Towards Semantic E-Learning Environment for Biological and Environmental Science – a Review

PIRJO KUITUNEN
University of Jyväskylä
FINLAND

ANNE HONKARANTA
University of Jyväskylä and
SYSOPENDIGIA Plc

KALEVI SALONEN
University of Jyväskylä
FINLAND

Abstract: - Biological and environmental science has remained a discipline in which e-Learning is not fully utilized. Traditionally the Learning Objects of biological and environmental science are Web pages, unstructured documents, numerical data files and numerical data in databases, numerical data being an important part of teaching and learning. This paper reviews research on metadata, ontologies and markup languages for ecological and environmental e-Learning. Two mostly commonly used metadata specifications for e-Learning; Learning Object Metadata (LOM) and Dublin Core (DC) are described as well as two potential markup languages for biological and environmental data, EML and GML. The paper contributes by illustrating the relationships and concepts for semantic e-Learning environments and by discussing the applicability of the specifications and markup languages to biological and environmental science e-Learning.

Key-Words: - E-learning, Markup languages, Environmental data, Metadata, Ontology

1 Introduction

Growing number of students at universities, International master's programs, distance learning, and the rationalization and improvement of the quality of teaching have addressed a need to bring a part of the academic teaching and learning to the Internet. In biological and environmental science there is also a need to bind teaching to the real environmental data. It is easier to understand phenomena in a way of exploration rather than by plain statements of facts.

E-learning is education via the Internet, network, or standalone computer. E-learning applications and processes include Web-based learning, computer-based learning, virtual classrooms and digital collaboration. Content is delivered via the Internet, intranet/extranet, audio or video tape, satellite TV, and CD-ROM. E-learning has offered previously unavailable access and presentation methods and thus proved to be an effective way of delivering materials to previously unreachable students under widely variable circumstances [19]. E-learning bridges the in time and space between teachers and those who are learning.

Biological and environmental data mostly consist of raw data, which are numerical measurements made on field and calculations and various summaries derived from those. For distance learning and as complementary material to classroom teaching, the data collected in databases offer a possibility for binding the teaching to real world and its phenomena. Use of data in databases requires that there are means to search and connect the data along with metadata that defines its features and context for e-Learning.

Research on structured documents and related metadata has developed new solutions for interoperability, reusability and durability of resources on the Internet. World Wide Web Consortium (W3C) has produced specifications, techniques and languages on that area. All this is aimed to find a 'lingua franca' to be used with resources on the Internet. Extensible Markup Language (XML) is designed for content management on the Internet. It is a meta-language on which different markup and metadata languages have been built.

This paper reviews metadata standards, ontologies and markup languages potentially usable for biological and environmental science e-Learning environments and examines possibilities to build semantic learning environments using concepts of both pedagogy and biological and environmental science

Section 2 reviews the concepts of metadata, semantics, and ontologies and describes the LOM and DC metadata standards. Section 3 describes the e-Learning resources, data and metadata for biological and environmental science. Section 4 describes potential markup languages and ontologies for e-Learning on the domain, and section 5 concludes the paper.

2 Metadata and Semantics of Learning Objects

There are multiple connotations about the concept of metadata. Gilliland-Swetland [12] defines metadata as "the sum total of what one can say about any information object at any level of aggregation". Haase [13] states that metadata is any data which conveys knowledge about an item without requiring examination of the item itself whereas Anido [3] defines metadata as information about information.

Metadata record consists of structured information about the resource it describes. It is structured in a manner that facilitates the management, discovery and retrieval of those resources [1].

Educational metadata extend the scope of description that can be included in a metadata record with information that has particular educational relevance. Metadata are used in networked systems for object and knowledge sharing [14]. That gives possibilities to connect and use resources located in distributed and heterogeneous repositories.

Learning Object Metadata (IEEE 1484.12.1-2002 Learning Object Metadata standard, LOM) support the use of Learning Objects and foster platform interoperability. The LOM data model describes a hierarchy of data elements. The LOM metadata is defined by utilizing XML and XML Schema schemas. Each LOM element contains a hierarchy of subelements. For example the <educational> element include <learningresourcetype> element, which denotes the specific type of Learning Object (e.g. Exercise, Simulation, Diagram etc.). In addition to the name, size (the number of values allowed) and values, also the value range and data types of elements may be defined. Value range is the set of values allowed, and the data type indicates one of the six allowed data types.

