. A metadata instance for a learning object describes relevant characteristics of the learning object to which it applies, such as the object's name (title), structure, language, version, size, etc. The characteristics are grouped into several categories (e.g., general, life cycle, educational, and technical), and the standard contains a large table describing all the characteristics in each category in detail. In web-based educational applications, such metadata help discover, manage, use, share and exchange learning objects on the web.

Note that the draft LOM standard builds upon the metadata work by the Dublin Core Metadata Initiative (DCMI), which is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models (http://dublincore.org/). DCMI Education Working Group develops proposals for the use of Dublin Core metadata in the description of educational resources. The draft LOM standard defines the mapping from its characteristics of learning objects to what is called unqualified Dublin Core Metadata Element Set.

LOM specifies a conceptual data model: it merely defines metadata of learning objects, but does not provide any specific encoding of metadata. On the other hand, W3C has created RDF (Resource Description Framework) language for representing metadata about web resources (http://www.w3.org/TR/rdf-primer/). The CID group (Center for user-oriented IT Design) of the Royal Institute of Technology, Stockholm, Sweden, has developed RDF binding of LOM metadata [16]. Essentially, the binding defines a set of RDF constructs that facilitates introduction of educational metadata into the semantic web. It is specified as a table defining the RDF property to use for each element in the draft LOM standard. This is very important for future web-based intelligent educational applications, since such a binding enables the RDF-based exchange of LOM instances between applications that implement the LOM data model. In collaboration with the University of Hannover, Germany, and Stanford University, USA, and other participants, the CID group has also developed Edutella, a metadata-based peer-to-peer system for handling educational resources [17].

The activities of the W3C Web-Ontology (WebOnt) Working Group are focused on the development of Web Ontology Language (OWL) [18] and its relationship with DAML+OIL, an earlier ontology development language [19]. Although not designed specifically for web-based educational applications, both languages provide powerful tools for developing ontologies and have already got significant momentum, hence they should be certainly considered as important enabling technologies in web-based education for the time being. The two languages are similar and both have layered structure, from versions supporting core features (e.g., OWL Lite) to full versions (OWL Full).

The ARIADNE Foundation (http://www.ariadne-eu.org/) develops tools and methodologies for producing, managing and reusing computer-based pedagogical elements. Its Educational Metadata Recommendation [20] is an application profile/implementation of the LOM specification that takes into account the specific needs and requirements of a community that is highly representative of European Higher Education and Continuing Professional Training. Similarly, EdNA (Education Network Australia) is an Australian national framework for collaboration on the use of the internet in education and training; it is organised around Australian curriculum and offers the DCMI-based EdNA Metadata Standard to support interoperability across all sectors of education and training in Australia in the area of online resource discovery and management (http://www.edna.edu.au/metadata/).

IMS Global Learning Consortium delivers a set of specifications pertaining to different issues in online learning, such as learning design, curriculum sequencing, content packaging, learner information, and question and test interoperability (http://www.imsproject.org/). The Advanced Distributed Learning (ADL) Initiative, sponsored by the Office of the Secretary of Defense (OSD), aims at establishing the interoperability of learning tools and course content on a global scale, anywhere and anytime (http://www.adlnet.org/). ADL best known product is SCORM (Sharable Content Object Reference Model), a collection of specifications adapted from multiple sources (such as IMS, IEEE, and ARIADNE) to provide interoperability, accessibility and reusability of web-based learning content, and is meant for content developers, learning designers and learning management system (LMS) vendors.

4 Conclusions

In order to be successful, an e-learning system must satisfy a number of requirements. These include learning effectiveness, sufficient coverage of the learning domain, adaptivity, personalisation capabilities, intelligence, various pedagogical issues, as well as some purely technological criteria, such as robustness, reliability, scalability, and stable roots in current internet technologies. There are a number of methods and techniques to measure and evaluate learning effectiveness and other features of e-learning systems, and it is important for designers of such systems to understand and apply such techniques when developing learning environments. Likewise, only those e-learning systems developed to comply with emerging learning-technology standards have good chances to survive competitive markets and stay out of isolation and brittleness. For these reasons, paying close attention to evaluation and standardisation issues is a must in further development of e-learning systems.

References

1 Devedžić, V. (2003) 'Next-generation web-based education', *International Journal for Continuing Engineering Education and Life-long Learning*, Vol. 11, No. 1/2, pp.232–247.

2 Johnson, W.L., Rickel, J. and Lester, J.C. (2000) 'Animated pedagogical agents: face-to-face interaction in interactive learning environments', *International Journal of Artificial Intelligence in Education*, Vol. 11, pp.47–78.

