The Sixth International Conference on e-Learning (eLearning-2015), 24-25 September 2015, Belgrade, Serbia

LRMI MARKUP OF OER CONTENT WITHIN THE BAEKTEL PROJECT

RANKA STANKOVIĆ
University of Belgrade, Faculty of Mining and Geology, ranka.stankovic@rgf.bg.ac.rs

DANIELA CARLUCCI
University of Basilicata, Matera, Italy, daniela.carlucci@unibas.it

OLIVERA KITANOVIĆ
University of Belgrade, Faculty of Mining and Geology, olivera.kitanovic@rgf.bg.ac.rs

NIKOLA VULOVIĆ
University of Belgrade, Faculty of Mining and Geology, nikola.vulovic@rgf.bg.ac.rs

BOJAN ZLATIĆ
University of Belgrade, Faculty of Mining and Geology, bojan.zlatic@rgf.bg.ac.rs

Abstract: *This paper outlines the approach to tagging of OER content with metadata within the BAEKTEL project with the aim of increasing the discoverability of OER materials by potential users. The approach is based on the Learning Resource Metadata Initiative (LRMI), which makes it easier to publish, discover, and deliver quality educational resources on the web using standard schemas for HTML pages markup in ways recognized by major search providers: Google, Yahoo, Microsoft Bing. LRMI is proposed as an extension to the Schema.org metadata vocabulary addressing the needs of the educational community. The main rationale for the approach is that the choice of a standardized learning object metadata vocabulary has valuable and beneficial implications. Within the BAEKTEL platform, machine readable information is inserted into the code of web pages, which helps search engines interpret the sense of the text on those pages, but also other metadata on OER content according to LRMI.*

Keywords: *E-Learning, Open Educational Resources, LRMI, Microtagging, Schema.org*

1. INTRODUCTION

Educational resources are shared in both physical and virtual communities, where more and more are published as Open Educational Resources (OER). Nowadays, online searches for educational materials are conducted more often than a few years ago. Although Internet searches are easy and speedy, they are not always efficient and retrieve lots of irrelevant results due to the huge amount of materials available. Through the use of metadata the Learning Resource Metadata Initiative (LRMI) offers ways to help pinpoint the right learning resources to address user needs. Applying metadata tags to educational materials makes them more easily discoverable through online searching [1]. Descriptors defined in LRMI include: physical attributes, pedagogical relevance and instructional strategies. Retrieval of resources on basis of the way the learner will be engaged supports differentiated strategies and individualized learning.

In order to make OER accessible and discoverable by potential users across the web, they should be annotated in such a way that the users can understand what specific learning objects are about, what is their learning content and prerequisites for their use, without even seeing them [2].

This paper aims to provide an overview of the ways that metadata are implemented in the LRMI and applied in BAEKTEL (Blending Academic and Entrepreneurial Knowledge in Technology Enhanced Learning, http://www.baektel.eu) project. The main motive for implementation of LRMI is to help students and teachers quickly find relevant educational content via regular search engine. Within the BAETEL project we have developed a metadata portal as the metadata and content repository, as we felt the need for a better way to describe the content and thus increase web visibility for OER resources published within BAEKTEL project.

In defining metadata for BAEKTEL resources several standards were analyzed and tree of them are being used: Dublin Core (DC) [3], IEEE Learning Object Metadata (LOM) [4] and the Learning Resource Metadata Initiative (LRMI) [5]. The focus of this paper is on LRMI metadata and their management in the context of BAEKTEL OER framework, which is described in more detail in section 2. In section 3 a review of semantic annotation implementation with examples of resource tagging is given. Section 4 of this paper outlines the key aspects of the LRMI standard for describing educational resources, including metadata schema and implementation of LRMI metadata. In section 5 is given short description and architecture of Open edX, as the open source platform that powers edX courses. The approach for semantic annotation, respectively markup of edX.BAEKTEL resources is given in section 6. Last section 7 gives conclusions, followed by expectation of benefit and future implementation plans.