For example the beginning of LOM document could be:

```
<lom>
<general>
<identifier>
<catalog>URI</catalog>
<entry>http://www.jyu.fi/documents/1234</entry>
</identifier>
<title>Kasviplanktonopas</title>
<language>fi</language>
</general>
…
</lom>
```

In addition to LOM the Dublin Core Metadata educational version, which has an element of "Audience" [29], [7] may be utilized. The "Audience" element has refinements "Mediator" and "Audience Education Level". Also DC element "Relation" has a new element-refinement, named "Conforms To", which is intended to refer to an established standard to which the resource conforms.

Mao et al. [18] argue that Metadata of LOM and DC focus only on the minimal set of elements and that their simple structure can't help students learn complex knowledge and relationships among topics. To complete a learning task, students not only require an understanding of what the learning materials talk about, but also the semantic connections and relationships between these materials.

IMS Learning Resource Meta-data Specification offers two mechanisms for adding semantics to LOM elements: 1) LOM element or value may be mapped to a more precisely defined element or term in a related schema or element set or 2) the meaning of the LOM element or value can be defined through reference to the best and common practices in the LOM community, or through the use and interpretation of definitions provided in the Oxford English Dictionary, Second Edition [15].

Sheth et al. [27] organize metadata into three types of semantics: 1) implicit semantics, which appear in unstructured text that has loosely defined and non-formal structure, 2) formal semantics, which appear when the data representation takes a more rigid form and 3) powerful semantics, which imply the combination of simple syntactic structures to represent the meaning of complex ones.

Ontology is an explicit specification of a conceptualization. It is a shared vocabulary plus a specification of its intended meaning [13]. The Semantic Web is a universe of metadata and ontologies expressed in machine readable format along with software tools, which allow the understanding of semantic relations among heterogeneous and distributed resources on the Web [8].

Among the current knowledge management techniques ontologies play a greater role than ever [31]. Current research on ontologies has shown that they facilitate the retrieval, interaction and management of resources (e.g. [17]). Ontologies allow teachers to organize major topics of the course. They also provide students opportunities to interact with Web-based courses and other educational systems and also support semantic information access to materials relevant to certain topics [19].

An example of an application of using semantic metadata and ontologies is the metadata project of Advanced Research in Intelligent educational Systems, Canada [6]. One of their approaches is to use WordNet [32] as a closed ontology from which learners and teachers select metadata vocabulary.

A controlled vocabulary is a set of unambiguous terms explicitly stated to be used in a specific domain [5]. Explicit conceptual relation occurs when there is at least one explicit class or entity representing a concept and related terms. The conceptualization is the product of a mental abstraction, which could be classification, aggregation or [4].

3 Biological and environmental e-Learning resources

According to Kaczorowska [17] "our academic community still does not embrace the opportunities offered by the Internet technologies and strategy of developing e-learning environment unfortunately falls far outside the mainstream of university education." Currently, there are many academic e-learning systems of which most influential ones are developed commercially [19].

3.1 E-Learning resources

Traditionally the digital learning resources for biological and environmental science are Web pages, unstructured documents, numerical data files and numerical data in databases. The contemporary Web pages contain both static XHTML-text with hypertext and interactive elements, implemented by some programming language (e.g. with Java, In Silico biology E-learning Environment, [16]).

Environmental and ecological research data consists of raw data and different kind of summaries made of that data. Data may be geospatial, i.e. related to a geographical location, and it may contain numeric or encoded observations stored in databases and tables as raw data. Interpreting the data needs understanding the types of variables, measurement units, potential biases in the measurements, sampling methodology and so on, i.e., the metadata. Environmental metadata may be seen as information about that data, the information required to understand data including data set contents, context, quality, structure, and accessibility [21]. Both data and metadata can be stored in local databases or they can be Web resources. Variables can be any quantities measuring water, air or land, e.g. water temperature, content of air pollutants etc.

Ecological or environmental metadata may provide multiple levels of support for e-Learning, it may 1) support data discovery, 2) facilitate acquisition, comprehension and utilization of data by humans, and 3) enable automated data discovery, ingestion, processing and analysis [20]. The last of these levels require comprehensive and structured metadata.