3 Hendler, J. (2001) 'Agents and the Semantic Web', *IEEE Intelligent Systems*, Vol. 16, No. 2, pp.30–37.

4 Fensel, D., Van Harmelen, F., Horrocks, I., Mcguinness, D.L. and Patel-Schneider, P.F. (2001) 'OIL: an ontology infrastructure for the semantic web', *IEEE Intelligent Systems*, Vol. 16, No. 2, pp.38–45.

5 Gilbert, J.E., Hübscher, R. and Puntambekar, S. (2001) 'Assessment methods in web-based learning environments & adaptive hypermedia – introduction', in *Proc. AIED 2001 Workshop on Assessment Methods in Web-Based Learning Environments & Adaptive Hypermedia*, 19 May, pp.1–2.

6 Mitrović, A., Martin, B. and Mayo, M. (2002) 'Using evaluation to shape ITS design: results and experiences with SQL-tutor', *User Modeling and User-Adapted Interaction*, Jan. Vol. 12, No. 2, pp.243–279.

7 Weber, G. and Brusilovsky, P. (2001) 'ELM-ART: an adaptive versatile system for web-based instruction', *Int. J. Artificial Intelligence and Education*, Vol. 12, pp.351–384.

8 Oliver, M., Macbean, J., Conole, G. and Harvey, J. (2002) 'Using a toolkit to support the evaluation of learning', *J. Computer Assisted Learning*, Vol. 18, pp.199–208.

9 Van Lehn, K. and Niu, Z. (2001) 'Bayesian student modeling, user interfaces and feedback: a sensitivity analysis', *Int. J. Artificial Intelligence in Education*, Vol. 12, pp.154–184.

10 Constantino-Gonzalez, M., Suthers, D.D., Escamilla De Los Santos, J.G. (2002) 'Coaching web-based collaborative learning based on problem solution differences and participation', *Int. J. Artificial Intelligence in Education* Vol. 13. To appear. [Online]. Available: http://www.cogs.susx.ac.uk/ijaied/members02/archive/Vol_13/constantino/paper.pdf, December 2002.

11 Devedžić, V., Jerinić, L.J., Radović, D. (2000) The GET-BITS model of intelligent tutoring systems', *J. of Interactive Learning Research (JILR*), Special Issue on Intelligent Systems/Tools in Training and Lifelong Learning, Vol. 11, No. 3/4, pp.411–434.

12 Smith, A.S.G. and Blandford, A. (2002) 'MLTutor: an application of machine learning algorithms for an adaptive web-based information system', *Int. J. Artificial Intelligence in Education*, Vol.13, 2002. To appear. [Online]. Available: http://www.cogs.susx.ac.uk/ijaied/members02/archive/Vol_13/smith/paper.pdf

13 De Bra, P., Brusilovsky, P. and Houben, G-J. (1999) 'Adaptive hypermedia: from systems to framework', *ACM Computing Surveys*, December, Vol. 31, No. 4es, pp.1–6.

14 Karagiannidis, C., Sampson, D. and Brusilovsky, P. (2001) 'Layered evaluation of adaptive and personalized educational applications and services', in *Proc. AIED 2001 Workshop on Assessment Methods in Web-Based Learning Environments & Adaptive Hypermedia*, 19 May, pp.21–29.

15 IEEE Learning Technology Standards Committee (LTSC) P1484.12 Learning Object Metadata (LOM) working group. (2002, July 15). Draft Standard for Learning Object Metadata [Online]. Available: http://ltsc.ieee.org/wg12/

16 Nilsson, M. (Ed) (2003) *IEEE Learning Object Metadata RDF Binding* [Online]. Available: http://kmr.nada.kth.se/el/ims/md-lomrdf.html, 15 March.

17 Nejdl, W. et al. (2002) 'EDUTELLA: a P2P networking infrastructure based on RDF', in Proc. *W3C-WWW2002* Conf., 7–11 May, Honolulu, Hawaii, USA. [Online]. Available: http://www2002.org/CDROM/refereed/index-bytitle.html, May.

18 W3C Web-Ontology (WEBONT) Working Group (2003) *Web Ontology Language (OWL) Abstract Syntax and Semantics* [Online]. Available: http://www.w3.org/TR/2003/WD-owl-semantics-20030203/, 3 February.

19 Connolly, D. et al. (2001) DAML+OIL Reference Description [Online]. Available: http://www.w3.org/TR/daml+oil-reference, March.

20 Ariadne Foundation, Ariadne Educational Metadata Recommendation - V3.2 [Online]. Available: http://www.ariadne-eu.org/en/publications/metadata/ams_v32.html February 2002.