2. THE BAEKTEL OER FRAMEWORK

The main goal of the BAEKTEL project is building an OER network for publishing educational materials by higher education (HE) institutions and best practice examples by enterprise experts. The resources can be published in different languages, with support for their translation [6].

The ICT solution for BAEKTEL OER framework comprises a central repository with OER resources and the BAEKTEL Metadata Portal (BMP). All project partners develop and publish their OER on http://edx.baektel.eu/, which is build using edX, an open-source online learning platform. BMP is providing all important information on the network resources, thus enabling their search and browse. [7]

The development of the ICT solution for the BAEKTEL metadata portal is based upon a metadata schema that contains elements taken from standard namespaces with guidelines for metadata creation. In metadata specification, the standard nomenclature was used, enabling learning resources to be described and shared in a common way, and thus enhancing their accessibility from other OER portals.

The approach to defining metadata within BAEKTEL combines Doublin Core and LOM Standard [8]. In development of the BMP model, compliance with these standards was obligatory, as the BMP metadata had to provide for sharing with other OER repositories.

BMP integrates also web services for terminological and linguistic support, such as query expansion, information retrieval, OER indexing, and the like. OER metadata include information such as resource title, author, subject, creation date and the like, which facilitates search, but also acquisition, use and reuse of learning objects.

3. SEMANTIC ANNOTATION

The next generation of the Web is denoted as Web 3.0, which is an umbrella term for customization, semantic contents, and more sophisticated web applications toward artificial intelligence, including computer-generated contents. While the conventional Web is the “Web of documents”, the Semantic Web includes the “Web of Data”, which connects “things” rather than documents. The machine-readable datasets of the Semantic Web are used in: search engines, data integration, resource discovery and classification, cataloging, intelligent software agents, content rating, Big Data processing, etc. [9]

The union of the structured datasets forms the Linked Open Data Cloud, the decentralized core of the Semantic Web, where software agents can automatically find relationships between entities and make new discoveries. Linked Data search engines crawl the Web of Data by following links between data sources and provide expressive query capabilities over aggregated data.

The key idea of the OER movement is that open educational content should be maximally shared. In [10] authors have shown that interoperability is an important aspect of open educational contents integration and reuse. The use of linked data with free open OCW (OpenCourseWare) data also fosters interoperability and enables re-use, remix, and adaptation of open educational tools.

Several markup standards are used to annotate information about products, people, organizations, events, within HTML pages. They serve the same purpose: providing a vocabulary for semantic markup. The most prevalent of these standards are Microformats, which use style definitions to annotate HTML text with terms from a fixed set of vocabularies; RDFa which is used to embed any kind of RDF data into HTML pages, and Microdata, a recent, currently the most deployed format developed in the context of HTML5. [11]

Microdata is an attempt to provide a simpler way of annotating HTML elements with machine-readable tags than the similar approaches of using RDFa and Microformats. Using attributes, nestable groups of name-value pairs of data, called microdata gives a whole new way to add extra semantic information and extend HTML5. [12]

Microdata is specified inside *itemtype* and *itemprop* attributes added to existing HTML elements; microformat keywords are added inside class attributes and RDFa relies on *rel*, *typeof* and *property* attributes added to existing elements. One example of such microdata tagging for a page containing an organization e.g. faculty with address is:

```
<div itemscope itemtype=
"http://schema.org/PostalAddress">
<span itemprop="name"> Faculty of Mining and
Geology </span>
<span itemprop="addressLocality">Đušina 7, Belgrade
</span>,
<span itemprop="addressCountry">Serbia </span>
</div>
```

HTML elements have new meta-syntax for meanings in HTML5 with microdata. A parser can extract the knowledge encoded in previous HTML code. It could be turned into the following JSON:

```
{
 "items": [
  {"type": [ "http://schema.org/PostalAddress" ],
   "properties": {
      "name": [ "Faculty of Mining and Geology" ],
      "addressLocality ": [ "Đušina 7, Belgrade" ],
      "addressCountry": [ "Serbia" ]}
  } ]
}
```

Schema exploits microdata in documentation, while microformats go into class names what can be seen in an example of Microformat use:

<li **class="org"**> Faculty of Mining and Geology
<li **class="tel"**>+381113219101

RDFa provides a set of attributes: *about, rel, rev, src, href, resource, property, content, datatype, typeof*, that can be used to carry metadata, like in example:

<div xmlns:dc="http://purl.org/dc/terms/">
My name is Ranka
and my last book that I read is <cite **property="dc:title"**> Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data </cite> by Leslie F. Sikos .published <em **property="dc:date" datatype="xsd:date" content="20150610"**>June 10th, 2015</div>

Schema.org provides a collection of microdata schemas, i.e. HTML tags that can be added to web pages to describe the things their pages are about. Schema.org was launched in 2011, as a joint initiative of Google, Yahoo, Microsoft Bing, Yandex and W3C. Schema.org provides a collection of schemas for HTML pages markup in ways recognized by major search providers and used for structured data interoperability [13].

The Schema.org metadata has the goal to improve the display of search results, thus making it easier for users to find the right web pages. The machine readable information inserted into the code of web pages helps search engines interpret the sense of the text on those pages.

4. LRMI METADATA MANAGEMENT

LRMI is an extension of schema.org. The main rationale for the approach fostered by IMS GLC is that the choice of a standardized learning object metadata vocabulary has valuable and beneficial institutional and pedagogical implications.

When a large amount of OER is reached, it is even more important that they are well described and tagged in a standard way in machine readable form. In that case, results returned by search engines are more relevant, and both educators and learners can find and compare learning materials that best suit their current needs more efficiently.

Implementation can be accomplished in the following four ways:
1. Use existing or create new metadata.
2. Map the metadata to the LRMI tag elements.
3. Use vocabulary that is common with everyone else's, such as the Common Core, plus additional terms from LOM and Dublin Core.

The metadata schema

Schema.org defines the metadata schema for basic information about material on the Web, such as a "Creative Work" in the form of a "Book," which has metadata like *author, publisher, datePublished*, etc. It also defines specific metadata for types like *Article, ImageObject, AudioObject, VideoObject*, and so forth.

The LRMI specification proposed classes *AlignmentObject* and *EducationalAudience*, as well as properties[1]: *alignmentType*, *educationalAlignment, educationalFramework, educationalRole, educationalUse, interactivityType, isBasedOnUrl, learningResourceType, targetDescription, targetName, targetUrl, timeRequired, typicalAgeRange, useRightsUrl*.

The educational role describes the target audience of the content, while educational alignment points out to an established educational framework (or other educational scheme). The educational use is the educational purpose of the resource, for example 'assignment' or 'group work'. The interactivity type describes the predominant mode of learning, like 'active', 'expositive' or 'mixed'.

If a learning resource is derived from another resource, property based on url is used. The learning resource type describes the predominant type e.g. 'presentation', or 'handout'. Similar to other scheme are other properties. Detailed scheme with all classes and properties is given in http://www.lrmi.net/the-specification, while LRMI Metadata Terms in RDF are given in http://dublincore.org/dcx/lrmi-terms/

Implementation of LRMI metadata

After the schema.org ontology expansion, the next step is to provide to publishers of OER to implement LRMI/schema.org metadata in content management and dissemination systems. The Learning Registry, which provides a mechanism for sharing information about learning resources is exploring the use of JSON-LD for sharing LRMI metadata as stand-alone records rather than embedded in the HTML of web pages so that it may be used in resource discovery by search services that are connected to the Learning Registry. Other OER projects, similar to BAEKTEL also can benefit from LRMI metadata tagging.

And at the end, users usually search for educational resources using big search engine providers: Google, Bing, Yahoo, Yandex, Baidu; LRMI has an advantage over others in enhancing educational resource discovery. Well described resources would help to overcome some of the problems that currently exist around the discovery of resources that meet specific individual learning requirements. [6 Nelson again]

Another type of additional data about courses are paradata. The term paradata itself is relatively new and it is used to refer to a particular kind of metadata about

[1] http://dublincore.org/dcx/lrmi-terms

usage data. Typical paradata are about how something is being used, the times of day it was used, how long the learning took, how many times it was accessed etc. In BAEKTEL project paradata are gathered by Google analytics.