E-Learning metadata in biological and environmental science may be attached to a repository of a document archive, schemas for databases and structured documents, metadata for Web resources and a vocabulary or ontology for the domain or discipline.

Figure 1 illustrates the aforementioned relationships of metadata and ontologies to the e-Learning resources of biological and environmental metadata. The resources and concepts are depicted by rectangles on the figure.

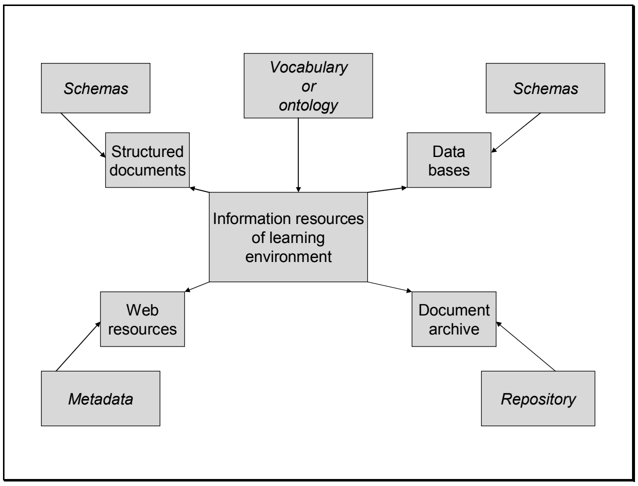

Figure 1. The resources of a learning environment and metadata (italic) connected to them.

3.2 Metadata Development – Related Work

There are several environmental and ecological metadata specifications and ongoing projects to produce widely accepted metadata and markup languages for environmental data. One of the projects is included in the work of ALTER-Net. Long Term Ecological Research (LTER) Network is creating LTER sites for Europe. ALTER-Net is developing the capacity for research and monitoring of the sites. The goal of the ALTER-Net WP I6 issue is the creation of a framework for sharing data, information and software tools amongst the

ALTER-Net partners in support of biodiversity research, policy and public understanding of science [26].

4 Markup Languages and Ontologies for Biological and Environmental Science

The need for markup languages for environmental metadata and data exchange has been recognized a long time ago. In the beginning of this century Arndt et al. [2] developed the Environmental Markup Language in Germany. However it never became widely adopted in academic world and the research of this topic slowed down [9]. This section reviews contemporary markup languages and ontologies that have potential for becoming more widely applied for biological and environmental science e-Learning.

4.1 Ecological Metadata Language (EML)

EML has been designed and developed by the ecological community to support data discovery, access, integration, and synthesis [10]. It is based on prior work done by the Ecological Society of America and associated efforts [21]. EML is implemented as a series of XML document types that can be used in a modular and extensible manner to document ecological data. Each EML module is designed to describe one logical part of the total metadata that should be included with any ecological dataset. EML was designed with many standards in mind. Those include Dublin Core, the Content Standard for Digital Geospatial Metadata (CSDGM from the US geological Survey's Federal Geographic Data Committee (FGDC)), the Biological Profile of the CSDGM (from the National Biological Information Infrastructure), the International Standards Organization's Geographic Information Standard (ISO 19115), the ISO 8601 Date and Time Standard, the OpenGIS Consortiums's Geography Markup Language (GML), the Scientific, Technical, and Medical Markup Language (STMML), and the Extensible Scientific Interchange Language (XSIL). EML is implemented in XML with schemas expressed as XML Schema schemas.

The EML module is a wrapper container that allows the inclusion of any metadata content in a single EML document. In EML, the definition of a resource comes from the Dublin Core metadata so that the top-level structure of EML has been designed to be compatible with the Dublin Core syntax. In EML, the definition of a "Data Package" is the combination of both the actual data and metadata for a resource. The metadata encoded by EML provides a formal description of data set contents. For example, all metadata files contain the name of the person who collected the data, where they were collected, the types of sampled organisms or systems, a description of the structure of the data set, the meaning of abbreviated variable names, the units of measurement, searchable key words, etc. In addition Ellison et al. [10] proposed that the current structure of EML would be expanded to allow for a specification of the computational methods or statistical models used to derive published data sets from raw data.

4.2 Geography Markup Language (GML)

The geographic information is the information that describes phenomena associated directly or indirectly with a location with respect to the Earth surface [22]. The geographic metadata describes the content, quality, condition and other characteristics of the data that allow a person to locate data and to understand them. The ISO-19115:2003 Geographic Information Metadata Standard is to most general standard available.

Geospatial applications are supported by database or file systems that can handle spatial data types. Spatial data objects have spatial attributes, such as location, geometry and neighborhood properties. GML documents nest spatial data types, permitting the effective representation of the various components of spatial data. GML targets both information storage and retrieval in its specifications.

GML includes various kinds of XML schemas for describing for features, geometries and topologies through a hierarchy of GML objects. The GML specification provides a series of schemas for describing geographic data in XML. The application schema applies the relevant features and types needed for the specific domain in question [28].

4.3 Towards Semantic E-Learning–Vocabulary and Ontology Development

Up to date generally no standards have been applied for environmental data resources, which can lead to problems in data exchange. It could also make it difficult to define a common ontology for the classification of data, which requires common concepts. Schentz et al. [25] have observed a strong interest in use of ontologies for data discovery as well as for harmonizing semantics for data transport.

Most initiatives in this area are using OWL, which has the advantage of integrating services (OWL-S) into ontology. This functionality would facilitate the integration of data with analysis and modeling tools.

In ALTER-Net the development of data sets for metadata system with associated core ontology is the first step towards an eventual information framework [25]. After a thorough study of available technologies, Schleidt and Schentz [27] have concluded that systems based only on XML and XML Schema schemas do not support semantics to the degree required. Relations between entries can not be annotated in a satisfactory way and there are no mechanisms for the extension of the schema. If an element must be extended, this is only possible by creating a new version of the entire schema. However, RDFS and OWL based systems cover most requirements, allowing for the creation of ontologies in order to neatly structure and annotate the environmental data. Object oriented metadata structure has advantages of inheritance and it allows information of new individual classes to be communicated without having to transfer the entire schema. Also the relations of ontologies link a data bit better to other data. As the American SEEK and US-LTER communities have started the development of ontologies for semantic annotation of ecological data, they strongly advised cooperation in this area.

A vocabulary of GEMET (GEneral Multilingual Environmental Thesaurus, [11]) has been developed as an indexing, retrieval and control tool for the European Topic Centre on Catalogue of Data Sources (ETC/CDS) and the European Environment Agency (EEA). It includes 22 different languages, e.g. Finnish, and it is divided to 40 different themes with 6562 terms. It defines a core of general terminology for the environment. Specific thesauri and descriptor systems (e.g. on Nature Conservation, on Wastes, on Energy, etc.) have been excluded and have been taken into account only for their structure and upper level terminology. RDF-files of the thesaurus are downloadable.

Semantic Web for Earth and Environmental Terminology (SWEET) offers ontologies, which are written in the OWL ontology language [30]. The SWEET project provides a common semantic framework for various Earth science initiatives.

Bermudez and Piasecki [5] showed that it is possible to extend and reuse metadata specifications and vocabularies distributed by OWL ontology language, by utilizing the language's flexibility to create restrictions on inherited properties. Also the domain specific ontologies can be used to harvest entries from distributed resources on the Internet.

5 Conclusions

In biological and environmental science the development of e-Learning has been slow and learning environments consist mostly of static Web-pages imitating non-digital learning material, or environments are used only for material delivery. The possibilities of e-Learning have still been only partly utilized. Interactivity with learning material and independence from time and place are unique for e-Learning. At the same time electronic material is free from organization as pages that should be turned in order, but remains to be freely organized in different ways for different purposes and even for different learners.

From the point of view of a learner one needs powerful means to orientate oneself on the information superhighway. Respectively, the teacher needs adequate tools to open up the information available. LOM and DC metadata about the resources has been used to facilitate the retrieval and use of learning materials. However, it has been pointed out that for understanding the semantic connections between learning objects semantic metadata is needed. Semantic metadata appends the metadata such as LOM or DC by providing explicit definitions of concepts and their semantic connections as well as the relationships between learning resources thus allowing teachers to organize and visualize major topics of the courses with respect to each other. Semantic metadata also provides students with opportunities to interact with Web-based material and educational applications, and support semantic information access to resources relevant to topics students are interested in. Using ontology makes learning resources reusable and searchable for learning and research. It also allows the use of multiple kinds of Internet resources such as data bases with metadata using the same vocabulary or learning objects from other learning environments.