5. EDX ARHITECTURE

For BAEKTEL project Open edX is used as a web-based platform for creating, delivering, and analyzing online courses. Figure 1 presents the current architecture of the platform. Most of the server-side code in Open edX is in Python, with Django as the web application framework, using Mako templates.

The browser-side code is written primarily in JavaScript with use of the Backbone.js framework. Open EdX uses Sass and the Bourbon framework for CSS code. In this paper we will outline just components used for metadata tagging.

- edX LMS with courses management in Mongo DB, with videos served from YouTube or Amazon S3 and student data in MySQL,
- edX Studio for course development in same Mongo DB that the LMS uses
- Course Browsing, as a simple front page for browsing courses (edx.org site home page and course discovery site is not open source).

Units of course are called XBlocks and users can create new types of XBlocks, but there are other ways to extend courseware behavior like: embedding an LTI (Learning Tools Interoperability) tools to integrate other learning tools into an Open edX course or JavaScript components integrated using JS Input. Also, courses can be exported and imported using Open Learning XML (OLX), an XML-based format for courses.

Other components of edX platform are course discussions, mobile application support, analytics, but they are not related to LRMI metadata tagging.

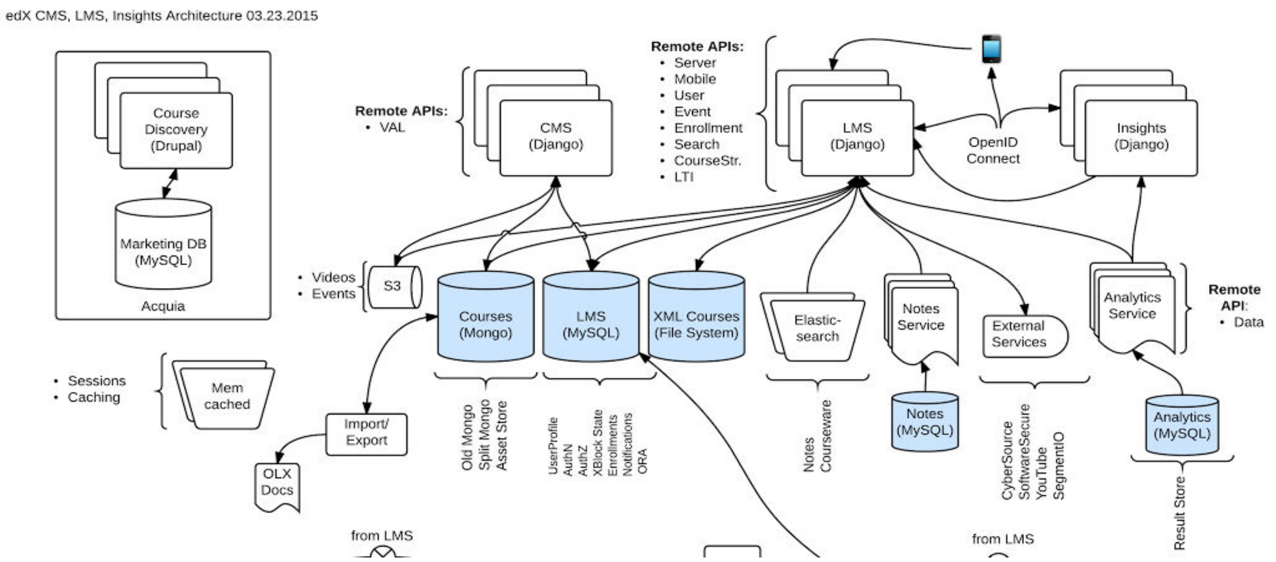

Figure 1: edX architecture (https://open.edx.org/contributing-to-edx/architecture)

6. MARKUP OF EDX.BAEKTEL RESOURCES

Integration of BMP portal for indexing and edX portal with contents is implemented with Open edX Platform REST APIs (https://open.edx.org/open-edx-rest-apis). It allows building of web, desktop, and mobile applications that work with local Open edX instances, use Representational State Transfer (REST) design principles and support JavaScript Object Notation (JSON) data-interchange format.