However connecting semantics to learning environments is a demanding task. Developing domain ontologies which are accepted widely and are extensible and applicable enough for different learning environments is a huge work.

Web Ontology Language (OWL) as an object orientated approach with classes, properties and inheritance provides a good basis for creating extensible ontologies. With such kind of ontologies it is possible to create learning objects and environments with learning and domain specific, semantic metadata. For the discipline of biological and environmental science EML, SEEK, US-LTER, GEMET and SWEET are a good basis for an ontology using the idea of reusing metadata specifications and vocabularies declared by OWL.

References:

[1] Al-Khalifa, H. S. and Davis, H. C., *The Evolution of Metadata from Standards to Semantics in E-Learning Applications*, Proceedings of the Seventeenth Conference on Hypertext and Hypermedia (Odense, Denmark, August 22 - 25, 2006). HYPERTEXT '06. ACM Press, New York, NY, 2006, pp. 69-72.

[2] Arndt, H.K., Bandholtz, T., Günther, O., Rüther, M. and Schütz, T., *EML – The Environmental Markup Language*, Workshop Symposium on Integration in Environmental Information Systems ISESS 2000. Zell am See, Austria May 30 to June 3, 2000, pp. 1-9.

[3] Anido, L., *An Observatory for E-learning Technology Standards*, Advanced Technology for Learning, Vol. 3, No. 2, 2006, pp. 99-108.

[4] Batini, C., Di Battista, G. & Santucci, G. *Structuring Primitives for a Dictionary of Entity Relationship Data Schemas*, IEEE Transactions on Software Engineering, Vol. 19, No. 4, 1993.

[5] Bermudez, L. & Piasecki, M., *Metadata Community Profile for Semantic Web*, Geoinformatica, Vol. 10, No. 2, 2006, pp. 159-176.

[6] Brooks, C., McCalla, G. and Winter M., *Flexible Learning Object Metadata*, 3rd International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL 05) held in conjunction with the 12th International Conference on Artificial Intelligence in Education (AIED 2005), July 18 – 22, 2005. Amsterdam, The Netherlands, 2005.

[7] DCMI., *DCMI Metadata Terms*, http://dublincore.org/documents/dcmi-terms/, retrieved 14 June 2007.

[8] Egenhofer, M.J., *Toward the Semantic Geospatial Web*, Tenth ACM International Symposium of Advanced in Geographic Information Systems, ACM Press: McLean, VA, USA, 2002.

[9] EIS, *Environmental Information Systems*, School of Business and Economics at the Humboldt-Universitaet zu Berlin. http://zope.wiwi.hu-berlin.de/Professuren/quantitativ/wi/forschung-en/uis, retrieved 13 June 2007.

[10] Ellison, A.M., Osterweil, L.J., Clarke, L., Hadley, J.L., Wise A., Boose, E. Foster, D.R., Hanson, A., Jensen, D., Kuzeja, P., Riseman, E. & Schultz, H.., *Analytic Webs support the Synthesis of Ecological data sets*, Ecology, Vol. 87, No. 6, 2006, pp. 1345-1358.

[11] GEMET, *GEMET Thesaurus version 1.0*, http://www.eionet.europa.eu/gemet, 2004, retrieved 30 July 2007.

[12] Gilliland-Swetland, A.J., *Defining Metadata*, in Introduction to Metadata. M. Baca (ed.), Getty Information Institute, Los Angeles, 1998.

[13] Gruber, T., *A Translation Approach to portable Ontology Specification*, Knowledge Acquisition, Vol.5, No.2, 1993, pp. 199-220.

[14] Hatala, M., Richards, G., Eap, T., and Willms, J., *The Interoperability of Learning Object Repositories and Services: Standards, Implementations and Lessons learned*, Proceedings of the 13th international World Wide Web Conference on Alternate Track Papers &Amp; Posters (New York, NY, USA, May 19 - 21, 2004). WWW Alt. '04. ACM Press, New York, NY, 2004, pp. 19-27.