The Course Info API exposes all the attributes of a course that describe a course, without exposing the underlying configuration or content. This can be used as a single source for About Pages, descriptions, and eventually additional front end clients. This API returns course information from all current Course documents available in edX-platform, including a comprehensive set of course information, including enrollment information, configurable marketing copy, and so on.

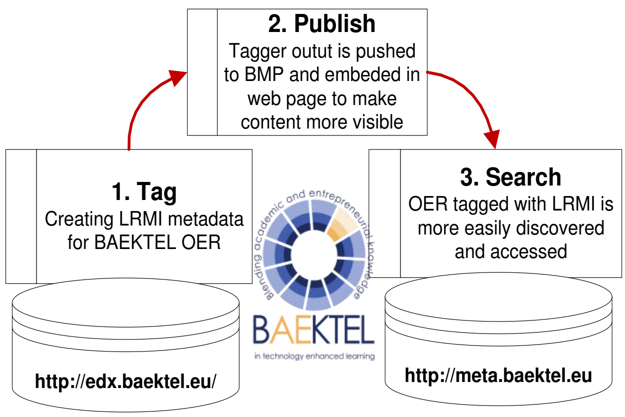

Figure 2: LRMI tagging in BAEKTEL

Edx platform is using Mongo NoSQL database for storing courses and custom APIs are used for retrieving metadata and structure for the courses. The use of these APIs is implemented with the GET request. Three most important types of interface are presented with examples of retrieving the aforementioned data.

- GET request for a list of courses is https://..../api/course_structure/v0/courses/ . In order to get the resulting list of courses as JSON data , it is necessary to set format request parameter to json (GET /api/course_structure/V0/courses/?format = json). The result will contain the number of courses, the url of previous and following page, as well as the list of JSON data containing information about the course: id, name, category, start time, end time, url course
- GET request api/course_structure/v0/courses/{course_id}/only returns metadata for the course course_id. The result in JSON is retrieved like in previous request
- The request GET /api/course_structure/v0/course_structures/{course_id}/ returns data related to the structure of the selected course. As in the previous examples with format parameter set to json (GET /api/course_structure/V0/course_structures/{course_id}/?format = json), the result will be JSON data containing: root_id and collection of blocks, which are component parts of the course. In the collection of retrieved data related to the blocks are: id of block, type (html, problem, video, discussions ...) and the collection of sub-blocks, the name of the block

Open edX Platform REST APIs is used for annotation of course static pages with microdata in compliance with LRMI standard. An example of extracted structured data:

rdfa-node relationship:
name: license
value: Attribution-NonCommercial 3.0 Unported
href: http://creativecommons.org/licenses/by-nc/3.0/
Item type: http://schema.org/creativework property:
datecreated: 1 September 2015
learningresourcetype: Course module
name: Leksicko prepoznavanje u obradi prirodnih jezika
author: Cvetana Krstev, Biljana Lazić
publisher: University of Belgrade
description: Kursom su obuhvaćene morfološka, leksička i sintaksička analiza u obradi prirodnih jezika. Neke od tema su upotreba regularnih izraza za prepoznavanje obrazaca u tekstu, konačni automati, transduktori, elektronski rečnici, kaskade i višečlane reči.
about: Unitex
about: Computational linguistics
about: Natural language procesing
about: Računarska linvistika
about: Obrada tekstova na prirodnom jeziku
about: elektronski rečnici
about: analiza teksta
about: konačni automati

Apart from edX resources, other OER published within BAEKTEL platform will be annotated as well in similar way.

7. CONCLUSION

The future of the web, what some call Web 3.0, is based on semantic search and algorithms that will help machines make sense of the data on web pages, interrelate it, and ultimately enable more adaptive learning environments for students. So the more material we tag with detailed microdata, the more "findable" our content and educational resources will be for teachers and students.

The solution for LRMI implementation within BAEKTEL metadata portal outlined in this paper enables efficient search and browse of OER content and provides the infrastructure for successful search engine optimization of education resources published within BAEKTREL project.

With variety of content and the expected growth of the number of resources, as well as different profiles of potential users, indexing of resources that enables their efficient location within the network is very important for dissemination and sustainability of BAEKTEL project. To that end a metadata vocabulary and data structure syntax based on DC and LOM were implemented within ResourceSpace, to offer a flexible and robust mechanism for indexing OER content and enabling the user to easily locate the resources of interest, but it is not enough because most of users search for educational resources using big search engines e.g. Google, Bing,...

For that purpose BAEKTEL project implements LRMI as a markup language, where its simplicity makes it useful and easily implemented convention for tagging content. Key benefits of using this approach is expanded access to descriptive data on educational resources, pooling knowledge about learning resources and providing tools and services to create applications that make use of big data about resources.

However, a lot of work still needs to be done before BAEKTEL enters full exploitation to the benefit of future and current university students, as well as university graduates working in enterprises. Namely, the population of the network with resources is now crucial for bringing the BAEKTEL into full function, thus providing usability to the features outlined in this paper.

The expectation and hope is that the implementation of LRMI within BAEKTEL project would lead to improved "discoverability" of published materials.

LITERATURE

[1] LRMI guide: The Content Developer's Guide to the Learning Resource Metadata Initiative and Learning Registry, 2014, Association of Educational Publishers, http://www.lrmi.net/wp-content/uploads/2013/03/lrmi_lr_guide.pdf.

[2] Velichová, D., *RLO in e-Learning Solutions.* In Proceedings of 9th International Conference Virtual

University Slovakia, 2008, http://virtuni.eas.sk/rocnik/ 2008/pdf/aid000262.pdf

[3] The Dublin Core Metadata Initiative (DCMI) http://dublincore.org/, ISO 15836:2009 Information and documentation - The Dublin Core metadata element set, 2012.

[4] IEEE Learning Technology Standards Committee (LTSC), 1484.12.1-2002 IEEE Standard for Learning Object Metadata, 2002.

[5] Friesen, N., *Technical Evaluation Report, 40. The International Learning Object Metadata Survey*, CanCore Initiative, 2004. http://www.irrodl.org/index.php/irrodl/article/view/195/277

[6] Obradović, I., Stanković, R., *Using technology for knowledge transfer between academia and enterprises* Proceedings of IFKAD, Matera, Italy, pp. 788-801, 2014.

[7] Carlucci, D., Kitanović, O., Rajlić, I., Vulović, N., Jevremović, Z., A platform for management of academic and entrepreneurial knowledge, In Proceedings of IFKAD, Bari, Italy, 2014.

[8] Stanković, R., Kitanović, O., Obradović, I., Linzalone, R., Schiuma, G., Carlucci, D., Using Metadata For Content Indexing Within An Oer Network. *In* Proceedings of The Fourth International Conference on e-Learning (eLearning-2014), Belgrade, Serbia, 2014.

[9] Sikos, L.F., Mastering Structured Data on the Semantic Web, From HTML5 Microdata to Linked Open Data, 2015, Springer Science+Business Media New York.

[10] Nelson, P., Edmundo, T., Ricardo C.P., Jorge, L.V., Janneth, A.C., "Consuming and producing linked open data: the case of OpenCourseWare", Program: electronic library and information systems, Vol. 48 Iss: 1, pp.16 - 40, 2014.

[11] Meusel, R., Petrovski, P., Bizer, C.: The webdatacommons microdata, rdfa and microformat dataset series. In: ISWC, 2014.

[12] Paulheim, H. What the Adoption of schema. org Tells About Linked Open Data, 2015, http://ceur-ws.org/Vol-1362/PROFILES2015_paper6.pdf

[13] Barker, P and Campbell, L.M. *Learning Resource Metadata Initiative: using schema.org to describe open educational resources*. In Proceedings of OpenCourseWare Consortium Global 2014: Open Education for a Multicultural World, 2014. http://conference.oeconsortium.org/2014/wp-content/uploads/2014/02/Paper_34-LMRI1.pdf, 2014.