[15] IMS, *IMS Meta-data Best Practice Guide for IEEE 1484.12.1-2002 Standard for Learning Object Metadata*, Revision: 31 August 2006, http://www.imsglobal.org/metadata/mdv1p3/imsmd_bestv1p3.html#1634949, retrieved 14 June 2007.

[16] ISee, *In Silico biology E-learning Environment*, http://www-helix.inrialpes.fr/projects/ISee/, retrieved 31 July 2007.

[17] Kaczorowska, A.-K., *Teaching on the Threats of Bioterrorism using an E-Learning Platform*, Acta Biochimica Polonica, 40th Meeting of the Polish Biochemical Society Lublin 19–23 September, 2005, Session 10: Didactics of biochemistry, molecular biology and biotechnology, p. 162.

[18] LOM, *Learning Object Metadata* , http://ltsc.ieee.org/wg12/index.html, retrieved 12 June 2007.

[19] Mao, M., Peng, Y. & He, D., *Ontology-based content Management and Access Framework for Supporting E-Learning Systems*, Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops) (WI-IATW'06), 2006.

[20] Michener, W.K., *Meta-information concepts for Ecological data Management*, Ecological informatics, Vol. 1, No. 1, 2006, pp. 3-7.

[21] Michener, W.K., Brunt, J.W., Helly, J., Kirchner, T.B. and Stafford, S.G., *Non-geospatial Metadata for the Ecological Sciences*, Ecological Applications Vol. 7, No. 1, 1997, pp. 330-342.

[22] Nogueras-Iso, J., Zarazaga-Soria, F.J., Lacasta, J., Béjar, R. & Muro-Medrano, P.R., *Metadata Standard Interoperability: Application in the*

Geographic Information Domain, Computers, Environment and Urban Systems, Vol. 28, No. 6, 2004, pp. 611-634.

[23] Polsani, P.R., *Use and Abuse of Reusable Learning Objects*, Journal of Digital Information, Vol. 3, No. 4, 2003.

[24] Polsani, P.R., *Why Learning Objects?*, Proceedings of the IEEE International conference of Advanced Learning Technolgies (ICALT'04), 2004, pp. 1-2.

[25] Schentz, H., Schleidt, K. & Lane, M., *A Long-Term Biodiversity, Functional Concept for a Biodiversity and Ecological Network*, A Long-Term Biodiversity, Ecosystem and Awareness Research Network, Project no. GOCE-CT-2003-505298, ALTER-Net, March 2006, pp. 1-75.

[26] Schleidt, K., Lane, M. & Schentz, H., *Requirements for an IT Framework for ALTER-Net Results from t he Requirements Analysis conducted by Work Package I6*, A Long-Term Biodiversity, Ecosystem and Awareness Research Network, ALTER-Net, Project no. GOCE-CT-2003-505298, WPI6_2005_01.

[27] Schleidt, K. and Schentz, H., *Metadata Required for and ALTER-Net Data and Application Network*, A Long-Term Biodiversity, Ecosystem and Awareness Research Network, Project no. GOCE-CT-2003-505298, ALTER-Net, February 2006, pp. 1-26.

[28] Smiths, P.C. and Friis-Christensen, A., *Resource Discovery in a European Spatial Data Infrastructure*, IEEE Transactions on Knowledge and Data Engineering, Vol. 19, No. 1, 2007, pp. 85-95.

[29] Sutton, S.A. & Mason, J., *The Dublin Core and Metadata for Educational Resources*, Proc, Int'l. Conf. on Dublin Core and Metadata Applications 2001, p. 25-31.

[30] SWEET, *Semantic Web for Earth and Environmental Terminology*, http://sweet.jpl.nasa.gov/sweet/, retrieved 30 July 2007.

[31] Tane, J., Schmitz, C., and Stumme, G., *Semantic Resource Management for the Web: an E-Learning Application*, Proceedings of the 13th international World Wide Web Conference on Alternate Track Papers &Amp; Posters (New York, NY, USA, May 19 - 21, 2004). WWW Alt. '04. ACM Press, New York, NY, 2004, pp. 1-10.

[32] WordNet, *WordNet, a Lexical Database for the English language*, http://wordnet.princeton.edu/, retrieved 14 June 2007.